



Mirja Tarnanen

ARVIOIJA VALOKEILASSA

suomi toisena kielenä -kirjoittamisen arviointia

ARVIOIJA VALOKEILASSA:
suomi toisena kielenä -kirjoittamisen arviointia

Mirja Tarnanen

**ARVIOIJA VALOKEILASSA:
suomi toisena kielenä -kirjoittamisen arviointia**

Jyväskylän yliopisto
Soveltavan kielentutkimuksen keskus
2002

**Arvioija valokeilassa
Suomi toisena kielenä -kirjoittamisen arviointia**

Jyväskylän yliopisto
Soveltavan kielentutkimuksen keskus
PL 35
40014 Jyväskylän yliopisto
<http://www.jyu.fi/solki>
solki-info@jyu.fi

© Mirja Tarnanen ja Jyväskylän yliopisto/Soveltavan kielentutkimuksen keskus

ISBN 951-39-1264-7 (nid.)
ISBN 978-951-39-3419-4 (PDF)

Jyväskylän yliopistopaino, Jyväskylä 2002

ABSTRACT

Tarnanen, Mirja

The rater in a spotlight: Rating Finnish as a second language writing

Jyväskylä: Centre for Applied Language Studies, University of Jyväskylä, 2002

ISBN 951-39-1264-7

This study explores rating behaviour by combining a quantitative and a qualitative research method. The purpose of the study is to investigate the way in which raters making judgements relate to the process and outcomes of assessment (ie., the nature of intra- and inter-rater consistency). The quantitative approach investigates this consistency numerically. The qualitative approach makes use of the arguments, beliefs and ideas reported by the judges. These comments relate to aspects in the rating process such as the type of scale used, the type of the task rated, and the feelings of the raters in the course of the rating process.

Writing performances were collected from adult learners of Finnish as a second language (N=247) representing different linguistic and cultural backgrounds. Each learner completed three tasks from a set of nine tasks in a matrix design such that one task was common to all writers. The raters (N=17) were teachers of Finnish as a second language, all of whom rated performances from at least 100 writers. There were three rating rounds. In the first round, the raters used a numerical scale from one to six, without any level descriptors. The purpose of this scale was to find out about the judges' perceptions about writing in Finnish as a second language. In the second rating round, the raters used a six-level scale *with* descriptors (based on the Council of Europe scale). The third rating round was a replication of the second. After the first and the second round, the raters were interviewed using a semi-structured interview. After the third rating round, the raters participated in a repertory grid interview. This procedure was completed by all the raters.

The numerical results of the study showed that scores were consistent across all three rating rounds. Such a high inter-rater consistency in the first rating round was unexpected. According to interview data from the first round, the most common criterion in the judges' scales were grammatical which may explain the high inter-rater consistency. The open scale used in the first rating round caused uncertainty, and brought forth comments on how difficult it was to apply one's own criteria systematically. The use of the scale with descriptors increased both intra- and inter-rater consistency to a degree. However, the raters found the scale too general and ambiguous. Especially, they criticized the lack of grammar descriptors. The third, repertory grid, interview worked as a bridge between perceptions and behaviour. It showed that the raters were aware of the different textual, linguistic and social-cultural features of the texts but, nevertheless, judged the same texts mainly on the basis grammar and vocabulary. As a whole, all interview data showed how complex, unpredicted and subjective an assessment process is, even though numerical data were fairly consistent. The methodological and educational implications of the study highlight the importance of combining a quantitative and qualitative approach to gain a deeper understanding of the assessment process and to make assessment more ethical and reliable.

Keywords: reliability, intra- and inter-rater consistency, assessment of writing, Finnish as a second language

ESIPUHE

Arviointi on kuin kamera, jolla voi ottaa kuvia osaamisen tasosta, oppimisprosessista tai opetuksen laadusta. Kuvat ovat usein luonteeltaan staattisia eivätkä ne paljasta koko dynaamista totuutta. Ne ovat kuitenkin kaikin tavoin mielenkiintoisia ja innostavia, koska niiden avulla saa tietoa osaamisesta, ne voivat ohjata paremmin ymmärtämään itseään oppijana tai muiden osaamisen arvottajana tai ne voivat antaa vihjeitä siitä, miten ja miksi opetusta pitäisi muuttaa. Arviointikuvien tulkinta ja käyttö edellyttävät kuitenkin eettistä ja koulutuksellista pohdintaa siitä, miten luotettavia kuvat ovat, miten valta ja vastuu jakaantuvat, kuka kuvista hyötyy ja millä tavalla ja miten kuvien tarkkuutta ja syvyyttä voitaisiin kehittää entisestään.

Käsillä oleva arviointikuva, kirja, on tutkimus arvioijien luotettavuudesta ja arviointiprosessin luonteesta. Vaikka tutkimuksen valokeilassa ovat erityisesti suomi toisena kielenä –opettajat, uskon, että valokeilaan osuneet käsitykset, uskomukset, tuntemukset ja kokemukset ovat jossakin määrin tuttuja kaikille arvioijille. Samoin toisen kielen kirjoittamisen arvioinnilla on erityispiirteistään huolimatta monia yhtymäkohtia myös muiden kielitaidon osa-alueiden ja äidinkielen kirjoittamisen arviointiin.

Tämä julkaisu perustuu samannimiseen monisteena tarkastettuun väitöskirjaani, minkä vuoksi tekstin rakenne on opinnäytämäinen. Toivon kuitenkin tekstin sisällön kuljettavan lukijaa tavoitteellisesti ja johdonmukaisesti – ja myös kiinnostavasti. Kiitänkin kaikkia ”esilukijoita”, jotka ovat lukeneet väitöskirjaani sen eri vaiheissa, hyödyllisistä ja ohjaavista kommentteista. On ollut opettavaista saada pohtia ja viedä esitettyjä kommentteja tekstiin myös prosessin tässä vaiheessa; omaa työtään on tullut arvioineeksi melko perusteellisesti ja monesta näkökulmasta, vaikka se ei aina näkyisikään tekstistä.

Kiitän lämpimästi Soveltavan kielentutkimuksen keskusta tutkimukseni julkaisemisesta. Sinikka Lampista kiitän editoidun version tarkastamisesta ja yhteistyöstä kannen suunnittelussa.

Kun tarkastelen mennyttä tutkimusprosessia arviointikameran läpi, näen lukuisia henkilöitä ja tahoja, jotka ovat edesauttaneet tutkimukseni edistymistä sen eri vaiheissa. Jos löydät itsesi tästä mentaalisestä kuvasta, haluan osoittaa kiitokseni myös sinulle.

Jyväskylässä kesäkuussa 2002

Mirja Tarnanen

SISÄLLYS

ABSTRACT	5
ESIPUHE	6
1 ALUKSI	11
1.1 Arvioinnin lähtökohtia	11
1.2 Tutkimuksen teoreettinen viitekehys	19
1.2.1 Kahden menetelmän yhdistäminen	19
1.2.2 Miten tutkimus ymmärtää: kaksi näkökulmaa merkityksen rakentamiseen?	22
1.2.3 Miten tutkimus selittää: miksi asiat tapahtuvat tietyllä tavalla?	25
1.3 Tutkimuskysymykset	26
1.4 Kirjan rakenteesta: vihjeitä lukijalle	28
2 TUTKIMUKSEN KULKU	30
2.1 Arvioitavan aineiston kerääminen	30
2.1.1 Kirjoitustehtävät ja arviointiasteikko	30
2.1.2 Suoritusten keruu	36
2.1.3 Kirjoittajat	36
2.2 Kirjoitustehtävien arviointi ja aineiston keruu	40
2.2.1 Arvioijat	40
2.2.2 Aineiston keruu	41
2.2.2.1 Keruusuunnitelma ja sen toteuttaminen	41
2.2.2.2 Haastattelujen tekeminen	43
2.2.2.3 Mikä ohjaa haastatteluaineiston analyysia?	45
3 KATSAUS ARVIOINTIKÄSITTEISIIN JA -KÄYTÄNTEISIIN	48
3.1 Arviointia erilaisissa konteksteissa	48
3.2 Arviointikriteerien tarkastelua testauksen näkökulmasta	52
3.2.1 Hyvän arvioinnin peruskriteerit	52
3.2.2 Validius: arvioidaanko sitä mitä on tarkoitus arvioida?	54
3.3 Reliaabelius: miten johdonmukaista ja yhdenmukaista arviointi on?	58
3.3.1 Yleistä reliaabeliudesta	58
3.3.2 Miten arvioijiin liittyvää reliaabeliutta arvioidaan?	63
3.3.3 Yhdenmukaisuuden ja johdonmukaisuuden tarkastelu tässä tutkimuksessa	65

3.3.4	Mikä edesauttaa ja mikä taas estää reliaabeliuden saavuttamista?	67
3.3.4.1	Johdonmukaisuus ja yhdenmukaisuus edellyttävät virheetöntä arviointia	67
3.3.4.2	Arvioijasta johtuvia mittausvirheitä	68
3.3.4.3	Arvioijakoulutus arvioinnin yhtenäistäjänä	71
3.3.4.4	Muita arvioinnin yhdenmukaisuuteen vaikuttavia tekijöitä	75
3.4	Kolme kirjoittamisen arviointitapaa	78
3.4.1	Arviointitapojen luokitteluperusteita	78
3.4.2	Piirrearviointi	79
3.4.3	Analyyttinen arviointitapa	81
3.4.4	Holistinen arviointitapa	84
4	KIRJOITTAMISEN ARVIOINTIPROSESSIN TARKASTELUA	90
4.1	Kirjoittamistaidon jäljillä	90
4.1.1	Käsityksiä kirjoittamisesta taitona	90
4.1.2	Kirjoittamisen tutkimuksen neljä näkökulmaa	97
4.1.3	Kirjoittaminen äidinkielessä ja vieraassa kielessä	100
4.1.4	Katsaus suomi toisena kielenä -kirjoittamiseen	103
4.1.5	Kirjoittamiskäsitys tässä tutkimuksessa	107
4.1.6	Miten käsitys kirjoittamisesta ja arvioinnista kohtaavat?	109
4.2	Kirjoittamisen arviointiprosessi	112
4.2.1	Arvottaminen reaktionä	112
4.2.2	Arvion antaminen päätöksentekoprosessina	114
4.2.3	Arviointiprosessi arviointikirjallisuudessa	117
4.2.4	Arviointiprosessi tässä työssä	122
4.3	Tutkimuksen kertomaa: mitä arvioija arvostaa toisen/ vieraan kielen kirjoittamisessa?	125
4.3.1	Arvioinnissa käytettyjä kriteereitä ja lukustrategioita	125
4.3.2	Lyhyesti arviointiprosessin ja arvion suhteesta	131
5	TULOKSET: ARVIOIJAT VIESTIVÄT NUMEROIN JA SANOIN	133
5.1	Mitä tasoarviot kertovat johdonmukaisuudesta ja yhdenmukaisuudesta?	135
5.1.1	Arvioijien sisäinen johdonmukaisuus	135
5.1.2	Arvioijien välinen yhdenmukaisuus eri tehtäväsarjojen arviointinissa	144
5.1.3	Arviointilinjan tarkastelua tehtäväsarjoittain	147
5.2	Teemahaastattelujen kertomaa	154
5.2.1	Ensimmäinen haastattelukierros: omaan taito- käsitykseen nojaaminen	154

5.2.1.1	Yleisvaikutelmasta liikkeelle	154
5.2.1.2	Arvioijien luomat asteikot	155
5.2.1.3	Millaista oman asteikon soveltaminen oli - tuntumaa arviointiprosessiin	168
5.2.1.4	Millaiset asiat nousevat pintaan arviointiprosessissa?	180
5.2.1.5	Ensimmäisen haastattelun satoa	190
5.2.2	Toinen haastattelu: määritellyn asteikon soveltaminen	192
5.2.2.1	Yleisvaikutelmia arvioinnista	192
5.2.2.2	Asteikosta kuultua	195
5.2.2.3	Asteikon soveltaminen: Millaisia tulkintoja asteikon tasokuvaukset herättivät?	197
5.2.2.4	Arvioija keskiössä	221
5.2.2.5	Toisen haastattelun satoa	226
5.2.3	Arvioija arvottavana lukijana	227
5.3	Valintaverkkohaastattelu	231
5.3.1	Millaiseen käsitykseen ja menetelmään valintaverkkohaastattelu pohjautuu?	232
5.3.2	Haastattelun 1. vaihe: tekstejä erottelevat konstruktit	238
5.3.3	Haastattelun 2. vaihe: suoritusten järjestäminen "ykkösestä yhdeksään"	249
5.3.3.1	Tekstien järjestäminen yksittäisten konstruktien mukaan	249
5.3.3.2	Tekstien järjestäminen niiden yleistason mukaan	251
5.3.3.3	Konstruktien vaikutus yleistason	253
5.3.4	Valintaverkkohaastattelun anti	255
6	KATSAUS TUTKIMUSPROSESSIIN	259
6.1	Tulosten kertomaa	259
6.1.1	Arvioinnin johdonmukaisuus ja yhdenmukaisuus	259
6.1.2	Kirjoittamiskäsitysten ja arviointiprosessin tarkastelua	261
6.1.3	Selittävän ja ymmärtävän aineiston vuoropuhelua	265
6.2	Tutkimuksen rajoituksia	270
6.2.1	Tutkimusasetelma	270
6.2.2	Asteikkojen puutteita	270
6.2.3	Laadulliset analyysimenetelmät ja niiden taustaoletukset	272
6.2.4	Tutkimuksen luotettavuus ja eettisyys	272
6.2.5	Tutkijan oma ääni – liian lähelläkö?	275
6.3	Tutkimuksen tarjoamia jatkolinjauksia	276
6.4	Loppusanat	279

LÄHTEET	283
Liitteet 1–11	
1. Kirjoittamisen arviointiasteikko	303
2. Kirjoitustehtävät	304
3. Kirjoittajien taustatietolomake	307
4. Arvioijien taustatietolomake	308
5. Esimerkki haastatteluaineiston analyysistä	309
6. Arvioijien antamien tasoarvioiden keskiarvot mittauskertakohtaisesti	314
7. Arvioijakohtaisten Pearsonin korrelaatioiden parittaisvertailut	316
8. Arvioijien samanmielisyysprosenttien parittaisvertailu	318
9. Arvioija- ja mittauskertakohtaiset taitotasofrekvenssit	323
10. Valintaverkkohaastattelussa käytetyt 9 tekstiä	328
11. Valintaverkkohaastattelun kokoamislomakkeet	336

1 ALUKSI

1.1 Arvioinnin lähtökohtia

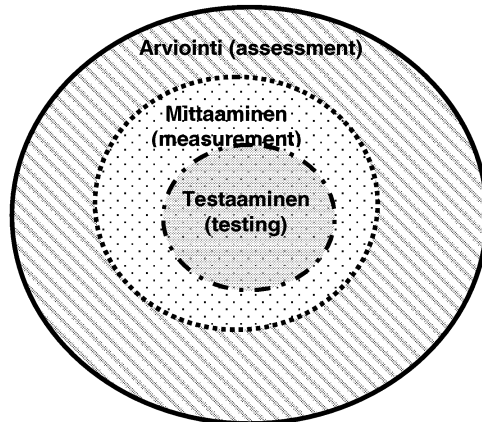
Oppimiseen liittyvällä arvioinnilla on monenlaisia tehtäviä: se muun muassa ohjaa oppimista, motivoi oppijaa, kontrolloi kehittymistä, kehittää oppimista ja ennustaa selviytymistä. Arvioinnin kohteena voi olla oppimisprosessin seuraaminen, tavoitteiden tai tuloksen saavuttaminen tai nykyisen osaamistason määrittäminen. Arvioijana voi toimia oppija itse, toinen oppija, opettaja tai ulkopuolinen arvioija. *Arviointi*-sanalla viitataan siis varsin monenlaisiin merkityksiin ja käyttöyhteyksiin. Arviointi liittyy muun muassa kokeisiin, testeihin, toisiin vertaamiseen, mittaamiseen, kriteerienvaraiseen luokitteluun, kritiikkiin, reflektointiin ja kannustavaan palautteeseen.

Arviointi on eräänlainen kattotermi kaikenlaiselle arvioinnille, se viittaa niin oppimisen, opetuksen, koulujen kuin koulutusjärjestelmienkin arviointiin. Yleensä ottaen arviointi sisältää monenlaisia tapoja kerätä systemaattisesti informaatiota osaamisesta ja/tai oppimisprosessista, tehdä havaintoja muutoksista ja arvioida muutosten määrää ja laatua (ks. esim. House 1995; Birenbaum 1996; Karjalainen 2001). Arvioinnissa on perimmältään kyse argumentoinnista, jonka eri muotoja käytetään eri tavalla arviointitilanteen ja kohderyhmän mukaan. Argumentoinnin hyödyllisyys ja käytettävyys edellyttävät puolestaan, että argumentointia voidaan pitää totuudellisena ja uskottavana. (House 1995, 23–25.)

Laadukas ja eettinen arviointi perustuu yhteisesti neuvoteltuihin arvostuksiin ja käsityksiin, joita voidaan kutsua myös yhteiseksi arviointiperustaksi (ks. Norris 1995, 37; Linnakylä 1998b, 176–177). Arviointitiedon keräämistä ohjaava arvioinnin viitekehys voi perustua sekä laadullisiin että mittaaviin menetelmiin, ja onkin tavallista, että arviointia ei enää ohjaa joko vain kokeellisiin tai vain laadullisiin menetelmiin perustuvat taustaoletukset (ks. House 1995). Toisaalta esimerkiksi kielitaidon arviointikirjallisuudessa menetelmien yhdistäminen ei ole itsestään selvää, vaan menetelmien yhdistämisestä ja siihen liittyvästä uskottavuudesta käydään yhä rajankäyntikeskustelua (vrt. Shohamy 2001a; Lynch 2001). Tämän keskustelun luonnetta havainnollistaa hyvin

kuvio 1, jossa arvioinnin sisällä erotetaan toisistaan (laadullisen) arvioinnin, testaamisen ja mittaamisen kehät.

Jos kuvion havainnollistamaa kielitaidon arviointia tarkastellaan laajemmasta näkökulmasta, siihen voisi lisätä vielä uloimmaksi kehäksi evaluaation (evaluation) eli koulutuksen arvioinnin kehän. Evaluaatiokehän paikka kuviossa riippuu tosin siitä, käsitetäänkö arviointi (assessment) tässä yleiseksi kattotermiksi, jonka alle voidaan sijoittaa yleensäkin arviointi tieteenalana sekä esimerkiksi meta-arviointi. Joka tapauksessa evaluaation puuttuminen kuvioista kaventaa kuvion välittämää kuvaa kielitaidon arvioinnin yhteiskunnallisista ja koulutuspoliittisista käyttöyhteyksistä ja -tarpeista.



KUVIO 1 Arviointi, mittaaminen ja testaaminen (Lynch 2001, 359).

Kuvion 1 keskeisimmässä ja sisimmässä ympyrässä esitetyt mittaaminen (measurement) ja testaaminen (testing) eroavat laadullisesta arvioinnista siinä, että arvioinnissa määritellään eri perustein se, mikä on arviointipäätösten ja itsearvioiden validius eli se, mitä on tarkoitus arvioida (Shepard 1993; Moss 1994; Lynch 2001).¹ Toisin sanoen *laadullisessa arvioinnissa* hyvän laadun edellytyksenä ei ole välttämättä psykometristen ehtojen täyttäminen, kuten *testaamisessa* ja *mittaamisessa* on. Laadullisessa

¹ Kuvion 1 ulommaisessa ympyrässä olevaa arviointia voidaan kutsua myös vaihtoehtoiseksi arviointiksi (alternative assessment), joka korostaa "arvioinnin" tarjoamaa vaihtoehtoisuutta mittaamisen ja testaamisen paradigmoille. Siitä, miten erilaisia arvioinnin, mittaamisen ja testaamisen käsitykset validiudesta ja reliabeliudesta ovat, ollaan eri mieltä. Joidenkin mielestä vaihtoehtoisessa arvioinnissa validiudelta vaaditaan eri asioita ja toisten mielestä taas ei (ks. Birenbaum 1996; Lynch 2001).

arvioinnissa huomionkohteena on oppija ja hänen oppimisprosessinsa, minkä vuoksi esimerkiksi portfolioarviointia ja itsearviointia käytettäessä arviointivastuuta on myös oppijalla itsellään. Näin tehdyn arvioinnin oletetaan tarjoavan työkaluja osaamisen ja sen kehittymisen reflektointiin.

Tämä ei kuitenkaan tarkoita sitä, että mittaaminen ja testaaminen eivät voisi pyrkiä samaan. Kyse on siis lähinnä siitä, miten ei-mitattavaa, kvalitatiivista informaatiota käytetään arvioinnissa hyväksi ja miten uskottavana sitä pidetään. Kun käytän myöhemmin joissakin yhteyksissä sanoja *testata* ja *testaaminen*, viittaan niillä ainoastaan testien avulla tehtävään mitattavaan arviointiin. *Arvioida-* ja *arviointi-*sanoilla viittaan sekä arviointiin yleensä että oppimiseen liittyvään arviointiin, jossa tarkastellaan osaamista jonkin sitä osoittavan näytön perusteella.

Suomen kielessä käytetään kuitenkin *arviointi*-sanan rinnalla joissakin yhteyksissä *arvostelua*. Mikä on sitten arviointia ja mikä arvostelua? Suomen kielen perussanakirja (1990) määrittelee arvon määrittämiseen liittyvän sanaperheen verbit seuraavasti:

- arvioida:* 1. määrittää, laskea, päätellä summittaisesti, tehdä arvio,
2. arvostella
- arvostella:* 1. määrittää jonkin arvo hyvien ja huonojen puolien perusteella, arvioida, 2. usein nimenomaan moittivasta suhtautumisesta, kritisoida, 3. arvioida julkisesti taiteen tai tieteen tuloksia varsinkin jossakin sanoma- tai aikakauslehdessä
- arvottaa:* sijoittaa arvoasteikkoon, antaa tietynsuuruinen arvo, arvioida
- arvostaa:* 1. pitää jotakin arvossa, antaa jollekin arvoa, kunnioittaa, 2. paremmin: määrittää jonkin arvo, arvioida, arvottaa

Kuten edelläkin kävi jo ilmi, perussanakirjan määritelmiin nähden viitataan tässä tutkimuksessa arvioida-verbillä varsin monenlaisiin merkityksiin. Toisaalta suomen kielessä arviointiin liittyvien lähikäsitteiden erottelu ei ole yhtä hienojakoista kuin englannin kielessä, joten *arvioida* ja *arviointi* näyttävät vakiintuneen yleistermeiksi. Perussanakirjan mukaan *arvioida* on siinä mielessä sävyllään neutraalein, että se ei toisaalta sisällä *arvostella*-sanan negatiivista latausta eikä *arvottaa*-sanan hieman yksipuolista merkitystä eikä myöskään *arvostaa*-sanan tunnelatautunutta positiivisuutta.

Erilaiset arviointiin liittyvät menettelytavat viestivät siis konkreettisesti erilaisista tavoista ja käytänteistä tehdä arviointia ja paradigma- tasolla erilaisista tietoon, osaamiseen ja olemassaoloon liittyvistä lähtö- oletuksista. Sanojen määrittelyssähän ei ole kyse mieltymyksistä sanoihin,

vaan ennen kaikkea erilaisista toiminnoista, joihin sanoilla viitataan. Oli pa menettelytapa mikä tahansa, arvioinnin odotetaan kuitenkin osoittavan osaamisen todellista laatua ja sisältöä ja sen odotetaan olevan myös teknisesti luotettavaa ja uskottavaa, oikeudenmukaista lähinnä tasavertaisuusmielessä ja eettistä (esim. Messick 1994; Linnakylä 1998b).

Vaikka arvioinnista puhuttaessa sen koulutuksellinen funktio korostuu, sillä on kuitenkin yhtä lailla taloudellisia, sosiaalisia ja poliittisia tehtäviä ja seurauksia (esim. Eggleston 1990; Filer 1995; Norris 1995; Shohamy 2001a). Se, miten yhteisö valitsee, luokittelee, jakaa ja arvioi koulutuksellista tietoa, kertoo myös vallan jakautumisesta ja sosiaalisesta kontrollista (Eggleston 1990, 59). Arviointi on keino edistää ja kontrolloida yksilön pätevyyttä, se sisältää kilpailua, kontrolloi yksilön valintoja sekä mahdollistaa järjestelmän toiminnan ja tuloksellisuuden seurannan (ks. Norris 1995; Linnakylä 1998b; Virta 1999). Sosiaaliset ja poliittiset vaikutukset eivät ole kenties yhtä läpinäkyviä kuin koulutukselliset, mutta niiden vaikutus voi olla tehokkaampaa ja kauaskantoisempaa sekä hyvässä että pahassa.

Arviointi muokkaa myös tietokäsitystämme: se osoittaa, millaista tietoa pidetään hyväksyttävänä. Tietynlainen arviointitapa ja -tehtävät sisältävät oletuksen siitä, ovatko tieto ja taito paloja vai kokonaisuuksia, ovatko ne luonteeltaan autonomisia vai kokonaisvaltaisesti osistaan riippuvaisia ja ymmärretäänkö tieto sosiaalisena konstruktiona, joka muotoutuu ihmisten välisessä vuorovaikutuksessa, vai kenties atomistisena ja yksilöllisenä eritasoisten suoritusten hierarkiana (ks. Kvale 1995). Oppimiskäsitysten muuttumisen myötä arviointikin on muuttunut yhä enenevässä määrin monimuotoiseksi, koska pelkästään käytettynä perinteiset kokeet ja standardoidut testit eivät ota riittävästi huomioon oppilaan omien tietorakenteiden kehittymistä ja aktiivisuutta oppimisprosessissa.² (Esim. Virta 1999; Hamp-Lyons 2000; Shepard 2000; Segers & Dochy 2001.)

Oppimisen ja osaamisen arvioinnin ongelmaksi, varsinkin kielitaidon testauksessa, onkin nähty se, että testien perimmäinen tavoite on "mitata" eli määritellä arvosanat ja asettaa oppijat paremmuusjärjestykseen. Ensisijaisena tavoitteena ei ole siis tarjota palautetta ja sen myötä mahdollisuuksia oppia entistä tehokkaammin. Testimäistä arviointia on kritisoitu myös siitä, että se saattaa ohjata yhä hanakammin tarkkailemaan reliaaбелиutta ja varmistamaan testin validiutta oppimisen

² Viitataan tässä laadullisella arvioinnilla arviointitapoihin, jotka nousevat kognitiivis-konstruktivistisesta oppimisenäkemyksestä. Laadullinen arviointi, esim. portfolioarviointi, ottaa huomioon oppijan yksilöllisyyden, kun taas perinteisessä arvioinnissa lähtökohtana ovat tietojen ja taitojen mitattavuus, kvantifioitavuus ja tilastollinen käsiteltävyys (ks. Linnakylä 1993; 1998b; Virta 1999).

ja opettamisen kustannuksella (Bachman 2000, 12; Lynch 2001, 317). Toisaalta testejä ja testijärjestelmiäkin on monenlaisia, ja ne voivat perustua erilaisiin käsityksiin tiedosta ja taidosta, mitä käsittelemme perusteellisemmin tämän työn kolmannessa ja osin neljännessäkin luvussa.

Ehkä testien, varsinkin testijärjestelmien, suurin ongelma ja toisaalta myös mahdollisuus liittyy niiden sosiaalisesti ja poliittisesti vahvaan asemaan. Tämä on havaittavissa muun muassa Opetushallituksen pääjohtajan Jukka Sarjalan kirjoituksessa (HS 12.8.2000), jossa hän ottaa vastustavan kannan peruskoulun päättökoetta koskevaan ehdotukseen: ”Peruskoulun valtakunnallinen päättökoe houkuttelisi kiinnittämään huomiota koulun tietotehtävään, joka sinänsä on tärkeä mutta vain yksi monista osa-alueista. Arvo- ja muu kasvatusta, valmentamista yhteisöllisyyteen sekä persoonallisuuden tuki ovat tietotehtävään rinnasteisia tehtäviä, joiden valtakunnallinen arviointi on vaikeaa. Arviointi ei saa korostaa sellaista osaamista, jota mitataan kokeissa.”

Ei liene epätavallista, että tällainen päättökoemainen tietopohjaisen osaamisen arviointi saattaa tuoda mukanaan oikeusturvaan, opetuksen yksipuolistumiseen ja diskriminointiin liittyviä ongelmia, joiden myötä kyse ei ole enää yksittäisen taidon ja tiedon oppimisesta vaan pahimmillaan yksilön mahdollisuuksista jatko-opiskella tai saada työpaikka. Seuraukset ovat toki myös riippuvaisia testin luonteesta ja laadusta. Yhtä hyvin voisikin kysyä, jos ja kun päättökoe tulisi, miksi sen pitäisi olla testi kielteisessä merkityksessä. Miksi päättöarviointi ei voisi olla portfolioarvioinnin/itsearviointin, suorituspohjaisen testin ja oman opettajan arvioinnin yhdistelmä (vrt. Shohamy 2001b)?

Jos testejä tarkastellaan oppijan näkökulmasta, on melko tavallista, että arkikokemuksissa standardoituihin kokeisiin tai arviointiin yleisimminkin liittyy mielikuvia huonommuuden tunteesta, osaamattomuudesta, kilpailusta ja alistamisesta (ks. myös Shohamy 1993, 2001a; Takala 1994; Hamp-Lyons 1997a; Virta 1999). Tosin päinvastaisiakin arkikokemuksia kuulee esimerkiksi siitä, miten latistava arviointi on itse asiassa ajanut näyttämään, että minäkin pystyn tai miten kannustavan arvioinnin muisteleminen on pitänyt yllä uskoa omiin kykyihin. Näissä arkikokemuksissa on kuitenkin mielenkiintoista se, että negatiivisia kokemuksia arvioinnista ei ole vain stereotyyppisesti huonosti menestyvillä oppijoilla vaan niitä voi olla myös kiltteinä ja moitteettomina pidetyillä oppijoilla. Tämän tyyppisiä kokemuksia heijastelevat seuraavat suomen kielen opiskelijoiden kokemukset:³

³ Olen kerännyt esimerkit kirjoittamisen kurssilta, jonka opiskelijat pohdiskelivat kirjallisesti käsitystään itsestään kirjoittajana.

”Huonoista numeroista tulee kirjoittamisen pelko . . . pitäisin kirjoittamisesta varmasti paljon enemmän, jos en olisi saanut niin paljon negatiivista palautetta aineistani lukiossa.”

”Jossain vaiheessa kirjoittaminen muuttui iloa tuottavasta vaivattomasti syntyvästä ajanvietteestä traumaattiseksi itseni ja kykyjeni epäilyksi. Tämä lienee tapahtunut lukiossa, jossa kirjoittamisestakin tuli suorittamista. Piti saada aineista hyviä numeroita, olla käyttämättä kuluneita kielikuvia jne.”

Myös tutkimusten mukaan arvioinnin yksilölliset vaikutukset näyttäisivät muokkaavan oppijan käsitystä omista taidoistaan (ks. esim. Laine & Pihko 1991; James & Garrett 1992).⁴ Yleensäkin oppijan itsetuntoon ja minäkäsitykseen vaikuttavat kaikki sanalliset ja sanattomat palautteet, joita hän saa oppimisestaan ja osaamisestaan (Korpinen 1993, 1996; Leimu1993; Virta 1999). On johdonmukaista, että kriittinen mutta kuitenkin kannustava ja hyväksyvä palaute ja arviointi vaikuttavat myös siihen, mitä asioita oppilas pitää tärkeinä ja itselleen merkityksellisinä.

Arvioinnin tehtävä on ymmärretty väärin tai sitä käytetään epätarkoituksenmukaisesti, jos se kategorisoi ja kapea-alaistaa opetusta ja käsitystä osaamisesta; arviointi voi parhaimmillaan innostaa ja motivoida sekä tarjota työkaluja edelleen kehittymiseen (ks. myös Välijärvi 1996). Arvioinninhan ei edes pitäisi olla pelkästään yksisuuntaista ’totuuden osoittamista’ vaan enenevässä määrin vuorovaikutuksellista ja reflektioivaa oman oppimisen pohdintaa. Myös testimäinen arviointi on muuttunut ja muuttumassa funktionaaliseksi, suorituspohjaiseksi, autenttisuutta tavoittelevaksi, ennalta sovittuihin kriteereihin perustuvaksi ja osaamista kannustavaksi arviointitavaksi. Mielenkiintoista on arviointikäytänteiden muuttumisen suhde käsitysten, arvostusten ja uskomusten muuttumiseen. Saman asian voisi myös muotoilla kysymykseksi: millaisin ehdoin muutoksia voidaan pitää *todellisina* muutoksina?

Arvioinnin historiassa on Guban ja Lincolnin (1989, 22–23) mukaan eroteltavissa neljä sukupolvea, joista ensimmäinen *mittaamisen* (measurement) sukupolvi käytti nimensä mukaisesti mittaamista arvioinnin synonyymina ja käsitti arvioinnin tehtäväksi mitata sen tiedon hallintaa, jonka yleisesti tiedettiin olevan totta. Ehkä tämä käsitys on juurtunut aika syvälle käsityksiin testiarvioinnin luonteesta ja myös itse arviointikäytänteisiin, koska se on edelleenkin havaittavissa arviointiin kohdistuvassa kritiikissä. Toista arviointisukupolvea Guba ja Lincoln (1989, 27–28) luonnehtivat *kuvailemisen* (description) sukupolveksi. Se korosti arvioinnin formatiivista luonnetta. Toisen sukupolven aikana

⁴ Tarkoitan tässä minäkäsityksellä sitä, miten oppija selittää ja tulkitsee itseensä liittyvät sosiaalisen ympäristön reaktiot. Minäkäsitykseen liittyvät todellinen kuva itsestä, ideaalikuva ja itsearvostus, joka säätelee muun muassa sitä, miten yksilö toimii ja miten hänen minäkuvansa muuttuu (esim. Laine & Pihko 1991; Keltikangas-Järvinen 1994; Atkinson ym. 2000).

arviointiin kehitettiin ja siinä käytettiin useita erilaisia arviointitapoja, joiden tavoitteena oli saada selville osaamisen tai lähinnä sen hallinnan vahvuudet ja heikkoudet.

Kolmatta sukupolvea Guba ja Lincoln (1989, 28) kutsuvat *arvostelu-*sukupolveksi (judgement). Sen mukaan arviointi tarvitsee ulkopuoliset standardit, joihin osaamista verrataan. Vaikka kolmannen sukupolven pyrkimyksenä oli tarkastella osaamista testiyksikköä laajemmasta näkökulmasta, mukana seurasivat edelleen mittaamisen ja kuvaamisen sukupolvet. Guban ja Lincolnin (1989, 34) mukaan kolmelle ensimmäiselle sukupolvelle on yhteistä se, että ne eivät ole moraalisesti vastuussa arvioinnin seurauksista, ja se, että ne eivät ole arvosidonnaisia – tai eivät ehkä tietoisesti harkitusti.

Guba ja Lincoln (1989) erottavatkin toisistaan neljättä sukupolvea luonnehtivan *konstruktivistisen* arvioinnin ja kolmea ensimmäistä sukupolvea luonnehtivan konventionaalisen eli positivistiseen näkemykseen perustuvan arvioinnin. Konstruktivistisen arvioinnin perustana on relativistinen käsitys tieteen ja todellisuuden suhteesta. Sen mukaan arvioinnissa tuotetaan tulkinta, jossa tosiasiat ja arvot liittyvät erottamattomasti yhteen. Itse asiassa asioiden arvottaminen on välttämätön osa arviointiprosessia. Arvioinnissa ei myöskään tavoitella mitään yhtä ainoaa totuutta, vaan totuuksia ja todellisuuksia on useita (ks. myös Scriven 1997; Shepard 2000; Lynch 2001). Tässä mielessä on syytä myöntää, että arviointi on arvosidonnaista ja aina jossakin määrin subjektiivista.

Tämä ei kuitenkaan tarkoita sitä, että arvioinnissa ei päästäisi yhteisymmärrykseen tai että toiminnan taustalla ei voisi olla yhteisiä käsityksiä arvioitavasta ilmiöstä. Itse asiassa lähtökohtana voidaan pitää sitä, että erilaiset arvostukset neuvotellaan yhteiseksi arviointiperustaksi, joka luo tietoisesti harkitun ja uskottavan pohjan arvioinnille (Norris 1995; Linnakylä 1998b). Arvioinnin voi nähdä pohjimmiltaan merkitysmaailman rakentamisena siinä mielessä, että arviot ovat tulkintoja arvioitavasta asiasta, oli kyse sitten oppilaan tai koulutuksen arvioinnista (evaluaatio). Tulkinta edustaa sekä kunkin yksilön omia että häntä ympäröivien sosiaalisten yhteisöjen arvoja ja käsityksiä (ks. Guba & Lincoln 1989; Messick 1989; McNamara 2001).

Niinpä kielitaidon arvioinnissakin myös opettajan omat käsitykset ohjaavat sitä, mitä arvioidaan ja millaisin arvostuksin. Henkilökohtaiset käsitykset ovat jossakin määrin läsnä myös silloin, kun arviointi perustuu opettajien yhteisesti käyttämään arviointiasteikkoon. Arviointiohjeet yhdenlaisina merkitysrakenteina eivät siis ohjaa ihmistä, vaan maailmaa jäsentäessään ja siinä toimissaan ihmiset käyttävät ja soveltavat merkitysjärjestelmiä, kulttuurisia jäsennyksiä, malleja, skeemoja tai tulkintarepertuaareja (vrt. Jokinen ym. 1993; Alasuutari 1999). Merkityk-

sen rakentumisen kautta arviointikin voitaisiin määritellä yhdenlaiseksi diskurssiksi muiden diskurssien joukossa.

Tämä tutkimus on osa sellaista arviointidiskurssia, jossa suomi toisena kielenä -opettajat arvioivat aikuisten kulttuuritaustoiltaan erilaisen suomen kielen oppijoiden tekstejä. Kyse on arvioijan toiminnan ja kirjoittamisen arvioinnin tutkimisesta. Tutkimuksen konkreettisena ja alkuperäisenä tavoitteena oli luoda itsearvioinnissa käytettävä kirjoittamisen maamerkistö DIALANG-nimiseen projektiin.⁵ Sittemmin aineiston laajahko keruusuunnitelma tarjosi mahdollisuuden selvittää myös sitä, millaisia merkityksiä arvioijien antamat arviot kantavat mukanaan. Tutkimuksen tavoitteena on siis 1) tutkia arvioijien välistä yhdenmukaisuutta ja sisäistä johdonmukaisuutta sekä arvioijien käsityksiä kirjoittamisesta ja arviointiprosessista ja 2) tarjota maamerkkiaineisto DIALANG-projektin käyttöön.

Mistä kirjoittamisen arviointiprosessissa on oikeastaan kyse? Sanaliitto ”kirjoittamisen arviointiprosessi” syntyy kolmesta sanasta: *kirjoittaminen*, *arviointi* ja *prosessi*. Käsitkset kirjoittamisesta ilmiönä ja taitona ovat tavallisesti joko tiedostettu tai tiedostamaton lähtökohta arvioinnille. Arviointi voi pohjautua tietoisesti harkittuun ja yhteisesti neuvoteltuun kirjoittamiskäsitykseen, tai sitten arviointi ikään kuin ohimennen viestii tietynlaisesta kirjoittamiskäsityksestä sen mukaan, mihin asioihin kiinnitetään huomiota. Sanaliiton viimeinen sana prosessi taas piirtää vuorovaikutuksellisen kaaren kirjoittamisen ja arvioinnin välille. Prosessi on jotakin, joka on dynaamista ja jonka kulku ja luonne eivät ole täysin ennustettavaa sen inhimillisen luonteen ja siinä vaikuttavien monien muuttujien vuoksi.

Kielitaidon arviointia koskevassa kirjallisuudessa on korostettu sitä, että meidän pitäisi tuntea paremmin arvioijan käyttäytymistä, päätöksentekoprosessia, joka johtaa arvion antamiseen (esim. Huot 1990; Bachman 2000; Smith 2000). Pelkkien arvioiden analysointi ja tulkinta numeerisesti ei riitä, koska arvioinnissa on kyse monimuotoisesta vuorovaikutuksesta oppijan, tehtävien ja arvioijan välillä. Filerin (1995, 32–33) mukaan opettajien luokkahuonestrategioiden tunteminen ja ymmärtäminen on ehto sille, että arvioinnin ’objektiivisuutta’ voitaisiin ymmärtää. Osana noita luokkahuonestrategioina tai oikeastaan niiden lähtökohtana voidaan pitää arvioijien käsityksiä arvioitavasta taidosta ja sen oppimisesta.

Arvioijan toiminnan tutkimisen lisäksi myös kirjoittamisen arvioin-

⁵ DIALANG-projekti on yhteiseurooppalainen 14 kielen hanke, jossa kehitetään diagnostista, verkkopohjaista arviointisysteemiä. Maamerkistön kehittäminen on osa ns. suoran kirjoittamisen arviointia, jossa oppija itsearvioi kirjoittamisen taitoaan siten, että hän vertaa omaa suoritustaan apukysymysten avulla esimerkki- eli maamerkkikirjoituksiin (ks. Tarnanen & Luoma 2001; Luoma & Tarnanen tulossa).

nin tutkimus tuntuu relevantilta käytännön opetuksen ja koulutuksen näkökulmasta. Cummingin (1990, 32) mukaan tiedämme liian vähän kirjoittamisen arvioinnista eli siitä, mitä kirjoittamistaidon arviointi vaatii ja miten kirjoittaminen taitona kehittyy (ks. myös Santos ym. 2000). Tiedon tarve on suuri varsinkin suomi toisena kielenä (S2) -opetuksessa ja arvioinnissa, koska S2-opetuksen tarve ja määrä ovat lisääntyneet eri koulutusasteilla ja työelämässä. Kyse ei ole vain aikuisista kielenoppijoista, vaan myös lapsista, mikä tuo omat haasteensa kielenopetukseen ja arviointiin. Kirjoittamisen osalta tulevaisuuden haasteita on muun muassa se, millaista kirjoittamisen taidon tasoa selviytyminen suomalaisessa yhteiskunnassa edellyttää, miten oppijan kirjoittamaan suomeen suhtaudutaan ja miten sitä ymmärretään erilaisissa kielenkäyttötilanteissa.

1.2 Tutkimuksen teoreettinen viitekehys

1.2.1 Kahden menetelmän yhdistäminen

Tätä tutkimusta voi pitää sekä kvantitatiivisena että kvalitatiivisena kuten arviointiakin. Viitataan tällä kahtiajaolla ontologiseen olettamukseen, jonka mukaan erityisesti kvantitatiivisiin metodeihin nojaavat paradigmat, kuten positivismin eri suuntauksat, lähtevät siitä, että inhimilliset ilmiöt eivät poikkea merkittävästi luonnonilmiöistä. Klassisen tai ehkä oikeammin loogisen positivismin mukaan on olemassa yksi objektiivinen todellisuus, josta tutkimus voi antaa yhdenmukaista tietoa.⁶ Maailma nähdään matemaattisesti järjestäytyneenä, joten tutkimuksen tekeminen edellyttää objektiivista ja määrällistä dataa. Siinä missä loogisen positivismin lähtökohtana on tieteellinen realismi, kriittisemmän postpositivismin mukaan maailma on havaittavissa vain osittain. (Guba & Lincoln 1994, 109; Raunio 1999, 221–224.)

Kvalitatiivisista metodologioista voidaan käyttää yhteisnimitystä konstruktivismi tai kriittinen paradigma, joiden mukaan todellisuuskäsitys on relativistinen (Guba & Lincoln 1994, 109–111). Niiden mukaan ei ole olemassa yhtä ainoata todellisuutta vaan sosiaalisia todellisuuksia on useita, minkä vuoksi tieteen tekemisessä tunnustetaan tutkijan subjektiivisuus ja arvosidonnaisuus (esim. Patton 1990; Kvale 1996; Tashakkori & Teddlie 1998). Tutkimustulokset eivät sinällään ole tutkijan löydöksiä ympäröivästä todellisuudesta, vaan ne syntyvät tutkimusprosessin tulok-

⁶ Viitataan tässä klassisella positivismilla vanhentuneeseen käsitykseen luonnontieteistä, mitä käytetään tavallisesti kritiikkinä positivismia vastaan, kun halutaan argumentoida subjektiivisten metodologioiden puolesta (vrt. Raunio 1999; Töttö 1999).

senä. Metodologialle on tyypillistä dialektinen vuorovaikutus ja prosessi-luonteisuus, mikä näkyy siinä, että tulokset ja tulkintojen merkittävyys tulevat ilmi erilaisten näkemysten ja niiden keskinäisen vertailun seurauksena. (Piirainen-Marsh & Huhta 2000, 80–81.)

Kvantitatiivisten ja kvalitatiivisten tutkimusmetodien epistemologisen ja metodologisen luonteen kahtiajakoa voitaisiin kuvata edellisen lisäksi muun muassa seuraavilla vastakohtaisilla adjektiivipareilla: objektiivinen–subjektiivinen, yleistävä–yksilöllinen, selittävä–ymmärtävä, ennustava–kuvaava, kokeellinen–hermeneuttinen, manipuloiva–dialektinen, kontrolloiva–suvaitseva. Sanaparit viestivät varsin voimakkaasta vastakkaisuudesta, joka voi olla seurausta siitä, että kvalitatiivisessa metodikirjallisuudessa lähtökohtana on argumentointi positivismia vastaan ja positivistisessa oman paradigman puolustaminen.

Käytännön tutkimuksen kannalta menetelmien näin jyrkkä kahtiajako ei kuitenkaan ole tarkoituksenmukaista, koska menetelmiä ei tarvitse nähdä toisiaan poissulkevinä. Kyse on siis ennemminkin jatkumosta menetelmien välillä kuin puhtaasta dikotomiasta (Kvale 1996, 10–11; Tashakkori & Teddlie 1998, 25–26). Keskustelua herättää myös, missä määrin ontologiset kysymykset ovat tutkimusta ohjaavia. Metodologinen fundamentalisti korostaa ontologisten kysymysten perustavaa merkitystä, kun taas metodologinen eklektikko pitää ensisijaisena metodisia ratkaisuja, jotka ovat tutkimustehtävän, tutkimuksen tavoitteiden ja käytettävissä olevien resurssien kannalta tarkoituksenmukaisia (Patton 1990, 39; Suoranta 1993, 40–43; Tashakkori & Teddlie 1998, 24).

Kielitaidon arviointikin on sekä kvalitatiivista että kvantitatiivista, ja on luultavasti oikein sanoa, että se ei ole koskaan vain jompaakumpaa (Pollitt & Murray 1996, 74). Niinpä arvioinnin tutkimuskin voi olla ilmiön ymmärtämisen kannalta tarkoituksenmukaisempaa, kun mukana ovat molemmat näkökulmat. Vain kvalitatiivisen tai kvantitatiivisen lähestymistavan käyttäminen ei ehkä tarjoaisi yhtä kattavaa ja monipuolista kuvaa arvioinnista, minkä vuoksi on perusteltua olettaa, että monimenetelmällisyys eli triangulointi voi heijastaa paremmin ja suoremmin tutkittavan ilmiön monimuotoisuutta (ks. esim. Miles & Huberman 1994; Kvale 1996; Chelimsky 1997; Eskola & Suoranta 1998).

Metodikirjallisuuden mukaan triangulointi ei ole kuitenkaan näin ongelmatonta. Astetta kriittisemmän näkökulman mukaan triangulaation käyttö nähdään kyllä hyödyllisenä mutta tutkimusmenetelmien yhdistämisestä aiheutuva teoreettinen ristiriita epäilyttää (esim. Brannen 1992; House 1995). Tätä vielä kriittisemmän näkökulman mukaan triangulaation käyttöä pidetään mahdottomana, koska on epäsoveliasta yhdistää tieteenfilosofisesti erilaisia ihmiskäsityksiä (vrt. Eskola & Suoranta 1998). Menetelmien yhdistäminen pakottaa siis hyväksymään toisilleen risti-

riittäisiä teoreettisia lähtökohtia ja tekee tutkimuksesta näin ollen moniselitteistä.

Monien menetelmien (mixed method studies) käyttö on kuitenkin perusteltua pragmaattisesta näkökulmasta, koska tutkimukselle ei voida antaa valmiita malleja, vaan käytäntö ja kokemus kertovat, mikä on milloinkin sopivinta (ks. Tashakkori & Teddlie 1998). Näin ollen pragmaatikolle myös ontologinen näkemys todellisuudesta on kaksijakoinen: toisaalta on olemassa meistä riippumaton ulkoinen todellisuus, mutta lopullisen totuuden löytäminen ei ehkä ole kuitenkaan mahdollista ainakaan silloin, kun tutkimuksen kohteena on ihmisen toiminta emotionaalisenä, rationaalisenä, herkisenä ja sosiaalisena yksilönä.

Käytännön näkökulmasta menetelmien yhdistäminen ei tunnu aina vaativan tiukkaa sitoutumista, esimerkiksi sellaisissa tapauksissa, joissa kvalitatiivista aineistoa kuvataan myös havaintoarvoilla tai korrelaatioita käytetään hyväksi ilman, että tutkija on paneutunut positivismiin tai konkreettisemmin tilastollisen tutkimuksen periaatteisiin. Lisäksi on syytä ottaa huomioon myös se, että ei ole olemassa yhtä ainoata kvantitatiivista tai kvalitatiivista menetelmää, vaan niitä molempia on lukuisia. Toki edellä on kuvattu päälinjauksia yleisesti, mutta esimerkiksi kvalitatiivinen menetelmä voi merkitä hyvin eri asioita ilmaisun käyttäjän mukaan. Vaikka tutkimussuunta, esimerkiksi diskurssianalyysi tai etnografia, olisikin ilmaistu tarkemmin, käsitteet ja menetelmälliset ratkaisut edellyttävät tutkimuksen tekijältä ja tutkimuksen lukijalta neuvottelua monenlaisista merkityksistä.

Lisäksi tilastollisesta tutkimuksesta tai tilastollisesta päättelystä puhutaan usein liian yksioikoisesti kvantitatiivisena tutkimuksena (vrt. Mäkelä 1995). Ero tilastollisen ja kvantitatiivisen tutkimuksen välillä näkyy lähinnä siinä, mikä merkitys teoriassa on operationalisoinnissa; tilastollinen tutkimus pelkistää teorian ja hypoteesit populaatioksi ja sitä koskeviksi väitteiksi, kun taas kvantitatiivinen tutkimus sallii käsitteiden johtamisen teoriasta (Bryman 1988, 20; Mäkelä 1995, 65). Mäkelän (1995, 62) mukaan tilastollinen päättely on johtanut käsitteelliseen empirismiin eli siihen, että tutkimuksen kohteena ovat käsitteet teorioiden sijasta. Näin ollen olisi syytä pohtia, missä määrin arviointitutkimuksen tai oikeammin testiteorian perusta lepää kvantitatiivisessa tutkimuksessa ja missä määrin tilastollisessa päättelyssä. Vaikka pohdinta on tässä yhteydessä retorista, haluaisin kuitenkin liittää tämän varauksen tekstini niihin kohtiin, joissa viitataan kvantitatiivisella tutkimuksella testiteoriaan.

Tutkimukseni menetelmällisten ratkaisujen vuoksi tunnustaudun tässä mielessä paradigmaattiseksi pragmaatikoksi tai jonkinasteiseksi metodologiseksi eklektikoksi. Ensinnäkin tutkimukseni tavoitteena on sekä selittää numeerisen aineiston avulla että ymmärtää haastattelujen

kautta arvioijan toimintaa, minkä vuoksi menetelmien yhdistäminen on perusteltua. En katso, ettei numeerisen aineiston ja hermeneuttisen tulkinnan välillä olisi mahdollista käydä ymmärrystä lisäävää dialogia (ks. esim. Guba 1990; Miles & Huberman 1994; Eskola & Suoranta 1998). Toiseksi kvantitatiivinen ja kvalitatiivinen menetelmä ovat pääasiassa työkaluja ja niiden hyödyllisyys punnitaan siinä, miten hyvin ne onnistuvat vastaamaan tutkimuskysymyksiin (Kvale 1996, 69). Kolmanneksi Niskasta lainatakseni (1995, 267) tutkimukseni metodit voivat kyllä vaihdella, mutta niiden päämäärä on sama.

1.2.2 Miten tutkimus ymmärtää: kaksi näkökulmaa merkityksen rakentamiseen?

Tarkastelen tutkimuksen laadullista aineistoa kahdesta näkökulmasta: etnografisesta ja fenomenografisesta. Kumpikin tarjoaa jotakin sellaista, mitä toinen ei tee. Lähdän siitä, että aineistoni lähtökohtana on arvioija yksilönä, joka toimii oman arvioija-opettajayhteisönsä jäsenenä. Tämä yhteisö on taas osa laajempaa yhteisöä. Tässä kerrostuneisuudessa etnografisuus viittaa arvioijayhteisöön, sen rakenteisiin ja toimintatapoihin yksilön kokemusten kuvaamina. Fenomenografisuuden tehtävänä taas on kuvata yksilön käsitysmaailmaa, sitä, miten yksilö ymmärtää ja käsitteellistää toimintaansa arvioijana.

Se, että menetelmiä yhdistetään näin kevyesti, voi herättää ristiriitaisiakin mielipiteitä. Toisaalta taas kummankin menetelmän taustaoletuksissa on samankaltaisuutta. Samanlaisuus näyttäisi liittyvän yleensäkin kvalitatiivisen tutkimuksen taustalla olevaan ihmiskäsitykseen, jonka mukaan ihminen on autonominen subjekti, joka pyrkii rakentamaan itselleen kuvan maailmasta (Holstein & Gubrium 1994, 263; Kvale 1996, 53). Yhteistä on myös kiinnostus yksilön kokemukseen ja vuorovaikutukseen merkityksen rakentajana (vrt. Lincoln & Denzin 1994).

Sanana *etnografia* tarkoittaa ihmisistä (*ethnos*) kirjoittamista, ihmisten kuvaamista (*graphein*). Etnografia on alun perin ollut antropologista kenttätutkimusta, mutta on sittemmin yleistynyt käsittämään varsin erilaisia lähestymistapoja (ks. Syrjälä & Numminen 1988; Vidich & Lyman 1994; Savage 2000). Vaikka etnografialle ei voida osoittaa yhtä epistemologista taustaoletusta tai tietokäsitystä, yleisesti ottaen etnografiaa voidaan pitää ihmisen toiminnan merkityksiä selvittävänä ja modernin maailman ymmärtämystä lisäävänä paradigmatena (Varto 1992; Hornberger 1994; Vidich & Lyman 1994; Savage 2000).

Tässä tutkimuksessa etnografisuus viittaa tulkintasääntöjen selvittämiseen ja niiden noudattamiseen eli siihen, miten arvioijat hahmottavat ja jäsentävät toisaalta toimintaansa arvioijana, osana arvioijajayhteisöä, ja toisaalta toimintatapoja, joihin he nojaavat arvioidessaan suorituksia (ks. Alasuutari 1999; Savage 2000). Tavoitteena on siis ymmärtää ihmismielen prosesseja ja toiminnan taustalla olevia yksilöllisiä ja yhteisöllisiä merkityksiä.

Metodologisesti etnografiaa pidetään monimetodisena, joten se suosii triangulaation käyttöä. Lisäksi etnografisessa näkökulmassa on tärkeää, että tutkija pystyy asettumaan tutkittavan asemaan eli ottamaan itse osaa tutkimukseen, koska mitä intensiivisempään ja avoimempaan vuorovaikutukseen tutkija pääsee tutkittavan kanssa, sen luontevampia tulkintoja tutkija voi tehdä (Ellis 1994, 568; Atkinson & Hammersley 1994, 249; Syrjälä & ym. 1994, 78). Toisaalta taas tästä syystä etnografinen metodi on varsin refleksiivinen tutkijan osalta, joten ymmärrys voidaankin saavuttaa vain silloin, kun tutkija itse on tietoinen niistä tekijöistä, jotka häntä motivoivat ja jotka hän on valmis kohtaamaan (vrt. Vidich & Lyman 1994; Savage 2000).

Tässä tutkimuksessa arvioin itsekin kaikki suoritukset yhtenä arvioijana, joten myös minulla on kokemus arviointiprosessista. Näin ollen voin tutkijana päästä ehkä lähemmäksi sitä, miten arvioija käyttäytyy prosessin aikana ja millaisia käsityksiä ja uskomuksia arviointiin saattaa liittyä. Tosin jos en ole tutkijana tietoinen omien motiivieni vaikutuksesta aineiston keräämis- ja analyysivaiheessa, siitä voi seurata ylitulkintaa, tulkitsematonta jättämisiä ja/tai oman näkökulman korostamista muiden näkökulmien kustannuksella.

Tämä tutkimus on metodiltaan etnografista tutkimusta. Pyrkimyksenä on lisätä ymmärrystä arviointiprosessista ja saada siitä aikaan avointa keskustelua (ks. esim. Suoranta 1999; Syrjälä ym. 1994; Savage 2000). Kulttuurisesta näkökulmasta tutkimus ei välttämättä täytä etnografisen tutkimuksen kriteereitä, vaikka arviointi voidaankin nähdä omana kulttuurisena kokonaisuutena, josta laajempänä ilmiönä puhuttaessa käytetään myös käsitettä *arviointikulttuuri* (vrt. Syrjälä & Numminen 1988; Suoranta 1999; Virta 1999). Toisaalta kulttuuri käsitteenä on liian helposti liimattavissa kaikenlaisten diskurssien päälle, toisaalta Suorannan (1999) mukaan keskeistä ei ole kulttuurikäsitteen määrittelyminen vaan yhteiskunnan merkitysvälitteisyyden korostaminen.

Fenomenografia puolestaan tutkii sitä, miten ympäröivä maailma ilmenee ja rakentuu ihmisen tietoisuudessa eli miten yksilö havaitsee, kokee, käsitteellistää ja ymmärtää erilaisia ilmiöitä ja häntä ympäröivää maailmaa (Marton 1994, 4426; Marton & Fai 1999, 1). Fenomenografia tulee kreikan kielen sanoista *phainemenon* 'esiintulo, ilmaantuminen' ja

graphiein 'kuvaus'. Fenomenografia sai alkunsa 1970-luvun alussa Ruotsissa Göteborgin yliopiston tutkimushankkeesta, jossa tutkittiin, miten yliopisto-opiskelijat oppivat lukemalla akateemisia tekstejä. Tutkimuksessa kävi ilmi, että opiskelijoiden oppimistulokset olivat kiinteässä yhteydessä heidän yksilölliseen tapaansa oppia ja heidän erilaisiin tapoihinsa kokea erityistilanteita (Marton & Fai 1999, 2–4).

Tämän tutkimuksen kannalta fenomenografinen ote tarkoittaa, että tutkimuksen kohteena ovat arvioijan käsitykset eli se, mitä jokin ilmiö, tässä tapauksessa esimerkiksi kirjoittaminen taitona, merkitsee arvioijalle tai miten arvioija ymmärtää ilmiön (vrt. Tesch 1990; Marton 1994; Syrjälä ym. 1994). Käsitykset ovat siis ajattelukonstruktioita, joiden pohjana on tavallisesti olioita ja tapahtumia koskeva ajattelu. Käsitykset ovat myös yksilöllisiä ja sisällöllisesti erilaisia, koska niiden viitetausta vaihtelee yksilöstä toiseen (Syrjälä ym. 1994, 119; Barnard ym. 1999, 217). Käsitykset ovat myös dynaamisia, sillä yksilö voi muuttaa käsityksiään useasti varsin lyhyessäkin ajassa (Marton 1994).

Omien käsitysten kuvaileminen edellyttää reflektiota ja tietoisuutta, mutta käsitysten ilmaiseminen onnistuu ilman reflektiotaikin: haastateltava voi ilmaista käsityksensä tiedostamatta sitä itse (Simoila 1993, 24, 26). Tietoisuutta ei kuitenkaan nähdä dikotomisena jakona tietoiseen ja eietoiseen tai tajuntaan ja alitajuntaan, vaan jotkut asiat tai aspektit ovat "etualalla" (foreground), ne ovat eksplisiittisiä ja tematisoituneita, ja jotkin toiset asiat ovat taas taka-alalla (background) ja ne ovat implisiittisiä ja tematisoitumattomia (Marton 1994, 4427). Me olemme tietoisia kaikesta kaiken aikaa ja me olemme tietoisia kaikesta eri tavalla kaiken aikaa. Fenomenografia siis olettaa, että ihmisen ajattelu on tajunnallinen ja aktiivinen prosessi ja että hänen toimintansa on tarkoituksellista.

Tavallisin tapa selvittää yksilön käsityksiä on haastatella ihmisiä ja luokitella sitten heidän käsityksensä niiden merkitysten perusteella. Käsitysten kuvaamisessa on tärkeää kuvata niitä niiden omista lähtökohdista käsin ja niiden omaa logiikkaa noudattaen; niitä ei siis esimerkiksi järjestellä paremmuusjärjestykseen (Marton 1994; Syrjälä ym. 1995). Erilaisia käsityksiä eli haastateltavien lausumia on tarkasteltava Martonin (1994, 4428) mukaan kahdessa eri kontekstissa. Toisen kontekstin muodostavat käsitykset, jotka johdetaan siitä, mitä tutkittavat ovat sanoneet yhdestä ja samasta asiasta, ja toisen yksilölliset rajat, jotka muodostuvat siitä, mitä yksi ja sama henkilö on sanonut muista asioista.

Käsityksiä on siis tarkasteltava sekä kollektiivisesta että yksilöllisestä näkökulmasta. Fenomenografisessa analyysissä tätä kutsutaan hermeneuttiseksi elementiksi (Marton 1994, 4428). Kun ilmaisut, haastateltavien lausumat, saadaan luokiteltua edellä kuvatulla tavalla, niistä muodostuu kuvauskategorioita (categories of description), jotka viestivät siitä, miten

tietty ilmiö koetaan, käsitteellistetään ja ymmärretään. Nämä kuvauskategoriat ovat fenomenografisen tutkimuksen päätuloksia (Marton 1994).

Käsitykset nähdään myös tässä tutkimuksessa toisaalta yksilöllisinä ja toisaalta sosiaalisina. Käsitykset eivät siis ole vain yksilöllisiä, vaan ne muotoutuvat myös osana sosiaalista diskurssia. Arvioijat kuuluvat myös arvioija- tai opettajayhteisöön, johon liittyvät omat odotuksensa tai käsityksensä siitä, millainen arvioija on ja toisaalta taas ei ole. Tässä mielessä arvioijien käsitysten ei oleteta olevan yhdensuuntaisia tai loogisia, vaan ne voivat heijastaa sekä yksilön että sosiaalisen yhteisön käsityksiä. Tämän vuoksi arvioijien käsityksiä on järkevintä tutkia itse arvioinnin yhteydessä eli sen suhteen kautta, mikä arvioijilla on arvioita-viin suorituksiin ja suoritusten kautta niiden kirjoittajiin.

Yhteistä näille molemmille tutkimusotteille on hermeneuttisuus, joka korostaa tosiasioiden ja arvojen välistä yhteyttä. Positivismia on taas kritisoitu nimenomaan siitä, että se sulkee arvot tieteellisen ajattelun ulkopuolelle. Positivismissa ilmiöt sidotaan sekä "leveysuuntaan" että "pituussuuntaan" muihin ilmiöihin ja tapahtumiin, ajallisesti samanaikaisiin tapahtumiin ja ilmiöiden kehitykseen; karkeasti ilmaisten tieteen tehtävä rajoittuu siis havaittavien ilmiöiden ja niiden välisten säännön-mukaisten yhteyksien kuvailuun. Tätä lähestymistapaa käsittelen seuraavassa alaluvussa.

1.2.3 Miten tutkimus selittää: miksi asiat tapahtuvat tietyllä tavalla?

Deduktiiviseen tieteenihanteeseen kuuluu se, että selitetään, miksi asiat tapahtuvat tietyllä tavalla. Selittämisessä on kyse selitysjärjestelmän ideaalista, jossa kaikki tieteelliset lauseet liittyvät toisiinsa deduktiivisesti. Deduktiivisessa järjestelmässä jonkin asian selittäminen merkitsee sen osoittamista, että tapahtuma on yleisen lain erityistapaus. Kun lait on keksitty, voimme tarkasti määritellä olosuhteet tai ehdot, joiden vallitessa lait pätevät, sekä eristää ja identifioida muuttujat, jotka vaikuttavat siihen, miten laki toimii. (Raunio 1999, 155–156.)

Siinä missä universaalit lait voivat sopia luonnontieteisiin, ihmis- ja sosiaalityeteissä universaalien lakien osoittaminen ei ole mahdollista eikä tarkoituksenmukaista. Sen sijaan Raunion (1999, 159) mukaan yhteiskuntatieteissä tulee sallia universaalien lakien sijaan tilastolliset säännön-mukaisuudet. Tilastolliset selitykset eivät kuitenkaan ota useinkaan kantaa selitettävän ilmiön johtamiseen tietystä selitysperstustasta, joka muodostuu laeista ja alkuehdoista. Kyse on tavallisimmin vain tiettyjen

ilmiöiden, asiaintilojen tai tapahtumien välisen riippuvuuden, yhteisvaihtelun, empiirisestä tutkimisesta (Cozby 1985, 36; Raunio 1999, 157).

Pelkkä tilastollisen yhteyden toteaminen ei kuitenkaan oikeuta oletamaan syy-vaikutusyhteyden olemassaoloa, koska kausaalista syytä ja sen vaikutusta ei voi perustella tilastollisella tekniikalla sellaisenaan. Yleensä kausaalisuhteen ja sen lähtökohtana olevan teoreettisen taustan analysointi edellyttää myös laadullisten piirteiden analysointia. Ihmistieteissä mittaaminen sellaisenaan ei siis riitä, jos ilmiötä ja siihen liittyviä mekanismeja halutaan selvittää perusteellisesti (vrt. Cozby 1985; Raunio 1999). Arvioijien toiminnan variaation yleisyydestä on kuitenkin mahdollista saada tietoa vain kvantitatiivisin menetelmin, joiden avulla pystytään tekemään ero sattuman ja systemaattisen vaihtelun välillä.

Numeerisen aineiston tehtävänä onkin osoittaa, millaista systematiikkaa arvioijan käyttäytymiseen liittyy ja miten yleisellä tasolla se heijastelee arvioijien toiminnan logiikkaa. Kvantitatiivisella otteella pyritään tekemään luotettavia päätelmiä ilmiön määrällisestä esiintymisestä. Siinä missä kvantitatiivisella otteella tähdätään menetelmälliseen täsmällisyyteen, tilastolliseen edustavuuteen ja säännönmukaisuuksien selittämiseen, kvalitatiivinen näkökulma pyrkii tavoittamaan sisällöllisesti olennaisia piirteitä ja niiden välisiä yhteyksiä (vrt. Brannen 1992; Sayer 1992).

Tässä työssä kvantitatiivisilla ja kvalitatiivisilla menetelmillä vastataan erilaisiin tutkimuskysymyksiin. Kysymykset ovat erilaisia, eikä menetelmien keskinäiseen paremmuuteen oteta kantaa. Taustalla on tässä mielessä eräänlainen metodinen pluralismi: erilaisilla menetelmillä tuotetut aineistot täydentävät toisiaan ja ovat periaatteessa yhtä tärkeitä tutkimuksen eri puolia tarkasteltaessa (vrt. Brannen 1992, 31; Raunio 1999, 346–348).

1.3 Tutkimuskysymykset

Käsillä olevan tutkimuksen aineisto koostuu määrällisestä ja laadullisesta aineistosta. Kvantitatiivinen aineisto on arvioijien antamia arvioita (1–6) suomi toisena kielenä -oppijoiden kirjoittamista teksteistä. Arvioijat arvioivat samat suoritukset kolme kertaa siten, että he käyttivät ensimmäisellä arviointikierröksellä pelkkää numeroasteikkoa, jonka sisällön he itse määrittelivät. Toisella kierroksella arvioijat arvioivat suoritukset 6-portaisella asteikolla, jossa oli tasokuvaukset. Kolmas kierros oli toisen kierroksen toisto. Kvalitatiivinen aineisto on taas peräisin kunkin arviointikierröksen jälkeisistä haastatteluista, myös

haastatteluja oli siis kaikkiaan kolme. Käsittelen tutkimuksen kulkua yksityiskohtaisemmin luvussa 2.

Tutkimuksessa etsitään vastauksia kirjoittamisen arviointia koskeviin kysymyksiin kvantitatiivisin ja kvalitatiivisin metodein. Kvantitatiivinen osa pyrkii kuvaamaan arvioijien välistä yhdenmukaisuutta ja johdonmukaisuutta kolmen eri arviointikierroksen ja kahden eri asteikon suhteen. Numeerisen aineiston keräämistä ja käsittelyä ohjaavat seuraavat kysymykset:

1. Miten *johdonmukaisesti* arvioijat arvioivat suorituksia
 - käyttäessään samaa asteikkoa kahdesti?
 - kahdella eri asteikolla?
2. Miten *yhdenmukaisesti* arvioijat arvioivat kirjoittajien suorituksia
 - käyttäessään 6-portaista asteikkoa, jonka sisällön he itse määrittelevät?
 - käyttäessään käyttäjä- ja tekstityyppilähtöistä asteikkoa?

Tutkimuksen laadullisessa osassa arviointia tarkastellaan selvittämällä arvioijien käsityksiä sekä arvioitavista teksteistä, arviointiprosessista ja omasta toiminnastaan. Haastatteluaineiston kautta vastauksia haetaan seuraaviin kysymyksiin:

3. Millaiset *konstruktit* (personal constructs) ohjaavat arvioijien työtä / millainen käsitys arvioijilla on kirjoittamisesta taitona
 - heidän itse luomansa asteikon / valmiista asteikosta esitettyjen kommenttien mukaan?
 - yleensäkin suorituksiin liittyvien kommenttien mukaan?
 - valintaverkko haastattelun (repertory grid) mukaan?⁷
4. Millaisia *tuntemuksia, vaikutelmia ja käsityksiä* arviointiprosessiin liittyy?

Työn viides tutkimuskysymys liittyy kvantitatiivisen ja kvalitatiivisen näkökulman keskinäiseen vuorovaikutukseen:

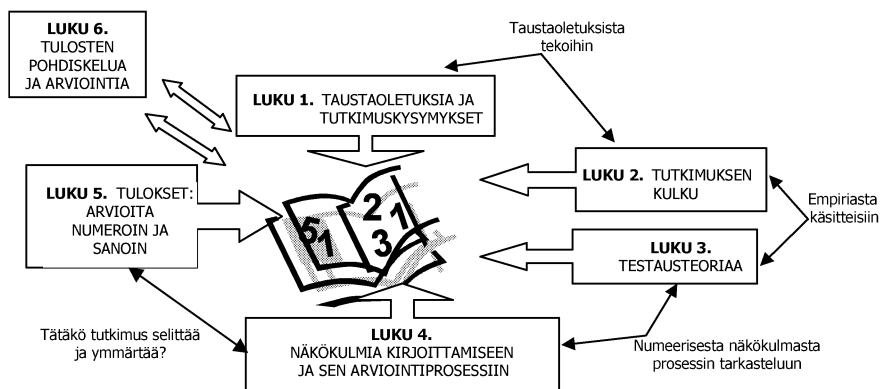
5. Miten
 - edustavasti ja systemaattisesti kvantitatiivinen ote kuvaa arvioijan toimintaa?
 - monipuolisesti ja syvästi kvalitatiivinen ote valottaa arvioijan toimintaa?
 - kokonaisvaltaisen kuvan kaksi otetta yhdessä luovat arvioijan toiminnasta?

⁷ Valintaverkko haastattelu eli repertory grid -haastattelu on strukturoitu haastattelumetodi, jonka avulla voidaan tarkastella haastateltavan tulkintoja ympäröivästä maailmasta. Tässä tutkimuksessa tulkinnoilla ympäröivästä maailmasta tarkoitetaan arvioijan havaintoja arvioitavista teksteistä.

1.4 Kirjan rakenteesta: vihjeitä lukijalle

Tutkimuksen sisältö rakentuu pääpiirteittäin kuvion 2 mukaisesti. Ensimmäisen luvun tarkoituksena on viestiä, millaiseen viitekehykseen tutkimus pohjautuu ja mihin kysymyksiin ja millaisilla metodeilla niihin etsitään vastauksia. Toisessa luvussa käyn läpi tutkimuksen kulkua eli esittelen kirjoittamistehtävät, arviointiasteikon ja työn päähenkilöt, niin arvioitavien tekstien kirjoittajat kuin arvioijatkin. Lisäksi esittelen tutkimusasetelman, jonka mukaisesti varsinainen tutkimusaineisto on kerätty. Käyn vaihe vaiheelta läpi eri arviointikierrokset ja niihin liittyvät haastattelut. Perustelen tässä yhteydessä myös haastatteluihin liittyviä valintoja.

Luvuissa 3 ja 4 tarkastelen tutkimuksen kannalta olennaista taustakirjallisuutta. Kolmas luku käsittelee testausteoriaa, jonka taustaoletuksena on pitkälti positivistinen realismi (ks. Hamp-Lyons & Lynch 1998; Lynch 2001). En kuitenkaan korosta taustaoletusta – en kriittisessä enkä kannustavassa hengessä – tietoisesti, vaan käsitellen tässä yhteydessä arviointia, nimenomaan testausarviointia, sille tyypillisistä näkökulmista. Tarkastelen arvioinnille asetettuja laatuksiteereitä, varsinkin arvioijan toimintaan liittyvää reliiabiiliuden käsitettä. Luvun lopuksi esittelen tyypillisimmät arviointitavat, jotta tutkimuksessa käytetty holistinen arviointitapa tulisi kuvatuksi suhteessa muihin, lähinnä analyttiseen arviointitapaan. Kolmannen luvun ensisijaisena tavoitteena on valottaa arvioinnin kvantitatiivista lähtöoletusta ja sitä, miten reliiabiiliuden käsite liittyy kirjoittamisen arviointiin.



KUVIO 2 Kirjan rakenne.

Neljännessä luvussa tarkastelen kirjoittamista taitona eri tutkimusparadigmojen näkökulmasta. Käsittelen myös äidinkielen ja vieraan kielen kirjoittamisesta esitettyjä käsityksiä ja pohdiskelen lyhyesti suomi toisena kielenä -kirjoittamista. Lisäksi tarkastelen arviointiprosessia, lähinnä sitä, millaisten tekijöiden vaikutuksesta ja miten itse vaikuttaen arvioija tekee päätöksiä eli antaa arvioita. Listaan lyhyesti myös holistiseen kirjoittamisen arviointiin liittyviä tutkimuksia, jotka havainnollistanevat taas sitä, mitä arvioija arvioi suorituksista ja millaisiin tekijöihin hän todennäköisesti perustaa arvionsa.

Viidennessä luvussa esittelen tutkimuksen kvantitatiiviset ja kvalitatiiviset tulokset. Raportoin ensin kvantitatiivisten analyysien tulokset. Tarkastelen arvioijien antamia arvioita sekä arvioijien sisäisen johdonmukaisuuden että heidän keskinäisen yhdenmukaisuutensa näkökulmasta. Haastattelujen tulokset esittelen arviointikierroksittain. Pidän tätä tarkoituksenmukaisena siitä syystä, että kussakin haastattelussa painotetaan hieman eri asioita: Ensimmäisen kierroksen jälkeisessä haastattelussa arvioijan itsensä laatimaa asteikkoa ja toisen kierroksen jälkeisessä haastattelussa annetun asteikon sovellettavuutta. Kolmannen kierroksen jälkeinen haastattelu perustuu taas täysin eri metodiin, valintaverkkoon (repertory gridiin). Esittelen valintaverkkometodin vasta tulosten yhteydessä. Luku voi tuntua varsin itsenäiseltä tästä syystä, mutta sillä on kuitenkin perustelunsa: näen valintaverkkometodin eräänlaisena siltana sen välillä, mitä arvioija sanoo arvioivansa ja mitä hän todella arvioi.

Viimeisessä luvussa on aika katsoa taaksepäin ja tarkastella sitä, millaista vuorovaikutusta kahden eri tutkimusotteen tuottamien tulosten välillä on. Puhuvatko tulokset samaa vai eri kieltä ja miten ristiriitaisilta ja/tai toisiaan täydentäviltä ne vaikuttavat? Koska käsittelen tätä menetelmien keskinäistä suhdetta omassa alaluvussaan, en juurikaan puutu siihen itse tulosten raportointia koskevissa alaluvuissa. Tutkimuksen rajoituksia ja luotettavuutta pohdittuani on tämänkin tutkimuksen päätteeksi mahdollista kysyä: "Entä sitten?". Vastaus löytyy viimeisestä luvusta.

2 TUTKIMUKSEN KULKU

2.1 Arvioitavan aineiston kerääminen

2.1.1 Kirjoitustehtävät ja arviointiasteikko

Kirjoitustehtävät valittiin tehtäväaineistosta, joka oli laadittu DIALANGin käyttöön. Kirjoittaminen on DIALANGissa jaettu suoraan ja epäsuoraan kirjoittamiseen. Suoralla kirjoittamisella tarkoitetaan tehtäviä, joissa kirjoittaja tuottaa tehtävänannon mukaisen tekstin. Tehtävänannon ohjeiden yksityiskohtaisuudella voidaan ohjata kirjoittamista, ja onkin tavallista, että ohjeessa mainitaan vähintään tekstityyppi (esimerkiksi kirje, viesti, tarina, hakemus), kuvataan tilanne (esim. vastaanottaja ja mahdollinen intentio kirjoittaa) sekä esitetään joitakin kysymyksiä tai aihealueita, joita tekstissä tulisi käsitellä (esimerkiksi syy kirjoittamiseen, kerro itsestäsi, mitä tapahtui).

Epäsuorassa kirjoittamisessa tehtävät mittaavat kirjoittajan tietoa kirjoittamisesta, kirjoittamisen konventioista, tekstin jäsentelystä ja rakenteesta sekä kielellisestä tarkkuudesta. Tehtävät ovat tavallisesti monivalinta- tai aukkotäydennystehtäviä. Kirjoittaja ei siis itse tuota varsinaista tekstiä. Jos tehtävä on avoin kysymys tai aukkotäydennys, kirjoittaja tuottaa vastaukseksi vain sanan tai kaksi. Tämän vuoksi epäsuora kirjoittaminen soveltuu hyvin tietokonepohjaiseen arviointiin. Sekä epäsuoran että suoran kirjoittamisen tehtävät on laadittu DIALANGin runkona olevan 6-portaisen asteikon mukaan, jolla taso 1 edustaa aloittelevan kielenoppijan taitotasoa ja taso 6 lähes syntyperäisen kaltaista tasoa (ks. liite 1).⁸

Tässä tutkimuksessa käytettiin vain suoran kirjoittamisen tehtäviä. Kirjoitustehtäviä oli kaikkiaan yhdeksän. Tehtävien vaikeustaso vaihteli sen mukaan, mille tasolle tehtävät oli niitä laadittaessa suunnattu. Tehtävän vaikeustaso ei kuitenkaan tarkoita sitä, ettei tehtävä toimisi myös muilla tasoilla. Tasomäärittelyllä on pyritty ennen kaikkea siihen, että kyseisellä tasolla oleva kirjoittaja pystyisi parhaalla mahdollisella tavalla osoittamaan osaamisensa. Esimerkiksi ostoslistatehtävä on varmasti liian helppo C-tason kirjoittajalle ja työpaikkahakemus taas turhaut-

⁸ 6-portaisesta asteikosta käytetään myös kirjaimia siten, että se jaetaan kolmeen osaan: tasoja 1 ja 2 kutsutaan A-tasoksi, tasoja 3 ja 4 B-tasoksi ja tasoja 5 ja 6 C-tasoksi.

tavan vaikea A-tason oppijalle, mutta tuttavallinen kirje sopii kaikille tasoille. Tutkimuksessa käytettävät yhdeksän tehtävää (ks. liite 2) olivat tasoittain, tekstityypeittäin ja funktioittain taulukon 1 mukaisia (ks. Council of Europe 2001, 126–127).

TAULUKKO 1 Kirjoitustehtävät tasoittain, tekstityypeittäin ja funktioittain.

TASO		TEHTÄVÄ	TEKSTITYYPPI	MAKRO-FUNKTIO(T)	MIKRO-FUNKTIO(T)
A-TASO	Tasot 1–2	1) Ostoslista	ohje/viesti	ohjaaminen, kuvaileminen	tiedon jakaminen
		2) Kirje ystävälle	tuttavallinen kirje	kuvaileminen, selittäminen	itsensä kertominen, asenteen ilmaiseminen, sosiaalinen sopeuttaminen
B-TASO	Tasot 3–4	3) Ajo-ohjeet	ohje	ohjaaminen, kuvaileminen, havainnollistaminen	tiedon jakaminen
		4) Kirje (materiaalina kirje-ilmoitus)	kirje tuntemattomalle	esitleminen, kuvaileminen	itsensä esittely, asenteen ilmaiseminen, sosiaalinen sopeuttaminen
		5) Lyhyt esittely	tiivistelmä/esittely	esitleminen, perusteleminen, kuvaileminen, raportointinen	tiedon jakaminen, vakuuttelu, sosiaalinen sopeuttaminen
C-TASO	Tasot 5–6	6) Mielenpidekirjoitus	lehtiteksti	perusteleminen, kommentointinen, kuvaileminen	mielipiteen ja asenteen ilmaiseminen, vakuuttelu
		7) Suruvalittelu	kirje	kuvaileminen, selittäminen	empatian ilmaiseminen, sosiaalinen sopeuttaminen
		8) Työpaikka-hakemus	virallinen kirje	perusteleminen, suostuttelemine	kuvaaminen, itsensä esittely, tiedon jakaminen, asenteen ilmaiseminen, vakuuttelu, sosiaalinen sopeuttaminen
KAIKKI TASOT		9) Tarina kuvasta	kertova teksti	kertominen, kuvaileminen, havainnollistaminen	asenteen ja tunteiden ilmaiseminen, vakuuttelu

Kaksi tehtävistä oli suunnattu taitotasoiille yksi ja kaksi, kolme tehtävää taitotasoiille kolme ja neljä ja kolme tehtävää tasoille viisi ja kuusi. Yksi tehtävistä oli tarkoitettu kaikille tasoille. Tehtävät instruktioineen ovat liitteessä 2. Tehtävien valinnassa on kiinnitetty huomiota siihen, että ne olisivat funktioiltaan erilaisia ja ainakin jossakin määrin relevantteja kirjoittajille. A-tason tehtäville on tyypillistä, että ne ovat lyhyitä ja sekä kielellisesti että tekstuaalisesti peruskielitaitoa vaativia. Niille on tyypil-

listä myös se, että niissä voi kirjoittaa hyvin konkreettisista ja kirjoittajaa itseään koskevista asioista.

B- ja C-tasolla tehtäviä on yksi enemmän kuin A-tasolla, koska näillä tasoilla sekä kirjoittamisen funktioiden vaatimukset kovenevat että oppijoiden kielitaitoprofiileissa on enemmän variaatiota. B-tasolla tehtävien sisältö on A-tasoa abstraktimpi ja tilanteet vaativampia, vaikkakin aiheet liittyvät pääasiassa kirjoittajan omaan elämään tai niitä on mahdollista käsitellä omasta näkökulmasta kirjoittajan niin halutessa. Ajo-ohje- ja kirjetehävä ovat tekstityypeiltään ja funktioiltaan perusasioita vaativia. Sen sijaan B-tason esittelytehtävän tarkoituksena on, että kirjoittaja pystyisi argumentoimaan valintojaan hiukan laajemmasta näkökulmasta. Välitasona B-tason tehtäville on tärkeää, että myös A-tason kirjoittaja pystyisi selviytymään niistä, esimerkiksi kirjoittamalla vain asialistan esittelytehtävään.

C-tason tehtävissä abstraktiotaso nousee, ja niissä edellytetään myös muodollisen rekisterin hallintaa. Lisäksi tehtävät ovat kirjoittamisen konventioiden puolesta vaativampia tai ainakin kirjoittamistilanteina spesifimpiä, joten niissä on mahdollista osoittaa hyvinkin idiomaattista kielitaitoa. Näissäkin tehtävissä voi tarkastella asioita omasta ja/tai konkreettisesta näkökulmasta, joskin ne tarjoavat myös laajemman perspektiivin ja jossakin määrin edellyttävät laaja-alaisempaa argumentointia.

Valitsin kuvapohjaisen tarinan kaikille tasoille yhteiseksi tehtäväksi lähinnä kahdesta syystä: tehtävä sopii kaikille tasoille ja tehtävää voi pitää vähemmän kulttuurisidonnaisena kuin muita tehtäviä. Jos kielitaito on vähäinen, voi kirjoittaja nimetä kuvasta tuttuja asioita, kuten *tyttö*, *sänky*, *lääkäri seisoo*, eli hän voi tuottaa tekstin vähäiselläkin kielitaidolla. Tutkimuksissa on myös osoitettu, että huonotasoiset kirjoittajat, jotka saivat kuvapohjaisen instruktioin, pystyivät tuottamaan paremmin jäsennellyn ja sidosteisemmän esseen kuin kielellisen instruktioin saaneet kirjoittajat (Quellmalz 1984, 30). Jos kielitaito taas on vahva, kirjoittaja voi viedä tarinaa kauemmaksi kuvasta, mutkistaa juonta haluamaansa suuntaan ja käyttää hyvin erikoistakin sanastoa. Lisäksi tarinan kerronnassa kirjoittajan koulutustausta ei välttämättä vaikuta niin paljon tehtävän suorittamiseen kuin esimerkiksi työpaikkahakemuksessa.

Vaikka kuva ja sen lukeminen voivat olla kulttuurisidonnaisia ja tarinan kertominen jossakin määrin persoonallisuusvaikutteista, narratiivisuutta voi kuitenkin pitää jossakin määrin kaikille kulttuureille yhteisenä ja yleisyytensä vuoksi luontevana ilmaisutapana (esim. Bruner 1986; Price & Takala 1988). Kuvapohjaisen tehtävän huonoihin puoliin voi lukea sen lapsellisuuden ja epärelevanttiuden. Koska kirjoittajat ovat aikuisia kielenoppijoita, kuvasta kirjoittaminen voi olla heidän mielestään

lapsellista. Myös tehtävän tarkoitus voi tuntua kirjoittajista turhauttavalta ja keinotekoiselta, siltä, että tehtävässä kirjoitetaan vain kirjoittamisen vuoksi. Tässä mielessä esimerkiksi kirje tai mielipidekirjoitus tuntuvat tarkoituksenmukaisemmilta.

Kirjoitustehtävät jaettiin kirjoittajille taulukon 2 mukaisesti. Taulukon 2 ylärivillä olevat T-alkuiset merkinnät viittaavat tehtäviin. Merkinnässä TA1 kirjain T tarkoittaa tehtävää yleisesti, kirjain A asteikon tasoja yksi ja kaksi, numero 1 nimenomaan ostoslistatehtävää edellä esitetyn tehtävälisan mukaisesti. Vastaavasti merkinnässä TB2 kirjain B tarkoittaa tasoja kolme ja neljä ja numero 2 kirjetehtävää. Kuten taulukosta 2 näkyy, tehtävät on jaettu neljään kolmen tehtävän tehtäväsarjaan, joista esimerkiksi kolmannen sarjan tehtävät suorittanut kirjoittaja on saanut kirjoitettavakseen esittelyn, mielipidekirjoituksen ja kuvapohjaisen tarinan. Jokainen kirjoittaja suoritti siis kolme tehtävää, joista yksi oli yhteinen kaikille kirjoittajille.

TAULUKKO 2 Kirjoitettavan aineiston asetelma.

Tehtäväsarja	TA1	TA2	TB1	TB2	TB3	TC1	TC2	TC3	TX
1. tehtäväsarja	X		X						X
2. tehtäväsarja		X					X		X
3. tehtäväsarja					X	X			X
4. tehtäväsarja				X				X	X

Tutkimusten mukaan arvioinnin ja kirjoittamisen taidosta tehtävien johtopäätösten luotettavuuden näkökulmasta on järkevää, että oppijat kirjoittavat useamman kuin yhden tekstin (esim. Cooper 1977; Quellmalz ym. 1982; Newkirk 1982; Huot 1990). Kirjoittajan kykyä kirjoittaa erimuotoisia tekstejä ja kykyä osoittaa eri funktioiden hallintaa on mahdotonta arvioida yhden tekstin perusteella. Monien kirjoittamisen taitoon liittyvien tekijöiden, kuten syntaksin, sanaston, kirjoittamisen konventioiden, hallinta vaihtelee paljonkin kirjoittamismuodosta ja jopa aiheesta toiseen (Lloyd-Jones 1977, 33; Huot 1990, 243–245; Reid 1990, 205). Useamman kirjoittamistehtävän teettämisen taustalla on siis ajatus siitä, että jos oppija osaa kirjoittaa hyvän kirjeen, ei asia välttämättä ole niin, että hän osaa kirjoittaa myös tarkoituksenmukaisen raportin tai tarinan.

Tehtävien määrällä ei tietenkään ole samalla tavalla merkitystä, jos ollaan kiinnostuneita vain yhden tekstityypin, esimerkiksi tuttavallisen kirjeen, kirjoittamisesta. Tällöinkin on tietysti otettava huomioon, miten edustava näyttö yksi teksti voi olla ja miten esimerkiksi aihe voi vaikuttaa tekstin laatuun (Reid 1990, 197). Kaiken kaikkiaan on tärkeää, että tehtävänanto on selkeä ja että kirjoittaja tietää, mitä hänen odotetaan kirjoittavan. Lisäksi lähes kaikissa kirjoittamistehtävää koskevissa tutkimuksissa on kiinnitetty huomiota siihen, että tehtävän pitäisi vastata niitä kirjoitustarpeita, joita kirjoittajalla on omassa elämässään (esim. Hamp-Lyons 1990; Matsuda 1998; Cushing Weigle 1999).

Tehtävän on oltava sellainen, että se yleensäkin mahdollistaa osaamisen osoittamisen eli että aiheesta sekä ohjeen sisällössä ja kieliasun muotoilussa otetaan mahdollisimman hyvin huomioon kulttuuritekijät. Myös tehtävän suorittamiseen varatun ajan on oltava realistinen ja tarkoituksenmukainen suhteessa tehtävänantoon ja tehtävän vaikeustasoon (ks. esim. Ruth & Murphy 1988). Kirjoittamiseen liittyvien olosuhteiden ja tehtävänantojen vaikutusta suoritukseen voidaan jonkin verran kontrolloida tutkimustiedon ja suunnitelmallisuuden avulla.

Sen sijaan tehtävääntoon sisällytettyä vaatimusta kirjoittamisen sisällöllisestä laadusta, haastavuudesta ja vaikeustasosta on vaikea erottaa muun muassa aiheen hallinnasta tai kirjoittajan motivoitumisesta, varsinkaan edistyneiden oppijoiden kohdalla. Vaikka tehtävän sisällön laatu ja haastavuus onnistuttaisiinkin määrittelemään etukäteen, tuo siihen senkin jälkeen oman lisänsä arvioijan tulkinta. Hamp-Lyonsin (1991c, 103) mukaan onkin syytä pohtia, missä määrin tehtävän haastavuus on tehtävässä itsessään ja missä määrin tehtävän vaikeuden määrittelee tekstin lukija.

Tehtävien arvioimiseen arvioijat käyttivät kahta arviointiasteikkoa siten, että ensimmäisellä arviointikierroksella arvioijat määrittivät itse 6-portaisen asteikon sisällön ja toisella ja kolmannella arviointikierroksella he käyttivät yhteistä valmiiksi määriteltyä 6-portaista asteikkoa.

Ennen ensimmäistä arviointikierrosta lähetetyssä kirjeessä arvioijat saivat ohjeen arvioida suoritukset asteikolla 1–6, jolla taso 1 on heikoin taso, 6 vahvin ja tasot 2–5 näiden välissä. Kirjeessä kerrottiin myös, että ensimmäisellä arviointikierroksella arvioijalla ei ole käytettävissään valmiita kriteereitä, vaan hänen tulee itse määrittellä, millainen suoritus edustaa taitotasoa 3 ja millainen suoritus esimerkiksi taitotasoa 5. Tämän enempää arvioijat eivät saaneet ohjausta arvioinnin aloittamiseen, itse arvioimiseen tai tasojen suhteuttamiseen suoritusten tasoon.

Toisella ja kolmannella arviointikierroksella arvioijat käyttivät 6-portaista asteikkoa, jonka mukaan DIALANGin kirjoitustehtävät on laadittu (ks. liite 1). Asteikko perustuu Euroopan neuvoston julkaisemaan

kielten oppimista ja opetusta koskevaan yleiseurooppalaiseen viitekehykseen (ks. Huttunen 1996; Council of Europe 1996, 2001). Asteikko keskittyy enimmäkseen kuvaamaan sitä, millaisia tekstityyppejä kirjoittaja pystyy tuottamaan kullakin taitotasolla. Asteikossa otetaan kantaa myös funktioihin ja tekstin sisältöön siltä osin, miten se liittyy tai ei liity suoranaisesti kirjoittajaan itseensä. Varsinaista kielen hallintaa asteikossa kuvataan yleisluonteisesti ja epäsystemaattisesti; kielen hallintaan otetaan kantaa lähinnä sen mukaan, miten se rajoittaa tai ei rajoita kirjoittajan ilmaisua tai lukijan ymmärtämistä. Yksityiskohtaisemmin kielen rakenteiden tietyn osa-alueen hallintaan viitataan oikeastaan vain taitotasolla 2.

Tässä mielessä asteikkoa voisi luonnehtia käyttäjakeskeiseksi eli se pyrkii ennen kaikkea antamaan informaatiota kielenoppijalle siitä, minkä tyyppisistä kirjoitustehtävistä hänen oletetaan selviytyvän kullakin taitotasolla. Alderson (1991, 72–74) jakaa kielitaitoa kuvaavat asteikot käyttäjakeskeisiin, arvioijakeskeisiin ja laatijakeskeisiin sen mukaan, kenelle asteikko on tarkoitettu. Käyttäjakeskeinen asteikko tarjoaa informaatiota tasolle tyypillisestä tai keskimääräisestä osaamisesta, arvioijakeskeinen taas ohjaa arviointia, se pyrkii vakiinnuttamaan arvioijiin liittyvää reliiabiliteettiä ja vahvistamaan validiutta. Laatijakeskeinen asteikko puolestaan ohjaa laatijoita tuottamaan oikeantasoisia tehtäviä kullekin taitotasolle.

Asteikossa on myös laatijakeskeisen asteikon piirteitä, sillä siinä viitataan taajaan yksittäisiin tekstityyppeihin. Arviointikeskeiseltä asteikolta kuitenkin vaaditaan, että se pystyy kuvaamaan yksityiskohtaisemmin tekstin laatua esimerkiksi kielen, jäsentelyn ja sisällön näkökulmasta (vrt. Pollitt 1991). Tässä käytettävän tyyppisiä asteikkoja käytetään myös arviointiasteikkona, joskin ne ovat saaneet kritiikkiä osakseen. Pollitt (1991, 87) kritisoi esimerkiksi hyvin samantyyppistä ASLPR:n (Australian Second Language Proficiency Ratings) arviointi- ja käyttäjäkohtaista asteikkoa siitä, että se ei tarjoa sen enempää informaatiota opettajille kuin oppijoillekaan.

Vaikka tässä käytettävä asteikko ei sovellukaan kovin hyvin arviointitarkoitukseen, sitä kuitenkin käytettiin siitä syystä, että DIALANGin tehtävät on laadittu tämän asteikon mukaan. Tämä asteikko eroaa kuitenkin oppijoille tarkoitettusta itsearviointiasteikosta siinä, että tasojen kuvaukset ovat laajempia kuin itsearviointiasteikossa, koska tässä asteikossa on yhdistetty yleisen kirjallisen tuottamisen ja tekstityyppi-kohtaisen kirjoittamisen kuvaukset (ks. liitteet Huttunen 1996). Toinen perustelu asteikon käyttämiselle on se, että se on sisällöltään lähellä itsearviointiasteikkoa, jota käytetään kehiteltävissä olevassa suorassa kirjoittamisessa. Suoran kirjoittamisen osuudessa kielenoppijoiden on

tarkoitus itse arvioida suorituksiaan vertaamalla niitä itsearviointi-asteikkoon ja tämän tutkimuksen perusteella valittuihin maamerkki-kirjoituksiin. Kolmanneksi DIALANGissa ei ollut käytössä vastaavaa nimenomaisesti arviointiin tarkoitettua asteikkoa, ja sellaisen kehittäminen olisi ollut oma tutkimuksensa. Asteikko saattaa kuitenkin puutteistaan huolimatta yhdenmukaistaa ja johdonmukaistaa arviointia sekä auttaa konkretisoimaan arvioijilta tulevan palautteen avulla tasojen ja suoritusten vastaavuutta.

2.1.2 Suoritusten keruu

Kirjoitustehtävien kerääminen aloitettiin ottamalla yhteyttä eri puolilla Suomea sijaitseviin oppilaitoksiin, joissa opetetaan suomea toisena kielenä. Kussakin vapaaehtoisesti mukaan tulleessa oppilaitoksessa oli yhteyshenkilö, useimmiten suomen opettaja, joka otti vastuun tehtävien kirjoittamisesta, taustatietolomakkeiden täyttämisestä ja niiden palauttamisesta. Tehtäviä lähetettiin kaikkiaan 359 kirjoittajalle, ja niistä palautuivat takaisin 288 kirjoittajan suoritukset. Tutkimuksen aineisto koostuu kuitenkin 248 kirjoittajan suorituksista, koska 40 kirjoittajan suoritukset saapuivat myöhässä eivätkä ehtineet tutkimukseen mukaan.

Oppilaitoksia, joissa tehtävät suoritettiin loppuun asti, oli kaikkiaan 17. Oppilaitokset sijaitsevat Rovaniemellä, Joensuussa, Tampereella, Lappeenrannassa, Helsingissä, Turussa ja Jyväskylässä. Oppilaitokset ovat erityyppisiä, mutta niistä tyypillisimmät ovat ammatillinen aikuis-koulutuskeskus, yliopiston kielikeskus, ammattioppilaitos ja aikuislukio.

Kirjoitustehtävien lisäksi opiskelijat täyttivät taustatietolomakkeen, jossa heitä pyydettiin taustatietojen lisäksi itsearviomaan kirjoittamisen taitoaan. Opiskelijat käyttivät itsearvioinnissa samantyyppistä 6-portaista Euroopan neuvoston asteikkoa, johon kirjoitustehtävät nojaavat, mutta asteikko oli kieleltään mukautettu helpommaksi ymmärtää. Opiskelijat suorittivat tehtävät tavallisesti oppilaitoksessa suomen tuntien aikana, vaikka kirjoittamispaikkaa ei sen kummemmin ohjeistettu. Tehtävät saattoi suorittaa yhtä hyvin kotonakin. Tehtävien suorittamisaikaa ei myöskään määritelty.

2.1.3 Kirjoittajat

Kirjoittajia oli kaikkiaan 248, mutta taustatietolomakkeen täytti 247 kirjoittajaa (ks. liite 3). Kirjoittajista 60 prosenttia on naisia ja 40 prosenttia miehiä. Kahden kirjoittajan sukupuolitieto puuttuu. Kirjoittajien ikä vai-

teli alle 20-vuotiaasta yli 55-vuotiaaseen, mutta suurin osa kirjoittajista oli iältään 20–40-vuotiaita. Kirjoittajien raportoiman mukaan äidinkieliä on kaikkiaan 38. Yleisin äidinkieli on venäjä, jonka puhujia kirjoittajista on 127 eli 52 prosenttia koko joukosta. Venäjänkielisten ryhmä on suuri verrattuna muihin kieliryhmiin, sillä toiseksi eniten kirjoittajista on äidinkieleltään englanninkielisiä (N=11) ja kolmanneksi eniten arabiankielisiä (N=10). Muiden äidinkielten, kuten yoruban, armenian ja thain, edustajien määrä on 1–3 puhujaa. Venäjänkielisten suuri määrä on sikäli odotettavaa, että he ovat Suomen suurin maahanmuuttajaryhmä (ks. Korkiasaari & Söderling 2001).

Kirjoittajilta kysyttiin myös, mitä muita vieraita kieliä he osaavat suomen lisäksi. 35 prosenttia kirjoittajista raportoi osaavansa äidinkieltään lisäksi vain suomea. 44 prosenttia kertoi osaavansa suomen lisäksi yhtä muuta vierasta kieltä ja 14 prosenttia kahta muuta vierasta kieltä. Suomen lisäksi kolmea, neljää tai viittä kieltä sanoi osaavansa 18 prosenttia kirjoittajista. Koulutukseltaan kirjoittajat eroavat toisistaan siten, että kolmella heistä ei ole minkäänlaista muodollista koulutusta, 30 prosenttia on suorittanut peruskoulua vastaavan oppimäärän, 34 prosentilla on ammatillinen koulutus ja 36 prosentilla on opisto- tai yliopistotasoinen koulutus. Koulutustaustaa koskeva tieto puuttuu viideltä kirjoittajalta.

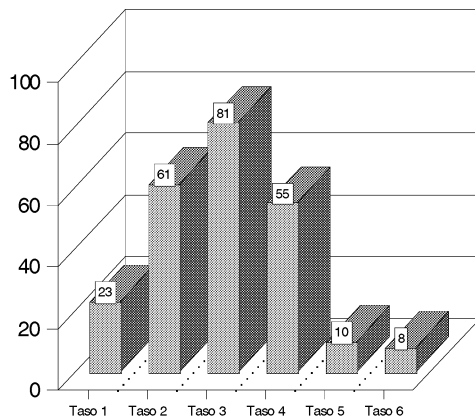
Kirjoittajien Suomessa oloaika vaihtelee parista kuukaudesta 17 vuoteen, ja keskimääräinen Suomessa oloaika on 2 vuotta 7 kuukautta. Kysymys siitä, millaisissa tilanteissa kirjoittajat käyttävät suomea, oli jaettuna kolmeen eri kontekstiin: käyttö ystävien kanssa, käyttö työssä ja käyttö asiointitilanteissa, esimerkiksi pankeissa, kaupoissa ja posteissa. Puuttuvien vastausten määrä on tässä kysymyksessä siinä mielessä mielenkiintoinen, että 108 kirjoittajaa jätti vastaamatta työkontekstia koskevaan kysymykseen, kun taas ystäviä koskevaan kysymykseen jätti vastaamatta vain 10 ja asiointitilanteita koskevaan kysymykseen 13 kirjoittajaa. Tämän perusteella näyttäisi siltä, että kirjoittajat vastasivat varsinkin työtä koskevaan kysymykseen sen perusteella, kuinka todellisia työtilanteet olivat heille. Kysymyksessä olisikin pitänyt olla yksi lisä-kategoria, jossa vastaaja olisi voinut osoittaa kysymyksen olleen epä-relevantti hänelle.

Suomen kielen käyttämiseen työssä vastasi kaikkiaan 138 kirjoittajaa. Heistä 56 raportoi käyttävänsä suomea työssään tuskin koskaan, 46 joskus ja 36 melkein aina. Sen sijaan 159 vastanneista ilmoitti, että he puhuvat suomea ystävien kanssa joskus, ja 139 raportoi käyttävänsä suomea melkein aina asiointitilanteissa. Toisin sanoen kirjoittajista yli puolet sanoi käyttävänsä useimmiten suomea asiointitilanteissa ja puhuvansa joskus suomea ystävien kanssa. Työssä käytettävän suomen

pienempi osuus voi johtua siitä, että vastaajat eivät ehkä käyneet työssä vastaamisaikana.

Suomen opiskelu on jaettu lomakkeessa kuuteen kategoriaan: 100 tuntia tai vähemmän, noin 200 tuntia, noin 400 tuntia, noin 800 tuntia, 1 000–2 000 tuntia, yli 2 000 tuntia. Kaikki kategoriat ovat edustettuina vastauksissa, mutta yleisimmin kirjoittajat (N=67) ovat saaneet muodollista suomen opetusta 200 tuntia. Toiseksi suurin muodollisen opetuksen määrä on 400 tuntia (N=60). Kirjoittajista alle 10 on saanut opetusta yli 2 000 tuntia. Tuntimäärien arviointiin on voinut vaikuttaa se, että niiden summittainenkin laskeminen voi olla vastaajan mielestä monimutkaista tai epärelevanttia; opiskeluaika voisi olla helpommin hahmotettavissa kurssityypeittäin tai vuosina.

Kirjoittajia pyydettiin taustatietolomakkeen lopuksi itse arvioimaan kirjoittamisen taitonsa 6-portaisella Euroopan neuvoston asteikolla. Asteikkoa mukautettiin kielellisesti vähän konkreettisemmaksi, jotta asteikon ymmärtäminen olisi mahdollista kaikentasoisille kirjoittajille. Kirjoittajat käyttivät asteikkojen kaikkia tasoja, kuten kuviosta 3 näkyy (ks. myös taulukko 3). Yksi kirjoittaja arvioi kirjoittamisen taitonsa alle taitotason yksi, vaikka sille ei ollutkaan osoitettu paikkaa erikseen asteikossa.



KUVIO 3 Kirjoittamisen itsearvioiden määrä (N=238) tasoittain.

Suurin osa kirjoittajista itse arvioi kirjoittamisen taitonsa tasolle 4 tai sitä huonommaksi. Itsearvioiden painottuminen asteikon puoliväliin tai sen alle heijastelee suoraan kirjoittajien keskimääräistä asumisaikaa Suo-

messä, mikä oli alle 3 vuotta. Suurin osa kirjoittajista valitsi taitotason 3, jonka mukaan 81 kirjoittajaa pystyy mielestään kertomaan asioita lähinnä itsestään tutuissa konteksteissa. Kirjoittajat valitsivat toiseksi useimmiten taitotason 2 ja lähes yhtä usein taitotason 4. Vähiten kirjoittajat käyttivät tasoja 5 ja 6, mikä osoittanee tasojen korkeaa vaatimustasoa tai/ja kirjoittajien realistista suhtautumista taitotasoonsa.

TAULUKKO 3 Itsearvioiden määrä ja itsearviointiasteikko kuvauksi- neen.

N / %	Taso	OSAAN KIRJOITTA...
23 / 10 %	T1	Osaan kirjoittaa itsestäni tietoja, esim. oma nimi, osoite, kansallisuus sekä hyvin lyhyitä ja yksinkertaisia viestejä, esimerkiksi postikortin.
61 / 26 %	T2	Osaan kirjoittaa lyhyitä viestejä, esim. lyhyen kirjeen ystävälle sekä kirjoittaa itsestäni, esimerkiksi perheestäni, harrastuksistani tai siitä mistä pidän ja en pidä.
81 / 34 %	T3	Osaan kirjoittaa erilaisia lyhyitä tekstejä tutuista aiheista, esimerkiksi kirjoittaa lyhyen ilmoituksen tai kertoa kirjeessä kokemuksistani tai mielipiteistäni.
55 / 23 %	T4	Osaan kirjoittaa erilaisia ja eripituisia tekstejä erilaisista aiheista, esimerkiksi kirjeen, mielipidetekstin tai muun lyhyen tekstin, jossa osaan esimerkiksi sekä esittää että perustella mielipiteeni tai kuvata selkeästi erilaisia tapahtumia.
10 / 4 %	T5	Osaan kirjoittaa erilaisia ja eripituisia tekstejä vaikeista ja monimutkaisistakin aiheista, esimerkiksi kirjoittaa virallisia kirjeitä, laajoja ja perusteltuja raportteja.
8 / 3 %	T6	Osaan kirjoittaa erityylisiä, erilaisiin tilanteisiin täysin sopivia tekstejä hyvin vaikeistakin ja minulle uusista aiheista, esimerkiksi erityyppisiä virallisia tekstejä.

Kirjoittajat esiintyvät tutkimuksessa pelkinä koodeina, joten he olivat arvioijille nimettömiä, maattomia ja tunnuksettomia. Kirjoittajien papereissa ja arviointilomakkeissa näkyi kunkin kirjoittajan koodi, joka oli viisinumeroinen numerosarja. Arvioijat eivät tunteneet koodien sisältöä, vaan ne merkitsivät arvioijille tunnistetta, jonka avulla he saattoivat tarkistaa, että suorituksen koodi ja arviointilomakkeen kyseinen kohta vastasivat toisiaan. Numerosarjassa on kolme ensimmäistä paikkaa tunnistenumeralle, neljäs paikka on itsearviolle ja viides paikka tehtävännumerolle (1–9). Esimerkiksi numerosarjassa 15929 on 159 kirjoittajan tunnistenumero, 2 itsearviointin taso ja 9 tehtävän numero (tarinasta kertominen). Kirjoittajien tekstit annettiin alkuperäisen näköisiin arvioijille, eli he lukivat alkuperäisistä otettuja kopioita.

2.2 Kirjoitustehtävien arviointi ja aineiston keruu

2.2.1 Arvioijat

Tutkimuksessa mukana olleet arvioijat ovat kaikki vapaaehtoisia, ja heille maksettiin työstä nimellinen korvaus. Noin viidellekymmenelle suomi toisena kielenä -opettajalle, arvioijalle ja/tai tutkijalle lähetettiin sähköpostiviesti, jossa heitä pyydettiin osallistumaan kirjoittamisen arviointitutkimukseen, jonka tarkoituksena oli koota maamerkistö DIALANG-nimistä arviointihanketta varten. Viestissä kerrottiin lyhyesti tutkimuksen luonteesta ja sen edellyttämästä työmäärästä. Viestiin vastanneita arvioijia oli alun perin enemmän mutta osa karsiutui ennen tutkimuksen varsinaista alkua. Ensimmäisen arviointikierroksen alkaessa arvioijia oli kaikkiaan 18, mutta toisen arviointikierroksen aikana yksi arvioijista lopetti kiireisiinsä vedoten.

Kaikki kolme arviointikierrosta suoritti 17 arvioijaa, joista yksi oli tämän tutkimuksen tekijä itse. Ensimmäisen ja toisen kierroksen haastatteluissa olivat mukana kaikki arvioijat eli 16 arvioijaa ja kolmannen kierroksen haastattelussa 15 arvioijaa. Arvioijista naisia oli 16 ja miehiä 1. Iältään kuusi oli alle 30-vuotiaita, seitsemän 31–40-vuotiaita ja neljä yli 40-vuotiaita. Kaikkien arvioijien äidinkieli on suomi. Kaikkien 17 arvioijan koulutus on korkeakoulupohjainen. Suomen kielen yliopistollisia arvosanoja arvioijat ovat suorittaneet siten, että yhdellä heistä on takanaan approbatur-opinnot, 3 on suorittanut cum laude -opinnot ja 12 laudatur-opinnot. Yksi arvioijista on valmistunut kieli-instituutista kääntäjäksi ja on suorittanut kääntämisopintoihin kuuluvia suomen kielen opintoja.

Kaikilla arvioijilla on kokemusta suomen opettamisesta vieraana kielenä, joskin kokemuksen määrä vaihteli alle vuodesta yli viiteen vuoteen. Kokeneimmat arvioijat olivat opettaneet suomea perustason ja edistyneen tason oppijoille yli viisi vuotta, ja vähiten kokenut oli opettanut pelkästään alkeistasolla alle vuoden. Eniten arvioijilla oli kokemusta alkeistason opetuksesta, sillä yli puolet arvioijista oli opettanut tätä tasoa vähintään kolme vuotta. Keskitason ja edistyneen tason opetuskokemusta oli vähintään kolme vuotta seitsemällä arvioijalla. Arvioijista, joilla oli vähemmän kokemusta suomen opettamisesta vieraana kielenä, oli kaksi vasta astumassa työelämään, yhdellä heistä oli pääasiassa kokemusta englannin opettamisesta, yhdellä alan tutkimuksesta ja yhdellä äidinkielen opettamisesta.

Kirjoittamisen arviointikokemusta on kaikilla arvioijilla. Yli puolet eli 12 arvioijaa on arvioinut kirjoittamista säännöllisesti ja 5 jonkin verran ja melko satunnaisesti. Arvioijat ovat arvioineet kirjoittamista sekä pelkin

kirjaimin (esim. approbatur–laudatur) että tasokuvatuin kirjaimin. Samoin he ovat arvioineet pistein sekä ilman kuvauksia että niiden kera. Vaikka kirjo pisteiden ja kirjaimien välillä kuvauksineen ja ilman vaihtelee arvioijittain, kaikki arvioijat ovat kolmea arvioijaa lukuun ottamatta arvioineet kirjoittamista myös ilman sanallisia kuvauksia. Yleisten kielitutkintojen arvioinnista on kokemusta 10 arvioijalla ja 2 ylioppilastutkinnon arvioinnista, toisella äidinkielestä ja toisella englannista.

Käytän tekstissä arvioijista pseudonyymeja, jotka ovat arvioijan sukupuolen mukaan keksittyjä nimiä. Tulosluvussa 5.1 käytän arvioijista myös tunnistenumeroita yhdestä seitsemääntoista. Seuraavassa listassa arvioijat on merkitty numeroittain ja pseudonyymeittäin:

1. Kaisu	5. Keskeyttänyt	9. Heidi	13. Pirkko	17. Sanna
2. Teija	6. Tiina	10. Pia	14. Martti	18. Pirjo
3. Tarja	7. Inkeri	11. Tuula	15. Merja	
4. Eeva	8. Raija	12. Päivikki	16. Tutkija	

2.2.2 Aineiston keruu

2.2.2.1 Keruusuunnitelma ja sen toteuttaminen

Tutkimuksen arvioitava aineisto koostuu kaikkiaan 248 kirjoittajan suorituksista, jotka jakaantuvat kirjoittajien ja arvioijien kesken taulukon 4 mukaisesti. Kirjoitelmien keruuta ja niiden arviointia koskevan asetelman suunnittelusta vastasi professori Fellynka Kaftandjjeva-Stoyanova Sofian yliopistosta.

Arviointiprosessin asetelmaa kuvaava taulukko 4 on jaettu kahteen osaan: vasemmalla puolella (sarakkeet TA1–TX) on kuvattu, kuinka monta kirjoittajaa suoritti kunkin tehtäväsarjan. Taulukon neljästä rivistä (TS1–TS4) kukin tarkoittaa tehtäväsarjaa (vrt. taulukkoon 1). Eniten kirjoittajia (N=101) oli kolmannessa tehtäväryhmässä, jonka tehtävät olivat esittelyteksti, mielipidekirjoitus ja kuvasta kertominen. Ensimmäisen ja neljännen tehtäväsarjan suoritti 50 kirjoittajaa ja toisen 57 kirjoittajaa. Yksittäisten tehtävien määrä voi vaihdella jonkin verran 248:sta, koska muutamat kirjoittajat eivät suorittaneet kaikkia tehtäviä.

TAULUKKO 4 Arviointiprosessin asetelma.

Tehtäväsarjat/kirjoittamistehtävät										Arvioijan tunnistenumero									
TS	TA1	TA2	TB1	TB2	TB3	TC1	TC2	TC3	TX	I	II	III	IV	V	VI	VII	VIII	IX	X
TS1	50		50						50	3	4	10	16	1	5	7	11	13	12
TS2		57					57		57	3	4	10	16	1	6	9	11	14	15
TS3					101	101			101	3	4	10	16	2	8	7	9	18	17
TS4				50				50	50	3	4	10	16	5	6	15	14	13	12

Taulukon 4 vasemmanpuoleinen osa kuvaa sitä, mitä tehtäväryhmiä kukin arvioija arvioi. Arabialaiset numerot tarkoittavat arvioijien tunnistenumeroita, esimerkiksi numero 4 eli Eeva on arvioinut kaikkien kirjoittajien suoritukset ja numero 17 eli Sanna vain kolmannen tehtäväsarjan kirjoittaneiden suoritukset. Arvioijat arvioivat vain niitä tehtäviä, joita heidän arvioimassaan tehtäväsarjassa oli. Tehtäväsarjoja jaettiin arvioijille sen mukaan, miten paljon he olivat valmiita arvioimaan. Tosin kaikilta arvioijilta edellytettiin kuitenkin vähintään sadan kirjoittajan suoritusten (noin 300 tekstiä) arviointia.

Taulukon 4 oikeanpuoleisen osan ylärivin roomalaiset numerot kuvaavat sitä, että jokainen suoritus sai arvion vähintään kymmeneltä arvioijalta. Esimerkiksi toisen tehtäväryhmän suorittaneiden kirjoittajien tuotoksia arvioivat 3, 4, 10, 16, 1, 6, 9, 11, 14 ja 15 eli Tarja, Eeva, Pia, tutkija, Kaisu, Tiina, Heidi, Tuula, Martti ja Merja. Arvioija numero 5 kuitenkin keskeytti ensimmäisen kierroksen jälkeen, joten tehtäväryhmän 1 ja 4 suoritukset saivat arvion yhdeksältä arvioijalta ja tehtäväryhmän 2 ja 3 kymmeneltä arvioijalta.

Taulukon 4 mukaista asetelmaa käytettiin kaikilla kolmella arviointikierroksella; arvioijat siis arvioivat samat suoritukset kolme kertaa. Ensimmäisen kierroksen jälkeen arvioijille ei kuitenkaan sanottu, että suoritukset tulevat olemaan samoja, mutta ei myöskään kerrottu, että ne olisivat eri suorituksia. Jokaisen arviointikierroksen alussa arvioijille lähetettiin suoritusten mukana kirje ohjeineen, arviointilomake, johon arvioijat merkitsivät tason rastiittamalla tasoa vastaavan numeron, kyselylomake ja laskutuslomake. Toisella ja kolmannella kierroksella mukana oli myös asteikko tasokuvauksineen.

Ensimmäisen arviointikierroksen arvioijat suorittivat noin kuukaudessa sen mukaan, miten paljon kullakin arvioijalla oli arvioitavaa. Arvioijat, jotka arvioivat kaikkien kirjoittajien suoritukset, tarvitsivat yli viikon enemmän aikaa kuin muut arvioijat. Arvioijilla oli mahdollisuus esittää

kysymyksiä ottamalla yhteyttä tutkimuksen tekijään, mikäli niin halusivat. He kuitenkin työskentelivät itsenäisesti. Pari arvioijaa varmisti, että he olivat ymmärtäneet ohjeet oikein ja että he voivat arvioida suoritukset omien kriteerien mukaan.

Ensimmäisen ja toisen arviointikierroksen väli oli noin kuukausi. Arviointien välinä haluttiin pitää vähintään kaksi viikkoa. Osa arvioijista tunsivat olevansa tauon tarpeessa ja osa taas halusi jatkaa mahdollisimman pian, mikä näkyi siinä, että pari arvioijaa teki toisen kierroksen heti kahden viikon jälkeen muutamassa päivässä ja osa taas piti kuukauden tauon tai teki arviointia jaksoittain pitemmällä aikavälillä.

Ajalliset erot toisen kierroksen aikana vaikuttivat myös siihen, milloin kukin arvioija aloitti kolmannen kierroksen arvioinnin. Toisen ja kolmannen kierroksen välinen tauko oli huomattavasti pitempi kuin ensimmäisen ja toisen kierroksen välinen. Pitkään väliin saattoi vaikuttaa myös se, että arvioijat tekivät työtä pitemmällä aikavälillä, tai ihan konkreettiset syyt, kuten syksyn opetuskiireet. Toisaalta on odotuksenmukaistakin, että arvioijat halusivat pitää pitemmän välin ihan senkin takia, että arviointi alkoi tuntua työlöältä ja osin tylsältäkin.

2.2.2.2 Haastattelujen tekeminen

Arvioijat haastateltiin kolme kertaa, ja haastattelut tehtiin kunkin arviointikierroksen päätyttyä. Toimin itse haastattelijana kaikissa haastatteluissa. Ensimmäisen ja toisen kierroksen jälkeinen haastattelu oli puolistrukturoitu teemahaastattelu siinä mielessä, että se oli ohjattu temaattisesti ja pääkysymysten sisältöjen osalta, mutta ei kuitenkaan kysymysten etenemisen osalta. Vaikka jokaisen haastateltavan kanssa käytiin läpi samat teema-alueet, itse haastattelukysymyksiin ja aiheisiin pureuduttiin eri järjestyksessä ja jossakin määrin eri laajuudessa eri haastateltavien kanssa (ks. Smith 1995; Kvale 1996; Eskola & Suoranta 1998). Keskustelua ohjasi pitkälti haastateltavan oma tarve raportoida kokemuksistaan ja näkemyksistään. Vaikka minulla haastattelijana olikin omat huomionkohteeni, yritin kuitenkin noudattaa haastateltavan omaa 'järjestystä': hänen tapaansa ja tarvettaan kertoa asioista ja palata niihin ehkä uudelleen eri kohdissa haastattelua.

Ensimmäisen ja toisen kierroksen haastattelurungot olivat haastateltavien seuraavanlaiset:

Ensimmäinen haastattelu (N=16):

- Miltä nyt tuntuu; millainen yleisvaikutelma sinulla on tästä arvioinnista?
- Millä tavalla lähdit liikkeelle, kun aloitit arvioinnin?

- Miten määrittelit käyttämäsi arviointiasteikon?
- Millainen on esimerkiksi taso 1?
- Millaisia ongelmia arviointiprosessiin liittyi?
 - Miten ratkaisit ne?
- Erosiko eri tehtävien arviointi jollakin tavalla?
 - Miten ja miksi?
- Millä tavalla näiden suoritusten arviointi erosi muusta tekemästäsi arvioinnista?
 - Huonompi/parempi; millä tavalla?

Toinen haastattelu (N=16):

- Mikä on yleisvaikutelmasi arvioinnista?
 - Miltä tuntuu?
 - Ehkä miksi?
- Millä tavalla lähdit liikkeelle tällä kertaa?
 - Miten ymmärsit asteikon/tulkitsit sitä?
 - Miten suoritukset istuivat tasoille?
- Millaisia ongelmia arviointiprosessiin liittyi ja miten ratkaisit ne?
- Erosiko eri tehtävien arviointi tällä kertaa?
 - Jos erosi, millä tavalla?
- Millä tavalla arviointi erosi ensimmäisen kierroksen arvioinnista?

Haastatteluissa siis käytiin läpi nämä etukäteissuunnitelman mukaiset kysymykset/aiheet mutta ei välttämättä edellä esitetyn listan mukaisesti kysymys kysymykseltä, jos asiat olivat haastateltavalle enemmän tai eri tavalla sisäkkäisiä tai jos ne tulivat hänen puheessaan eri järjestyksessä. Toisin sanoen haastattelun ensisijainen tavoite oli luoda suhde vastaajaan ja suoda haastattelulle vapaus liikkua alueilla, jotka nousivat haastattelussa esiin.

Haastattelijana vastuullani oli myös saada kysymyksiini riittävän laajat vastaukset tekemällä jatkokysymyksiä tai palaamalla takaisin johonkin asiaan, joka oli tullut esille esimerkiksi jonkin toisen aiheen mukana aiemmin keskustelussa. Moni arvioijista nosti haastattelussa esille asioita ennen kuin niitä ehdittiin häneltä kysyä, jolloin niitä oli luontevampaa käsitellä kyseisessä kohdassa. Haastateltavat saattoivat haluta puhua myös sellaisista asioista, joita en itse ollut tullut ajatelleeksi etukäteen. Haastattelijana tehtävänäni oli ennen kaikkea saada haastateltavan ääni kuuluviin.

Kahden ensimmäisen arviointikierroksen jälkeiset haastattelut pyrittiin suorittamaan niin nopeasti kuin mahdollista arviointikierroksen

päätyttyä. Kaikki arvioijat käyttivät ensimmäisellä kierroksella arviointiin melkein yhtä paljon aikaa, koska kaikkien muiden – paitsi kaikkien kirjoittajien suoritukset arvioineiden – haastattelut suoritettiin kesäkuun kahden viimeisen viikon aikana vuonna 1998. Loput ensimmäisen kierroksen haastattelut tehtiin heinäkuun ensimmäisellä viikolla. Sen sijaan toisella kierroksella arviointiin meni aikaa paljon vaihtelevammin, sillä ensimmäiset toisen kierroksen haastattelut tehtiin heinäkuun loppupuolella ja viimeiset lokakuussa. Tosin 16 haastattelusta 10 tehtiin elokuun aikana.

Yleensä haastattelut suoritettiin paikkakunnalla, jossa arvioija asui. Arviointikierroksen loppumisen ja haastattelun väli oli yleensä pari päivää, joskus vähemmänkin. Haastattelujen kesto vaihteli yli puolesta tunnista yli tuntiin. Haastattelut nauhoitettiin audiokaseteille. Olen litteroinut kaikki haastattelut karkeahkosti, millä tarkoitan sitä, että litteroinneissa on mukana kaikki haastateltavan ja haastattelijan sanoma, myös naurahdukset, epäröinnit ja palautteen annot, kuten ”mm”, ”joo”, ”aa tota”. Sen sijaan esimerkiksi taukojen pituuksia, puheen painotuksia ja äänensävyjä ei ole eritelty tai merkitty sen vuoksi, että tässä yhteydessä haastattelujen analyysissa on ensisijaista käsitysten sisältö eikä niinkään esimerkiksi kielen tarkasteleminen niiden välittäjänä.

Kolmannen arviointikierroksen haastattelu ei ollut samalla tavalla sidoksissa arvioinnin päättymiseen kuin ensimmäinen ja toinen haastattelu, koska se oli luonteeltaan erilainen. Kolmas haastattelu oli valintaverkkotyypinen haastattelu, jonka avulla pyrittiin selvittämään, millaisia konstruktia arvioijilla (N=15) on ja miten ne ohjaavat heidän päätöksentekoaan.⁹ Kolmannessa haastattelussa käytetty aineisto koostettiin toisen arviointikierroksen tulosten pohjalta. Valintaverkkohaastattelut suoritettiin pääasiassa maaliskuussa 1999. Haastattelut kestivät keskimäärin tunnin, ja ne nauhoitettiin audiokaseteille.

2.2.2.3 Mikä ohjaa haastatteluaineiston analyysia?

Aineiston analysoinnin peruseriaatteena on löytää ja kehittää sellaisia käsitteitä, jotka auttavat ymmärtämään, mistä arviointiprosessissa on kyse. Varsinaista analyysia voisi kutsua yleisesti laadulliseksi analyysiksi, jonka taustaoletusta ja metodista lähtökohtaa voisi luonnehtia etnografisfenomenografiseksi. Analyysi noudattelee seuraavia periaatteita: ensin muodostetaan pääkategorioita, sitten niiden alle kytketään niihin liittyviä

⁹ Valintaverkkohaastattelun metodia ja sen soveltamista tässä tutkimuksessa käsitellään tarkemmin luvussa 5.3.

tyologioita ja lopuksi niistä muodostetaan käsitteitä. Analyysia voisi kuvata myös tyypittelyksi, koska sen lähtöoletuksena on etsiä samankaltaisuuksia eri arvioijien välillä ja toisaalta taas arvioijakohtaisesti (vrt. Marton 1994; Syrjälä ym. 1994; Savage 2000).

Tyypittelyn taustalla on tai sitä oikeastaan ohjaa aineiston teemoittelu, joka taas nojaa lähinnä haastattelun teemoihin. Teemoittelua seuraa aineiston tyypittely sen mukaan, millaisia merkityksiä arvioijat viestivät haastattelun aikana. Toisin sanoen tyypittelyn taustalla ei ole sellaista ajatusta, että vain yleisimmät tyypit saisivat huomiota tai että vain ne olisivat kategorisoitavissa ja edelleen raportoitavissa. Myös epätyypilliset ajatukset ja käsitykset ovat mukana kategorisoinnissa.

Merkitysten tyypittelyllä pyrin luomaan kuvaa siitä, millaisia käsityksiä ja asenteita haastateltavat viestivät arvioinnista. Toisin sanoen analyysin tarkoituksena on ymmärtää sitä, mitä pelkät reliaabeliuskertoimet eivät pystyisi kuvaamaan: miten arvioijat käsittävät arvioinnin ja millaiset tekijät vaikuttavat heidän työhönsä arviointiprosessin aikana? Vaikka analyysin tuloksena syntyneekin säännönmukaisuuteen perustuvia luokkia, voi sen seurauksena ilmetä yhtä lailla poikkeuksia, toisilleen ristiriitaisia käsityksiä niin haastateltavien välillä kuin haastateltava-kohtaisestikin. Käsitykset voivat siis muuttua nopeasti, ja ne voivat olla toisilleen ristiriitaisia saman keskustelun sisällä (Marton 1994).

Tärkeintä on, että se, mitä haastateltava sanoo, on merkityksellistä hänelle ja sillä on jonkinlainen, ei välttämättä transparentti, yhteys hänen käsityksiinsä (vrt. Smith 1995; Baker 1997; Kvale 1996). Haastattelu itsessään korostaa tiedon konstruktivista ja sosiaalista luonnetta; tieto luodaan osin haastattelukeskustelun aikana, kun merkityksistä ja niiden luonteesta neuvotellaan (Kvale 1996, 11). Haastattelu on siis tavallaan pääsytie arvioijan käsityksiin, kokemuksiin ja tietoon. Merkitykset nähdään tässä eräänlaisena hermeneuttisena kehänä: osasten merkitys voi olla ristiriidassa kokonaismerkityksen kanssa ja yksittäiset merkitykset voivat muuttaa kokonaismerkitystä jne. Kehämäisyys voidaan nähdä spiraalina, joka viestii siitä, että merkityksen ymmärtämistä voi syventää jatkuvasti (Kvale 1996, 47–48).

Tämän tutkimuksen aineiston analyysin kannalta tämä tarkoittaa sitä, että haastattelulitteraatioiden ensimmäisellä lukukerralla pyrin luomaan yleisen käsityksen teksteistä. Sitten siirryin tarkastelemaan tiettyjä teemoja ja ilmauksia ja aukaisemaan niiden mahdollisia merkityksiä. Teemat liittyivät osin haastattelurungossa oleviin teemoihin ja osin asioihin, jotka toistuivat systemaattisesti joko yhden haastateltavan puheessa tai useimpien haastateltavien puheessa. Tämän jälkeen palasin takaisin kokonaismerkitykseen selvittääkseni sitä, mikä yksittäisten teemojen ja ilmausten suhde oli muun muassa lähtöoletuksiin. Sitten

tarkastelin haastateltavia uudelleen ja yritin keskittyä kaikkeen sanottuun riippumatta sen toistuvuudesta, minkä jälkeen palasin takaisin jo syntyneisiin kategorioihin (ks. liite 5). Merkityksen rakentaminen oli siis vuorovaikutusprosessi aineiston eri osasten ja tutkijan välillä. Luonnehdin haastattelujen käsittelyn perustavoitteita seuraavasti:

- Tavoitteena on kuvata, miten haastateltavien puhumat asiat jäsentävät arviointidiskurssia.
- Tarkoituksena on katsoa, mitä luokkia haastatteluaineistot sisältävät, miten ne pyrkivät sisältönsä puolesta ryhmittymään; kyse ei ole siis tutkijan ennalta hahmottamista luokista, joskin haastattelukysymysten ohjaavuus on syytä ottaa huomioon.
- Pyrkimyksenä on selittää ilmiötä, syventää ymmärrystä siitä.

Käytin ensimmäisen ja toisen arviointikierroksen haastatteluaineiston analyysissa ATLAS/ti-nimistä kvalitatiivisen aineiston käsittelyohjelmaa. ATLAS on siinä mielessä kätevä työkalu, että se mahdollistaa jatkuvan ja päivittyvän vuorovaikutuksen aineiston ja sen tulkinnan välillä. Ohjelman positiivisiin puoliin kuuluikin se, että analyysin tekijänä minun oli mahdollista täydentää ja muuttaa luokittelua helposti sekä tehdä tulkinnoistani muistiinpanoja. Myös koodia vastaavien aineistonäytteiden hakeminen, selaaminen ja päällekkäiskoodausten tekeminen ovat käteviä toteuttaa (Silvonen 1999). Kuvaan liitteessä 5 esimerkein haastattelujen analyysin eri vaiheita ja havainnollistan samalla, miten ATLAS/ti-ohjelma toimii analyysin apuvälineenä. Teemahaastattelujen analyysien tuloksia esittelen luvussa 5.2.

3 KATSAUS ARVIOINTIKÄSITTEISIIN JA –KÄYTÄNTEIHIN

3.1 Arviointia erilaisissa konteksteissa

Tämä työ käsittelee testityyppistä arviointia, mutta monet periaatteet, kuten arviointiprosessi itsessään, voisivat koskea myös esimerkiksi opetukseen liittyvää arviointia. Toisaalta käytännössä eri arviointikontekstit eroavat toisistaan kuitenkin niin paljon, että erojen esiin tuominen on edellytys kunkin arviointikontekstin tarkoituksenmukaiselle käsittelylle. Tämän alaluvun tavoitteena on kuvata kielitaidon arvioinnin roolia erilaisissa ympäristöissä ja osoittaa, miten testiarviointi suhteutuu muihin arviointikonteksteihin. Yleisesti ottaen kaikki kielitaidon arviointi liittyy läheisesti sekä kielen oppimiseen että opettamiseen tai oikeastaan se on osa sitä kiinteästi tai lyhyemmin sen mukaan, mikä on arvioinnin tarkoitus ja käyttöyhteys.

Arviointi voi olla otantapohjaista kuten survey-tyyppiset arviointitutkimukset tai yksilökeskeistä kuten testiarviointi. Otantapohjaisessa arvioinnissa arvioidaan esimerkiksi kansallista tasoa jossakin opittavassa taidossa tai verrataan eri maiden oppimissaavutuksia, minkä vuoksi arvioidut valitaan otantamenetelmällä niin, että he ovat mahdollisimman edustava joukko kokonaisuudesta tai jotenkin muuten määritellystä joukosta (ks. Ross 1988). Yksilöarvioinnissa yksilö edustaa ensisijaisesti itseään ja arvioinnin tavoite on antaa palautetta yksilölle. Arviointi voi näiden lisäksi perustua myös itsearviointiin, jolloin vastuu arvioinnista on pääasiassa oppijalla itsellään. Itsearviointin tavoitteena on syventää oppijan tietoisuutta osaamisensa vahvuuksista ja heikkouksista, mikä asettaa esimerkiksi arvioinnissa käytettäville kriteereille haasteita. Tavallisesti itsearviointia käytetään lähinnä luokkahuoneopetuksessa, esimerkiksi portfolioarvioinnissa, joskin itsearviointi voi kuulua myös testiarviointiin. Esimerkiksi DIALANGissa itsearviointi on osa oppijan kielitaidon diagnosointia.

Arviointityyppensä voidaan siis luokitella monella eri tavalla, itse jaan tässä yhteydessä arvioinnin kuitenkin kahteen pääkategoriaan: opetukseen liittyvään arviointiin ja testiarviointiin. Käytän opetukseen liittyvästä arvioinnista käsitettä luokkahuonearviointi ja testeihin liittyvästä arvioinnista käsitettä taitotasoarviointi (vrt. Huhta & Takala 1999). Kahtiajako

taitotasoarviointiin ja luokkahuonearviointiin on hyvä siinä mielessä, että se valottaa arvioinnin taustalla olevien teoreettisten viitekehysten välistä eroa ja arvioinnin suhdetta opetuksen sisältöön ja/tai sen tavoitteisiin. Taitotasoarvioinnin tarkastelun lähtökohtana on aina jossakin määrin sen mittauksellinen viitekehys, kun taas luokkahuonearvioinnin taustalla voi olla taitotasoarviointia diskurssiivisempi käsitys taidon hallinnasta esimerkiksi siten, että arviointi on osa oppimisprosessia ja oppijalla on aktiivinen osuus arviointiprosessissa (esim. Linn & Gronlund 1995; Shepard 2000). Opetukseen liittyvä arviointi palvelee ennen kaikkea oppimista ja parhaimmillaan se tukee sitä monin eri tavoin. Testiarviointi voi taas olla täysin irrallaan jostakin tietystä oppimäärästä tai kurssista; se voi tarjota lähinnä informaatiota senhetkisen osaamisen tasosta ja auttaa asettamaan jatkotavoitteita opiskelulle.

Testiarvioinnin mittauksellinen tausta näkyy muun muassa siinä, miten testin tarkoitus määritellään, miten testin luotettavuus varmistetaan, minkä verran oppijan tuotosta täytyy olla, jotta arviointi on yleensäkin mahdollista, miten mittauksen tarkkuuteen suhtaudutaan, missä määrin testitehtävien ja arvioitavan taidon välistä yhteyttä tarkkaillaan ja kehitetään, millä perusteella arviointitehtävät valitaan, miten ne pisteitetään ja miten arviointituloksia tulkitaan ja millaisia johtopäätöksiä niistä tehdään (esim. Henning 1987; Bachman & Palmer 1996; AERA ym. 1999).

Taitotasoarvioinnissa edellä mainitut kysymykset liittyvät sekä itse arviointiin että sen laadun tarkkailuun ja kehittämiseen. Tasoarvioinnissa pyritään siis vastaamaan kysymykseen, mitä oppija osaa ja voi tehdä jossakin kielenkäyttötilanteessa. Luokkahuonearvioinnissa halutaan selvittää tämän lisäksi, mitä oppija haluaa tehdä ja pystyy tekemään, mikä ei koske vain taidon hallintaa vaan myös hänen asennettaan, motivaatiotaan, persoonallisuuspiirteitään ja sosiaalista ympäristöään – ainakin opetussuunnitelmatasolla. Luokkahuonearvioinnin lähtökohta voi olla oppijakeskeisempi ja -lähtöisempi, joskin myös testiteoreettisesta näkökulmasta epäluotettavasti subjektiivisempi.

Pintatasolla kahtiajako ei ole kuitenkaan läheskään aina selvärainen, sillä esimerkiksi testiarviointi voi olla osa luokkahuonearviointia, kuten taulukosta 5 näkyy. Taulukkoon 5 on koottu taitotasoarvioinnille ja luokkahuonearvioinnille tyypillisiä arvioinnin käyttötarkeitä (ks. myös Linn & Gronlund 1995; Wiggins 1998; Huhta & Takala 1999). Kouluopetuksessa ja luokkahuonearvioinnissa on tavoiteltu monia muutoksia 1990-luvulla. Suurin muutos liittyynee yleensäkin tietokäsitykseen, opittavien asioiden konstruimiseen ja niiden hallintaan; oppimista ei tarkastella produktina vaan ennen kaikkea prosessina, joka on sekä yksilöllinen että kokemuksellinen (esim. Rauste-von Wright 1994; Segers

& Dochy 2001). Muutoksen kokeneissa luokkahuoneissa oppiminen ja arviointi ovat tiiviissä suhteessa toisiinsa; arviointi on monimuotoista, prosessin kulkua seuraavaa ja yksilöllistä (esim. Linnakylä 1994; OPH 1994; Apajalahti 1996; Frisk & Räisänen 1996; Takala 1996; Virta 1999; Shepard 2000).

TAULUKKO 5 Taitotaso- ja luokkahuonearviointi.

	Arvioinnin tarkoitus		Arvioinnin tavoite		Keinot arvioida	Kuka arvioi
Taitotaso- arviointi	1) <i>Todistuksia jakavat systeemit</i>		Taitotason, pätevyyden, kelpoisuuden osoitus		<ul style="list-style-type: none"> • testitehtävät 	<ul style="list-style-type: none"> • testausorganisaatiot • luvan tms. saaneet yksilöt
	2) <i>Diagnostinen arviointi</i>		Lähtötason, oppimongelmien yms. kartoitus, jatko-opiskelun suunnittelu		<ul style="list-style-type: none"> • testitehtävät • haastattelut 	<ul style="list-style-type: none"> • opettaja • opetuksen ulkopuolinen tah
	3) <i>Tutkimus</i>		Kielitaidon, sen oppimisen ja siihen liittyvien tekijöiden selvittely, testien kehittäminen		<ul style="list-style-type: none"> • testitehtävät • haastattelut • kyselyt 	<ul style="list-style-type: none"> • testausysteemit • tutkijat
Luokkahuone- arviointi	1) <i>Formatiivisen tiedon hankkiminen</i>	Suoritus- pohjainen/ autenttinen arviointi	Palaute oppimisen edistymisestä, oppimisvaikeuksien yms. paikantaminen	Opittujen asioiden soveltaminen ja oppimisprosessin seuraaminen	<ul style="list-style-type: none"> • harjoitusteht. • projektityöt • observointi • itsearviointi • testit/kokeet 	<ul style="list-style-type: none"> • opettaja • oppilaat
	2) <i>Summatiivisen tiedon hankkiminen.</i>		Todistusten ja arvosanojen antaminen eri vaiheissa		<ul style="list-style-type: none"> • testit/kokeet • itsearviointi, portfolio • ns. jatkuva näyttö 	<ul style="list-style-type: none"> • opettaja • oppilaat
	3) <i>Opetusohjelmien arviointi, evaluointi</i>		Opetuksen laadun ja tehokkuuden selvittely		<ul style="list-style-type: none"> • testit • kyselyt • haastattelut • itsearviointi 	<ul style="list-style-type: none"> • opettajat • vanhemmat • viranomaiset • tutkijat

Taitotasoarviointi ja luokkahuonearviointi näyttäisivätkin eroavan toisistaan erityisesti arviointikeinoiltaan: taitotasoarviointi perustuu pääasiassa testitehtäviin, kun taas luokkahuonearvioinnissa arviointikeinojen kirjo on paljon monipuolisempi. Erilaiset arviointikeinot viestivät myös siitä, että valta ja vastuu arvioinnista jakaantuvat eri tavalla: luokkahuonearvioinnissa oppijalla on huomattavasti paljon enemmän vastuuta oman osaamisensa arvioinnista. Tähän vaikuttavat tietysti ennen kaikkea näiden kahden arviointikontekstin erilaiset tavoitteet; luokkahuonearviointi

seuraa oppimista enemmän prosessina ja suhteessa aikaan ja taitotaso-arviointi taas lähinnä tietynhetkisenä tilana, jolla kuitenkin nähdään ennustettavuusarvoa.

Toisaalta näiden kahden arviointikontekstin liiallinen polarisointi saa ne näyttämään toistensa ääripäiltä, joiden välillä ei ole minkäänlaista liukuma-aluetta. Arkikeskusteluissa se näkyy muun muassa niin, että laadullisten arviointikeinojen, kuten oppimispäiväkirjojen ja portfolion, käyttö mielletään ensisijaisemmin arvioinniksi ja taitotasoarviointi taas mittaamiseksi ja vertaamiseksi. Koska mittaaminen perustuu kvantifiointiin, sitä pidetään ehkä turhan yksioikoisestikin pelkästään objektiivisuutta tavoittelevana ja luonteeltaan diskriminoivana. Testi voi kuitenkin olla myös oppijakeskeinen, tavoitella autenttisuutta ja sisältää arvioitavalle relevantteja tehtäviä. Testiin voi myös sisältyä itsearviointiosuus, joten taitotasoarvioinnissa ja luokkahuonearvioinnissa voi sekä lähtökohtien että arviointikeinojen osalta olla toisaalta paljonkin yhteistä. Lisäksi erilaisten arviointikeinojen käyttö itsessään ei tee arvioinnista tavoitteet täyttävää; pelkkä tekninen toteutus ei siis riitä.

Arviointi voidaan jakaa kahteen luokkaan myös sen mukaan, millainen vaikutus arvioinnilla on esimerkiksi oppijan valintoihin ja tulevaisuuteen (esim. Bachman & Palmer 1996; Davidson ym. 1997). Arviointi voi olla high-stakes-tyyppistä, jolloin sillä on välitön vaikutus oppijan tulevaisuuteen, kuten ylioppilastutkinnolla ja joidenkin maiden kansalaisuutta koskevilla kielitesteillä. Low-stakes-arvioinnissa, johon kuuluu muun muassa luokkahuonearviointi, yhdellä arviointitapahdumalla ei ole puolestaan niin suurta tai välitöntä vaikutusta oppijan mahdollisuuksiin esimerkiksi jatko-opiskelun kannalta.

Arvioinnin tiukka kahtiajako vaikuttaa keinoitekoiselta ja liian yleistävältä, koska arvioinnin vaikutukset ovat niin monitasoisia ja monisyisiä ei vain konkreettisten jatkomahdollisuuksien vaan myös esimerkiksi oppijan minäkuvan kannalta. Yhtä lailla esimerkiksi peruskoulun päästötodistuksella on vaikutuksensa; se perustuu edelleenkin vuosien kuluessa tehtyyn arviointiin tai siitä syntyviin vaikutelmiin, joita sitten erityisesti viimeisenä kouluvuonna tehty luokkahuonearviointi vahvistanee suuntaan tai toiseen. Tässä mielessä koulun kokeet, testit ja myös muut arviointimuodot ovat myös ainakin jossakin määrin high-stakes-arviointia (ks. myös Clapham 2000; Passman 2001).

Kaiken kaikkiaan mittaamisen periaatteet liittyvät siis taitotasoarviointiin, mutta ne ovat läsnä myös luokkahuonearvioinnissa. Mittaaminen toimii mittaamisen periaatteiden mukaisesti, joten se mikä on mittaamisen näkökulmasta tarkoituksenmukaista, ei välttämättä ole sitä, kun käytetään ei-mittaavia arviointitapoja, joita voitaisiin kutsua myös diskurssiivisiksi, prosessiluonteisiksi tai laadullisiksi. En kuitenkaan

puutu tähän rajankäyntiin seuraavissa alaluvuissa, vaan tarkastelen kieli- taidon arviointia mittaamisen periaatteiden mukaisesti.

Tarkoitan tässä tutkimuksessa mittaamisella prosessia, jossa kuvataan numeeristen kuvaajien avulla sitä, missä määrin oppijalla on joitakin taidon piirteitä (Linn & Gronlund 1995, 6). Mittaaminen vastaa siis kysymykseen: "Kuinka paljon?" Testi taas tarkoittaa systemaattista tapaa mitata taitoa kysymällä yhdenmukaisia kysymyksiä (Linn & Gronlund 1995, 6). Testi vastaa periaatteessa siinä, missä jokin muukin arviointimuoto, kysymykseen, kuinka hyvin oppija osaa jotakin.

3.2 Arviointikriteerien tarkastelua testauksen näkökulmasta

Käsittelen seuraavassa ensin arvioinnin peruskriteereitä ja sitten validiutta ja reliabeliutta, tosin kohdentaen enemmän huomiota tämä tutkimuksen kannalta olennaisempaan käsitteeseen reliabeliuteen. Tarkastelen reliabeliutta sekä käsitteenä että konkreettisena tutkimus- alueena, varsinkin arvioijaan liittyvää reliabeliutta. Lopuksi tarkastelen piirrearviointia, holvistista ja analyttistä arviointia sekä niiden mahdollisuuksia ja rajoituksia.

3.2.1 Hyvän arvioinnin peruskriteerit

Hyvän arvioinnin peruskriteereitä on useita, mutta käytetyimmät ja ehkä myös tärkeimmät ovat validius, reliabelius, vertailtavuus ja eettisyys. Lyhyesti määritellen *validius* viittaa siihen, mitaako testi sitä, mitä sen on tarkoitus mitata. *Reliabeliudella* taas tarkoitetaan mittarin eli testi- tehtävien ja niiden arvioinnin luotettavuutta toistettavuusmielessä. *Vertailtavuus* liittyy siihen, miten saman testin eri versiot ovat vertailtavissa keskenään. Testiä voidaan pitää *eettisesti hyväksyttävänä* ja *oikeudenmukaisena* (fair/equal), mikäli testin validius ja reliabelius nojaavat eettisesti hyväksyttäviin periaatteisiin ja mikäli myös testin suorittajan näkökulma on otettu huomioon. (Esim. Messick 1994; Hamp- Lyons 1997b; Kunnan 1997; AERA ym. 1999.)

Messickin (1994, 13) mukaan validius, reliabelius, vertailtavuus ja eettisyys kuuluvat kaikkeen arviointiin, koska ne eivät ole vain mittaamisen periaatteita vaan myös sosiaalisia arvoja. Niiden merkitys ja valta ulottuvat myös mittaamisen ulkopuolisiin tilanteisiin, itse asiassa kaikkialle, missä tehdään arviointiin liittyviä ratkaisuja ja päätöksiä. Tässä mielessä validius ja reliabelius eivät siis ole pelkästään psykometrisiä käsitteitä.

Muita hyvän arvioinnin kriteereitä ovat muun muassa tehokkuus ja käytännöllisyys, joilla tarkoitetaan sitä, miten laaja testin pitää olla tai miten laaja sen on järkevää olla suhteessa siihen, miten paljon aikaa joudutaan käyttämään sen suunnitteluun, järjestämiseen, pisteittämiseen ja tulosten prosessointiin (ks. esim. Weir 1993). Tehokas ja käytännöllinen tehtävä tai testipatteristo tarjoaa mahdollisimman paljon informaatiota mahdollisimman vähällä vaivalla.

Bachman ja Palmer (1996, 17–18) puhuvat myös testien hyödyllisyydestä (usefulness), jonka he määrittelevät seuraavalla tavalla:

$$\text{Hyödyllisyys} = \text{reliaabelius} + \text{käsitevalidius (construct validity)} + \\ \text{autenttisuus} + \text{vuorovaikutuksellisuus} + \text{vaikutus} + \\ \text{käytännöllisyys}$$

Heidän mielestään hyödyllisyyden laatukriteerit ovat itsenäisesti määriteltävissä, mutta ne ovat kuitenkin kiinteästi toisiinsa vaikuttavia. Arvioinnin näkökulmasta hyödyllisyys kasvaa samassa suhteessa sen kanssa, miten yksittäisten osatekijöiden vaikutusta saadaan maksimoitua. Hyödyllisyyden varmistamiseksi kielitaitotestejä pitäisi kehittää siten, että testin tarkoitus, testiin osallistujat ja kielen funktiot tulisivat otetuiksi huomioon. Bachman ja Palmer (1996, 19) painottavat kuitenkin sitä, että testin hyödyllisyyttä ei voi määritellä yleisesti kaikkia testejä koskevaksi, vaan sitä on tarkasteltava ja sen toteutumista arvioitava aina kustakin testistä ja sen käyttöyhteydestä käsin.

Arviointia ei siis tulisikaan nähdä staattisena, vaan dynaamisena prosessina, joka on helposti sovellettavissa ja uudelleen määriteltävissä (Henning 1991, 279). Päätelmiin laadukkaasta arvioinnista vaikuttavat monenlaiset tekijät, joten hyvälle arvioinnille voidaan kyllä esittää kriteereitä, mutta itse 'hyvyys' voi olla riippuvainen esimerkiksi arviointitilanteesta ja sen ajankohdasta, arviointiin liittyvistä yksilöistä ja käytetyistä metodeista. Arvioinnin dynaamisuudesta seuraa myös se, että eri kriteereiden vaikutusta on vaikea ennustaa.

Arvioinnin kriteerit ovat keskenään vuorovaikutuksessa, joskin näkökulman ja kriteerin laadun mukaan joko löyhästi tai jopa kausaalisesti. Reliaabeliudeltaan hyvä arviointi ei ole välttämättä validia mutta reliaabeliudeltaan heikko arviointi on tuskin validia ja eettistä, varsinkaan silloin kun arviot vaikuttavat oppijan mahdollisuuksiin esimerkiksi päästä opiskelemaan haluamaansa oppilaitokseen.

Arviointikriteerien täyttämisen näyttäisi siis olevan monimutkaista ja käytännön näkökulmasta katsottuna edellyttävän kompromissien tekemistä. Mistä ja kenen näkökulmasta arviointi on hyödyllistä ja miten reliaabeliuden takaaminen voi onnistua muuttuvissa olosuhteissa?

Katsoipa arviointia lähes mistä näkökulmasta tahansa, sen käsitteellinen määrittely ja empiirinen kuvaaminen kietoutuvat aina tavalla tai toisella validiuteen ja reliaabeliuteen, jotka esiintyvät poikkeuksetta arvioinnin tai testauksen kehittämistä koskevissa yhteyksissä (ks. esim. Weir 1993; Messick 1994; Alderson ym. 1995; McNamara 1996).

Kriteereistä eettisyyden voi taas nähdä kaiken leikkaavana yleiskriteerinä, jonka pitäisi vaikuttaa testausprosessin kaikissa vaiheissa, niin taidon operationaalistamisessa, tehtävien laatimisessa ja pisteittämisessä kuin testien tutkimuksessakin (Kunnan 1997, 94–95). Itse testin ja sen eri vaiheiden tutkiminen ei kuitenkaan riitä, vaan eettisyyden ja oikeudenmukaisuuden tavoittelemisen edellyttää myös testin aiheuttamien seurausten huomioon ottamista ja kriittistä arviointia (Lynch 1997, 316–317).

Eettisyyteen liittyy läheisesti myös vastuullisuuden käsite (accountability), joka tarkoittaa sitä, millaisia seurauksia testeillä on yksilölle ja miten testisysteemi tai koulutusviranomaiset ottavat seurauksista vastuun (Norton 1997, 314–315; Linn 2000, 4). Käytännössä eettisyys näkyy muun muassa siinä, käytetäänkö testiä tietoisesti esimerkiksi koulutuksellisista tai taloudellisista syistä kontrollointikeinona, miten validiutta tarkastellaan, millaisia tulkintoja tuloksista tehdään tai miten testin suorittajien yksilöllisyys ja heiltä tuleva palaute otetaan huomioon. Eettisyys ei myöskään rajoitu pelkästään arviointikontekstiin ja sen välittömiin seurauksiin, vaan se liittyy yhtä lailla arvioinnin kaikkinaisiin yksilöllisiin ja yhteisöllisiin vaikutuksiin, jotka voivat liittyä välillisestikin itse arviointiin.

3.2.2 Validius: arvioidaanko sitä, mitä on tarkoitus arvioida?

Validiuden tarkastelussa on ensisijaisesti kyse tulkintojen validiudesta: emme siis pääse käsiksi itse validiuteen, vaan erilaiset tutkimustulokset ja niistä tehtävät johtopäätökset testin validiudesta ovat kukin erilaisia tulkintoja siitä. Validius on siis varsin monikasvoinen tulkinnallisena käsitteenä: itse asiassa lähes kaikki materiaali, joka kerätään testin kehittämiseen tai sen käyttämiseen liittyen, on jollakin tavalla yhteydessä validiuteen (Anastasi 1986, 3; Angoff 1988, 29). Hieman kapeammin katsottuna validiutta arvioidaan suhteessa pistemäärien ja arvioiden pohjalta tehtyihin tulkintoihin eli siihen, missä määrin teoria ja aineisto tukevat pistemääristä tehtyjä tulkintoja (Messick 1989, 13; AERA ym. 1999, 11).

Kirjoittamisen arvioinnissa validius viittaa esimerkiksi siihen, missä määrin mittaamisvälineet, joita ovat kirjoitustehtävät, asteikkokuvaukset, numeeriset asteikot tai muut pisteitystavat ja niiden käyttö, mittaavat sitä

kirjoittamisen taitoa, mitä niiden oletetaan mittaavan (esim. Henning 1991). Kirjoittamistehtävät voivat tosin tarjota erilaisia validiustulkintoja eri tarkoituksissa esimerkiksi siten, että saman tehtävän ja sen suorittamisesta annetun arvion validius voi vaihdella sen mukaan, millaisia tulkintoja arviolle halutaan antaa.

Validiutta voidaan siis tarkastella eri näkökulmista ja tarkastelukulmien mukaisesti sitä on useampaa eri tyyppiä, esimerkiksi *käsitevalidius* (construct), *sisältövalidius* (content), *ilmeisvalidius* (face), *seurausvalidius* (consequence) ja *ennustevalidius* (predictive).¹⁰ Vuosien mittaan validiuskäsitteiden määrä on kasvanut huomasti. Angoff (1988) viittaa validiuden historiaa käsittelevässä artikkelissaan kuuteentoista validiuden alakäsitteeseen (ks. myös Cumming 1996). Eri validiustyypit voidaan nähdä erilaisina metodeina tarkastella ja arvioida validiutta (esim. Anastasi 1986; Angoff 1988; Messick 1989; Bachman 1990; AERA ym. 1999).

Vaikka validiuden eri tyyppjä on käsitelty psykometrisessa kirjallisuudessa aina 1930-luvulta lähtien, näyttäisi *käsitevalidius* kuitenkin saaneen eniten huomiota. Itse asiassa käsitevalidiuden vakiinnuttua 1950-luvulla sitä on alettu pitää eräänlaisena validiuden pää- tai peruskäsitteenä (Anastasi 1986, 4; Angoff 1988, 26). Käsitevalidius sisältää muut validiuskäsitteet tai on niille eräänlainen kattokäsite. Käsitevalidiudessa on tavallaan kyse koko arvioinnin perusolemuksen tarkastelusta.

Mitä käsitevalidiudella sitten oikeastaan tarkoitetaan? Ebelin (1979, 307) mukaan käsitevalidiudessa on kyse siitä, miten testin tulokset tai arviot kuvaavat hypoteeseja taidosta, joita oletamme niiden kuvaavan. Mitä paremmin arviot vastaavat oletuksia, sitä validimpi testi on. Ebel ei kuitenkaan pidä käsitettä (construct) yksinkertaisen mitattavana ilmiönä vaan osana ihmisen käyttäytymistä, joka ei ole koskaan suoraan mitattavissa tai observoitavissa. Erilaiset kielitaitotehtävät kattavat vain jonkin osan taidosta, ja tehtävien perusluonteeseen kuuluu, että ne voivat toimia eri tavalla eri aikoina ja eri oppijoiden kohdalla.

Joidenkin tutkijoiden mukaan käsitevalidius viittaa lähemmin siihen, missä määrin testi edustaa sen taustalla olevaa teoriakäsitystä eli kielitaitokäsitystä (esim. Henning 1987, Bachman & Palmer 1996; AERA

¹⁰*Sisältövalidius* viittaa siihen, kuinka riittävän kattava ja edustava testin sisältö on suhteessa siihen, mitä sen oletetaan mittaavan. *Ilmeisvalidius* (face) liittyy siihen, miten uskottavaksi testin suorittaja kokee testin. Tämä validiuden muoto on sikäli impressionistinen, että sitä ei voida osoittaa millään korrelaatiokertoimella, koska siinä on kyse lähinnä hyväksyttävyydestä (acceptability). *Seurausvalidius* (consequence) koskee nimensä mukaisesti arvioinnin seurauksia yksilölle ja liittyy myös arvioihin arvioinnin eettisyydestä. (Angoff 1986; Henning 1987; Weir 1990.)

Ennustevalidius tarkoittaa sitä, miten luotettavasti esim. sisäänpääsykoe ennustaa selviytymistä opinnoista tai kielitaitotesti tietyistä kielenkäyttötilanteista. Ennustettavuutta kuvataan korrelaatiokertoimilla. (Angoff 1986; Henning 1987.)

Käsitevalidiutta käsitellään itse tekstissä myöhemmin.

ym. 1999). Bachman ja Palmer (1996, 261–262) määrittelevät käsitteen (construct) siten, että se tarjoaa pohjan tai perustuksen testille tai testitehtävälle ja myös tulkintatavan pistemäärille. Mittari eli tehtävät ja arviointiaisteikot itsessään eivät vielä riitä käsitevalidiuden osoittamiseksi, vaan siihen tarvitaan pistemääriä (scores). Yhteys arvioiden ja käsitteen välillä on olennainen.

Tosin arviotkaan eivät heijastele ainoastaan tai suoraan käsitettä vaan myös muita tekijöitä, esimerkiksi testin suorittajan fyysistä olotilaa, liian lyhyttä vastausaikaa tai sitä, miten suorituksen arvioija on tulkinnut tehtävän ja sen taustalla olevan taitokäsityksen ja mukauttanut oman käsityksensä sen mukaiseksi. Tämän vuoksi käsitevalidiutta rakennettaessa ja varmistettaessa on tärkeää pystyä minimoimaan ulkopuolisten tekijöiden vaikutus pistemäärään.

Vaikka validius itsessään olisikin staattinen, siitä tehtävät tulkinnat eivät sitä ole. Validiustulkinnat eivät siis ole yksi yhteen sama asia kuin käsitevalidius ominaisuutena. Tässä mielessä validius on enemmänkin prosessinomainen, se on ikään kuin alati liikkeessä (ks. esim. Anastasi 1986, 4). Samaa tietysti voi olettaa itse taidostakin, sillä se voi olla käsitteellisellä tasolla mallinnettavissa taitokomponenteiksi ja jonkinasteiseksi edistymishierarkiaksi, mutta käytännössä taito tai siitä tehtävät tulkinnat elävät jatkuvasti yksilöittäin ja yksilölläkin monien muuttujien vaikutuksesta, esimerkiksi vireystilan, kielenkäyttötilan ja -ajan, tunnetilan ja kokemusten mukaan.

Validiutta on aina pidetty psykometriikassa tärkeimpänä ja keskeisimpänä käsitteenä (Angoff 1988, 19). Validius ei kuitenkaan yksinään riitä, vaan sen eriasteinen olemassaolo edellyttää testitehtävien reliabeliutta. Jos esimerkiksi arvioijan tulkinnan ja mittarin taustalla olevan konstruktin välillä on ristiriita, testitehtävän laatu ei viesti testitehtävän validiudesta vaan ensisijaisesti arvioinnin reliabeliudesta. Arviointikeskustelukaan ei ole siis välttynyt muna-kanakeskustelulta: kumpi, käsitevalidius vai reliabelius, on suhteessa toiseen ensisijainen tai voiko kumpikaan toteutua, jos toinen ei toteudu (ks. esim. Moss 1994)?

Käytännössä reliabeliuden ja validiuden yhteys näkyy esimerkiksi siinä, mitä tekijää tai tekijöitä arvioija painottaa arviota antaessaan. Miten hyvin arvioija seuraa kirjoitustehtävän taitokäsitystä, jos hän esimerkiksi perustaa päätöksensä pääasiassa kielen rakenteiden hallintaan, vaikka se ei olisikaan olennainen kriteeri kirjoitustehtävästä suoriutumisen kannalta tai rakenteiden hallinnan painoarvo ei olisi niin vahva arviointikriteereissä? Tällaisessa tapauksessa syy ei olekaan tehtävässä tai siinä, että se ei mittaa sitä, mitä sen oletetaan mittaavan. Ristiriita on arvioijan ja mittarin/suorituksen välillä, joiden näkemykset eroavat toisistaan. Tässä mielessä käsitevalidiutta ja reliabeliutta ei voida erottaa

toisistaan, vaan ne liittyvät kiinteästi toisiinsa ja niiden erotteleminen toisistaan voi olla keinotekoisia (ks. esim. Moss 1994, 7).

Viime vuosikymmenen aikana pohdiskeleva keskustelu validiuden ja reliiabieliuden suhteesta ja keskinäisestä painoarvosta arviointitutkimuksessa on lisääntynyt. Lisäksi 1990-luvulla on kirjoitettu paljon validiuskäsitteestä siltä kannalta, miten testit vaikuttavat opettajiin ja oppilaisiin (ks. esim. Alderson & Wall 1993; Messick 1996; Hamp-Lyons 1997a; Shohamy 1993 ja 1997). Tätä vaikutusta kutsutaan takaisinvaikutukseksi (washback). Testien vaikutus ilmenee muun muassa siinä, mitä asioita ja millä tavalla opetetaan ja mitä asioita oppijat motivoituvat oppimaan (esim. Shohamy 1992, 513; Alderson & Wall 1993, 117, 121). Takaisinvaikutus voi olla sekä negatiivista että positiivista (esim. Shohamy 1992; Messick 1996; Bailey 1997).

Negatiivisimmillaan varsinkin high-stakes-tyyppiset testit voivat ohjata ja kaventaa opetusta merkittävästi; arviointia voidaan myös käyttää pahimmillaan vallan välineenä, esimerkiksi uhkailumielessä (ks. esim. Aldersson & Wall 1993; Messick 1996; Shohamy 2001a). Kommunikatiivisilla ja autenttisilla testeillä on taas toivottu olevan positiivinen vaikutus varsinkin siihen, mitä opetetaan. Esimerkiksi kielten opetuksessa sen jälkeen, kun puhumista on alettu arvioida testeissä, sen opetus on myös lisääntynyt tai jopa saanut vasta alkunsa.

Takaisinvaikutus ei ole kuitenkaan näin yksiselitteinen validiuskäsitteenä esimerkiksi Aldersonin ja Wallin (1993, 116) ja Messickin (1996, 242) mukaan, koska takaisinvaikutuksen toteutuminen edellyttää, että testi saa opettajat ja oppilaat tekemään sellaisia asioita, joita he eivät muuten tekisi. He painottavat myös tutkimuksen, empiirisen aineiston, tärkeyttä, jotta tietäisimme paremmin, miten takaisinvaikutus toimii. Joissakin yhteyksissä takaisinvaikutus liittyy läheisesti eettisyyden ja oikeudenmukaisuuden käsitteisiin ja sillä viitataan epäsuoremmin yleensä testauksen aiheuttamiin vaikutuksiin riippumatta siitä, tehtäisiinkö kyseisiä asioita ilman testausta vai ei (ks. Bailey 1997).

Validiuskäsitteiden määrään ja määrittelyjen moninaisuuteen viitaten toinen kysymys on tietysti se, tarvitaanko opetuksen ja arvioinnin välisen suhteen kuvaamiseen omaa validiuskäsitettä. Tosin validiuskäsitteiden lukumäärä kertoo sekin jotakin; todistaaksemme testin olevan uskottava emme selviä yhdellä validiuskäsitteellä emmekä myöskään yhdellä näkökulmalla. Se, mikä on validia yhdestä näkökulmasta, ei välttämättä olekaan sitä toisesta. Arvio tai oikeaan osunut tulkinta arvioinnin validiudesta edellyttää argumentointia, jossa monet eri näkökulmat ovat relevantteja.

Arviointi näyttäisi olevan prosesseiltaan niin monimutkaista ja monitasoista ja seurauksiltaan vaikuttavaa, että sitä täytyykin lähestyä

useammasta näkökulmasta, jotta arvioinnin monitahoisuus tulisi huomioon otetuksi. Validiuden tarkastelu kantaa aina mukanaan tietynlaista ihmis-, taito- ja tietokäsitystä. Olisi hyvä tunnistaa, tuntea ja tunnustaa ne arvot ja vaikuttimet, jotka ohjaavat tutkimusten taustalla olevaa ajattelua, toimintaa ja valintoja. Ei voida myöskään ohittaa yksilöllisiä ja yhteiskunnallisia seuraamuksia, koska arvioinnin vaikutukset ovat pelkkää testikontekstia tai yhtä arviointikertaa laajemmat. Esimerkiksi suomi toisena kielenä -arvioinnissa tulisi suhtautua harkiten ja vakavasti arviointiin, johon nojautuen tehdään yksilön kannalta ratkaisevia päätöksiä.

Cronbachin (1990, 185–189) mukaan olisikin parempi puhua evaluatiivisesta perustelusta kuin validointitutkimuksesta. Testin arvioita (score) tulisi siis tarkastella mahdollisimman monesta näkökulmasta, minkä vuoksi tutkimuksessa pitäisi tukeutua sekä kvantitatiiviseen että kvalitatiiviseen aineistoon. Perustelu- ja tulkintaluonteiseen tarkasteluun viittaisi sekin, että validointiprosessin kannalta kaikenlainen aineisto on tärkeää, myös ristiriidat esimerkiksi laadullisen ja numeerisen aineiston välillä (esim. Angoff 1988; Luoma 2001). Testin käsitevalidointi on jatkuvaa vuorovaikutusta, sillä mikään tulkinta arviosta ei ole lopullinen tai vakiintunut.

Millainen taitokäsitys on sitten tässä tutkimuksessa käytettävien kirjoitustehtävien taustalla? Tässä yhteydessä kirjoittaminen taitona määritellään pääosin Euroopan neuvoston julkaiseman yleiseurooppalaisen viitekehyksen¹¹ mukaan, koska sitä käytetään DIALANG-hankkeen perustana ja tutkimuksessa käytetyt tehtävät on laadittu sen perusteella laadittujen ohjeiden mukaan. Myös tehtävien arvioinnissa käytetty asteikko perustuu samaan viitekehykseen. Luvussa 4.1.5 käsitellään yksityiskohtaisemmin sitä, millaista käsitystä kirjoittamisen taidosta perusteet ja asteikko heijastelevat ja millaiseen kirjoittamisen taitokäsitykseen DIALANG pohjautuu.

3.3. Reliaabelius: miten johdonmukaista ja yhdenmukaista arviointi on?

3.3.1 Yleistä reliaabeliudesta

Jokapäiväisessä elämässä reliaabeliuden käsite liittyy läheisesti toistettavuuteen ja sen seurauksena syntyvään odotuksenmukaisuuteen. Esimerkiksi leipominen sujuu koko lailla odotuksenmukaisesti, jos leipoja

¹¹ Yleiseurooppalainen viitekehys on käänös alkuperäisestä Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference -asiakirjasta.

ei poikkea tutuksi tulleen reseptin työvaiheista eikä muuta aineiden määrää totutusta. Hiustenkuivaaja, joka käynnistyy vain joka toinen kerta ilman, että sitä täytyy hakata lavuaarin reunaan, ei taas ole johdonmukaisuudessaan luotettava. Sitä voi pitää johdonmukaisesti toimivana kuivaajana, jos se käynnistyy poikkeuksetta joka käyttökerta.

Reliaabelius ei kuitenkaan välttämättä edellytä täydellistä käyttäytymistä tai toistettavuutta, vaan siinä voi olla erilaisia tasoja: jokin voi olla johdonmukaisempaa tai luotettavampaa kuin jokin toinen, vaikka se ei olisikaan 'täydellistä'. Esimerkiksi kakku voi olla matalampi kuin edellisellä kerralla mutta silti maistua yhtä hyvältä, tai oppija, joka kirjoittaa kymmenestä sanasta johdonmukaisesti kahdeksan oikein, on parempi oikeinkirjoituksessa kuin sellainen kirjoittaja, joka kirjoittaa lähes poikkeuksetta neljä sanaa kymmenestä oikein (Traub & Rowley 1990, 171).

Reliaabelius viittaa siis mittaamisen johdonmukaisuuteen. Sillä tarkoitetaan arviointikontekstissa sitä, miten toistettavaa yksilön tai ryhmän arviointi on eri aikoina saman testin tai sen eri tehtäväversioiden perusteella (esim. Ebel 1979; Henning 1987; Feldt & Brennan 1993; Linn & Gronlund 1995). Reliaabelius ilmenee lähinnä pistemääristä tehtävien analyysien ja niiden tulkintojen johdonmukaisuutena (Crocker & Algina 1986, 105; Feldt & Brennan 1993, 105–107). Se voidaankin liittää mitattavien pistemäärien tarkkuuteen, johdonmukaisuuteen, riippuvuuteen ja myös oikeudenmukaisuuteen (Henning 1987, 74; Cronbach 1996, 257–258).

Tulkinnat reliaabeliudesta ovat pistemäärien kautta yhteydessä sekä testitehtäviin että suoritusten arvioijiin. Testitehtävien reliaabelius tarkoittaa sitä, missä määrin tietyt tehtävät mittaavat taitoa samalla tavalla eri aikoina. Arvioijiin liittyvä reliaabelius viittaa taas siihen, miten yhdenmukaisesti arvioijat keskenään arvioivat ja miten johdonmukaisesti kukin arvioija arvioi. Jos testitehtävät mittaavat tarkoituksesta eri taitoa eri testikertoina tai jos arvioijat arvioivat subjektiivisten mielty mysten mukaan, arvioinnin tuloksia ei tietenkään voida pitää luotettavina.

Nämä kaksi reliaabeliusnäkökulmaa eivät kuitenkaan ole välttämättä toisistaan riippuvaisia. Jos arvioijien välinen reliaabelius on korkea, se ei tarkoita sitä, että myös arviointitehtävien reliaabelius on korkea tai toisin päin (Ebel 1979, 277). Arviointi ei ole myöskään luotettavaa yleensä, vaan se on sitä suhteessa ajanjaksoon, eri tehtäviin tai arvioijiin (esim. Feldt & Brennan 1993; Linn & Gronlund 1995). Tulkinnat reliaabeliudesta perustuvat käsitykseen, jonka mukaan yksilön ja ryhmien käyttäytyminen on vakaata ja ainakin jossakin määrin ennustettavaa. Näin ollen voidaan olettaa, että sekä yksilön että ryhmän käyttäytymisen mittaaminen

voidaan toistaa lähes samanlaisena eri mittauskertoina, kun mittaukseen vaikuttavat häiriötekijät otetaan huomioon.

Erot yksilön käyttäytymisessä, suorituksissa ja vastauksissa siis vaihtelevat tilanteesta toiseen, vaikka olosuhteet olisivat täysin kontrolloituja. Joidenkin yksilöiden arviointituloksissa voi olla vähemmän variaatiota kuin joidenkin toisten. Kenenkään arviointitulokset eivät ole kuitenkaan täysin johdonmukaisesti samanlaisia eri arviointikertoina. (Bachman 1990; Feldt & Brennan 1993; Linn & Gronlund 1995; AERA ym. 1999.) Vaihtelu liittyy yksilön sisäisen vaihtelun lisäksi eri yksilöiden välisistä eroista johtuvaan vaihteluun, jos reliabeliutta tarkastellaan esimerkiksi ryhmäkohtaisesti. Vaihtelu ei koske vain arviointitehtävien suorittajia vaan yhtä lailla suoritusten arvioijakin. Johtuipa vaihtelu millaisesta tekijästä tahansa, tulokset heijastelevat sitä eli se näkyy nimenomaan arvioinnin tuloksessa eli pisteissä tai tasoarviossa.

Vastaus siihen, millä tavalla reliabeliutta sitten mitataan, ei ole yksioikoinen; reliabeliuden määritelmiä on useita. Klassisessa testiteoriassa arviointitulos sisältää käsitteellisesti virhekomponentin, minkä mukaan on mahdollista olettaa, että arviointiin sisältyy hypoteettinen virheetön arvio. Tätä virheetöntä arviota kutsutaan yksilön todelliseksi pistemääräksi (true score), jonka yksilö voisi saada, jos mittaaminen ei sisältäisi ollenkaan virhearvioita (ks. esim. Crocker & Algina 1986, 109; Henning 1987, 74; AERA ym. 1999, 25).

Tilastollisena terminä todellinen pistemäärä tarkoittaa yksilön henkilökohtaista parametria, jota sitten tehdyt arviot, esimerkiksi pisteet tai taitotasokuvaukset, edustavat tai kuvaavat (esim. Feldt & Brennan 1993, 106; AERA ym. 1999, 25). Spearmanin viime vuosisadan alussa kehittäneen ja siitä edelleen elaboroidun klassisen mallin mukaan mitkä tahansa testistä saadut pistemäärät voidaan jakaa kahteen hypoteettiseen komponenttiin eli $X = T + E$, jossa X on saatu pistemäärä, T todellinen pistemäärä ja E satunnainen virhekomponentti (Crocker & Algina 1986, 107). Virhekomponentti voi olla joko positiivinen tai negatiivinen, esimerkiksi vastauksen oikein arvaaminen olisi positiivinen ja kysymyksen väärin lukeminen negatiivinen.

Klassisen testiteorian todellista pistemäärää on joissakin yhteyksissä kritisoitu siitä, että sitä pidetään absoluuttisena arviona tai että sillä väitetään pystyttävän kuvaamaan oppijan todellista osaamista. Crockerin ja Alginan (1986, 109) mukaan on tärkeää tehdä ero psykologisen ja fyysisen tai biologisen muuttujan välillä. Esimerkiksi diagnoosi jalan murtumisesta on absoluuttinen sinä mielessä, että jalka joko on poikki tai ei ole. Tosin tämäkin diagnoosi tai niin kutsuttu todellinen pistemäärä voi sisältää virhekomponentin, jos esimerkiksi röntgenkuvia tulkitaan eri tavalla eri mittauskertojen yhteydessä.

Sen sijaan psykologinen todellinen pistemäärä on täysin riippuvainen käytetystä mittausprosessista (Crocker & Algina 1986, 110; Feldt & Brennan 1993, 106). Itse asiassa mikä tahansa mittausväline tai -tapa voi saada useanlaisia reliabeliuskertoimia, koska olosuhteet, muuttujat, muuttuvat mittauskerrasta toiseen (Feldt & Brennan 1993, 106–107). Tämän tulkinnan mukaan kritiikki todellisen pistemäärän absoluuttisuudesta on turhaa, koska todellinen pistemäärä on yksinkertaisesti määriteltävissä henkilön keskimääräisesti saamaksi pistemääräksi eri aikoina toistetuista testeistä (Crocker & Algina 1986, 110; Traub & Rowey 1990, 173). Todellisessa pistemäärässä ei ole siis kyse absoluuttisesta arviosta vaan keskimääräisestä arviosta.

Klassisesta testiteoriasta eriytyvän paradigman, jota voitaneen kutsua moderniksi testiteoriaksi, mukaan reliabelius viittaa mittariin arvioinnin työkaluna eikä sinänsä itse mittariin (ks. esim. Linn & Gronlund 1995). Kuten validiudenkin kohdalla, tässäkin on kyse lähinnä reliabeliustulkinnoista. Niinpä modernien testiteoreetikkojen mukaan mille tahansa mittausvälineelle voidaan laskea useita erilaisia reliabeliusarvoja mitattavan ryhmän ja mittausolosuhteiden mukaan. Reliabeliusarvot voivat vaihdella jopa sellaisessakin tapauksessa, jossa mitattavien joukko on tarkasti rajattu. (Feldt & Brennan 1993, 107.)

Arviointi ei ole siis koskaan *yleensä* reliabelia, vaan se on sitä suhteessa esimerkiksi tiettyyn aikaan, erilaisiin tehtäväyhdistelmiin tai tiettyihin arvostelijoihin. Jos arviointi on johdonmukaista esimerkiksi tehtävien suhteen, se ei ole sitä välttämättä arvostelijoiden tekemän arvioinnin kannalta. Niinpä reliabeliudessa on kyse, kuten jo edellä kävi ilmi, johdonmukaisuudesta tietyn komponentin osalta. Esimerkiksi jos arvioinnin tehtävänä on diagnosoida oppijan lähtötasoa, silloin kiinnitetään huomiota erityisesti tehtävien reliabeliuteen. Saavutusarvioinnissa opitun lisäksi on tärkeää myös aikasuhte eli se, miten oppiminen on edistynyt suhteessa aikaan.

Testiteoriassa reliabelius nähdään ensisijaisesti tilastollisena terminä eli ennen kaikkea mitattavana ilmiönä. Oletukset ja päätelmät esitetään matemaattisin käsittein, ja reliabeliuden tarkastelu on mahdotonta – päinvastoin kuin validiuden – ilman numeerista testidataa (Dunn 1989, 235; Feldt & Brennan 1993, 83). Yksinkertaisesti kyse on siitä, miten yksimielisesti tulkitaan tuloksia, jotka on saatu samalla menettely- tai mittaustavalla. Arvioijien on siksi arvioitava samojen oppijoiden suorituksia tai tehtävien on oltava ainakin osittain samoja ja tietyn oppijaryhmän suorittamia, jotta niiden johdonmukaisuuden tarkastelu olisi mahdollista.

Edellä esitettyä tilastollisiin teorioihin ja käsitteisiin perustuvaa reliabeliutta pidetään ennen kaikkea psykometrisena entiteettinä (esim.

Moss 1994, 7; Lumley 2000, 51). Missä määrin sitten esimerkiksi arvot ja eettisyys kuuluvat reliabeliuteen? Miten mitattava ilmiö reliabelius oikeastaan on? Tai miten prosessiluonteisuus liittyy siihen? Onko näiden kysymysten esittäminen testiteoreettisesta näkökulmasta katsottuna epätarkoituksenmukaista? Mikäli kysymyksiä halutaan käsitellä, on tehtävä ero psykometrisen ja hermeneuttisen reliabeliuden välillä. Hermeneuttinen näkökulma ottaa arviointituloksen lisäksi huomioon myös prosessin, jonka tutkimiseen tarvitaan myös kvalitatiivisia metodeja.

Suurimmat erot psykometrisen ja hermeneuttisen reliabelius-käsityksen välillä liittyvät siihen, miten näkökulma suhtautuu toisaalta arviointiprosessissa vaikuttavien tekijöiden keskinäisiin suhteisiin ja toisaalta taas siihen, miten kokonaisuus otetaan huomioon ja miten arvioija saadaan tai voidaan ohjata päätyämään haluttuun lopputulokseen. Psykometrinen lähestymistapa painottaa sitä, että arvioijat arvioivat suoritukset itsenäisinä ilman taustatietoa kirjoittajista. Itse asiassa kaikki reliabeliuden analysointiin tarvittava informaatio sisältyy arvioon tai pistemäärään, joka viittaa kriteeriin tai johonkin normiryhmään. (Moss 1994.) Tällöin oletetaan, että arvioijat toimivat ohjeiden mukaisesti ja jakavat saman käsityksen mittarin kanssa.

Hermeneuttinen näkökulma on holistisempi ja pyrkii ymmärtämään kokonaisuutta sen osasten valossa. Arvioinnin ei tarvitse olla riippumatonta eikä sitä tarvitse karsia muuttujista. Monilla arviointiin vaikuttavilla tekijöillä on informaatioarvonsa. Alustava erimielisyys arvioijien välillä ei sekään invalidisoi arviointia, vaan se voi tarjota hedelmällisen sysäyksen keskusteluun tai väittelyyn, joka voi taas lisätä ymmärrystämme siitä, mistä ja miten suorituksesta syntyvä vaikutelma eli arvio muotoutuu. (Moss 1994.) Tällaista keskustelua käytäneenkin monien testisysteemien arvioijakoulutuksissa, ja se on usein hedelmällinen tapa etsiä yhteistä linjaa arvioijien välille.

Hermeneuttisen näkökulman mukaan reliabeliutta, kuten validiutakin, voidaan tarkastella epistemologisena ja eettisenä kysymyksenä kvalitatiivisten metodien avulla. Käytännön kysymyksenä eettisyys liittyy muun muassa siihen, missä määrin me voimme yleistää esimerkiksi kirjoittamisen taidon muutamien oppijoiden kirjoittamien tekstien ja niistä tehtyjen arviointien perusteella niin, että arviointi olisi oikeudenmukaista (Moss 1994). Eikä reliabeliuskaan riitä eikä se takaa oikeudenmukaisuutta, mikäli tehtävät eivät ole hyväksyttävästi valideja (esim. Linn & Gronlund 1995).

Klassisen testiteorian mukaan validiuden mahdollinen estimaatti on sitä suurempi, mitä suurempi on reliabeliuden estimaatti (Henning 1987, 90). Toisaalta heikko reliabeliuden estimaatti implikoi kyllä heikkoa

validiuden estimaattia, mutta korkea reliaaбелиuden estimaatti ei kuitenkaan varmista korkeaa validiuden estimaattia (Linn & Gronlund 1995, 82). Validiuden ja reliaaбелиuden keskinäisestä suhteesta on esitetty erilaisia näkemyksiä: se on nähty muun muassa sillä tavalla ehdottomaksi, että validiuden olemassaolo edellyttää reliaaбелиutta, joskaan reliaaбелиus ei ole riittävä peruste validiuden olemassaololle (Henning 1987, 89).

Asetettaessa laadukkaan arvioinnin kriteereitä järjestykseen validius tulee yleensä kriteerilistassa ensimmäisenä ja reliaaбелиus toisena (ks. esim. Linn & Gronlund 1995). Näiltä osin käsitteiden keskinäinen järjestys näyttäisi olevan melko selvä. Toisaalta on vaikea sanoa, miten mielekästä väittäely näiden kahden käsitteen ensisijaisuudesta on, koska molemmat kriteerit ovat kiinteässä yhteydessä siihen, miten onnistuneesti arviointi vastaa sille asetettuja vaatimuksia. Keskustelua kriteerien ensisijaisuudesta voidaan käydä loputtomiin. Tärkeintä lienee kuitenkin arvioinnin päämäärä eli se, minkä vuoksi oppijoiden taitoa arvioidaan ja miten arviointi palvelee käyttötarkoitustaan.

3.3.2 Miten arvioijiin liittyvää reliaaбелиutta arvioidaan?

Arvioijien työtä kuvaavaa reliaaбелиutta on mitattu yleisimmin yksimielisyyden prosentteina (percentage of agreement) tai korrelaatioina (ks. esim. Crocker & Algina 1986; Engelhard 1996; Burry-Stock & Shaw 1996; Hayes & Hatch 1999). Molemmilla pyritään kuvaamaan sitä, missä määrin tulosten suhteellinen järjestys on sama arvioijasta toiseen (Linn & Gronlund 1995, 91). Yksimielisyyden prosentti näyttäisi ainakin aikaisemmin olleen yleisempi tapa tarkastella reliaaбелиutta. Yhtenä syynä sen taajaan käyttöön on pidetty sitä, että se on käsitteellisesti yksinkertaisempi ja laskennallisesti helpompi kuin korrelaatio.

Yksimielisyysprosentin käyttöä on kuitenkin kritisoitu siitä, että se ei osoita reliaaбелиutta teoreettisessa mielessä. Sitä pidetään ongelmallisena myös tulkinnan näkökulmasta. Hayesin ja Hatchin (1999, 355) mukaan yksimielisyyden prosentit ovat paljon vaikeammin tulkittavia kuin korrelaatioarvot, koska johdonmukaisuuden muutokset eivät tule automaattisesti lasketuiksi prosenteissa, kuten ne tulevat korrelaatioissa. Korrelaatio ottaa huomioon sattumaan perustuvan yksimielisyyden (chance agreement) jo perusoletuksessaan, joka lähtee nollayksimielisyydestä. Vaikka esimerkiksi kappa, Spearmanin järjestyskorrelaatio ja Pearsonin tulomomenttikorrelaatiot (PM) (product-moment) lasketaan eri tavalla, niissä on kaikissa oletuksena vaihtelu eli ne ottavat huomioon systemaattiset ja samansuuntaiset erot arvioijien välillä.

Yksimielisyyttä osoittavissa prosentteissa vaihtelu otetaan huomioon siten, että arvioijien välisen vaihtelun numeroarvoilla ilmaistuun suuruuteen kiinnitetään huomiota eli tehdään ero selkeän ja vähäisen erimielisyyden välillä. Tosin yksimielisyyssprosentti voi osoittaa isompaa erimielisyyttä kuin se käytännössä on, koska arvioijien väliset erot voivat olla pieniä, esimerkiksi yhden tason suuruisia. Yhdellä tasolla poikkeavaa, lähes yksimielistä arviota kutsutaan vierekkäiseksi arvioksi (adjacent score) (Hayes & Hatch 1995, 357; Penny ym. 2001, 145). Tosin korrelaatioiden puolesta puhuu se, että Spearmanin järjestyskorrelaatio ja Pearsonin korrelaatio ottavat automaattisesti huomioon arvioiden läheisyyden.

Yksimielisyyttä osoittava prosentti voi olla riippuvainen myös arvioitavien määrästä ja skaalasta, jolla taitoa mitataan (Dunn 1989, 19). On eri asia, mitataanko kirjoittamisen taitoa 3-, 6- vai 10-portaisella asteikolla (esim. Penny ym. 2001). Abedin ym. (1995, 18) tutkimuksen mukaan yksimielisyyden prosenttiin vaikuttaa myös arvioijien määrä, sillä mitä suurempi arvioijien määrä on, sitä alhaisempi on yksimielisyyden prosentti. Korrelaatiokertoimen Cronbachin alfan kohdalla vaikutus on taas päinvastainen: mitä enemmän arvioijia on, sitä korkeampi on alfa (ks. myös Cortina 1993).

Toisaalta taas yksimielisyysestimaatit ovat käytännöllisiä, koska ne ovat nopeasti laskettavia ja siksi hyödynnettävissä jo paikan päällä esimerkiksi arvioijakoulutuksessa. Yksimielisyyttä osoittava indeksi on myös sovellettavissa helposti erilaisissa yhteyksissä esimerkiksi luokkahuone- ja portfolioarvioinnissa (Burry-Stock & Shaw 1996). Toisaalta korrelaatiokertoimien ja yksimielisyyttä osoittavien estimaattien eroa tai oikeastaan niiden molempien vahvuuksia ja heikkouksia kuvaa hyvin se, miten eri tavalla ne mittaavat arvioijien välistä yksimielisyyttä.

Jos esimerkiksi kaikki opiskelijat kirjoittavat erittäin hyvän tekstin vieraalla kielellä ja tekstit arvioineet arvioijat antavat kaikille opiskelijoille parhaan mahdollisen arvion, yksimielisyyssprosentti on sata, mutta PM-korrelaatio voi olla nolla, koska arvioijien välillä ei ole vaihtelua. Mutta jos taas esimerkiksi kahdesta arvioijasta toinen on ankarampi ja antaa systemaattisesti yhden tason alempia arvioita kuin toinen arvioija, yksimielisyyssprosentti olisi tässä tapauksessa nolla ja PM-korrelaatio lähellä ykköstä, koska arviot vaihtelevat samansuuruisesti ja samansuuntaisesti. (vrt. Abedi ym. 1995.)

Hayesin ja Hatchin (1999, 359) mielestä korrelaatio toimii tarkoituksenmukaisemmin kuin yksimielisyyssprosentti ja se pitäisi vakiinnuttaa perusmenetelmäksi tarkasteltaessa reliabeliutta. Burry-Stock ja Shaw (1996, 253) taas painottavat sitä, että on tärkeää tulkita oikein

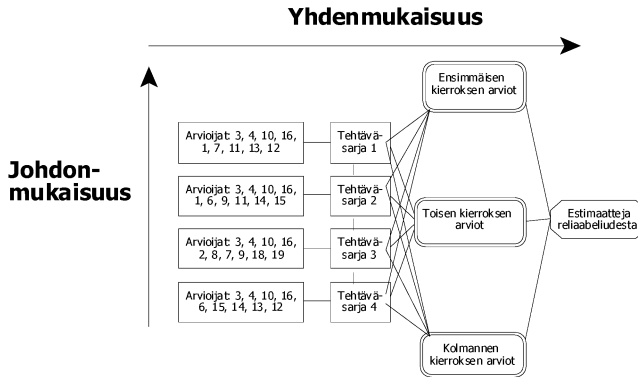
korrelaatioita ja yksimielisyyttä osoittavia estimaatteja, koska ne kertovat eri asioita.

Näyttää siis siltä, että molempia mittaustapoja tarvitaan. Koska arvioijien välisen reliabiliuden arvioimiseen vaikuttavat niin monenlaiset tekijät, esimerkiksi arvioijien määrä, aineiston koko ja arvioitaviin liittyvät taustekijät, tuntuisi tarkoituksenmukaiselta käyttää tutkimuksessa erilaisia mittaustapoja. Tämän lisäksi Abedi ym. (1995) korostavat sitä, että mittaustavan valinnan tulisi perustua muun muassa tutkimuksen aineistoon ja sen luonteeseen.

3.3.3 Yhdenmukaisuuden ja johdonmukaisuuden tarkastelu tässä tutkimuksessa

Tässä tutkimuksessa arvioijien sisäistä johdonmukaisuutta tarkastellaan eri arviointikierrosten avulla ja heidän keskinäistä yhdenmukaisuuttaan arviointikierröksittain, kuten kuvioista 4 näkyy. Tehtävät (N=9) oli jaettu neljään kolmen tehtävän sarjaan (ks. myös luku 2.2.2.1). Yksi tehtävistä oli sama kaikissa tehtäväsarjoissa. Arvioijat (N=17) arvioivat tehtäväsarjoja siten, että kukin suoritus sai arvion yhdeksältä tai kymmeneltä arvioijalta. Neljä arvioijista arvioi kaikki tehtäväsarjat eli kaikkien kirjoittajien suoritukset.

Arviointikierröksia oli kolme, joista ensimmäisellä arvioijat arvioivat suoritukset käyttäen 6-portaista asteikkoa ilman tasokuvauksia. Tällä kierroksella arvioijat määrittelivät itse asteikon sisällön. Toisella kierroksella arvioijat arvioivat suoritukset holistisesti käyttäen heille annettua 6-portaista asteikkoa, jossa oli tasokuvaukset. Kolmas arviointikierrös oli toisen arviointikierröksen toisinto. Arvioijat arvioivat kaikilla kolmella kierroksella samojen kirjoittajien samat suoritukset.



KUVIO 4 Johdonmukaisuus ja yhdenmukaisuus arviointikierroksittain ja arvioijittain.

Johdonmukaisuutta tarkastellaan siis arviointikierrosten välisen analyysin avulla: miten johdonmukaisesti arvioijat toimivat varsinkin toisella ja kolmannella arviointikierroksella? Mielenkiintoista on tarkastella myös sitä, miten ensimmäinen arviointikierros eroaa toisesta ja kolmanteen kierroksesta eli muuttaako annettu asteikko arviointia ja eroavatko toinen ja kolmas kierros ensimmäisestä kierroksesta systemaattisesti. Ensimmäinen kierros on tässä mielessä tavallaan ankkuri arvioijan omaan asteikkoon, hänen käsityksiinsä teksteistä ja kirjoittamisesta taitona.

Arvioijien välistä yhdenmukaisuutta taas on järkevintä tarkastella tehtäväsarjoittain, koska kunkin tehtäväsarjan arvioijat ovat arvioineet samojen kirjoittajien suoritukset samoista tehtävistä. Tehtäväsarjojen tehtäviin voi tutustua tarkemmin luvussa 2.1.1. Yhdenmukaisuuden odotus liittyy lähinnä toiseen ja kolmanteen kierrokseen ja myös siihen, miten käsikkäin yhdenmukaisuus ja johdonmukaisuus kulkevat näillä kahdella kierroksella. Asetelma siis valottaa myös sitä, miten arvioijien sisäinen johdonmukaisuus vaikuttaa yhdenmukaisuuteen. Ensimmäisellä arviointikierroksella yhdenmukaisuuden tarkkailussa on kiinnostavaa tutkia arvioijien antamien arvioiden perusteella sitä, miten lähellä tai kaukana toisistaan arvioijien itsensä määrittelemät asteikot ovat.

3.3.4 Mikä edesauttaa ja mikä taas estää reliabiliuden saavuttamista?

3.3.4.1 Johdonmukaisuus ja yhdenmukaisuus edellyttävät virheetöntä arviointia

Arvioijien sisäinen johdonmukaisuus edellyttää, että arvioijien antamat arviot perustuvat systemaattisesti samoihin kriteereihin suorituksesta toiseen vaihtelevista muuttujista huolimatta. Lienee kuitenkin aika selvää, että arvioijat tuskin tulkitsevat ja soveltavat arviointikriteereitä täysin samalla tavalla toisaalta suorituksesta toiseen ja toisaalta suhteessa toisiinsa, kun he lukevat ja arvioivat tahollaan samojen oppijoiden kirjoittamia tekstejä.

Arvioijien antamat arviot heijastelevat sekä arvioijien subjektiivista arviointityyliä että heidän puolueellista suhtautumistapaansa (bias) kuhunkin arvioitavaan tuotokseen (Thorndike & Hagen 1977). Näistä ensin mainittu viittaa arvioijan pysyvään piirteeseen, esimerkiksi siihen, että hän on toistuvasti ankara tai lempeä arvioija suhteessa muihin arvioijiin. Jälkimmäinen taas vaihtelee kunkin arvioitavan suorituksen ja sen aiheuttamien mielikuvien mukaan, esimerkiksi sen mukaan, mitä käsiala viestii kirjoittajasta tai miten kirjoittajan tapa käsitellä aihetta joko miellyttää tai ärsyttää arvioijaa.

Arvioijan toimintaa koskevan reliabiliuden saavuttamiseksi pyritäänkin ennen kaikkea siihen, että vähennetään arvioijien erilaisuudesta johtuvaa vaihtelua eli niitä tekijöitä, jotka aiheuttavat sen. Psykometrisessä arvioinnissa erimielisyydet tai konkreettisemmin poikkeamat arvioijien arvioinneissa nähdään mittausvirheinä (measurement errors), jotka aiheutuvat yksilön käyttäytymisen vaihtelusta (esim. Everitt & Hay 1992; AERA ym. 1999; Xitao & Chen 2000). Arvioinnin johdonmukaisuuteen liittyviä virheitä jaottelevat eri tutkijat hieman eri tavoin. On kuitenkin tavallista, että mittausvirheen ja systemaattisen virheen välillä tehdään ero.

Varsinainen mittausvirhe syntyy hypoteettisesta erosta todellisen pistemäärän ja annetun pistemäärän välillä. Esimerkiksi kaksi arvioijaa arvioi samat suoritukset ja heidän antamiensa arvioiden väliseksi korrelaatioksi saadaan .80. Korrelaatiokerroin viestii tämän tulkinnan mukaan siitä, että 64 prosenttia annetuista pistemääristä selittyy todellisella pistemäärällä ja 36 prosenttia vaihtelusta on virheen aiheuttamaa, mikä taas johtuu kahden arvioijan keskinäisestä epäyhdenmukaisuudesta.

Systemaattiset virheet liittyvät taas esimerkiksi testin vaikeustasojen vaihteluun, jolloin virhe on valmiina mittarissa. Myös arvioitavalla henkilöllä voidaan katsoa olevan systemaattisia virhevaikutuksia, jotka

voivat johtua esimerkiksi hänen vireytyksensä, motivaatiostaan tai kokemuksestaan. Arvioijakeskeisen näkökulman mukaan arvioitavasta henkilöstä johtuvien virhevaikutusten ei kuitenkaan katsota vaikuttavan johdonmukaisuuden ja yhdenmukaisuuden vaihteluun. Reliaabeliuden tarkastelemisessa huomio kiinnitetäänkin vain arvioijista johtuvaan vaihteluun, jota pyritään minimoimaan eri tavoin.

3.3.4.2 Arvioijasta johtuvia mittausvirheitä

Tavallisimmat arvioijasta johtuvat virheet luokitellaan yleensä neljään tai viiteen ryhmään. Kirjallisuudessa esiintyvät tavallisimmin seuraavat viisi kategoriaa: 1) arvioijan ankaruus tai lempeys, 2) sädekehävaikutus (halo effect), 3) asteikon keskivälin painottaminen (central tendency), 4) samojen tasoarvioiden tai pistemäärien käyttäminen (restricted variability) ja 5) arvioijien välisen yhdenmukaisuuden tai yksimielisyyden puute (interrater reliability/agreement) (ks. esim. osin Thorndike & Hagen 1977; Saal ym. 1980; Linn & Gronlund 1995; Fox & Caspy 1996; Engelhard & Stone 1998). Joissakin tapauksissa kolmas ja neljäs kategoria on yhdistetty ja joissakin arvioijien välistä yhdenmukaisuutta ei erotella omaksi kategoriakseen.

Ensimmäisessä ryhmässä on kyse yleistämismvirheestä eli siitä, että *lempeä arvioija* käyttää toistuvasti lähinnä asteikon yläpäättä ja *ankara arvioija* taas sijoittaa suoritukset asteikon alemmille tasoille. Lempeän arvioijan asteikko alkaa käytännössä suunnilleen asteikon puolestavälistä tai vähän sen alapuolelta, ja ankaran arvioijan asteikko taas päättyy puoleenväliin tai vähän sen yläpuolelle. Arvioijat eivät kuitenkaan ole joko täysin ankaria tai täysin lempeitä, vaan ankaruuden ja lempeyden ilmeneminen voi vaihdella vähäisestä taipumuksesta asteikon subjektiiviseen uudelleen tulkintaan. Huomattava asteikon uudelleen tulkinta tarkoittaa käytännössä sitä, että esimerkiksi matala arvio voi muuttua keskitason arvioksi tai toisin päin.

Sädekehävaikutuksessa on kyse siitä, että arvioija ei kiinnitä huomiota kaikkiiin niihin tekijöihin, joita arviointiasteikko häneltä edellyttää, vaan jokin yksittäinen piirre johtaa korkean tasoarvion antamiseen. Sädekehävaikutuksesta on kyse myös silloin, kun arvioijan toimintaa ohjaavat hänen omat henkilökohtaiset mieltymyksensä: jos arvioijan asenne on positiivinen arvioitavaan suoritukseen tai siihen henkilökuvaan, joka suorituksesta välittyy, arvioija antaa yleensä korkeamman arvion. Jos asenne taas on negatiivinen, vaikutus on päinvastainen. Tällaisia tekijöitä voisivat olla kirjoittamisen arvioinnissa esimerkiksi käsiala tai tekstin pohjalta syntyvät vaikutelmat kirjoittajan

taustatekijöistä, kuten sukupuolesta, äidinkielestä, kulttuurista tai koulutuksesta.

Thorndike & Hagen (1977, 452) määrittelevät sädekehävaikutuksen arvioijan taipumuksesi arvioida yleisvaikutelman mukaan. Tällöin arvioija ei erittele taitoa tiettyihin sovittuihin piirteisiin tai sitten hän sallii, että arvioitavasta henkilöstä johtuvat reaktiot värittävät yksittäisten piirteiden arviointia. Sädekehävaikutus liittyy siis kiinteästi arvioijan asenteisiin, ja siksi on epätodennäköistä, että siitä voitaisiin koskaan päästä täysin eroon. Lisäksi on melko vaikeaa tietää, missä määrin arvio perustuu arviointikriteereihin ja missä määrin taas sädekehävaikutukseen.

Käytännössä sädekehävaikutus ilmenee esimerkiksi silloin, kun arvioijan pitäisi arvioida kriteeriviitteisesti mutta hän arvioikin holistisesti (Engelhard 1994, 98–99). Vaikutus näkyy arvioiden samanlaisuutena jokaisen kriteerin kohdalla. Joskus arvioiden samanlaisuus kriteereittäin voi pitää paikkaansa, mutta silloin kyse on *todellisesta sädekehävaikutuksesta* (true halo), joka ei ole mittausvirhe, vaan oppijat ovat todella tasaisia eri kriteereillä tarkasteltuna. Sen sijaan *kuvitteellinen sädekehä* (illusory halo) on mittausvirhe, joka ilmenee tietyn oppijan saaman arvion laadussa. Kuvitteellinenkaan sädekehä ei ole suoraan pääteltävissä tilastojen perusteella, vaan erikseen on tapauskohtaisesti tarkasteltava, onko kyseessä todellinen vai kuvitteellinen sädekehä.

Sädekehän vaikutusta on yritetty pienentää muun muassa peittämällä suorittajan henkilöllisyys, mutta usein suoritus itse puhuu puolestaan. Arvioija pystyy tekemään monenlaisia päätelmiä kirjoittajasta pelkän tekstin perusteella. Tässä tutkimuksessa suoritukset ovat nimettömiä eikä niitä ole järjestetty kirjoittajittain. Arvioija voi kuitenkin tunnistaa samat käsialat, mutta suoritukset ovat niin hajallaan ja niitä on niin paljon, että tarkempi kirjoittajien erittely vaatisi melkoista viitseliäisyyttä.

Asteikon keskivälin ja samojen tasoarvioiden tai pistemäärien käyttäminen yhdistetään monissa tutkimuksissa samaan kategoriaan tai kategorioiden nimityksiä käytetään toistensa synonyymeina, mutta Saal ym. (1980) erottavat ne toisistaan. Heidän (1980, 417) mukaansa asteikon keskivälin painottamisessa on nimenomaan kyse siitä, että arvioija käyttää vain asteikon keskivälin tasoja eikä ollenkaan asteikon kumpaakaan ääripäätä. Kyse on heidän mukaansa arvioijan virheestä, asteikon tarkoituksellisesta virhekäytöstä.

Samojen tasoarvioiden tai pistemäärien käyttäminen tarkoittaa taas sitä, miten arvioijien antamat arviot eroavat toisistaan. Tasoarviot voivat siis kasaantua asteikon mihin kohtaan tahansa, mutta ne ovat ryhmittyneet johonkin tiettyyn kohtaan. Tällä tavalla annetut arviot eivät

tietenkään erottele yksilöiden välisiä eroja (Saal ym. 1980; Engelhard 1994). Itse asiassa kaikki edellä mainitut virheen aiheuttajat tai häiriötekijät voivat johtaa rajoittuneeseen tai samantyyppisten arvioiden käyttöön. Esimerkiksi arvioija, joka on ankara, käyttää vain asteikon alapään tasoja, mikä edelleen näkyy negatiivisesti vinona jakaumana.

Viides kategoria liittyy siihen, miten *yhdenmukaisia tai yksimielisiä* arvioijat ovat suhteessa toisiinsa. Yhdenmukaisuuden ja yksimielisyyden mittaamista pidetään kaikkein yleisimpänä tapana arvioida reliaaabeliutta, joten sitä ei monissakaan yhteyksissä erotella omaksi virhelähteekseen (ks. esim. Engelhard 1996). Saal ym. (1980, 419) korostavat kuitenkin sitä, että korkeaa yhdenmukaisuutta pidetään laadukkaana arvioinnin tunnusmerkkinä, vaikka todellisuudessa se voi olla seurausta jostakin muusta tekijästä. Hyvä reliaaabelius ei siis välttämättä implikoi hyvää validiutta, eli se ei välttämättä aiheudu todellisesta pistemäärästä (true score), vaan se voi olla yhtä hyvin seurausta arvioijien samansuuntaisesta puolueellisuudesta (bias).

Jos arvioijien välinen yhdenmukaisuus tai yksimielisyys on heikko, se voi olla seurausta siitä, että jokin viidestä edellä mainitusta virhelähteestä toteutuu. Esimerkiksi jos joku tai jotkut arvioijista ovat lempeitä tai käyttävät vain asteikon keskiväliä, he erottuvat epäyhdenmukaisina arvioijina yleisestä linjasta. Vaikka arvioijien välinen samanmielisyyys tai korrelaatio olisikin korkea, saattaa siitä huolimatta arvioijien välillä olla eroja siinä, miten helposti he antavat keskimäärin minkäkin korkuisia arvioita (Linn & Gronlund 1995; Doyle 1998). Mutta tämäntasoisista eroista ei välttämättä olla kiinnostuneita, jos reliaaabeliuskerroin ylittää hyväksytyyn rajan eli on tarpeeksi korkea (esim. > .80).

Monet kategorioista näyttäisivät tavalla tai toisella kietoutuvan toisiinsa ja jopa niin, että niillä on keskinäinen kausaalisuhte. Esimerkiksi arvioija voi arvioida subjektiivisen mieltymyksensä mukaan jonkin tietyn ryhmän tekstejä, minkä vuoksi hän antaa tälle ryhmälle tavallista parempia tasoarvioita ja tämän vuoksi hänen arvionsa kasaantuvat asteikon yläpäähän. Tämä taas aiheuttaa sen, että hän poikkeaa muiden arvioijien linjasta 'lempeänä' arvioijana. Kategorioiden väliset rajat näyttäisivät kyllä olevan tunnistettavissa ja määriteltävissäkin, mutta toisaalta ne ovat veteen piirrettyjä viivoja, joista on vaikea sanoa, mikä johtuu mistäkin tekijästä.

Toinen kategorioille yhteinen piirre näyttäisi olevan se, miten ne ovat yhteydessä epäsuorasti arvioinnin validiuteen. Esimerkiksi jos arvioija antaa pääasiassa samoja tasoarvioita jonkin arviointivirheen vuoksi, hän ei tule erotelleeksi yksilöiden välisiä eroja taidon hallinnassa. Tällöin arviointi ei ole reliaaabelia eikä myöskään validia, koska arviointi ei mittaa sitä, mitä sen pitäisi mitata. Foxin ja Caspyn (1994, 46) mukaan

tilastolliset perusteet osoittavat, että molemmat, sekä arvioijan ankaruus/lempeys että sädekehä, liittyvät suoraan arvioinnin validiuteen. Toisaalta tutkimukset eivät kuitenkaan ole vielä pystyneet osoittamaan, kasvattaako lempeyden/ankaruuden tai sädekehän vaikutusten vähentäminen suoraan arvioinnin validiutta.

Mittausvirheiden tarkastelu yleensä ja niiden paikantaminen arvioijalähtöiseksi viestii käsityksestä, jonka mukaan erot arvioijien välillä johtuisivat pelkästään arvioijasta itsestään (vrt. Henning 1991, 286–287). Sama ajatus käänteisesti ajateltuna viestii siitä, että reliaabeliutta on kyllä kvantitatiivisesti tarkasteltu useista eri näkökulmista, mutta ne ovat tavallisesti liittyneet mittauksen virhelähteen paikantamiseen ja huomioon ottamiseen (Stansfield & Ross 1988, 173). Miten laajasti arvioijien välisiä eroja tulisi tarkastella ja miten laajaa tarkastelua psykometrinen näkökulma edellyttää? Miten mittausvirheen paikantaminen lisää ymmärrystä prosessista ja millaisia eväitä se tarjoaa esimerkiksi arviointikoulutuksen sisällöiksi? Riittääkö vain sen tosiseikan toteaminen, että arvioijan lempeys heikentää arvioinnin reliaabeliutta?

Arvioinnin reliaabeliutta voivat heikentää tai vahvistaa edellä mainittujen viiden kategorian lisäksi muut arvioijan taustasta johtuvat tekijät, esimerkiksi arvioijien kokemus, heidän saamansa koulutus tai sen puuttuminen, arvioijan vireystila, varsinkin väsymys sekä arviointiskaala itsessään (esim. Charney 1984; Davis ym. 1987; Shohamy ym. 1992; Cushing Weigle 1994; Linn & Gronlund 1995). Tiukasti ajatellen monet arvioijien välisistä eroista voidaan kuitenkin palauttaa edellä mainittuihin viiteen kategoriaan. Esimerkiksi jos opettajan kielitaitokäsitys on hyvin virhelähtöinen, hän saattaa siitä syystä olla ankara arvostelija. Vastaavasti arvioija, joka ajattelee 'oppilaidensa parasta', voi olla liian lempeä. Jos taas asteikko on liian monitulkintainen, arvioija - joka ei halua asteikon huonouden vaikuttavan oppijoiden tuloksiin ja antaa systemaattisesti vähän paremman arvion - tulee myös aiheuttaneeksi virheen lempeyden tai asteikon yksipuolisen käytön vuoksi.

3.3.4.3. Arvioijakoulutus arvioinnin yhtenäistäjänä

Testien kehittämissuunnitelmissa reliaabeliuden saavuttamista pyritään tukemaan siten, että määritellään huolellisesti arviointikriteerit, laaditaan niiden soveltamiseen tarkoitettuja ohjeistuksia sekä valitaan ja koulutetaan tietty joukko arvioijia. Koulutuksen tavoitteena on keskustellen ja suorituksia yhdessä arvioiden saada arvioijat jakamaan samanlainen käsitys tietynlaisesta suorituksesta ja vakiinnuttaa tällä tavalla arvioijien keskinäinen yhdenmukaisuus (esim. Henning 1991; Cumming 1997). Arvioijakoulutus nähdäänkin yhtenä tärkeimmistä ja tehokkaimmista

keinoista vaikuttaa arvioijien keskinäiseen yhdenmukaisuuteen ja arvioijan sisäiseen johdonmukaisuuteen (esim. Homburg 1984; Huot 1990; Cushing Weigle 1994; Moss 1994; Engelhard 1994).

Arvioijakoulutus on sitä tehokkaampaa, mitä monipuolisemmin se pureutuu arvioijalähtöisten virheiden lähteisiin. Arvioinnin kehittämisen kannalta on tärkeää, että arvioijien työtä seurataan jatkuvasti ja koulutusta räätälöidään sen mukaan, millaisista arvioijalähtöisistä virheistä analyysit raportoivat. Arviointityön seurantaan liittyy myös se, että mikäli poikkeavia arviointeja löytyy, poikkeavia arvioita antaneiden arvioijien tekstit arvioidutetaan vielä jollakin muulla arvioijalla. Tehokkaaseen arvioijakoulutukseen kuuluvat lisäksi palautteen saaminen ja arviointiprosessista keskusteleminen muiden arvioijien kanssa.

Koulutuksella ei ole kuitenkaan kaikissa tutkimuksissa ollut kovin merkittävää vaikutusta reliabiliuteen. Esimerkiksi Shohamy ym. (1992) selvittivät tutkimuksessaan sitä, miten arvioijan tausta ja arvioijakoulutus vaikuttavat kirjoittamisen arvioinnin reliabiliuteen. Arvioijat jaettiin tutkimuksessa neljään ryhmään: 1) arvioijakoulutusta saaneet englanti toisena kielenä -opettajat, 2) englanti toisena kielenä -opettajat, jotka eivät saaneet arvioijakoulutusta, 3) tavalliset syntyperäiset englannin puhujat, joilla ei ollut opetuskokemusta ja jotka saivat arvioijakoulutusta ja 4) tavalliset syntyperäiset englannin puhujat, joilla ei ollut opetuskokemusta ja jotka eivät saaneet arvioijakoulutusta. Arvioijakoulutus viittaa tässä yhteydessä nimenomaan kirjoittamisen arvioijakoulutukseen.

Kaikki arvioijat olivat syntyperäisiä englannin puhujia. Jokaisessa neljässä ryhmässä oli viisi arvioijaa. Arvioinnissa arvioijat käyttivät kolmenlaisia asteikoita: 1) holistista asteikkoa, 2) kommunikatiivista asteikkoa, joka painotti viestivyyttä, tekstin jäsentelyä ja asioiden esittämistapaa ja 3) kieliasteikkoa (accuracy of written samples), joka koski rakenteiden ja sanaston hallintaa. Kukin arvioija arvioi 50 satunnaisesti valittua tekstiä käyttäen kaikkia kolmea asteikkoa. (Shohamy ym. 1992, 28.)

Tulosten mukaan arvioijien välinen reliabilius vaihteli .80:n ja .93:n välillä. Koulutettujen arvioijien korrelaatio oli hiukan muita korkeampi. Vahvin korrelaatio oli arvioijakoulutusta saaneiden opettajien ja heikoin taas arvioijakoulutusta saamattomien maallikoiden. Opettajilla koulutus näytti vaikuttaneen eniten kieliasteikolla tehtyihin arvioihin. Arvioijien sisäistä johdonmukaisuutta tutkittiin siten, että arvioijakoulutusta saaneiden opettajien ryhmä (N=5) arvioi kymmenen satunnaisesti valittua tekstiä uudelleen kolmen viikon kuluttua. Korrelaatio vaihteli Spearman-Brownin kaavaa käyttäen .76:n ja .96:n välillä, mitä tutkimuksen tekijät pitivät suhteellisen vakaana. (Shohamy ym. 1992, 30.)

Tutkimuksen johtopäätösten mukaan arvioijat pystyvät arvioimaan luotettavasti riippumatta taustoista ja koulutuksesta, joskin koulutuksen avulla reliäabeliutta pystyttiin vielä parantamaan (Shohamy ym. 1992). Toisaalta korrelaatioihin on varmasti vaikuttanut arvioijien pieni määrä. Viiden arvioijan perusteella ei voitane vetää vielä kovin yleistäviä johtopäätöksiä, vaikka korrelaatiot olisivatkin tilastollisesti merkitseviä. Toinen seikka, joka vaikuttanee opettaja- ja maallikkoryhmän korrelaatioiden pienehköihin eroihin, on se, että maallikot olivat yliopisto-opiskelijoita ja itsekkin tottuneita kirjoittamaan. Tutkimus vaikuttaa kuitenkin mielenkiintoiselta kokeilulta ja laajennettuna se voisi tarjota tarpeellista ja hyödyllistä tietoa arvioijien, myös maallikkojen, toiminnasta ja arvioijakoulutuksen sisältöjen ja määrän kehittämistarpeista.

Arvioijakoulutukselta edellytetään edellä esiin tulleiden seikkojen lisäksi jatkuvuutta, kertaluonteisuus ei siis riitä. Lumley ja McNamara (1995) selvittivät puhumisen arviointia koskevassa tutkimuksessaan arvioijien sisäistä johdonmukaisuutta ja arvioijien yksilöllisiä linjaeroja kolmen arviointikerran välillä. Arvioijia oli ensimmäisellä kierroksella 55, toisella 32 ja kolmannella 4. Kaikki arvioijat saivat arvioijakoulutusta kahden ensimmäisen arviointikerran yhteydessä. Osa arvioijista ja kaikki arvioitavista puhumisen suorituksista olivat uusia kolmannella arviointikierroksella, mutta uudet suoritukset oli ankkuroitu vaikeustasoltaan ensimmäisen arviointikerran suorituksiin ja arvioijista tehdyt analyysit verrannettiin siten, että kierrosten välinen vertailu oli mahdollista. Arviointidata analysoitiin FACETS-ohjelmalla.¹² Ensimmäisen ja toisen arviointikierroksen välissä oli 2 vuotta ja toisen ja kolmannen välissä 1–2 kuukautta.

Lumleyn ja McNamaran (1995) tutkimustulosten mukaan arvioijista vain yksi oli huomattavan johdonmukainen kaikilla kolmella arviointikierroksella. Arvioijakoulutus ei siis näyttänyt vaikuttavan arvioijan ankaruuteen tai lempeyteen sitä eliminoivasti tai edes vähentävästi. Toisen ja kolmannen kierroksen analyysit osoittivat myös, että arvioijakoulutuksen yhden- ja johdonmukaisuutta lisäävä vaikutus ei kannata ajallisesti kovin pitkälle, ei edes arvioijakoulutusjakson loppuun. Tutkimuksen mukaan arvioijakoulutuksen tarve on jatkuvaa. (Lumley & McNamara 1995.) Tosin tutkimuksesta ei käy ilmi, millaista arvioijakoulutus oli luonteeltaan ja millaista palautetta arvioijat saivat arvioinneistaan ja kuinka heidän arviointejaan seurattiin koulutuksen aikana. Koulutuksen tehottomuus voi viestiä myös sen huonosta laadusta eikä niinkään arvioijakoulutuksen merkityksestä yleensä.

¹² FACETS on tietokoneohjelma, joka perustuu Raschin malliin. Ohjelma mahdollistaa monien muuttujien tarkastelemisen yhtä aikaa. Esimerkiksi suoritusarviointissa on mahdollista ottaa huomioon arvioitavan taito, tehtävän vaikeustaso, arvioijan piirteet, asteikon vaikutus ja muut tekijät, jotka vaikuttavat testitulanteessa.

Lumleyn ja McNamaran (1995) tutkimustulos panee kuitenkin pohtimaan yleisimminkin arvioijakoulutuksen sisältöjä ja kykyä ohjata arvioijan toimintaa. Arvioijakoulutusta on kritisoitu muun muassa siitä, että siinä kiinnitetään pääasiassa huomiota pintatason tekijöihin. Arvioijat voidaan esimerkiksi kirjoittamista koskevassa koulutuksessa kouluttaa olemaan yhtä mieltä teksteistä, mutta on tavallista, että yhteinen käsitys koskee lähinnä vain tekstin pintatekijöitä, kuten käsialaa tai oikeinkirjoitusta, eikä tekstin varsinaisia sisällöllisiä tekijöitä (Charney 1984, 75–77; Hamp-Lyons 1991b, 143–144). Itse asiassa reliaaabeliuden korostaminen ajaa usein siihen, että kirjoituksista arvioidaan lähinnä rakenteiden tarkkuutta eikä niinkään tekstuaalisia seikkoja, kuten koheesiota ja tekstin jäsentelyä (ks. esim. Cohen & Cavalcanti 1990; Vaughan 1991).

Koulutuksen yksipuolisuus ei tietenkään koske kaikkea arvioijakoulutusta. Reilussa kymmenessä vuodessa arvioijakoulutus on varmasti kehittynyt, vaikka toisaalta Charneyn ja Hamp-Lyonsin huoli ei liene aiheeton nykyäänkään. On syytä korostaa myös sitä, että arvioijakoulutuksesta ei voi tehdä mitään yhtä ideaalista sisältömallia, vaan senkin kehittäminen edellyttää erilaisia koulutusmalleja erilaisiin arviointitarpeisiin. Tutkimustulokset siitä, miten eri koulutusmallit toimisivat suhteessa toisiinsa, mihin palautteessa kannattaa kiinnittää huomiota ja millainen arvioijan työn seuranta tukisi parhaiten hänen työtään, olisivat tarpeellisia. Yleensäkin arviointiprosessin parempi tunteminen auttaisi myös asettamaan koulutukselle tarkoituksenmukaisia tavoitteita.

Opettajat ja arvioijat voivat arvioida suorituksia myös pitkälti oman kirjoittamiskäsityksensä mukaan eivätkä niinkään arviointiskaalaan nojaten, vaikka pintatasolla skaala ohjaisikin arviointia. Tällöin saattaa käydä niin, että saavutettu reliaaabelius voi esimerkiksi heijastella enemmän opettajien yhteisiä käsityksiä kuin asteikon sisältöä. Jossakin määrin arviointi heijastaa aina myös arvioijan omia painotuksia, vaikka hän arvioisikin jonkin tietyn testisysteemin viitekehyksen ja asteikkojen ohjaamana. Subjektiiivisuus on aina läsnä, tosin vaihtelevissa määrin, joten sen tiedostaminen ja ohjaaminen olisi tärkeää.

Koulutuksen näkökulmasta arvioijakoulutuksen puute näkyy ennen kaikkea siinä, miten paljon arvioijien omat odotukset ja käsitykset siitä, mitä arviointi ja arvioitava taito ovat, vaihtelevat ja vaikuttavat arvioinnin johdonmukaisuuteen ja yhdenmukaisuuteen (ks. myös Cushing Weigle 1998). On selvää, että mitä homogeenisempi arvioitavien joukko on ja mitä yhteneväisemmän käsityksen arvioijat jakavat sekä arviointiasteikosta että arvioitavien suoritusten laadusta, sitä todennäköisemmin reliaaabeliuskertoimet ovat korkeita.

Käsitysten yhteneväisyyttä korostavaan johtopäätökseen päätyivät myös Tyndal ja Kenyon (1996) englanti toisena kielenä -kirjoittamisen arviointia koskevassa tutkimuksessaan. Englantia toisena kielenä opettavat (N=10) arvioivat suoritukset (N=400) holistisesti 7-portaisella asteikolla. Varsinaiseksi tutkimusdataksi valittiin näistä 246 suoritusta, joista oli pudotettu pois tason 7 saaneet. Näin ollen tutkimuksessa analysoitavat suoritukset sijoittuivat 6-portaiselle asteikolle. Suorituksista 70 prosenttia kaksoisarviointiin ja 28 prosenttia kolmoisarviointiin. Reliaabelius (Raschin analyysi) oli .85, ja vain 5,3 prosenttia arvioinneista (N=28) ei ollut yhteensopivia.

Tutkijoiden mukaan asteikko toimi hyvin, koska arvioinnissa asteikon kuvaukset ja se, miten arvioijat sisäistivät asteikon, kohtasivat toisensa. Myös asteikon kuvaukset liittyivät suoraan siihen, mitä kyseisen yliopiston englannin kursseilla opetetaan. Asiaan saattoi vaikuttaa sekin, että opettajat olivat olleet mukana kehittämässä asteikkoa, minkä vuoksi asteikko heijasteli vahvasti opettajien omia käsityksiä kirjoittamisesta. Neljä kymmenestä arvioijasta raportoikin luottaneensa jossakin määrin annettua asteikkoa enemmän omaan sisäiseen asteikkoonsa. Korkeaan reliaabeliuteen lienee vaikuttanut sekin, että opettajat olivat hyvin tottuneita lukemaan tässä tutkittujen tyyppisiä tekstejä.

3.3.4.4. Muita arvioinnin yhdenmukaisuuteen vaikuttavia tekijöitä

Reliaabeliuden virhelähteenä voi joidenkin käsitysten mukaan olla myös jokin muu kuin arvioija itse. Tarpeeksi korkean yleistettävyyden saavuttamista on myös perusteltu eräänlaisen mikrotason toiminnan kautta. Sitä ei voida saavuttaa, jos arvioija arvioi vain yhden suorituksen kultakin opiskelijalta (Cooper 1977; Takala 1985). Suorituksia tulisi siis olla useita, kultakin kirjoittajalta vähintään kaksi, jotka on kirjoitettu mielellään eri aikoina, esimerkiksi kahtena eri päivänä. Syyksi tähän Cooper (1977, 19) mainitsee sen, että kirjoittajien taso vaihtelee eri tehtävissä ja eri aikoina. Testiarvioinnissa on helpompi toteuttaa usean tekstin vaatimus kuin eriaikaisuus, koska tekstit kirjoitetaan tavallisesti peräkkäin tunnin tai kahden sisällä.

Lisäksi kullakin *suorituksella* tulisi olla vähintään kaksi arvioijaa, jotta suorituksen arviointi ei jäisi yhden arvioijan näkemyksen varaan. Cooperin (1977, 19) mukaan arvioijien väliset erot voivat olla suuria siitä huolimatta, että he ovat koulutettuja. Arvioijan sisäisen johdonmukaisuuden varmistamiseksi arvioijan pitäisi koulutuksen yhteydessä saada yksityiskohtaista tietoa omasta linjastaan. Arvioinnin aikana arvioijan taas tulisi tarkistaa linjansa säännöllisesti noin joka 25. suorituksen jälkeen (Cooper 1977, 20).

Koulutuksen ja muun edellä mainitun lisäksi mittausvirheiden vähentäminen tai niiden vaikutusten minimoiminen voi edellyttää myös *arviointiasteikon* ja *-ohjeiden* uudelleenmäärittelyä. Jos asteikko kuvaa taitoa liian ylimalkaisesti ja monitulkintaisesti tai jos asteikon kuvaama taito on ristiriidassa tehtävien edellyttämän taidon kanssa, asteikon soveltamisesta tuskin seuraa yhdenmukaisia arvioita. Monitulkintaiset asteikot jättävät yleensä liikaa tilaa arvioijien omille tulkinnoilla ja voivat lisätä arviointiin liittyvää epävarmuutta. Yhtä lailla asteikon tukena toimivat arviointiohjeet voivat mahdollisesti olla epäselviä ja monitulkintaisia tai ne voivat olla jopa ristiriidassa asteikon kanssa.

Reliaabeliuteen voi vaikuttaa heikentävästi myös *oppijoiden erilaisuus*. Toisin sanoen virhelähde ei olekaan arvioijan toiminnassa vaan siinä, että oppijat ovat keskenään heterogeeninen joukko. Oppijoiden suuret yksilölliset erot heijastuvat edelleen arvioitavien tekstien erilaisuutena, mikä voi aiheuttaa monitulkintaisuutta arvioijien keskuudessa (Swartz ym. 1999, 499). Erilaisuus voi vaikuttaa myös arvioijan sisäiseen johdonmukaisuuteen esimerkiksi siten, että arvioija soveltaa kriteereitä eri tavalla eri oppijoiden kohdalla. Tai sitten arvioon vaikuttaa se, minkä tasoisia suorituksia arvioija arvioi peräkkäin: esimerkiksi monen heikon suorituksen jälkeen vahvempi suoritus voi tulla arvioiduksi vielä vahvemmaksi tai toisin päin. Edellisten suoritusten tason vaikutus voi olla tuntuva varsinkin silloin, kun arvioidaan suuria määriä kerrallaan. Tällaisia tulkintoja tehdessä on tietysti vaikea sanoa, mikä on syy ja mikä on vaikutus ja mistä virhe oikeastaan aiheutuu, koska arvio palautuu ensi kädessä kuitenkin arvioijaan ja hänen päätökseensä, joka on taas peräisin arvioijan ja tekstin välisestä vuorovaikutuksesta.

Vaikka tehtävien reliaabeliutta tarkastellaan arvioista käsin, mutta erillään arvioijiin liittyvästä reliaabeliudesta, voi tehtävien laadulla olla vaikutuksensa siihen, miltä arvioijien välinen yhdenmukaisuus näyttää. Jos arviot kasaantuvat asteikon toiseen päähän, kyse voi olla *tehtävien vaikeustasosta* eikä niinkään arvioijasta virhelähteenä. "Kattoefekti" (ceiling effect) ilmenee, kun tehtävä on liian helppo, esimerkiksi jos suurimman osan arviot ryppääntyvät asteikon yläpäähän. "Lattiaefekti" esiintyy taas päinvastaisessa tapauksessa, jolloin tehtävä on liian vaikea ja arviot kasaantuvat asteikon alapäähän. (Ks. esim. McDonald 1999.) Tehtävä, joka on vaikeustasoltaan sopiva, näyttäisi noudattavan normaalijakaumaa, eli hyviä ja huonoja suorituksia on vähän ja suurin osa suorituksista on asteikon keskivälillä.

Taulukkoon 6 on koottu niitä tekijöitä, joiden on todettu vaikuttavan arvioiden reliaabeliuteen joko heikentävästi tai vahvistavasti. Tavallisin tapa tarkastella arvioijien sisäistä johdonmukaisuutta ja keskinäistä yhdenmukaisuutta on paikantaa mittausvirheiden lähteet, jotka ovat

psykometrisen arviointikirjallisuuden perusteella pääasiassa arvioijalähtöisiä. Mittausvirheiden paikantamisella ja analysoinnilla pyritään selittämään, miksi numeerinen data eli pistemäärät tai tasoarviot eivät ole toistettavia.

TAULUKKO 6 Reliaabeliuteen vaikuttavia tekijöitä.

	Heikentävästi ↓	Vahvistavasti ↑
Arvioijasta johtuvat	<ul style="list-style-type: none"> • ankaruus • lempeys • sädekehävaikutus • asteikon keskivälin käyttäminen • samojen arvioiden käyttäminen • arvioijien välinen epäyhdenmukaisuus 	<ul style="list-style-type: none"> • koulutus • kokemus/tottuneisuus • arviointiasteikon samanlainen tulkinta • yhteisesti jaettu taitokäsitys • oman toiminnan reflektointi
Arvioitavasta johtuvat	<ul style="list-style-type: none"> • oppijoiden erilaisuus - tekstien erilaisuus ja eritasoisuus - peräkkäisten tekstien erilainen taso 	<ul style="list-style-type: none"> • oppijoiden samanlaisuus
Arviointiasteikosta ja/tai -ohjeista johtuvat	<ul style="list-style-type: none"> • monitulkintaisuus • ylimalkaisuus • epäselkeys • huono sovellettavuus 	<ul style="list-style-type: none"> • yksiselitteisyys • selkeys ja ymmärrettävyys • suoritusten ja asteikon sisällöllinen yhteensopivuus
Arviointitehtävästä johtuvat	<ul style="list-style-type: none"> • vaatimustaso liian korkea tai matala suhteessa suorittajien tasoon tai asteikon kuvauksiin • suoritettavia tehtäviä, esim. kirjoitustehtäviä on vain yksi 	<ul style="list-style-type: none"> • vaatimustaso erottelee suorituksia tarkoituksenmukaisesti • useampi kuin yksi tehtävä kultakin oppijalta

Taulukkoon 6 kootut asiat saavat pohtimaan sitä, miten virhelähteiden paikantaminen ottaa huomioon arviointiprosessin ja arvioijan inhimillisen käyttäytymisen, johon vaikuttavat monenlaiset erilaiset muuttujat, esimerkiksi affektiivisuus, motivaatio ja vireystila. Missä määrin arviointiprosessin ymmärtäminen on psykometrisestä näkökulmasta relevanttia? Missä määrin sen tarvitsee vastata esimerkiksi kysymykseen, miksi yksi arvioija käyttää vain asteikon yläpään tasoja ja miksi toinen arvioija antaa työille parempia arvioita kuin pojille?

Vaikka edellä esitetyt virhelähteet ovat jossakin määrin osoitettavissa lähes kaikista arviointiyhteyksistä ja niiden tunnistaminen ja korjaaminen on kiistatta edellytys reliaabeliuden saavuttamiselle, eivätkö sittenkin monet niistä ole paikannettavissa viime kädessä yhteisen ja yhtenäisen taitokäsityksen puuttumiseen tai siihen, että arvioija ei pysty ohjaamaan omaa toimintaansa arviointiprosessin aikana? Siihen, miksi arvioijat ovat epäjohtonmukaisia ja epäyhdenmukaisia, millaiset tekijät suorituksissa tai arvioijan omassa päätöksenteossa vaikuttavat ja mihin suuntaan, emme saa varmastikaan vastausta tarkastelemalla arvioita pelkästään numeroina.

Joidenkin näkemysten mukaan reliaabeliusongelma näyttää melko väistämättömältä, koska esimerkiksi Hamp-Lyons (1990) epäilee, jakavatko asiantuntijatkaan samasta tekstistä samanlaista käsitystä. Toisin sanoen asiantuntijoidenkin käsitykset kirjoittamisesta taitona eroavat jossakin määrin toisistaan, ainakin silloin kun kirjoittamista tarkastellaan tekstien tasolla. Toisaalta se ei estä arvioinnin reliaabeliuden tavoittelemista, koska yhteisestä kirjoittamiskäsityksestä voidaan neuvotella kussakin kontekstissa erikseen muun muassa tavoitteiden ja tarpeiden mukaan.

Kyse näyttäisi olevan enemmänkin metodisista ja paradigmaattisista vaikutuksista: pitäisikö systemaattisten virheiden etsimisen lisäksi ottaa entistä painokkaammin huomioon diskurssi tai konteksti, jonka mukaan arvio määritellään? Tarkoitan tässä diskurssilla sekä testitehtävän ja sen suorittajan vuorovaikutusta että arvioijan, arvioitavan suorituksen ja arviointiskaalan välistä vuorovaikutusta (ks. Upshur & Tuner 1999). Numeerinen data kuvaa yleisiä tendenssejä, mutta arvioinnin laadullinen tarkastelu toisi mukaan myös selittävän aspektin, jolloin kyse ei olisi enää niinkään virhelähteiden paikantamisesta vaan kokonaisvaltaisemmin arvioijan käyttäytymisen ymmärtämisestä. Käsitelen arvioijan toimintaa yksityiskohtaisemmin neljännessä luvussa.

3.4 Kolme kirjoittamisen arviointitapaa

3.4.1 Arviointitapojen luokitteluperusteita

Kirjoittamisen arviointitapoja voidaan arviointiasteikon analytyttisyyden ja tarkoituksen perusteella jakaa erilaisiin ryhmiin. Yleisimmän jaottelun mukaan arviointitavat voidaan erotella kahteen ryhmään: 1) ne, jotka tuottavat yhden arvion eli tavallaan yhdistävät kirjoittamisen erilaiset komponentit ja 2) ne, jotka tuottavat monia arvioita eli erikseen arvioivat useampaa kuin yhtä kirjoittamisen komponenttia (esim. Huhta 1993;

Feldt & Brennan 1995; Bachman & Palmer 1996; Kroll 1998). Toisin sanoen skaalat voidaan jakaa *holistisiin* ja *analyyttisiin* skaaloihin. Yhteistä skaalatyypeille on se, että ne molemmat pyrkivät kuvaamaan hyvää kirjoittamista. Arvioinnin näkökulmasta kyse on siitä, että mitä enemmän tekstissä on hyvän kirjoittamisen piirteitä, sitä korkeamman arvion se saa.

Analyyttisen ja holistisen arviointitavan lisäksi kolmantena arviointitapana käytetään *piirrearviointia* (*primary trait scoring*). Arviointitapojen jakoa piirrearviointiin, holistiseen ja analyyttiseen arviointiin voidaan pitää eräänlaisena perusjakona (esim. Cooper & Odell 1977; Purves ym. 1988; Huot 1990; Shohamy ym. 1992). Käsittelen seuraavassa kutakin näistä kolmesta arviointitavasta, mutta kuitenkin niin, että käyn yksityiskohtaisemmin läpi holistista ja analyyttistä arviointia, koska kirjoittamista arvioidaan tavallisimmin joko holistisesti tai analyyttisesti. Vaikka nämä kolme arviointitapaa jaetaan omiin luokkiinsa, jako ei ole piirrearvioinnin osalta aina kovin selvärajainen, koska piirrearviointi voi soveltamistavan mukaan olla joko holistista tai analyyttistä. Esittelen ensin piirrearvioinnin, sitten analyttisen ja lopuksi tässäkin tutkimuksessa käytetyn holistisen arvioinnin.

3.4.2 Piirrearviointi

Piirrearvioinnissa on lähtökohtana, että kirjoittamisella on aina jokin funktio, esimerkiksi vaikuttaminen, opastaminen, viihdyttäminen tai tiedottaminen. Tällöin arvioinnin tehtävänä on arvioida sitä, miten hyvin kirjoitettu teksti täyttää tarkoituksensa. Lloyd-Jonesin (1977, 37–38) mukaan piirrearvioinnissa arviointia tarkastellaan diskurssin näkökulmasta. Itse asiassa kirjoittaminen nähdään diskurssin synonyyminä (Lloyd-Jones 1977, 33). Tässä yhteydessä diskurssi määritellään tavoitteellisenä toimintana, joka liittyy kielen funktioihin ja niiden piirteisiin.

Piirteet liittyvät yleensä kirjoittamisen tarkoitukseen, tekstin lukijaan ja kirjoittamistehtävän vaatimuksiin. Esimerkiksi työpaikkahakemuksessa voidaan korostaa sen puhuttelevuutta työnantajan näkökulmasta ja sitä, miten hakemus noudattaa tekstinä kyseisen kulttuurin kirjoittamis-konventioita. Piirteitä voitaisiin luonnehtia myös kielen toisistaan eroteltavissa oleviksi elementeiksi ja tehokeinoiksi, jollaisia ovat esimerkiksi sanasto, luettavuus ja tekstimäisyys.

Ollakseen informatiivista arvioinnin tulisi pystyä kuvaamaan suoritusta piirre piirteeltä. Piirrearvioinnissa arvioidaan siis yhtä tai useampaa sellaista piirrettä, joka on relevantti kyseisessä kirjoittamistehtävässä (Huot 1990, 238). Piirrearvioinnin tehtäväkohtaisuus ottaa huomioon kunkin diskurssityypin vaatimukset ja pyrkii siten kirjoittami-

sen autenttisuuteen. Piirrearviointin autenttisuutta ja tavoitelähtöisyyttä konkretisoi kuvio 5 (Lloyd-Jones 1977, 40).



KUVIO 5 Yksinkertaistettu diskurssimalli (Lloyd-Jones 1977, 40).

Kuviossa 5 esitetyn mallin tarkoituksena on auttaa sijoittamaan kukin teksti sille kuuluvaan retoriseen kontekstiin. Kunkin tehtävän suorittamista ja sen arvioimista koskeva kokonaisuus sisältää itse kirjoitustehtävän, maininnan tehtävän ensisijaisesta retorisesta piirteestä, selityksen siitä, miten tehtävänannon kunkin osan odotetaan vaikuttavan oppilaiden suoritukseen, varsinaisen arviointiohjeiston ja maamerkki-kirjoitelmia, jotka havainnollistavat eri arvioita arvioineen ja perusteluneen.

Kuvio 5 on varsin yksinkertainen, mutta se viestii hyvin piirrearviointin tavoiteorientoituneisuudesta ja osoittaa, millaisin kielellisin ja tekstuaalisin keinoin tavoitteisiin voidaan päästä. Kaiken kaikkiaan piirrearviointissa näyttäisi olevan kyse siitä, että arvioinnissa käytettävät kriteerit ovat määriteltävissä etukäteen kunkin kielenkäyttö-tilanteen tai diskurssityypin mukaan.

Piirrearviointin huonoja ja hyviä puolia

Arviointitapana piirrearviointi on sekä suunnittelun että itse arviointi-prosessin kannalta hidasta ja kallista, koska esimerkiksi erilaisia tehtäviä lukevan arvioijan tulee tutustua uusiin kriteereihin jokaisen tehtävän kohdalla ja kunkin piirteen arvioimiseen aina yhdessä tehtävässä voi mennä yhdestä kahteen minuuttiin. Kalleuteen vaikuttaa myös se, että kullekin tehtävälle on laadittava oma arviointiasteikkonsa piirre piirreeltä. Lisäksi piirrearviointi ei ole osoittautunut kovinkaan luotettavaksi tavaksi arvioida.

Piirrearviointia käytetään lähinnä tutkimuksen tarpeisiin, esimerkiksi NAEP:ssä (National Assessment of Educational Progress), joka on Amerikassa toimiva kansallinen tutkimusprojekti (Huot 1990). Vaikka NAEP:ssä käytetäänkin piirrearviointia suurten tutkimusaineistojen arvioimiseen, se sopii joidenkin näkemysten mukaan paremmin pienen aineistojen arviointiin.

Piirrearvioinnin hyvänä puolena on pidetty sen autenttisuutta ja mahdollisuutta ottaa huomioon kunkin kirjoittamistilanteen erityispiirteet ja -vaatimukset. Positiivista piirrearvioinnissa onkin sen pyrkimys välttää holistisen ja analyttisen arvioinnin rajoituksia. Piirrearviointi ei pyri yleistämään osaamista yhden tai kahden tehtävän perusteella, eikä arvioinnissa käsitellä taitoa, esimerkiksi kirjoittamista, yhtenä yleistettävänä kokonaisuutena vaan diskurssi- ja tehtäväkohtaisena.

3.4.3 Analyttinen arviointitapa

Analyttisessä arvioinnissa lähtökohtana on kirjoittaminen taitona. Kirjoittaminen nähdään kompleksisena ja monimuotoisena ilmiönä, tai ainakin oletetaan, että kirjoittamista täytyy tarkastella sen monista osasista käsin (Huot 1990, 238; Hamp-Lyons 1991d, 247–248). Analyttisen arvioinnin etuna pidetään sitä, että se kuvaa kirjoittamista tarpeeksi tarkasti ja yksityiskohtaisesti, koska kriteerit voidaan valita sen mukaan, millaisia komponentteja kirjoittamisen taitoon katsotaan kuuluvan. Analyttinen arviointi edellyttääkin kirjoittamistaidon yksityiskohtaista analysointia ja määrittelyä eli taidon purkamista kriteereiksi. Tämän lisäksi kukin kriteeri on sijoitettava taidon etenemistä kuvaavalle asteikolle (Bachman & Palmer 1996, 211).

Taitoa kuvaavat kriteerit muodostetaan perusteellisesti harkitun taitokäsityksen pohjalta, joskin kriteerien valintaan vaikuttavat monet muutkin tekijät. Esimerkiksi kriteerien määrän tulee olla arvioinnin kannalta käytännöllinen, kukin kriteeri pitää pystyä operationaalistamaan ja kriteerien tulisi olla tarpeeksi konkreettisia. Kriteerien valinta ja sisällöllinen määrittelemine ei ole kuitenkaan yksinkertaista, nopeaa eikä vaivatonta. Kriteerityöskentely vaatii vankkaa näkemystä taidosta ja sen oppimisesta sekä neuvottelua ja kompromisseja, koska kriteerit heijastelevat laatijoidensa käsitystä kirjoittamisesta taitona. Esimerkiksi sanasto, rakenteet ja ymmärrettävyys kriteereinä piirtävät omanlaisensa kuvan kirjoittamisesta taitona määrittelijänsä tai määrittelijöidensä mukaan.

Mikäli kriteerit edustavat monipuolisesti kirjoittamisen taitoa, niiden informaatioarvo on, esimerkiksi diagnostisessa mielessä, parempi kuin holistisessa arvioinnissa. Analyttinen arviointi tarjoaa yksityiskohtaisempaa palautetta kirjoittajan heikkouksista ja vahvuuksista ja vihjeitä niistä asioista, joissa vielä tarvittaisiin lisää harjoitusta. Kriteerikohtainen palaute luo oppijan taidosta myös eräänlaisen profiilin, jonka perusteella voidaan ymmärtää paremmin taidon kokonaisuutta. Jos esimerkiksi teksti on sanastoltaan ja syntaksiltaan hyvin kohdekielen omaista mutta oikeinkirjoitus hyvin epätarkkaa, oikeinkirjoituksesta

johtuvia puutteita saatetaan vahvuudet huomioon ottaen tarkastella ja kehittää tarkoituksenmukaisemmin.

Analyyttisen arvioinnin tulos voidaan raportoida kolmella tavalla: 1) kriteeri kriteeriltä, 2) yhtenä arviona, joka voi perustua aritmeettiseen tai painotettuun keskiarvoon tai vaikutelmaan yleistasosta, tai 3) sekä kriteerikohtaisina arvioina että yleisarviona (Huhta 1993, 150–151). Hamp-Lyonsin (1991d, 249) mukaan olisi suositeltavaa, että arvio raportoitaisiin kriteeri kriteeriltä varsinkin silloin, kun arvioinnin funktio on diagnostinen. Yhden arvion antamista vastaan sotii muun muassa se, että eri kriteerien keskinäinen painoarvo ei ole yksiselitteinen eikä tietyn kriteerin painottaminen tunnu taas tekevän oikeutta analyyttisen arvioinnin perusidealle. Vaikutelmaan perustuva kokonaisarvio ei taas välttämättä eroa kovinkaan paljon holistisesta arviosta, koska se ei tarjoa yhtään enempää informaatiota kriteerien sisällöistä ja vaikutuksesta tekstin kokonaisarvioon.

Mikäli eri kriteerien arviot kuitenkin yhdistetään, Hamp-Lyonsin (1991d, 249) mielestä olisi järkevintä painottaa jokaista kriteeriä saman verran. Mikäli kriteereitä painotetaan eri verran, painotus voi johtaa yhä lähemmäksi holistista ja yleensäkin vaikutelmapohjaista arviointia. Painotukseen päätyminen viestii jonkin yhden muita vahvemman ja merkittävemmän komponentin olemassaolosta. Tavallisin tapa raportoida tulokset näyttäisi kuitenkin olevan yhden yleisarvion antaminen, ainakin isommissa testijärjestelmissä. Käytännöt vaihtelevat kuitenkin tilanteen mukaan: se mikä on tavallista arviointijärjestelmissä, ei välttämättä ole tarkoituksenmukaista ja toimivaa esimerkiksi opetuksessa tai oppilaitosten omissa kokeissa.

Opettajan näkökulmasta analyyttinen arviointi tarjoaa haastavan mahdollisuuden kehittää tuottamistaitojen arviointia. Arviointikirjallisuudessa onkin esitetty yhä enemmän kritiikkiä siitä, että arviointitavat ja -asteikot ovat liiaksi testinkehittelijöiden tuottamia, joilla itsellään ei ole välttämättä kosketusta tai edes kokemusta vieraan kielen opettamisesta. Esimerkiksi Hamp-Lyons (1991d, 248) korostaa nimenomaan analyyttisessä arvioinnissa arvioijan sitoutumista ja sitouttamista arvioinnin kehittämiseen, jotta arviointi toimisi paremmin käytännössä. Analyyttisessä arvioinnissa pidetään tärkeänä myös koulutusta. Vaikka analyyttinen arviointi voi vaikuttaa luotettavalta ja ohjatulta sen yksityiskohtaisen luonteen vuoksi, kriteerien tulkintaan ja niiden soveltamiseen liittyvät tekijät saattavat kuitenkin aiheuttaa yksilöllisiä tulkintaeroja.

Joissakin kirjoittamisen arviointia vertailevissa tutkimuksissa on päädytty siihen, että analyyttinen arviointi on kaikkein luotettavin suoran kirjoittamisen arviointitavoista (esim. Veal & Hudson 1983). Toisaalta taas on myös tutkimuksia, joissa analyyttinen arviointi ei ole osoittautunut sen

luotettavammaksi kuin holistinenkaan arviointitapa tai kriteerien käyttö itsessään ei ole ohjannutkaan niiden yksimieliseen tulkintaan (esim. Lehman 1994; Waltman ym. 1998). Itse asiassa väittely siitä, kumpi on luotettavampi arviointitapa, ei näytä olevan tutkimusten mukaan kovinkaan hedelmällinen. Sen sijaan järkevämmältä tuntuisi kysyä, kumpi arviointitapa vastaa paremmin arvioinnin tarpeita ja on informaatioarvoltaan ja luonteeltaan tarkoituksenmukaisempaa kussakin tilanteessa.

Analyyttiseen arviointiin liittyviä ongelmia

Analyttisen arvioinnin uskotaan helposti olevan luonteeltaan objektiivisempää kuin holistisen arvioinnin, joka perustuu lähinnä yleisvaikutelmaan. Vaikka analyttinen arviointi tarjoaa yksityiskohtaisemmat ja tarkemmat työkalut, se ei kuitenkaan ole automaattisesti objektiivisempää (ks. esim. Davis ym. 1987). Itse arviointitapa ei siis riitä, vaan kriteerien sisällöistä ja niiden soveltamisesta tulee olla hyvin tietoinen. Kriteerien soveltamisen pitäisi olla samanlaista sekä eri arvioijien välillä että saman arvioijan arvioidessa eri suorittajia. Analyttisestä arvioinnista esitetyn kritiikin mukaan sen yksityiskohtaisuus on myös sen haitta, koska se pirstoo taidon niin pieniin osiin, että kokonaisuuden tarkastelu niiden läpi on usein mahdotonta ja arvioija voi usein katsoa tekstejä jonkin tietyn yksityiskohdan ohjaamana.

Toinen kriteereihin liittyvä ongelma on niiden tulkinta. Kriteerit voidaan määritellä hyvinkin yksiselitteisesti, ja käsitys niistä voi olla samansuuntainen eri arvioijilla, vaikka käytännössä jokaisen suorituksen kohdalla kriteereitä sovelletaan erilaisessa ympäristössä. Tässä mielessä kriteerien soveltaminen ei ole toistonomaista tai merkitykseltään vakiintunutta. Esimerkiksi suunnilleen samantasoinen rakenteiden hallinta voi näyttää hyvinkin erilaiselta sellaisessa kirjoituksessa, jossa on looginen jäsentely ja sisältö tehtävänannon mukainen, kuin sellaisessa kirjoituksessa, jossa ei ole käytetty lainkaan välimerkkejä ja jossa on paljon oikeinkirjoitusongelmia. Vaikka arvioija arvioisikin joka kerta kriteeri kriteeriltä, tuntuu epäuskottavalta väittää, ettei muu konteksti vaikuttaisi esimerkin mukaisesti rakenteiden arviointiin.

Kriteerien sisältöön vaikuttaa sekin, että ne ovat yksityiskohtaisuudesta huolimatta kokonaisuuksia, jotka taas muodostuvat erilaisista osatekijöistä (Huhta 1993, 153–154). Esimerkiksi sanastoa voidaan tarkastella merkityksen, tarkkuuden, varioivuuden ja idiomattisuuden näkökulmasta. Samoin rakenteiden hallinnassa on tehtävä päätös, millainen merkitys esimerkiksi verbien taiputuksella ja frekventtien sijamuotojen hallinnalla on suhteessa toisiinsa. Pelkkä ilmaus *rakenteet* on itsessään varsin laaja, joten jos rakennekuvauksessa ei oteta tarkemmin kantaa

niiden hallintaan, voi arvioija painottaa tietyn rakenteen osa-alueen hallintaa oman käsityksensä ja mieltymystensä mukaisesti.

Sen lisäksi, että kriteerit ovat laajoja, ne ovat myös varsin abstrakteja. Joskus on vaikea osoittaa tarpeeksi konkreettisesti, miten esimerkiksi kykyä jäsentää tekstiä tai koheesiota arvioidaan eri tasoilla. Joistakin taidon komponenteista on siis vaikea osoittaa toisaalta progressiota ja toisaalta taas niiden läsnäoloa. Toisin sanoen mitä tarkoittaa esimerkiksi taitotason 3 koheesio verrattuna taitotason 4 koheesioon konkreettisesti, millaiset kielelliset piirteet ja niiden puutteet tai vahvuudet osoittavat eron koheesiohallinnassa? Jos taas esimerkiksi käsi-alaa tai tekstin pituutta käytetään kriteerinä, ero on usein tehtävissä ennemminkin dikotomiana epäselvä–selvä tai lyhyt–pitkä kuin asteittaisena.

Ongelmallista on sekin, minkä verran arvioija ehkä tiedostamattaankin arvostaa eri kriteereitä, vaikka hän lukisikin tekstin kriteerit kriteeriltä. Jos arviointiin ei kuulu lopuksi tekstin lukeminen kokonaisuutena, mitä pirstalemainen lukeminen jättää huomioimatta? Toisaalta jos taas halutaan painottaa tekstiä kokonaisuutena, täytyy usein painottaa jotakin kriteeriä muita enemmän. Esimerkiksi kielitaidon alemmilla tasoilla ymmärrettävyyttä pidetään usein ensisijaisena kriteerinä ja kielen idiomaattisuutta taas ylempillä tasoilla.

Analyttisen arvioinnin suunnittelun näkökulmasta vaikeuksia aiheuttaa kriteerien lukumäärästä päättäminen. On todella vaikea valita järkevä määrä sellaisia kriteereitä, jotka kuvaisivat kirjoittamista tarpeeksi laajasti ja syvällisesti sekä tekisivät oikeutta sille taitona. Käytännössä kriteerien valinta perustuu usein kompromissiin. Hyvä kompromissi edellyttää kuitenkin, että kriteerit kuvaavat tarpeeksi monipuolisesti taitokäsitystä ja ovat sovellettavissa hyvin taidon arviointiin. Kriteerien kykyyn kuvata tai paremminkin edustaa kirjoittamista taitona on hyvä kuitenkin suhtautua varovaisesti, sillä kirjoittaminen on enemmän kuin kriteeriensä summa (ks. White 1986).

3.4.4 Holistinen arviointitapa

Cooper (1977, 4) määrittelee holistisuuden minä tahansa arviointina, jossa ei pysähdytä erikseen laskemaan lingvistisiä, tekstuaalisia tai muita vastaavia tekijöitä tekstistä. Holistinen arviointi voi nojautua siihen, että arvioija perustaa arvionsa yleisvaikutelmaan tekstistä, tai siihen, että hän arvioi jokaisen kriteerin erikseen, mutta häntä ei koskaan vaadita erikseen laskemaan tai pisteittämään tekstin puutteita ja vahvuuksia (esim. Charney 1984; Hamp-Lyons 1991d). Jos arvioija tukeutuu kriteereihin, hän käyttää kriteerilistaa vain yleisenä ohjeena, joka ohjaa häntä

kiinnittämään huomiota olennaisiin seikkoihin, mutta hän ei lue tekstiä kriteeri kriteeriltä.

Hamp-Lyons (1991d, 243) haluaa kuitenkin tehdä eron holistisen pisteittämisen (scoring) ja holistisen metodin välillä, koska niitä käytetään erheellisesti arviointikirjallisuudessa toistensa synonyymeina. Holistinen pisteittäminen on vain yksi holistisista metodeista, joihin kuuluu muun muassa myös piirrearviointi. Itse tarkoitan holistisella arviointitavalla holistista pisteittämistä, josta käytän ilmaisua holistinen arviointi. Holistisessa arvioinnissa lukuprosessi on siten holistinen, että arvioija voi arvioida suorituksen myös piirrekohtaisesti, mutta arviointiprosessin aikana hän ei pysähdy erikseen merkitsemään piirrekohtaisia arvioita. Holistinen arviointi antaa informaatiota lähinnä taitotasosta yleensä eikä jonkin yksittäisen piirteen hallinnasta.

Hiukan abstraktimmasta näkökulmasta katsottuna holistinen arviointi nojaa näkemykseen, jonka mukaan kirjoittaminen koostuu sille luontaisista ominaisuuksista, jotka ovat enemmän kuin kirjoittamisen osataitojen summa. Toisin sanoen kirjoittaminen ei ole yhtä kuin esimerkiksi sanasto + rakenteet + ymmärrettävyys + tekstuaalisuus = kirjoittamisen taito. Holistisen näkemyksen mukaan kirjoittamiselle taitona tehdään parhaiten oikeutta arvioimalla sitä holistisesti, jolloin kaikilla kriteereillä – myös niillä, joita ei voida tiivistää yksilöityihin kriteereihin – on mahdollisuus vaikuttaa arvioon.

Holistinen arviointi heijastelee arvioijan yleistä käsitystä tai mielikuvaa arvioitavan tekstin tasosta, sillä arvioijaa rohkaistaan luottamaan yleisvaikutelmaansa ja lukemaan tekstit melko nopeasti läpi. Holistista arviointia kuvataankin joissakin yhteyksissä silmäilevänä arviointitapana. Tekstistä syntyvä yleisvaikutelma ei ole kuitenkaan subjektiivinen, vaan se perustuu arviointiasteikossa tai -ohjeissa esitettyihin vaatimuksiin. Vaatimukset on tavallisesti kuvattu sanallisesti taso- tai pistemääräkohtaisesti.

Holistisille asteikoille on kuitenkin ominaista se, että niissä asteikkoväli on yleensä suppeahko; asteikot ovat yleensä joko 4- tai 6-portaisia (White 1986, 125; Faigley ym. 1989, 106; Kroll 1998, 227). Holistisessa taso- tai pistemääräkohtaisessa kuvauksessa viitataan johonkin taidon osa-alueeseen, esimerkiksi rakenteiden hallintaan tai tekstimäisyyteen. Arvioija ei kuitenkaan painota mitään osa-aluetta toista enempää vaan arvioi tekstiä yleisellä tasolla. Hän päättää pistemäärän tai tasoarvion sen mukaan, mikä taso- tai pistemääräkuvauksista vastaa parhaiten arvioitavaa suoritusta.

Holistista arviointia voidaan periaatteessa tehdä niinkin, että käytetään pelkkiä numeroita tai sanallisia kuvauksia, kuten *tyydyttävä – hyvä – erinomainen*, jolloin arvio perustuu täysin arvioijan subjektiiviseen

käsitykseen tyydyttävästä tai hyvästä suorituksesta (esim. Huhta 1993). On kuitenkin vaikea uskoa, että holistista arviointia käytettäisiin näin mielivaltaisesti missään virallisessa testisysteemissä. Holistisessakin arvioinnissa on siis oltava jonkinlainen yhtenäinen kriteeristö sille, mitä suoritukselta edellytetään kullakin taitotasolla tai kunkin pistemäärän kohdalla.

Charneyn (1984, 69) mukaan ne, jotka käyttävät holistista arviointia, pitävät sitä luotettavana, jos se täyttää seuraavat vaatimukset:

- sekä arvioijakoulutus että arviointitilanne ottavat huomioon ne seikat, joita reliabeliuden saavuttaminen edellyttää
- arvioijat ovat päteviä ja heillä on samanlainen tausta
- arvioijat on koulutettu tulkitsemaan ja soveltamaan arviointi-asteikkoa/-kriteerejä samalla tavalla
- kriteerit ovat tarkoituksenmukaisia
- arvioijat työskentelevät nopeasti ja ohjauksessa.

Charneyn esittämät vaatimukset vaikuttavat varsin tarkoituksenmukaisilta ja järkeviltä, joskin myös osin hiukan epärealistisilta. Ne ovat järkeviä siksi, että muun muassa koulutuksen, yhteisten jaettujen taitokäsitysten ja kriteerien relevanttiuden on tutkimusten mukaan todettu parantavan arvioinnin luotettavuutta (ks. myös taulukko 6). Itse asiassa holistisen arvioinnin toimivuuden edellytyksenä pidetään sitä, että huolellisesti valitut ja koulutetut arvioijat lukevat ja arvioivat tekstit (ks. Hamp-Lyons 1990). Toisaalta sellaista arviointia ei kai olekaan, minkä onnistumisen edellytyksenä ei pidettäisi arvioijien saamaa koulutusta ja asiantuntijuutta.

Vaatimukset vaikuttavat epärealistisilta taas siksi, että arvioijan työskentelynopeutta on vaikea ohjeistaa ja valvoa, ja siksi, että työn ohjaaminen käytännössä on yleensä hankalaa ja kallista. Lisäksi arviointiprosessin näkökulmasta kriteerit eivät toimi listana, vaan ne ovat keskenään kiinteästi vuorovaikutuksessa. Toisin sanoen yhden kriteerin ontuminen tai sen hyvä toimivuus tuntuu tavallisesti jossakin muussa kriteerissä ja edelleen koko prosessissa.

Vaikka holistinen arviointi ei joidenkin tutkimusten mukaan näytäkään olevan yhtä luotettavaa kuin analyttinen arviointi, sen arviot kuitenkin korreloivat sen verran hyvin analyttisen arvioinnista saatujen tulosten kanssa, että sitä voidaan pitää uskottavana vaihtoehtona analyttiselle arvioinnille (esim. Cooper 1977; Cumming 1990). Kaiken kaikkiaan erot arviointitapojen vertailussa ovat jääneet pieniksi. Joidenkin tutkimusten mukaan holistinen arviointi on jopa osoittautunut luotettavammaksi kuin analyttinen arviointi (ks. esim. Waltman ym. 1998).

Myös Cooperin (1977, 3) mukaan holistinen arviointi on luotettavin tapa, jos arvioijat ovat sitoutuneet arviointiin ja heillä on siihen riittävästi aikaa. Kunkin kirjoittajan suoritus voi saada .90 korrelaation holistisella asteikolla, vaikka arvioija ei käyttäisi enempää kuin kaksi minuuttia kuhunkin suoritukseen. Toisaalta holistinen arviointi saattaa vaatia kunkin suorituksen lukemista useampaan kertaan ainakin silloin, kun arvioija ei ole kokenut. Holistinen arviointi on kuitenkin yleisesti ottaen nopeampaa ja edullisempaa kuin analyyttinen arviointi, minkä vuoksi lähinnä amerikkalaisten tutkimusten mukaan se on käytetyin arviointitapa (ks. esim. White 1986; Davis ym. 1987; Connor 1991; Huhta 1993).

Holistisen arvioinnin rajoituksia

Holistisen arvioinnin rajoitukset liittyvät sekä arviointitavan lähtöoletuksiin että arviointikäytänteisiin. Ydinkysymys on, voidaanko kirjoittamisen taitoa yleensäkin määritellä holistisesti (vrt. Huhta 1993). Taidon määrittelemisen holistisesti voi olla vaikeaa, koska tuntuu mahdottomalta olla olematta analyyttinen. Tämä näkyy holistisen arvioinnin kuvauksista, joista tulee usein ainakin osittain analyyttisiä. Taidon kuvaaminen yleiselläkin tasolla pakottaa ottamaan kantaa siihen, mistä taito koostuu.

Esimerkiksi kirjoittamisen holistinen taitokuvaus tasolla 1 voisi kuulua seuraavasti: "Pystyy tuottamaan lyhyen, irrallisista sanoista koostuvan tekstipätkän, jota on vaikea ymmärtää." Kuvauksessa ovat mukana ainakin seuraavat taitokomponentit tai kriteerit: *pituus, lauserakenne, sanasto ja ymmärrettävyys*. Vaikka arvioija lukeekin arvioitavaa tekstiä holistisesti, hänen huomionsa kiinnittyy myös mainittuihin kriteereihin, ainakin jollakin tasolla. Miten arvioija pystyisi olemaan holistinen, jos sitä ei voi olla tasokuvauskaan?

Holistista arviointia on kritisoitu myös siksi, että sen diagnostinen informaatioarvo ja yleensäkin sen palautearvo on varsin alhainen, koska arvio itsessään ei kerro, mistä matala tai korkea arvio johtuu (esim. White 1986; Davis ym. 1987; Kroll 1998). Matala arvio voi olla merkki esimerkiksi huonosta rakenteiden hallinnasta, epäselvästä käsialasta tai tehtävänannon ymmärtämisestä väärin. Yhtä hyvin korkea arvio voi kertoa, että teksti on rakenteellisesti hyvää mutta se on sisällöltään tylsä. Holistisessa arvioinnissa ei välity myöskään yksityiskohtaisempaa tietoa esimerkiksi pintapuutteiden ja sisältöosaamisen välisestä erosta (Haswell & Wyche-Smith 1994, 228). Arvioinnin epävarmuutta lisää myös se, että holistinen arviointi jättää liikaa tilaa erilaisille arviointiin vaikuttaville tekijöille, ja se, että mahdollisista vaikuttavista tekijöistä ei tiedetä tarpeeksi (esim. Charney 1984; Huot 1990; Connor & Carrell 1993). Holistinen arviointi ei siis paljasta tarpeeksi avoimesti sitä, mihin arvioija on kiinnittänyt huomiota arvioinnissaan.

Hamp-Lyons (1991d, 243–245) tuntuu suhtautuvan kaikkein kriittisimmin holistiseen arviointitapaan, sillä hänen mielestään arvioijat eivät voi edes keskustella holistisesta arvioinnista ja jakaa siitä yhteisiä käsityksiä, koska perustelut eivät ole yhteisiä. Toisin sanoen arviointia ei ikään kuin katsota yhdessä sovittujen linssien läpi. Hamp-Lyonsin (1991d, 245) mielestä holistinen arviointi voi johtaa jopa siihen, että perusteluksi riittää se, että arvioija toteaa ”arvio on nelonen, koska minä sanon niin”, mikä on tietysti epätarkoituksenmukaista ja epäeettistä sekä käytännön että teorian kannalta.

Joidenkin näkemysten mukaan holistisesta arviosta viestittyy ennen kaikkea se, miten teksti sijoittuu suhteessa toisiin kirjoittajiin (Charney 1984; White 1986). Tästä seuraa myös se, että holistisessa arvioinnissa yliarvioidaan reliabiliuden merkitystä. Whiten (1986, 29) mukaan tällöin ei tulla kiinnittäneeksi tarpeeksi huomiota arvion tarkkuuteen eli siihen, miksi teksti saa kyseisen arvion, vaan tärkeämpää on, että teksti saa saman arvion eri arvioijilta. Korkean reliabiliuden tavoittelu on tuskin yksin holistisen arvioinnin tavoite tai sen ainoa tavoite.

Korkean reliabiliuden saavuttaminen voi olla helpompaa holistisessa kuin analyyttisessä arvioinnissa, mikä voi olla osaltaan syy siihen, että holistista arviointia pidetään reliabiliuskeskeisenä (ks. esim. Waltman ym. 1998). Toisaalta analyyttisessä arvioinnissa arviointia voidaan perustella pitkälti kriteerien avulla, kun taas holistisessa arvioinnissa perusteluna tai arvioinnin uskottavuuden vakuutteluna ehkä käytetään herkemmin arvioinnin luotettavuutta.

Vaikka holistinen arviointi on nopeampaa tehdä ja hinnaltaan edullisempaa kuin analyyttinen arviointi ja sitä näin suositaan käytännön syistä, sen valinta arviointitavaksi heijastelee muutakin, jotakin syvempää käsitystä siitä, millaisena taitona kirjoittaminen halutaan tietoisesti tai tiedostamatta nähdä. Sama koskee tietysti myös analyyttisen arviointitavan valintaa. Tiivistetysti sanoen holistisen näkemyksen mukaan kirjoittaminen taitona, kuten ihmisen toiminta yleensäkin, tulee hahmottaa ja ymmärtää kokonaisuutena, koska se on jotakin enemmän kuin osiansa summa (White 1986, 33).

Kumpi arviointitapa on sitten järkevämpää valita, jos siihen on mahdollisuus? Tutkimustuloksethan puhuvat molempien puolesta ja molempia vastaan (ks. esim. Charney 1980; White 1986; Huot 1990; Hamp-Lyons 1991d). Arvioijat itse suosivat holistista arviointia silloin, kun arviointiin on vähemmän aikaa ja kun arvioinnin tarkoitus on tarjota yleisluonteinen arvio. Analyyttistä suositaan taas vastakkaisista syistä: sitä pidetään parempana vaihtoehtona, kun arviointiin on enemmän aikaa ja kun testauksessa tai opetuksessa halutaan tarjota yksityiskohtaisempaa palautetta osaamisesta sekä opettajan että oppilaan näkökulmasta

(Waltman ym. 1998). Opettajien omat perustelut arviointitapojen käyttötarkoituksista kuvaavat hyvin sitä, miten keinotekoisista on taistella siitä, kumpi on parempi, toimivampi tai uskottavampi, koska molemmilla arviointitavoilla on rajoituksensa ja etunsa.

Tässä tutkimuksessa arvioijat käyttävät holistista asteikkoa, joka perustuu Kielten opetuksen ja oppimisen yleiseurooppalaiseen viitekehykseen (Council of Europe 1996; 2001; Huttunen 1997). Holistisen asteikon käyttöä ohjasi DIALANG-projekti, jonka käyttöön kirjoittamistehtävät on laadittu ja jossa eri osataitojen, lukemisen, puheen ymmärtämisen, kirjoittamisen asteikot ovat holistisia. Itse arviointiprosessin näkökulmasta, kuten olen osin edellä pohdiskellut analyttisen ja holistisen arvioinnin yhteydessä, arviointi ei ole kuitenkaan vain jompaakumpaa, vaan lukemisprosessi ja tekstin arvottaminen ovat luonteeltaan sekä holistista että analyttistä: arvioija päätyy oma-kohtaisesti tulkittuun yhteen tekstiä holistisesti kuvaavaan tasoarvioon tekstin eri piirteitä arvottamalla.

Tämän tutkimuksen kirjoittamistehtävien arvioinnissa on myös piirrearvioinnin piirteitä, koska tehtävät edustavat eri tekstityyppisiä ja kirjoittamisen eri funktioita. Asteikon kuvauksissa painotetaan tekstityyppien hallintaa (ks. liite 1), joten arvioija joutuu holistisesti arvioidessaankin ottamaan kantaa siihen, miten hyvin teksti täyttää tekstityypin ja sen edellyttämien tai edellyttämien funktioiden hallinnan, vaikka sitä ei asteikossa suoraan sanotakaan. Tältä osin tässä tutkimuksessa käytettävä holistinen asteikko edellyttää arvioijan arvioivan sellaisten piirteiden hallintaa, joiden tarkastelemista ei eksplisiittisesti ohjata asteikossa.

On siis oletettavaa, että tämän tutkimuksen asteikko jättää tulkinnanvaraa arvioijille. Tämä tulkinnanvara voidaan tulkita joko sattumanvaraisuutta ja epävarmuutta lisääväksi tekijäksi tai holistisuudeksi sen varsinaisessa merkityksessä. Tutkimuksen toinen ja kolmas arviointikierron osoittavat, miten yhdenmukaisesti ja johdonmukaisesti arvioijat arvioivat holistisella asteikolla, ja toinen haastattelu kuvaa taas sitä, miten arvioijat itse kertovat tulkinneensa ja soveltaneensa holistista asteikkoa.

4 KIRJOITTAMISEN ARVIOINTIPROSESSIN TARKASTELUA

4.1 Kirjoittamistaidon jäljillä

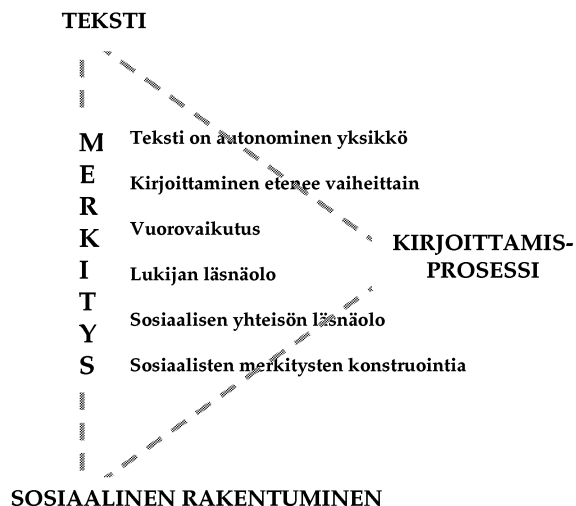
Tämä luku koostuu kolmesta asiakokonaisuudesta, joista ensimmäinen käsittelee kirjoittamista taitona sekä tutkimuksellisesta että Euroopan neuvoston yleiseurooppalaisen viitekehyksen näkökulmasta. Tässä yhteydessä nostan esiin myös ensimmäisen ja toisen/vieraan kielen kirjoittamisen välisiä eroja ja käsittelen lyhyesti suomi toisena kielenä - kirjoittamista.

Toisessa kokonaisuudessa tarkastelen kirjoittamisen arviointiprosessin luonnetta eli sitä, millaisia mekanismeja ja toimintamalleja arvostukseen liittyy arvion antamiseen eri näkemysten mukaan.

Kolmas kokonaisuus tuo esiin toisen kielen kirjoittamisen arviointiin liittyviä konkreettisia arvostuksia: mitä arvioija oikeastaan arvostaa kirjoittamista arvioidessaan? Luvun lopuksi pohdiskelen lyhyesti reliaabeliuden ja arvioijan toiminnan välistä yhteyttä. Neljännen luvun tarkoituksena on valottaa työni kvalitatiivista osaa: siinä missä kolmas luku on johdatusta arvioijien välisen yhdenmukaisuuden ja sisäisen johdonmukaisuuden tarkasteluun, neljäs luku luo pohjaa haastatteluille ja niiden tulkinneille.

4.1.1 Käsitteitä kirjoittamisesta taitona

Kirjoittamisen tarkastelun voidaan karkeasti katsoa muodostuvan kuvion 6 mukaisesti kolmiosta, jonka kärjissä ovat *teksti, sen tuottaminen* eli *kirjoittamisprosessi* ja *sosiaalinen rakentuminen* (social construction). Kolmion kärjet ovat olleet lähtökohtana tai viitekehyksenä sille, millaisena taitona toisen kielen kirjoittaminen on ymmärretty eri aikoina; painopiste on vaihdellut kontrastiivisesta tekstianalyysistä kirjoittamisprosessin kautta sosiaalisen yhteisön aseman korostamiseen (ks. esim. Grabe & Kaplan 1996; Cumming 1998). Käsitteet kirjoittamisesta taitona näyttäisivät seuranneen, kuten kielentutkimus yleensäkin, tieteen yleisiä kehityslinjoja.



KUVIO 6 Kirjoittamisen tutkimuksen kolme kärkeä.

Jos lähdetään liikkeelle kolmion yläkärjestä, tekstistä, tekstianalyysi on ollut pitkälti produktipohjaista eli tekstejä on tarkasteltu ja analysoitu lauseiden ja virkkeiden muodostamina itsenäisinä kokonaisuuksina (Flower & Hayes 1980, 32, Raimes 1991, 408–409; Luukka 1995, 51). Teksti on siis nähty formalistisesta näkökulmasta lausejaksojen ja sanaston kokoelmana, lähinnä kielellisenä tuotteena (Silva 1990, 13). Toisen kielen kirjoittamisessa on ehkä korostettu äidinkielistä kirjoittamista enemmän mikrotason eli *rakenteiden tarkkuutta* pragmatiikan kustannuksella (Kern & Schultz 1992, 1–2). Toisen kielen tutkimus on edelleenkin jossakin määrin formalistista, esimerkiksi silloin, kun lasketaan pelkästään passiivien ja pronomien määriä ja verrataan niiden käyttöä eri kulttuuritaustaisten välillä (Reid 1990).

Toisen kielen kirjoittamisen tarkastelu onkin 1960-luvulta lähtien ollut vahvasti virhe- ja kontrastiivisten analyysien tekemistä (esim. Johns 1990; Raimes 1998; Leki 2000). Kontrastiivisen retoriikan alulle pannut Kaplan (1966) havaitsi eri kulttuureista tulleiden koululaisten teksteistä, että ne erosivat amerikkalaisesta kirjoitusnormista ja että erot olivat yhdenmukaisia suhteessa kirjoittajan kulttuuritaustaan. Kontrastiivisen kielen vertailun ja kontrastiivisen retoriikan on katsottu hyödyntävän erityisesti pedagogisia tarkoituksia: tieto kulttuurisista ja kielellisistä eroista auttaa suuntaamaan toisen kielen opetuksen sisältöjä ja strategioita tarkoituksenmukaisella tavalla ja ennakoimaan oppijan

tulevia ongelmia (Kaplan 1988; Connor 1996; Sajavaara 1999). Kontrastiivisen retoriikan käyttö laajimmassa merkityksessään voi siis toimia tiedostamisen työkaluna (Raimes 1991, 418). Sitä pidetään myös yhtenä varteenotettavana lähestymistapana kirjoittamisesta keskustelemiseen (Connor 1996, 24).

Kontrastiivinen näkemys kirjoittamisesta on kuitenkin alusta alkaen herättänyt kriittistä keskustelua. Vastakkaisten näkemysten mukaan kontrastiivisuus edustaa liian muotopainotteista ja osoittelevaa näkemystä kielestä ja kirjoittamisesta. Kontrastiivinen metodi on myös vaikea hallita siitä syystä, että tutkijan on pystyttävä käsittelemään lukuisia erilaisia muuttujia pystyäkseen tunnistamaan erottelevat tekijät ja samalla varottava tekemästä liian karkeita yleistyksiä (Purves 1988, 15). Leki (1997, 239) mukaan kontrastiivinen näkökulma lisäksi yksinkertaistaa ei vain kulttuureja ja kirjoittamisen retoriikkaa vaan myös kohdekieltä eli hänen tapauksessaan englantia. Sama koskee oppijan lähtökieltä, joka tulkitaan kontrastiivisessa näkökulmassa helposti esteeksi, kun se päinvastoin toimii voimavarana (vrt. Cumming 1989, 127–128).

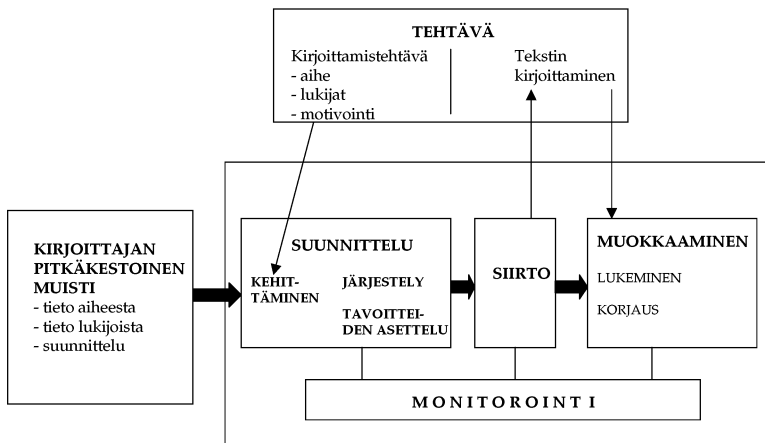
Kritiikkiä tarkasteltaessa on kuitenkin otettava huomioon, että kontrastiivinen tutkimus on muuttunut vuosien saatossa ja siitä on kehittynyt itse asiassa varsin monialainen tutkimusalue, jossa ollaan tekstitasoa laajemmin kiinnostuneita diskursseista (ks. esim. Connor 1996; Mauranen 2000). Kyse ei ole siis enää kaikilta osin pelkästään muotopohjaisesta kontrastoinnista. Tekstejä ei siis tarkastella tuotteina vaan mentaalisisina prosesseina; tekstiin liittyvät ominaisuudet eivät ole tekstin ominaisuuksia vaan kirjoittajan ajattelun tulosta. Näin ollen sekä tekstien analyyseissa että kirjoittamisen opettamisessa lähtökohtana on tekstin tuottamis- ja ymmärtämisprosessi (esim. Nystrand 1987; Raimes 1991; Kern & Schultz 1992; Leki 1993; Luukka 1995).

Kun kirjoittamisen tutkimuksen näkökulma on muuttunut 1970-luvun loppupuolelta tekstilähtöisestä kirjoittajalähtöiseksi, kirjoittamisessa ei niinkään ole enää korostettu tarkkuutta (accuracy) vaan merkityksen ilmaisemista, prosessia ja keksimistä. Kirjoittajalähtöinen näkökulma korostaa kirjoittajan kognitiivista kykyä käsitellä informaatiota kirjoittamisprosessin aikana ja jäsentää sitä tekstiksi erilaisten strategioiden avulla. Odellin ja Cooperin 1970-luvulla esittämän kirjoittamisprosessin määritelmän mukaan kirjoittaminen on vaativa, ennustamaton ja monimutkainen vuorovaikutusprosessi kirjoittajan ja tekstin välillä:

Kirjoittaminen on aiheen tutkimista ja miettimistä: suunnittelua, alkuun pääsemistä, tunteiden, arvojen tai ajatusten ilmaisemista; päätösten tekemistä sanavalinnoista, syntaksista ja retorisisista keinoista suhteessa

merkitykseen ja merkityksen muotoon; kirjoitetun editointia ja vielä kirjoitettavan ennakointia ja läpikäymistä; korjailemista ja uudelleenmuotoilua; pysähtymistä, lopullisen version tarkastelemista ja ehkä vielä lopuksi sen korjaamista. Tätä kompleksista, ennustamatonta ja vaativaa toimintoa kutsumme kirjoittamisen prosessiksi. (Odell & Cooper 1977: xi.)

Odellin ja Cooperin (1977, xi) mukaan kirjoittaminen on hyvin kirjoittajakeskeistä. Samanlaista näkemystä edustavat myös Hayes ja Flower (1980, 11), joiden mukaan kirjoittamisprosessi kompetentin kirjoittajan kohdalla voi koostua kolmesta vaiheesta: suunnittelu-, siirto- ja muokkaamisvaiheesta (ks. kuvio 7). Suunnitteluvaiheeseen kuuluvat kehittämisen, järjestelemisen ja tavoitteiden asettamisvaihe. Kehittämisellä, jota voidaan kutsua myös hakemiseksi, tarkoitetaan sitä, että kirjoittaja tukeutuu pitkäkestoiseen muistiinsa hakiessaan sieltä aiheeseen sopivia asioita. Sillä tarkoitetaan myös sitä, että hän suhteuttaa asiat ja tulevan tekstin lukijoille sopiviksi.



KUVIO 7 Kognitiivinen kirjoittamisen prosessointimalli Hayesin ja Flowerin (1980, 11) mukaan.

Järjesteleminen, jota kutsutaan myös organisoinniksi, tarkoittaa pitkäkestoisesta muistista haettujen asioiden järjestelyä ja kirjoittamisjärjestyksen laatimista. Kirjoittamisprosessin siirtovaiheessa kirjoittaja ottaa muistista materiaalia kirjoittamissuunnitelman mukaisesti ja muotoilee sen kirjoitettavan kielen mukaisiksi lauseiksi. Lauseiden muotoilu on välttämätöntä, koska muistissa olevat asiat eivät välttämättä ole kielellisiä vaan propositioita, joilla tarkoitetaan toisiinsa suhteessa olevia käsitteitä. Kolmas vaihe eli muokkausvaihe liittyy kirjoitetun

tekstin lukemiseen ja korjailemiseen, ja sillä pyritään kirjoitetun tekstin laadun parantamiseen.

Eri tutkimuksissa on yritetty selvittää sitä, minkä verran aikaa eri vaiheet vievät kirjoittajalta ja missä määrin ja millä tavalla kirjoittajat syvenyvät niihin kirjoittamisprosessin aikana. Näkökulmana on ollut esimerkiksi se, miten hyvät ja huonot kirjoittajat eroavat toisistaan prosessin eri vaiheissa (suunnittelu, siirto, muokkaaminen) (esim. Bereiter & Scardamalia 1987; Levy & Ransdell 1995).

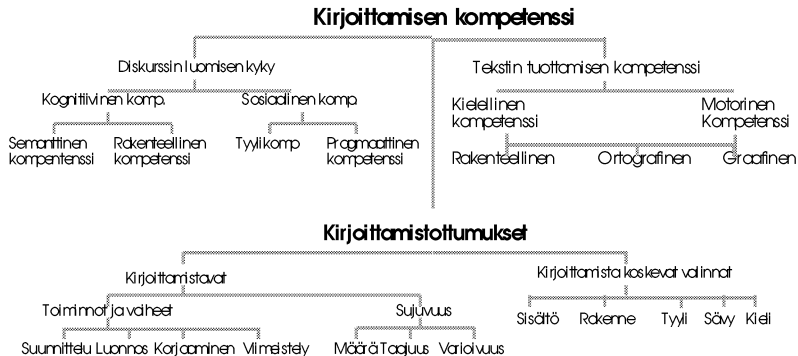
Kognitiivisesta kielentutkimuksen näkökulmasta on joidenkin mielipiteiden mukaan pystytty kuvaamaan kirjoittamisen prosessia kaikkein yksityiskohtaisimmin, vaikka esimerkiksi Hayesin ja Flowersin kuviossa 7 esitetty malli ei koskekaan kaikkia kirjoittamisprosesseja. Tämä johtuu muun muassa siitä, että Flowerin ja Hayesin (1980, 70) esittämän itsekritiikin mukaan kirjoittamista koskevat teoriat ja mallit ovat alustavia. Mallien yleistettävyyttä vaikeuttaa myös se, minkä Flower (1987, 4) itekin toteaa, että kirjoittamisprosessit ovat yksilöllisiä ja monenlaisten muuttujien alaisia. Yksilöllisyys ja yksilöllisen prosessin mallintaminen ovat kognitiivisessa tutkimuksessa johtaneet kuitenkin siihen, että malleissa on pystytty kuvaamaan vain tehokkaita ja sujuvia kirjoittajia eli jonkinlaisen ideaalikirjoittajan prototyyppiä.

Yksilöllisen prosessoinnin korostaminen on johtanut myös siihen, että useimmat kognitiiviset mallit ovat merkitys- tai tarkoituslähtöisiä ja niissä kirjoittaja kääntää tai muuntaa merkitykset tai tarkoitukset tekstiksi kirjoittamalla (esim. Nystrand 1986, 23). Prosessi siis alkaa kirjoittajasta ja hänen ajattelustaan ja päättyy tekstiin. Tosin myöhemmin esimerkiksi Flower (1989, 1994) on laajentanut näkökulmaansa sosiokognitiiviseen suuntaan, joka ottaa yksilön lisäksi huomioon kontekstin ja sen vaikutuksen prosessin kulkuun ja luonteeseen.

Miten yksilöllistä kirjoittaminen sitten oikeastaan on ja miten erillisenä, autonomisena taitona sitä voidaan pitää? Englanninkielisessä kirjallisuudessa käytetään literacy-käsitettä, jonka mukaan lukemista, kirjoittamista ja puhumista ei voida erottaa toisistaan, vaan ne kaikki nähdään sosiaalisina teksteihin liittyvinä toimintoina (esim. Rubin 1988; Venezky 1991; Barton & Hamilton 1998). Suomen kielessä literacyn vastineena on käytetty muun muassa tekstitoiminta- ja tekstitaidot-käsitettä (ks. Pitkänen-Huhta 1999; Luukka 2001; Leiwo & Luukka 2002). Ottamatta sen enempää kantaa suomenkielisten vastineiden osuvuuteen käytän tekstissänini rakentaidot-käsitettä viitatessani literacyyn.

Sosiaalisesti rakentuvassa kontekstissa kirjoittaja muodostaa kirjoittamisesta mentaalisen representaation sosiaalisena kontekstina, ja kirjoittaminen sosiaalisena prosessina voi luoda edelleen uusia sosiaalisia konteksteja. Kirjoittaessaan kirjoittajat siis luovat uusia tekstejä kollektiiv-

visesti yhteisön muiden jäsenten kanssa, lataavat kirjoittamiseen sitä myötäileviä arvoja ja siten konstruoivat sosiaalisia merkityksiä (esim. Faigley 1985; Rubin 1988; Luukka 1995; Beaufort 2000).



KUVIO 8 Purvesin (1990, 94) malli kirjoittamisen taidosta.

Purvesin (1990, 94) mukaan oppijan täytyy hallita kuvion 8 mukaisia taitoja, jotta hän tulisi kompetentiksi kirjoittajaksi. Ensisijaisena taitokomponenttina voisi pitää diskurssin luomisen kykyä, jotta kirjoittaja pystyisi ilmaisemaan haluamansa merkitykset tekstissään. Kuvioissa 7 ja 8 on yhteistä se, että kirjoittajan tulee pystyä suunnittelemaan sekä muokkaamaan ja järjestämään uudelleen muun muassa havaintojaan, ajatuksiaan ja mielikuviaan. Kognitiivinen kyky ei pelkästään riitä, vaan kompetenssin lisäksi kirjoittajan tulee pystyä käyttämään kieltä sosiaalisesti hyväksyttävästi ja tehokkaasti. Kirjoittamiseen liittyy myös tekstuaalinen kompetenssi, jolla tarkoitetaan niitä elementtejä, esimerkiksi rakenteita, oikeinkirjoitusta ja kirjoitussysteemin hallintaa, jotka tekevät tekstistä luettavan.

Purvesin (1990, 92–93) itsensä mielestä hänen mallinsa on varsin yleinen. Yksityiskohtaisempi esitys, esimerkiksi normien kuvailu, ei olisi kuitenkaan mielekäs, koska kukin sosiaalinen yhteisö luo ne tahollaan. Normistoja tai oikeastaan yhteisöjä on niin monia, että niiden kokoaminen yhteen malliin ei ehkä olisikaan mahdollista tai tarkoituksenmukaista, koska yhden ihmisen on mahdotonta olla niistä kaikista täysin tietoinen. Normistot ovat lisäksi dynaamisia ja myös neuvoteltavissa, joten ne voivat muuttua koko ajan tilanteittain.

Opetuksen näkökulmasta tekstitaidot-käsite tarkoittaa sitä, että kirjoittaminen, samoin kuin puhuminen tai lukeminenkin, ovat kiinteästi sidoksissa kulttuuriarvoihin sekä käsityksiin identiteetistä, persoonallisuudesta ja sosiaalisista suhteista. Kirjoittaminen ei siis ole mikään tekninen tai neutraali taito, vaan se on aina sidoksissa valtaan, ideologioihin ja niin henkilökohtaisiin kuin kollektiivisiinkin kulttuuriarvoihin (Nystrand 1990; Street 1994; Barton 1994; Beaufort 2000). Jos tätä käsitystä vertaa kuvioon 7, vaikuttaa näiden kahden näkökulmien ero varsin selvältä. Näkökulmat eivät välttämättä sulje pois toisiaan mutta ne painottavat kirjoittamisen tai kirjoittamisprosessin eri puolia hyvin eri tavalla. Tässä mielessä kirjoittamista voidaan katsoa kolmesta eri ikkunasta eli luvun alussa mainitun kolmion kulumista: tekstiä, yksilöä ja hänen prosessiaan tai sosiaalista yhteisöä painottaen.

Kolmio valottaa myös sitä, missä eri tutkimussuuntausten mukaan merkitys – ainakin epäsuorasti ilmaistuna – on. Painottaessaan tekstiä ominaisuuksineen kontrastiivinen näkökulma antaa ymmärtää, että merkitys on itse tekstissä, vaikkakin tietysti lukijan ymmärtämänä. Merkityksen tekstikeskeinen tai pelkästään tekstiin rajoittuva tulkinta viestii jossakin määrin sellaista mielikuvaa, että kussakin tekstissä on yksi oikea tulkinta, jonka eri lukijat tulkitsevat samalla tavalla. Äärimmillen vietynä merkitys ei ole edes tekstissä vaan kielessä, joka sitten siirtyy staattisena koodina kirjoittajan kautta myös hänen kirjoittamaansa tekstiin. Merkitys on jollakin tapaa ennustettava, ja se on löydettävistä nimenomaan tekstistä itsestään.

Kirjoittajalähtöinen näkökulma taas painottaa sitä, että merkitys on nimenomaan peräisin kirjoittajasta ja kieli on enemmänkin väline merkitysten välittämiseen. Merkityksen ilmaiseminen on mahdollista valtaosin Hayesin ja Flowerin (1980, 11) mallin mukaisesti. Epäselväksi jää, mikä lukijan osuus on merkityksen ymmärtämisessä. Kirjoittajan strategioiden ja kirjoittamiseen liittyvien kognitiivisten ja fysiologisten mekanismien painottaminen viestii siitä, että merkitys on kirjoittajalla ja hän muokkaa sen edelleen tekstiksi. Tekstin onnistuminen on taas riippuvainen siitä, miten tehokkaita ja toimivia kirjoittajan tuottamis-mekanismit ovat.

Toisaalta (socio)kognitiivisissa malleissa tuodaan esiin myös lukija (esim. Flower 1994). Niissä kirjoittamista kuvataan vuorovaikutuksena tekstin, kirjoittajan ja lukijan välillä, joskin kirjoittamisen kompetenssi on kiinni kirjoittajan kognitiivisista toiminnoista. Tekstin laatuun tai käsitykseen kirjoittajasta ei katsota vaikuttavan esimerkiksi sen, miten erilaiset lukijat voivat ymmärtää eri tavalla kirjoittajan tekstin. Tekstin kirjoittamista ja tekstin tulkintaa tarkastellaan vuorovaikutuskaavion eri päissä, irrallaan toisistaan.

Sosiaalisen näkökulman mukaan merkitys on enemmän yhteisökuin kirjoittajalähtöinen. Merkitystä ei tavallaan ole ilman sosiaalista yhteisöä. Merkitys ei siis ole olemassa kirjoittajalla autonomisena yksikkönä eikä se ole myöskään tekstissä, koska sekä lukijat että kirjoittajat toimivat osana sosiaalista yhteisöä – sen ohjaamina ja tietysti myös itse sitä ohjaten (esim. Nystrand 1990; Barton 1994; Luukka 1995). Yhteisön normit ja odotukset, joiden kautta merkityksiä tulkitaan, eivät ole vain kirjoittajan luomia eivätkä myöskään ainoastaan lukijan tulkitsemia, vaan yhteisön muodostavat kirjoittaja, teksti ja lukija kaikki yhdessä (Rafoth 1988; Nystrand 1990; Pitkänen-Huhta 1999). Normit ohjaavat ja suuntaavat merkityksiä mutta niiden vaikutus on joskus vaikeasta havaittavissa.

4.1.2 Kirjoittamisen tutkimuksen neljä näkökulmaa

Kirjoittamisen teoriaa on rakennettu neljästä eri näkökulmasta: prosessin, vuorovaikutuksen, sosiaalisen ja sosiokulttuurisen rakentumisen näkökulmista, joissa kaikissa tarkastellaan kirjoittamis- ja/tai lukemishetkeä siinä vaikuttavine tekijöineen eikä niinkään kirjoittamisen tulosta eli tekstiä (esim. Johns 1990; Raimes 1991; Hansman & Wilson 1998). Kokoamassani taulukossa 7 esitetyt suuntaukset ovat myös jonkinasteisessa aikajärjestyksessä, vaikkakin ne liukuvat osittain päällekkäin. Aikajana lähtee 1970-luvulta ja etenee tähän päivään.

Kuten taulukosta 7 näkyy, prosessiteoriassa voidaan erottaa toisistaan ekspressiivinen eli ilmaisullinen ja kognitiivinen suuntaus, joista ensimmäinen viittaa yksilölliseen ajatusten ilmaisemiseen ja liittyy näin ollen luovaan ja taidemaiseen kirjoittamiseen. Kyseisen suuntauksen mukaiset kirjoittamistehtävät perustuvat itsensä tutkimiseen ja painottavat kirjoittamisen vapautta. Kognitiivinen näkemys taas, kuten edelläkin kävi ilmi, määrittelee kirjoittamisen lähinnä ongelmanratkaisuksi, jossa avainasemassa ovat erilaiset kirjoittajan käyttämät strategiat. Kognitiivisella suuntauksella on ollut paljon vaikutusta myös toisen kielen, lähinnä englannin, kirjoittamisen tutkimukseen (Johns 1990, 25–26; Sasaki 2001, 62).

TAULUKKO 7 Kirjoittamisen tutkimuksen neljä pääsuuntausta.

Näkökulma		Mitä kirjoittaminen on?	Avainsanoja
P r o s e s s i	Ekspressiivinen	luovaa toimintaa, itsetutkiskelua, kirjoittamisprosessi yhtä tärkeä kuin teksti itse	kirjoittaja itse, luovuus, kirjoittamisen vapaus
	Kognitiivinen	ongelmanratkaisua, yksilöllinen prosessi, ei-lineaarinen, strateginen taito	ajatteleminen, mentaalinen prosessi, strategiat, tietoisuus omasta prosessista
Interaktiivinen / vuorovaikutuksellinen		dialogia lukijan kanssa, molemmilla on vastuu tekstistä, lukija on otettava huomioon kirjoittamisen kaikilla tasoilla (aihe, jäsentely, argumentit jne.)	lukijan huomioonottaminen = vastuu, lukijälähtöisyys, dialogisuus
Sosiaalinen		sosiaalista toimintaa, joka tapahtuu aina tietyssä kontekstissa tiety(i)lle lukija(o)ille, ympäröivän yhteisön ohjaamaa	yhteisöllinen diskurssi(t), sosiaalisuus, jäsenten samanmielisyyden (community of like-minded peers)
Sosiokulttuurinen		yhteisöllistä, toiminnallista konstruktivisesti, peilaa yhteisön ja kulttuurin uskomuksia, tietokäsityksiä jne.	kulttuuriarvot, identiteetti, sosiaaliset suhteet, tieto- ja uskomusjärjestelmät

Vuorovaikutuksellisen suuntauksen mukaan teksti on sitä, mitä yksilö luo dialogissa muiden vuorovaikutuksessa mukana olevien kanssa. Toisin sanoen sekä kirjoittaja että lukija ottavat vastuun tekstin ymmärrettävyydestä (Johns 1990, 30; Reid 1993, 47). Kirjoittajan kannalta vastuun ottaminen edellyttää aiheen, perustelujen ja jäsentelyn muokkaamista lukijaystävälliseksi. Kirjoittaminen on siis jatkuvaa neuvottelua lukijan tai lukijoiden kanssa merkityksistä, jotka eivät ole staattisia ja valmiiksi luotuina tai varastoituneina kirjoittajan pitkäkestoisessa muistissa (ks. esim. Nystrand 1990; Lähteenmäki & Dufva 1998; Hansman & Wilson 1998). Kirjoittaminen on jatkuvaa merkityksen luomista vuorovaikutuksessa kirjoittajan ja sosiaalisen ympäristön välillä.

Sosiaalisen rakentumisen näkökulman mukaan kirjoittaja on luova yksilö, jonka kognitiiviset prosessit ja rooli kirjoittajana ovat tutkimuksen kohteena. Näkemyksen mukaan kirjoittaja on fyysisen todellisuuden ympäröimä eli kirjoittamista ohjaavat esimerkiksi ympäröivä yhteisö arvoineen, asenteineen, odotuksineen ja normeineen. Yhteisö lukijoineen siis määrittää tiedon, kielen ja diskurssin luonteen (Johns 1990, 27; Barton 1994, 34–35, 57; Hansman & Wilson 1998, 23). Yhteisöllisiä diskursseja on monenlaisia, ja niiden luonne vaihtelee yhteisöittäin,

esimerkiksi ammattialojen mukaan. Toisen kielen oppijan näkökulmasta kirjoittamisen oppiminen tarkoittaa tietyn sosiaalisen diskurssin omaksumista. Monet toisen kielen kirjoittamisen tutkimukset käsittelevät esimerkiksi sitä, miten toisen kielen oppijat omaksuvat kohdekielisen akateemisen diskurssin (esim. Hamp-Lyons 1991; Mauranen 1993; Ventola & Mauranen 1996; Spack 1997; Lilis & Tuner 2001). Osa niistä pohtii myös sitä, miten mielekästä sen omaksumista on odottaa toisen kielen oppijoilta.

Sosiokulttuurinen näkökulma on näistä näkökulmista kaikkein lavein siten, että se lähestyy kirjoittajaa ja kirjoittamisen välityksellä konstruoituja merkityksiä sosiokulttuurisesta näkökulmasta. Kirjoittaminen ja tekstit potentiaalisine merkityksineen ovat osa diskurssikäytänteitä, jotka taas ovat osa sosiokulttuurisia käytänteitä (esim. Fairclough 1995; Pietikäinen 2000a).¹³ Näin ollen tekstien luomisen ja tulkitsemisen analysointi edellyttää myös uskomus- ja tietojärjestelmien, sosiaalisten suhteiden ja identiteettien tarkastelemista; jokainen teksti kaiuttaa ja muotoilee osaltaan sosiaalista yhteisöä ja kulttuuria (Halliday 1994; Fairclough 1995). Vieraan kielen kirjoittajan näkökulmasta kirjoittamisen oppiminen saattaa edellyttää toisenlaisen uskomusjärjestelmän ja tietokäsityksen omaksumista sekä oman identiteetin laajentamista.

Jos kirjoittamiskäsityksessä lähdetään liikkeelle kirjoittamisen monitahoisesta, vuorovaikutuksellisesta ja dynaamisesta luonteesta, ei yksikään suuntaus pysty yksin kuvaamaan prosessin kompleksista luonnetta muuttujineen. Vaikutelma kirjoittamisesta näyttää sitä täydemmältä, mitä useampi näkökulma tarkasteluun saadaan mukaan; kirjoittaminen ei ole vain kognitiivista, sosiaalista tai vain sosiokulttuurista, vaan se on niitä kaikkia (Williams 1993, 563; Eskey 1993, 224; Barton 1994, 166–168).

Arvioinnin näkökulmasta interaktiivinen ja sosiaalinen näkökulma voivat tuntua uhkaavilta: jos teksteillä on useita eri tulkintoja ja jos merkitys on lukijan ja kirjoittajan neuvottelun yhteinen tulos, voivatko arvioijat yksilöinä ymmärtää tekstit siinä määrin samalla tavalla, että yhdenmukaisuuden kriteeri voi toteutua? Esimerkiksi toinen arvioija voi pitää tekstin tyyliä loistavana ironiana ja toinen mauttomana tyylikkeenä (vrt. Kalliokoski 1996). Kysymyksiä herättää myös kirjoittamisen sosiokulttuurinen luonne: miten arvioinnissa on mahdollista tehdä ero esimerkiksi kulttuuristaustoista johtuvien erilaisten tietokäsitysten ja kielen hallinnasta johtuvien puutteiden välillä. Missä määrin

¹³ Käytän tässä sosiokulttuurinen näkökulma -käsitettä varsin laveassa merkityksessä; jätän ottamatta kantaa siihen, missä merkityksessä ja missä määrin tässä yhteydessä voitaisiin puhua kielenkäytön tutkimuksesta ja opettamisesta esimerkiksi kriittisen diskurssianalyysin näkökulmasta tai systeemifunktionaalista näkökulmasta (vrt. Pietikäinen 2000b; Luukka 2000).

kulttuuristen erojen huomioiminen on tarkoituksenmukaista ja oppijaa tukevaa ja missä määrin niiden korostaminen ehkä päinvastoin vahvistaa asenteita ja stereotyyppioita tai jopa luo niitä (vrt. Cazden 2000)?

4.1.3 Kirjoittaminen äidinkielessä ja vieraassa kielessä

Edellä esitetyt kirjoittamiskäsitykset tai näkökulmat kirjoittamiseen koskevat niin äidinkielen kuin vieraankin kielen kirjoittamisen tutkimusta. Olen tekstissäni jo edellä viitannut joihinkin äidinkielen ja vieraan kielen kirjoittamisen eroista ja yhtäläisyyksistä tehtyihin havaintoihin. Käsittelen kuitenkin tässä vielä lyhyesti sitä, millaisia käsityksiä äidinkielen ja vieraan kielen kirjoittamisen suhteesta on nostettu esiin sekä kirjoittamisen oppimisen että kirjoittamisprosessin näkökulmasta.

Käsitykset siitä, missä määrin ja millä tavalla ensimmäisen ja toisen/vieraan kielen kirjoittamisprosessit eroavat toisistaan, ovat jopa toisilleen vastakkaisia. Joidenkin käsitysten mukaan eroja on ja ne ovat suuria, ja toisten käsitysten mukaan eroja ei juurikaan ole (vrt. esim. Kaplan 1988; Cumming 1989; Raimes 1991; Silva 1993). Näkemykset eroavat myös sen mukaan, millaisesta paradigmasta käsin kirjoittamista tarkastellaan: siinä missä kontrastiivisen näkemyksen lähtökohtana on erojen havaitseminen, prosessitutkimuksen lähtökohtana on kaikelle kirjoittamiselle yhteisen mallin laatiminen.

Kirjoittamisen tutkimukseen kohdistuneen kritiikin mukaan tutkimus on painottanut liikaa sitä, miten tieto ja taito ensimmäisessä kielessä edesauttavat kehittymistä vieraassa kielessä. Äidinkielen ja toisen kielen välinen suhde ei ole välttämättä kausaalinen, vaikka niiden välillä onkin selvä korrelaatio: äidinkielessään sujuva kirjoittaja on todennäköisemmin sitä myös toisessa kielessä (Raimes 1985, 231; Bell 1995, 688). Esimerkiksi kirjoittamisstrategioiden, tekstin suunnittelun, kirjoittamiseen käytetyn ajan ja tekstin korjaamisen on osoitettu siirtyvän ensimmäisestä kielestä toiseen kieleen (Zamel 1983; Raimes 1985; Cumming 1989). Tosin näidenkään taitojen siirtyminen ei ole näin yksioikoista esimerkiksi silloin, kun kahden kirjoituskulttuurin välillä on suuria eroja (vrt. Cumming 1998, Sasaki 2000).

Silvan (1993) tutkimus osoittaa, että päällisin puolin äidinkielen ja vieraan kielen kirjoittamisprosessit näyttäisivät olevan vaiheiltaan hyvin samantyyppisiä, mutta erojakin on ja niitä on useanlaisia. Silvan (1993, 668) mukaan toisen kielen kirjoittajat yleensä kirjoittavat lyhyempiä tekstejä ja tekevät enemmän virheitä. Diskurssitasolla he eivät pysty pitämään selvästi toisista erillään argumentatiivisia, narratiivisia ja selittäviä tekstejä. Akateemiset tekstit, essee ja tekstipohjainen kirjoitus,

eroavat selkeästi syntyperäisten kirjoittamista teksteistä ja ovat tehottomampia. Lukijan huomioonottaminenkaan ei ole aina tarkoituksemukaista ja hyväksyttävää. Erojen vähentämiseksi niiden huomioonottaminen on tärkeää sekä opetuksessa että arvioinnissa (Silva 1993, 670–671).

Toisen kielen kirjoittajat tarvitsevat siis harjoittelua ei vain kielessä vaan myös strategioiden ja retoriikan hallinnassa (esim. Raimes 1985; Cumming 1989; Silva 1993; Reid 1993; Kalliokoski 1996, 2002). Myös toisen kielen opettajien ja oppijoiden genre-tietoisuuden lisääminen on tärkeää, jotta oppijoiden kirjoittamat tekstit täyttäisivät kunkin diskurssi-yhteisön vaatimukset (Silva 1990, 20; Ramanathan & Kaplan 2000, 180). Grabe (2001, 45–46) on tiivistänyt ensimmäisen ja toisen kielen kirjoittamistutkimuksissa esiin tulleita eroja seuraavalla tavalla:

1. Epistemologiset tekijät (esim. kulttuuri, arvot, uskomukset)
2. Kirjoittamisen funktiot
3. Kirjoitusaiheet (vrt. esim. yksilö- ja yhteisökeskeiset kulttuurit)
4. Tietopohja (äidinkieleen perustuva tieto voi monimutkaistaa vieraan kielen kirjoittamista)
5. Lukemisen pohjalta kirjoittaminen (lukemisen taito vieraassa kielessä vaikuttaa)
6. Tietoisuus lukijoista
7. Tekstuaaliset tekijät (kontrastiivinen näkökulma)
8. Kopiointi ja tekijänoikeudet (kuka omistaa sanat)
9. Muistaminen, jäljentäminen, lainaaminen
10. Oppijoiden oikeus omaan kielimuotoon/välikieleen (kenen "vieras" kieli on oikein?)

Lisäisin itse vielä listaan kirjoitusjärjestelmän omaksumisen. Graben luettelemat erot eivät tietenkään ole säännönmukaisia, eivätkä ne ole aina helposti havaittavissa. Ne oikeastaan vain osoittavat, minkä luonteisia erot voivat olla. Ne viestivät myös siitä, että vieraan kielen kirjoittajille on tavallista tehdä kirjoituksissaan luettelossa mainittuja 'virheitä' tai että virheiden esiintyminen on yleensäkin odotuksenmukaista ja luonteenomaista. Arvioinnin ja opetuksen näkökulmasta systemaattisempi tieto ja kokemus eroista auttaisivat niiden tunnistamisessa ja paikantamisessa sekä tarkoituksemukaisessa tulkinnassa. Toisaalta eroja voidaan helposti myös yleistää liikaa tai ylitulkita.

Vieraan kielen ja äidinkielen kirjoittamisen välisille yhtäläisyyksille ja eroille on tyypillistä, että ne ovat erisuuruisia ja niiden vaikutukset ovat erilaisia. Jonkin eron tiedostaminen ja siitä poisoppiminen voi olla hyvin olennaista ja aikaa vievää, kun taas ehkä jonkin toisen eron kohdalla on enemmänkin kyse lukijan suvaitsevaisuudesta kuin varsinaisesta

'ongelmasta'. Jokin eroavuus voi olla myös sellainen, että kirjoittaja ei halua luopua siitä säilyttääkseen identiteettinsä. Esimerkiksi kirjoittaja voi käyttää äidinkielelleen tyypillistä kohteliasta tyyliä, joka taas ei ole kohdekielessä tarkoituksenmukaista ja tehokasta.

Erojen tarkastelussa ja niistä tehtävissä tulkinnoissa on syytä olla tarkkana siinä, mikä on varsinaisesti kahden eri kielen kirjoittamisesta johtuvaa ja mikä taas seuraa jostakin muusta, esimerkiksi vaikeuksista sopeutua uuteen kulttuuriin tai yleisestä motivaation puutteesta. Eroille on melko tyypillistä myös se, että niitä ei ole kovin helppo havaita ja tunnistaa (Grabe 2001, 45). Lisäksi Raimesin (1985, 231) mukaan on yleensäkin vaikea tehdä eroa ensimmäisen ja toisen kielen kirjoittamisen välillä, koska tutkijoillakaan ei tunnu olevan yksimielisyyttä siitä, mitä *harjaantumattomalla* (unskilled) kirjoittajalla tarkoitetaan.

Jos kirjoittaja taas hallitsee kirjoittamisen, missä kulkee raja, jonka jälkeen hän on tarpeeksi taitava vieraassa kielessä? Millaista taitoa ja kenen taitoa tavoitellaan? Millainen on se syntyperäisen kirjoittajan prototyyppi, jonka mahdolliseen taitoon oppijan tekstiä verrataan? Arvioinnissa koetaan ongelmallisena usein syntyperäiseen vertaaminen, koska tällaista kompaktaa prototyyppiä ei oikeastaan löydy. Syntyperäisetkin kirjoittajat ovat varsin epätasaisia taidoiltaan. Toinen vertaamisen synnyttämä harha voi olla se, että syntyperäinen kirjoittaja kohoo kaikkien kirjoittamisen oppaiden ilmentymäksi. Hyvä kirjoittaja on siis sellainen, jonka kaltaisen taidon saavuttaa harva syntyperäinenkin. Tämä koskee kuitenkin vain hyvin edistyneen tason kirjoittajia, tai oikeastaan mitä parempi toisen kielen kirjoittaja on, sitä tiukemmaksi seula voi muuttua.

Joskus kuulee sanottavan, että vertaaminen on helppoa virheiden laadun perusteella: syntyperäiset kirjoittajat eivät tee tietynlaisia virheitä. Tällöin kirjoittaminen kuitenkin rajoitetaan helposti pelkkään kielelliseen, lähinnä kielen rakenteiden hallinnan tarkkuuteen. Lisäksi syntyperäisetkään eivät noudata täysin samaa kirjoitetun kielen tarkkuusnormistoa; esimerkiksi suomen kielessä poikkeuksia normin mukaisesta käytöstä ilmenee lauseenvastikkeissa, pronominiin viittaussuhteissa, joissakin rektioissa ja sanaston käytössä. Poikkeuksia normistosta siis esiintyy siitäkin huolimatta, että tietyntasoinen kaikille kirjoittajille yhteinen tarkkuus ja idiomaattisuus ovat osoitettavissa äidinkielisten kirjoittajien teksteistä.

Erojen tarkastelussa olisi siis tarkoituksenmukaista ottaa huomioon se, että syntyperäiset kirjoittajat ovat itsekin sekä kielen että kirjoittamisen oppijoita koko ikänsä – ainakin jossakin määrin. Samoin vieraan kielen kirjoittaja ei vain joko osaa tai ei osaa, vaan osaamisessa on liukumia eli hän voi jo hallita jonkin alueen, joskin epäsystemaattisesti. Itse asiassa

oppija voi olla oppimisen aloittamisesta hamaan tulevaisuuteen asti niin kutsutussa välikielivaiheessa eli matkalla kohti parempaa, systemaattisempaa ja automaattistuneempaa taitoa (ks. Selinker 1992). Joillakin vieraan kielen oppijoilla voi taidon edistyminen kielen rakenteissa myös pysähtyä eli fossiloitua (ks. Selinker 1992). Tästä huolimatta kielenkäyttäjää voi viestiä myös kirjallisesti ymmärrettävästi, vaikka teksti ei täyttäisikään esimerkiksi kielen rakenteiden tarkkuuden kriteeriä.

Edellä käsiteltyihin eroihin vaikuttavat tietysti myös oppijan lähtökieli ja kohdekieli. Tutkittujen erojen löytämisellä on kuitenkin yksi yhteinen implikaatio: tulokset osoittavat, että opetuksen ja arvioinnin näkökulmasta on tärkeää tietää, minkä luonteisia ja laatuksia erot ovat, mihin kirjoittamisprosessin vaiheisiin ne voivat liittyä ja mistä ne voivat johtua. Esimerkiksi kielellisiksi tulkitut virheet voivat johtua sosiaalisesta tietämättömyydestä eli kirjoittajalla ei ole tietoa siitä, miten eri genret ohjaavat sosiaalisia toimintoja (Ramanathan & Kaplan 2000, 182). Myös sisällön epäloogiselta tuntuva käsittely voi yksinkertaisesti johtua erilaisesta tavasta jäsentää asioita.

Prosessiluonteisen lähestymistavan lisäksi opetuksessa ja arvioinnissa tarvittaisiin enemmän tietoa siitä, miten kirjoittamisen tai yleensä tekstitaitojen eri komponenttien oppiminen etenee: missä järjestyksessä asioita opitaan ja mikä taito on toista relevantimpi ja miten esimerkiksi lukemis- ja kirjoittamistaito tukevat toisiaan. Lukemisen ja kirjoittamisen keskinäisestä vaikutuksesta on vallalla erilaisia näkemyksiä siitä, onko vaikutussuhde suora, epäsuora vai lähinnä vuorovaikutuksellinen (Carson 1990, 93). Näyttää kuitenkin siltä, että esimerkiksi hyvä lukemistaito saattaa edesauttaa edistyneiden strategioiden käyttöä (esim. Stotsky 1987, 386; Carson 1993, 98). Hyvät lukijat tunnistavat myös esimerkiksi paremmin sanoja ja kirjoittavat kompleksisempia lauseita ja yleensäkin sidosteisempia tekstejä (Gajdusek & vanDommelen 1993; 208). Varsinkin vieraan kielen opetuksen kehittämisen kannalta eri taitojen tarkoituksenmukainen integrointi olisi tärkeää.

4.1.4 Katsaus suomi toisena kielenä -kirjoittamisen ja sen arvioinnin tutkimukseen

Suurin osa toisen kielen kirjoittamisen tutkimuksesta liittyy englanti toisena kielenä -tutkimukseen, jonka tulokset eivät aina ole sovellettavissa muiden kielten kirjoittamiseen. Suomen kirjoittamista toisena/vieraana kielenä ei ole kuitenkaan kovin paljon tutkittu. Tutkimuksen kohteena ovat lähinnä olleet ruotsinkielisten koululaisten kirjoittamat aineet, joista on tehty peruslause- ja virhetyyppianalyysejä sekä tarkasteltu sanaston kehittymistä, verbisysteemin omaksumista sekä tempusten ja modusten

käyttöä (ks. esim. Andelin 1991; Grönholm 1991, 1992, 1993; Inkinen 1992; Lähdemäki 1995; Korkman 1995). Näkökulma on edellä mainituissa tutkimuksissa lähinnä kontrastiivinen, joten tekstejä on tarkasteltu pääasiassa suhteessa lähtökieleen.

Esimerkiksi Grönholm (esim. 1991; 1993) on tutkinut ruotsinkielisten peruskoululaisten 4., 6. ja 8. luokan oppilaiden suomen kielen kirjoittamia kertovia ja argumentoivia ainekirjoituksia. Grönholm (1991, 102) keskittyy erityisesti verbien morfologiseen ja leksikaaliseen omaksumiseen, ja hänen lähtökohtanaan onkin, että verbien käytön kehitys on kiinteässä yhteydessä kielenoppimisen tason kanssa. Vaikeuksia ruotsinkielisille oppijoille tuottavat Grönholmin (1993, 411) mukaan muun muassa pitkien ja lyhyiden äänneiden kuuleminen, oikeinkirjoitus ja suurtaajuisten sanojen käyttö.

Lähdemäki (1995, 197–198), joka on tutkinut ruotsinkielisten ylioppilaskokelaiden ja pieneen kielikokeeseen osallistuneiden tekemiä virheitä, listaa tyypillisimmiksi virheiksi muun muassa äänneiden kestoihin liittyvien vaikeuksien lisäksi sanavalintavirheet, inkongruenssityypit, ortografiset virheet ja nominatiivimuotoisen nesessiivirakenteen subjektin. Lähdemäen mukaan suurin osa virheistä johtuu puhekielen vaikutuksesta. Puhekielen merkitystä virheiden selittäjänä on pohtinut myös Korkman (1995), joka on tutkinut kaksikielisten peruskoululaisten aineita.

Mutta missä määrin edellä esitetyt virhetyypit ovat virheitä? Ovatko ne vain ruotsinkielisille kirjoittajille tyypillisiä virheitä? Tekeekö aikuinen toisen kielenoppija vastaavia virheitä? Kalliokoski (1996, 2002) osoittaa mielenkiintoisesti, että kirjoittamisen oppiminen ei ole vaikeaa vain toisen kielen oppijalle, vaan se on sitä myös äidinkieliselle. Kirjoitetun ja puhutun kielimuodon erot eivät ole selkeitä äidinkielisille kirjoittajillekaan (Kalliokoski 1996, 110). Leinon (1992) mukaan äidinkielisten ylioppilaskokelaiden teksteissä puhekieleen viittaavat muun muassa ääntämisen ja kirjoituksen vastaamattomuus, subjektin ja predikaatin kongruenssi, omistusliite ja yhdyslauseet.

Puhutun ja kirjoitetun kielen erot osoittavat, että suomen ääntämyksen ja ortografian suhde ei ole aivan yksi yhteen, puhutun ja kirjoitetun kielen normit ovat monin paikoin ristiriidassa keskenään ja erilaiset käsitykset normien soveltamisesta eri tilanteissa aiheuttavat tulkintaongelmia kirjoittajan ja lukijan kannalta (Kalliokoski 1996, 110). Toisen kielen oppijalle kirjoittamisen oppiminen on siis tavallaan kolmas kielimuoto äidinkielen ja suomen kielen puhutun kielimuodon lisäksi.

Puhutun ja kirjoitetun kielen erojen huomioonottaminen edellyttää myös virheiden uudelleen tulkintaa. Oppijan tekemistä virheistä ei ole kuitenkaan aina helppo päätellä, mikä johtuu toisen kielen omaksumi-

sesta ja mikä kirjoitetun kielimuodon normin rikkomisesta. Toisaalta virheetkään eivät ole pelkästään kielen rakenteiden tarkkuuteen liittyviä, vaikka toisen kielen oppijan tekstissä saattaa olla helpointa puuttua kielen rakenteellisiin puutteisiin suomen kielen synteettisen luonteen vuoksi. Virheiden, joita oikeastaan kutsuisin mieluummin sopimattomiksi tai puutteellisiksi ilmauksiksi, luonne on paljon monikerroksisempi. Kirjoittamisen ja tekstin ymmärtämisen kannalta olennaisia ovat myös tekstin jäsentelyyn, sisällön ilmaisemiseen, tyyliin, eri tekstityyppien funktioiden ja diskurssien hallintaan liittyvät ongelmat. Näiden tekijöiden opettaminen ja arviointi muuttuvat oletettavasti sitä monimutkaisemmiksi, mitä useammista lähtökulttuureista ja äidinkielistä on kyse. Diskurssin tai eri genrejen hallintaan liittyviä tekijöitä on kuitenkin tutkittu vähän suomi toisena kielenä -kirjoittamisesta.

Aikuisen oppijan oppimiskonteksti ja oppimisen tavoitteet eroavat myös jonkin verran koululaisesta. Aikuisten opetusryhmät ovat usein hyvin heterogeenisiä, ja kielen oppimisen tavoitteena on tavallisesti oppia sellainen kielitaito, jolla selviää auttavasti arkielämän tilanteista. Kirjoittamisen oppiminen voi olla toissijainen tavoite. Lisäksi kirjoittamisen opettamiseen ja oppimiseen vaikuttavat muun muassa oppijan lähtökieli, kirjoitus- ja lukutaito, koulutus ja kyky omaksua kirjoitetun kielen normisto sekä hänen lähtökielensä kirjoitusjärjestelmä. Kirjoittamisen opettamiseen, oppimiseen ja arviointiin liittyy siis monenlaisia tekijöitä ja haasteita.

Maahanmuuttajien, peruskoulun päättövaiheessa olevien ja aikuisten, kielitaitoa on selvitetty kahdessa Opetushallituksen hankkeessa. Suni (1996) on tutkinut peruskoulun päättövaiheessa olevien maahanmuuttajien suomen kielen taitoa ja Pälli ja Latomaa (1997) ovat selvittäneet aikuisten maahanmuuttajien kielitaitoa ja arvioineet maahanmuuttajakoulutusta.

Sunin (1996) tutkimukseen osallistui 90 maahanmuuttajaoppilasta 18 yläasteelta eri puolilta maata. Tulosten mukaan maahanmuuttajaoppilaista neljännes saavutti Yleisten kielitutkintojen kirjoittamisessa taitotason 3, mikä tarkoittaa, että oppijat pystyvät kirjoittamaan esimerkiksi kirjeitä ja viestejä tuttavilleen ja nämä viestit ovat pääosin ymmärrettäviä (Sunin 1996, 64–65). Heikoimmin menestyneiden oppijoiden kielitaidon ongelmia olivat sanaston ja rakenteiden hallinnan heikkous ja ortografian vieraus. Heikoimmat oppijat kuitenkin selviytyivät paremmin kirjoittamisesta kuin tekstin ymmärtämisestä, mikä voi johtua siitä, että kirjoittamisessa oppija voi itse kontrolloida kielen käyttöä. Kaiken kaikkiaan maahanmuuttajaoppilaiden suullinen kielitaito oli kuitenkin keskimäärin kirjallista vahvempi (Sunin 1996, 67).

Pällin ja Latomaan (1997) tutkimukseen osallistui 90 aikuista maahanmuuttajaa 7 eri oppilaitoksesta ympäri Suomea. Tutkimuksen mukaan aikuisista maahanmuuttajista yli puolet sai Yleisten kielitutkintojen kirjoittamisesta taitotason 3 tai 4, mikä tarkoittaa taitotasolla 4, että teksteissä ilmenee jo sidoksisuutta ja rekisterien hallintaa ja tekstit ovat pääosin ymmärrettäviä (Pälli & Latomaa 1997, 52–53). Myös Pällin ja Latomaan (1997, 55) tutkimuksessa aikuisten kielitaidon vahvin osa-alue oli puhuminen.

Molemmat tutkimukset osoittavat, että kirjoittamisen oppimiseen liittyvistä haasteista huolimatta kirjoittamisessa voidaan saavuttaa sellainen taso, jolla selviää ainakin lyhyiden ja henkilökohtaisten tekstien kirjoittamisesta. Arviointi on tosin tehty sekä Sunin että Latomaan ja Pällin tutkimuksissa Yleisten kielitutkintojen testillä, mikä on otettava huomioon kielitaidosta tehtävissä päätelmissä. Pelkästään jo testitilanne tai tekstityyppien ja aiheiden variaatio kyseisessä testissä voivat vaikuttaa tuloksiin. Yleisten kielitutkintojen tehtävien etu on kuitenkin se, että tehtäviä on useampi kuin yksi ja ne edustavat eri tekstityyppisiä ja edellyttävät erilaisten funktioiden hallintaa.

Jos testitulos kertoo oppijalle, että hän pystyy kirjoittamaan ymmärrettäviä viestejä, voiko hän luottaa siihen, että hänen tekstiään ymmärtää kuka tahansa suomalainen? Entä miten arvioijat suhtautuvat suomi toisena kielenä -teksteihin; ymmärtävätkö he 'liikaa' vai edellyttävätkö he teksteiltä enemmän kuin ymmärtämisen kannalta olisi tarpeellista? Kirjoittamisen arvioinnilla on pitkä historia finskan osalta; sen sijaan eritasoisten taustoiltaan heterogeenisten suomen oppijoiden tekstien lukemisella ei ole vielä pitkiä perinteitä. Nyt olisikin hyvä aika pohtia, miten suomen oppijoiden teksteihin suhtaudutaan, millainen ja minkä tasoinen kirjoitettu suomen kieli on hyväksyttävää ja kenen näkökulmasta.

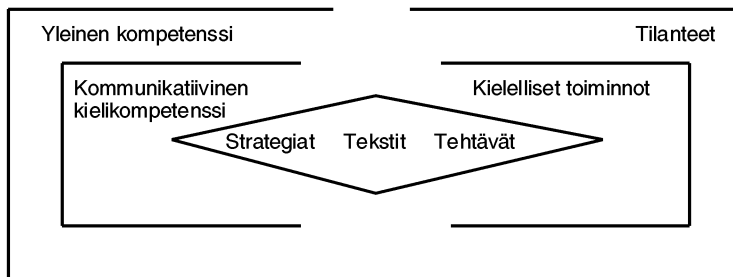
Enqvistin (1990, 26) mukaan tekstin tulkinta on riippuvainen tilanteesta ja siitä, missä määrin lukijan ja kirjoittajan käsitykset maailmasta kohtaavat. Käsityksissä on oltava jotakin ristiriitaista, koska toisen kielen oppijoiden teksteihin tunnutaan suhtauduttavan eri tavalla kuin äidinkielisten teksteihin? Lukijoiden huomiota kiinnittävät esimerkiksi tyylin rohkea käyttö, sanajärjestys, kielikuvat ja kirjoitetun kielen normeista poikkeavat kielimuodot (Kalliokoski 1996, 117; Latomaa 1996, 102).

Suomi toisena kielenä -kirjoittamisen arvioinnin yhteydessä joutuu pohtimaan, mikä kirjoittamisessa on yleensäkin kielellistä hallintaa, mikä kirjoitetun kielimuodon hallintaa, mikä sosiokulttuurista tietoa, mikä universaalialia, mikä oppijan äidinkielestä johtuvaa, mikä kirjoittamistottumusten tai -tottumattomuuden aiheuttamaa, mikä eri testilajien

hallintaa ja mikä ehkä taas arvioijan subjektiivisesta tulkinnasta, arvostuksista, odotuksista ja uskomuksista johtuvaa. Kysymyksiin vastaaminen ja niiden huomioon ottaminen kirjoittamisen opetuksessa ja arvioinnissa edellyttävät nykyistä monipuolisempaa tutkimusta suomi toisena kielenä -kirjoittamisesta.

4.1.5 Kirjoittamiskäsitys tässä tutkimuksessa

Käsillä olevan tutkimuksen kirjoittamiskäsitys perustuu Euroopan neuvoston yleiseurooppalaiseen viitekehykseen, koska kirjoittamistehtävät ja toinen arvioinnissa käytetyistä asteikosta perustuvat siihen. Viitekehys pohjautuu vahvasti käsitykselle, jonka mukaan kielen käyttö ja oppiminen ovat toimintaorientoituneita (action-oriented), eli kielenkäyttö edellyttää tiettyä kommunikatiivista kompetenssia ja strategioiden käyttöä, jotta kielenkäyttäjä pystyy toimimaan tilanteen edellyttämällä tavalla ja siten saavuttamaan tavoitteensa (ks. kuvio 9).



KUVIO 9 Toimintaorientoitunut lähestymistapa (Council of Europe 1996, 10).

Kuvion 9 mukainen *yleinen kompetenssi* viittaa tietoihin ja taitoihin, jotka mahdollistavat yleensäkin sosiaalisen intension eli tarpeen toimia. Tähän liittyvät niin persoonallisuuspiirteet, asenteet, kokemukset kuin opittu tietokin. *Kommunikatiiviseen kompetenssiin* kuuluvat sosiolingvistiset, kielelliset ja pragmaattiset komponentit, joiden avulla sosiaalisen intension täyttäminen kielellisesti on mahdollista (Council of Europe 2001, 9–10).

Näistä kahdesta kompetenssista sosiolingvistinen kompetenssi on linkki muihin kompetensseihin, eli se on hyvin vahvasti yhteydessä

kulttuuriin ja sosiaalisiin konventioihin. Sen vaikutus kulttuurien välisessä viestinnässä on huomattava, vaikka se on usein myös se alue, jonka vaikutuksesta kielenoppijat eivät ole tietoisia. Yhteisön ja kulttuurin vaikutusta on myös vaikea erottaa kielellisistä ja pragmaattisista tekijöistä, koska raja on usein varsin häilyvä. (Council of Europe 1996; Council of Europe 2001.)

Kuvion 9 *kielelliset toiminnot* kuvaavat kommunikaation luonnetta, sitä, onko se reseptiivistä, produktiivista vai vuorovaikutuksellista. Kielelliset toiminnot voivat olla kirjallisia, suullisia tai molempia. *Tilanteet* taas kuvaavat sosiaalisen elämän erilaisia alueita, jotka ovat niitä konteksteja, joissa kielenoppija viestii. Kielenopetuksen näkökulmasta viitekehys jakaa tilanteet neljään kategoriaan: julkisiin, henkilökohtaisiin, koulutuksellisiin ja ammatillisiin. Kielellisten toimintojen luonne ja sisältö vaihtelevat eri tilanteiden mukaan.

Vinoneliön sisässä olevasta kolmesta termistä *tekstit* viittaavat tässä yhteydessä mihin tahansa puhuttuihin tai kirjoitettuihin diskursseihin, jotka liittyvät aina johonkin tiettyyn sosiaaliseen tilanteeseen. *Strategiat* taas mahdollistavat tehtävän mielekkään, tarkoituksenmukaisen ja organisoidun toteuttamisen, eli strategioiden avulla oppija pystyy tuottamaan esimerkiksi kirjeen sen mukaan, miten automaattistunut hänen kirjoittamistaitonsa on. Kommunikointiin ja yleensäkin oppimiseen liittyy monenlaisia toimintoja eli *tehtäviä*. Tehtävät ovat sellaisia toimintoja, joiden avulla kommunikoija saavuttaa intentionsa. Tällaisia tehtäviä voivat olla esimerkiksi kirjan kirjoittaminen, pelien pelaaminen tai ruuan tilaaminen ravintolassa.

Vieraan kielen oppija on siis tekemässä monenlaisten muuttujien kanssa, kun hän kommunikoi vieraskielisessä ympäristössä. Miten viitekehys sitten lähestyy kirjoittamista edellä kuvatusta lähtöoletuksesta käsin? Kielitaito jaetaan sekä produktiivisiin että reseptiivisiin taitoihin, joista edelliseen kirjoittaminen kuuluu. Kirjoittaminen määritellään varsin mekaanisesti siten, että tavoitteena on tekstin tuottaminen ja tuotetulla tekstillä on yksi tai useampi lukija. Kirjalliseen tuottamiseen kuuluvat muun muassa lomakkeiden täyttäminen, muistiinpanojen tekeminen, artikkelin kirjoittaminen, fiktiivinen kirjoittaminen ja kirjeiden kirjoittaminen. Kirjoittamisessa on siis tärkeää se tarkoitus, johon kirjoittaja pyrkii tekstillään, esimerkiksi informaation antamiseen lomaketta täyttäänsä tai vaikuttamiseen artikkelia kirjoittaessaan.

Myös viitekehyksessä esitetyt tietyn kielitaitotason saavuttamista kuvaavat kirjoittamisen asteikot lähestyvät kirjoittamista lähinnä tekstityypeistä käsin. Liitteessä 1 kuvatusta kirjoittamisen taitotasosteikosta voi havaita, että vähäisellä taidolla selviytyy lyhyiden ja yksinkertaisten viestien kirjoittamisesta ja hyvällä kielitaidolla taas

raporttien ja argumentoitujen tekstien kirjoittamisesta. Kirjoittaminen määritellään asiakirjassa ja asteikossa paljon mekaanisemmin kuin kuviossa 9. Kommunikoinnista ja kielenkäytöstä syntyy moniulotteisempi ja kompleksisempi mielikuva taustaoletusten kuin itse kirjoittamis-kuvauksen mukaan. Viitekehysten mukaan kirjoittamisen taitokäsitys näyttäisi pohjautuvan lähinnä tekstityyppien tuntemukseen sekä virheiden määrään ja kirjoittamisen tekniikan hallintaan.

4.1.6 Miten käsitys kirjoittamisesta ja arviointi kohtaavat?

Kirjoittamisen arvioinnin niin kuin toisen kielen opetuksenkin näkökulma on ollut pitkälti rakennekeskeinen (Fathman & Whalley 1990, 178; Raimes 1995, 519; Filer 1995, 28). Kirjoittamista on arvioitu tekstikeskeisesti eli produktilähtöisesti. Produktilähtöisyyden yhtenä syynä on voinut olla se, että tekstistä on helpompi arvioida konkreettisesti havaittavia yksiköitä kuin näkymättömältä tuntuva prosessia (vrt. Luukka 1995, 51–52). Produktilähtöisyyttä on perusteltu myös siten, että kyse on suorasta näytöstä eikä tulkinnan tulkinnasta. Prosessista ei voida sanoa mitään, koska siitä ei ole suoraa näyttöä. Toisaalta prosessin arviointi ei onnistuisikaan produktilähtöisesti, vaan se edellyttää prosessikeskeisiä työ- ja arviointitapoja.

Pelkkä produktilähtöisyys ei tee kuitenkaan oikeutta kirjoittamisen arvioinnille, koska kirjoittamisen arviointi on ensisijaisesti lukijakeskeistä; tekstin laadun ja kirjoittamisen taidon arviointi on kiinni siitä, miten arvioija lukijana tulkitsee ja arvottaa tekstiä. Kun lukija ymmärtää tekstin, silloin tekstin ja kirjoittajan perustarkoitus on täytynyt eli merkitys tulee merkitykselliseksi (ks. esim. Nystrand 1986). Ymmärtäminen tai ymmärrettävyys toteutuu silloin, kun teksti luo tasapainon kirjoittajan ilmaisutarpeiden ja lukijan ymmärtämistarpeiden välille (Nystrand 1986, 41–45). Nystrandin (1986, 40) mukaan kirjoittaminen ei olekaan usein esitettävien väitteiden mukaisesti vähemmän vuorovaikutuksellista kuin puhuminen.

Kirjoittamisen arvioinnissa on toki otettava huomioon se seikka, että kirjoittaminen arviointitilanteessa on aina jossakin määrin keinotekoista. Tehtävät, vaikka ne olisivatkin autenttisia, kirjoitetaan arviointia varten eikä todellisen tarpeen vuoksi. Onnistuneisuutta ei myöskään arvioi todellinen vastaanottaja, vaan opettaja tai joku ulkopuolinen tuntematon arvioija. Tämä voi johtaa siihen, että tekstin, esimerkiksi valituskirjeen, onnistuneisuutta ei arvioida todellisen vastaanottajan tai sisällön näkökulmasta vaan sen mukaan, miten se kielellisesti ja muodollisesti täyttää tehtävän. Oppijoiden teksteihin tottuneet opettajat ja arvioijat

voivat ymmärtää ja yliymmärtää myös sellaisia tekstejä, joita maallikko ei ymmärtäisi.

Vaikka testiyhteyksissä arvioija ei tavallisesti tunne kirjoittajaa, kirjoittaja on kuitenkin läsnä tekstin välityksellä ja arvioijalle syntyy hänestä erilaisia mielikuvia. Tämä tulee esiin esimerkiksi seuraavantalaisissa tämän tutkimuksen arvioijien kommentteissa: ”tää kirjottaja on varmasti näitten virheiden perusteella venäläinen”, ”tää teksti on varmasti lääkärin kirjottama ku se osaa tälläsiä sanoja”. Itse asiassa jokainen kirjoittaja tuo persoonallisuutensa kirjoittamistehtävään, mikä on kirjoittajan ja tekstin mukaan eriasteisesti havaittavissa. Kukin kirjoittaja on kompleksinen yhdistelmä muun muassa kokemuksia, käsityksiä, tunteita ja mielipiteitä, mikä heijastuu myös hänen kirjoittamisensa teksteissä.

Flowerin (1994, 243–250) mukaan kirjoittajaa voidaan tarkastella affektiivisesta, kognitiivisesta ja kontekstuaalisesta näkökulmasta. Kirjoittajan affektiivisuus eli asenteet, tunteet, motivaatio ja minä-käsitys säätelevät pitkälti sitä, miten kirjoittaminen käynnistyy ja miten prosessi kulkee eteenpäin eli miltä kirjoittaminen kirjoittajasta tuntuu. Sen sijaan kognitiiviset tekijät, kuten tavoitteet, strategioiden käyttö ja metakognitiivinen tietoisuus, liittyvät kirjoittamiseen ajatteluprosessina, päätöksen tekemisenä ja ongelmanratkaisuna. Tilanne eli kirjoittajan kokemukset, oletukset ja kirjoittamisen funktio taas vaikuttavat arvioijan päätelmiin kirjoittamisesta sosiaalisena toimintana.

Kirjoittamistehtävät toimivat kirjoittajan tuottaman tekstin kautta tavallaan välittäjinä kirjoittajan ja lukijan välillä sekä äidinkielen ja toisen/vieraan kielen välillä. Jokaisen kirjoittajan on sovittava oma maailmansa tehtävän maailmaan, ja hänen on tulkittava tehtävä itsensä kannalta järkeväksi, ennen kuin hän voi vastata siihen. Mitä kapeampi ja yksitulkintaisempi tehtävä on ja mitä paremmin tehtävä sopii kirjoittajan käsityksiin, sitä helpommin hän pystyy suoriutumaan tehtävästä. Toisin sanoen kirjoittajan pitäisi pystyä sitoutumaan eli motivoitumaan tehtävän kirjoittamiseen. Niinkin voi tietysti käydä, että kirjoittaja kirjoittaa motivoituneesti, mutta ei siitä huolimatta täytä tehtävänantoa odotuksenmukaisella tavalla, jos hän ei pysty mukautumaan tehtävän edellyttämään kulttuuriseen ja sosiaaliseen normistoon.

Jotta tehtävän suorittaminen sujuisi mahdollisimman hyvin, jokainen kirjoittaja tarvitsee ohjeen siitä, mikä tehtävässä on tärkeää ja mitä siitä arvioidaan (esim. Hamp-Lyons 1991; Connor 1996). Käytännön näkökulmasta tehtävä ei kuitenkaan pysty olemaan koskaan niin täysin ohjeistettu, ettei se aiheuttaisi monitulkintaisuutta ja sallisi tässä mielessä kulttuurivariaatiota. Toisaalta tämä ei ole perimmäinen tarkoitukseen. Hyvä tehtävä on hyvin ja ymmärrettävästi ohjeistettu, mutta jättää

kuitenkin tilaa kirjoittamisen luovuudelle. Kirjoittamista arvioitaessa tulisi myös arvioida tehtävää ja sen mielekkyyttä ja sopivuutta.

Kirjoittajan näkökulmasta tekstin tuottamiseen vaikuttavat monenlaiset tekijät. Vaikka kirjoittaja on ennen kaikkea yksilö, hän on myös jonkin kulttuuritaustan edustaja. Kulttuuritekijät ovat siinä mielessä laaja-alaisia, että ne eivät vaikuta vain kielen tai tekstin tasolla erilaisina jäsentelyinä tai ilmaisun ja aiheen käsittelyn suoruusasteiden vaihteluina, vaan ne vaikuttavat myös siihen, mitä pidetään loogisena ja tarkoituksenmukaisena ajatteluna. Varsinkin puhevaltaisesta kulttuurista tulevalle kielen oppijalle kirjoitetun kielen tapa jäsentää ja ilmaista asioita voi tuntua vieraalta (Ivanič & Moss 1991, 217–218; Ballard & Glanchy 1991, 23). Kirjallinen ajattelutapa saattaa edellyttää puhekuulttuurin edustajalta jopa uudenlaisen ajattelutavan omaksumista (Cook-Gumpertz & Gumperz 1981, 91–93). Scribnerin ja Colen (1981, 251–252) mukaan psykologiset erot, kuten logiikan käyttö, abstrahointi, muistaminen ja viestivyys, jakavat kansat kirjallisiin ja oraalisiin. Tosin on muistettava, että kirjallinen kulttuurikaan ei ole mikään myötäsyttyinen ilmiö, vaan se on tietoisesti rakennettu pitkän ajan kuluessa.

Kirjoittamisen arviointi testitilanteissa kertoo sekin kulttuurisesta kirjoittamiskäsityksestä, siitä, millaista ja miten suoritettua kirjoittamista arvostamme. Testiarvioinnin taustalla on usein käytännön syistä produktiivaisempi näkemys kirjoittamisesta, koska kirjoittaja kirjoittaa testipaikassa tietyn ajan kuluessa yhden tai kaksi tekstiä, jotka sitten yksi tai kaksi arvioijaa arvioi annettujen ohjeiden mukaisesti. Testitilanteessa pidetään esimerkiksi kirjoittamisprosessin seuraamista mahdottomana ja tekstien monitulkintaisuutta negatiivisena seikkana, vaikka niin ei tarvitsisi olla. Luokkahuonearvioinnissa on mahdollista hyödyntää erilaisia arviointimuotoja, kuten portfolioarviointia, johon on jo valmiiksi sisäänrakennettuna esimerkiksi prosessiluonteisuus, monenlaisten tekstien kirjoittaminen ja oman kirjoittamisen reflektointi (ks. Linnakylä 1993, 1994; Huot 1997; Thaiss & Zawicki 1997).

Kirjoittamiskäsityksen perusteellinen ja analyttinen pohdinta on kuitenkin kaikenlaisessa arvioinnissa tärkeää, sillä jokainen kirjoittamistehtävä ja sen arviointitapa heijastelevat joko tietoisesti tai tiedostamatta jonkinlaista kirjoittamiskäsitystä. Mitä analyttisemmin on perehdytty kirjoittamiseen taitona ja mitä konkreettisemmin taitokäsitys osataan perustella, sen vankemmalle pohjalle kirjoittamisen opetus ja arviointi perustuvat.

4.2 Kirjoittamisen arviointiprosessi

4.2.1 Arvottaminen reaktiona

Arvioinnissa on kyse arvottamisesta: jotakin pidetään vahvempana tai parempana suhteessa johonkin toiseen. Arvottaminen ei koske vain tietynlaisia valintatilanteita, jollainen myös arviointitilanne on, vaan se on ihmissielun perusreaktio, joka esiintyy lähes kaikkien asioiden yhteydessä (Turunen 1997, 95–97). Se tekee kohdatut asiat arvokkaiksi tai tietyksi joskus epäarvokkaiksikin. Arvottamista voi kuvata myös sanannoilla “antaa arvoa” tai “pitää arvokkaana” ja vastaavilla kielteisillä ilmaisuilla.

Filosofian näkökulmasta arvostusten voimaa kuvaa se, että yksilöt tekevät valintoja pääasiassa, ehkä lopulta syvimmiltään yksinomaan, arvostusten perusteella. Arvostusten jokapäiväisyydestä huolimatta – tai juuri siksi – niistä ei ole aina helppo tai edes mahdollistakaan olla tietoinen. Joskus on vaikea osoittaa, millaiseen arvostukseen valinta liittyy, koska arvostukset voivat osin olla niin vakiintuneita ja tavanomaisia, ettei niiden läsnäoloa huomaa (Turunen 1997, 99–100).

Millainen reaktio arvostaminen sitten on? Niiniluodon (1994, 177) mukaan arvostaminen on suhtautumistapa tai akti, jossa joku arvosubjekti pitää jotakin kohdetta eli arvo-objektia A arvokkaana. Arvoasetelmien perusrakenne noudattaa siten kaavaa:

“X pitää A:ta arvokkaana” tai “A on arvokas X:lle”.

Tällöin X voi olla yksilö, ryhmä tai kokonainen kulttuuri; arvo-objektissa taas voidaan erottaa arvona kantaja (konkreettinen olio, esine tai tapahtuma) ja arvo-ominaisuudet (arvostuksen perustana olevat arvo-objektin piirteet).

Kielitaidon arvioinnissakin on kyse arvostuksesta, ja arvioinnin kohteena ovat ne taidon piirteet, joita pidetään suoritusta vahventavina tai heikentävinä. Arviointi ei ole kuitenkaan sattumanvaraista, vaan yleensä se perustuu yhteisesti sovittuihin normeihin, esimerkiksi taitotasokriteereihin. Tässä mielessä arviointipäätökset ovat perusteltavissa ja selkeämmin osoitettavissa kuin jokapäiväisen elämän valintatilanteiden päätökset, koska Niiniluodon (1994, 178) mukaan arvostukset näkyvät luotettavimmin niissä valinnoissa, joita ihmiset joutuvat tekemään todellisissa tilanteissa eikä niinkään sanoissa, koska todellisia arvostuksia on vaikea ilmaista sanallisesti.

Arviointiin liittyvät päätökset ja valinnat ovat perusteltavissa ainakin silloin, kun niistä sovitaan ja neuvotellaan ja kun ne hyväksytään

yhdessä. Tosin arvion valintaan liittyy myös paljon sellaista, mistä emme arvioijina ole tietoisia tai mitä emme muuten pysty kielellisesti perustelemaan yhteisistä kriteereistä huolimatta. Yleensäkin osassa toiminnoistamme kielentäminen ei ole mahdollista tai se ei anna oikeaa kuvaa toiminnasta. Myöskään yhteys oman toiminnan ja sen välillä, miten uskomme toimivamme, ei ole aina kovin suora tai selkeä.

Jos arviointi on arvottamista, millainen prosessi sitten arvottaminen on arvioijan näkökulmasta? Arvottaminen on tavallisesti monikerroksinen mentaalinen tapahtuma, johon liittyy kaikkien sielun osien – siis järjen, tunteen ja tahdon – käyttöä (Niiniluoto 1994, 40–45). Eri teorit kiinnittävät kuitenkin huomiota tämän kokonaisuuden eri puoliin eri tavalla; arvottaminen nähdään ennen kaikkea esimerkiksi tahdon ilmentymänä tai sitä käsitellään rationaalisena päätöksentekoprosessina.

Arvottamista koskevissa teorioissa ei ole myöskään aina tehty eroa arvojen alkuperän ja niiden perustelujen välillä, vaikka yhteiskuntatieteilijät ovat osoittaneet, että ihmiset usein omaksuvat arvonsa eirationaalista tietä – elämäntavan tai tradition pohjalta tai auktoriteettien tai vanhempien ohjaamina – mikä ei kuitenkaan tarkoita sitä, etteikö arvoväittämien tueksi tai eettisten konfliktien ratkaisemiseksi voitaisi esittää järkeenkäypiä argumentteja. (Niiniluoto 1994, 58.)

Niin sanotusti asiayhteyksiin liittyvän arvioinnin pitäisi oletusten mukaan perustua ennen kaikkea rationaalsiin arvoihin, jotka ovat Niiniluodon (1994, 58–59) mukaan moraalin alueella sosiaalisia konstruktioita; ne syntyvät yhteisöllisen neuvottelun, suostuttelun ja argumentaation kautta ohjaamaan ihmisten yhteistä toimintaa, suhdetta muihin ja omaan itseen. Myös Turusen (1997, 100) mukaan päätökset eivät ole vain yksilöllisiä vaan niihin liittyy myös ympäröivän toimintaympäristön, opetusyhteisön, ihanteet, tavoitteet ja ennen kaikkea käsitykset, esimerkiksi arvioija- ja kielikäsitteet. Yksilöiden ja yhteisöjen ratkaisuja ohjaavat lisäksi erilaiset ihanteet, mutta perusaineksen muodostavat kuitenkin arvostukset.

Rationaaliset arvot konstruoidaan siis prosessien kautta, joita ovat mm. julkinen harkinta ja järkipäiväinen keskustelu (Niiniluoto 1994, 59). Julkisuus voisi viitata kielitaidon arvioinnissa yhteisölliseen keskusteluun ja yhteisiin sopimuksiin ja järkipäiväisyys taas siihen, että keskustelussa voidaan päästä yhteisiin tulkintoihin yhteisen ammattikielen kautta, mikä taas edelleen takaa sen, että toiminta ei ole täysin sattumanvaraista.

Miten rationaalisena arvioijan toiminta voidaan nähdä, jos tunne, järki ja tahto ovat kaikki läsnä? Niiniluodon (1994, 42–44) mukaan vastakkainasettelu sen suhteen, onko ihminen järjellinen, emotionaalinen vai vapaavalintainen, on keinotekoinen, koska ihmisen toiminnassa on aina mukana järkeä, tunnetta ja tahtoa. Selvää ei ole sekään, onko jokin

näistä toista ensisijaisempi yleensä tai jossakin tietyssä tilanteessa. Arvioinnin näkökulmasta olennaisinta lienee, että tunnustamme ja tunnistamme kolminaisuuden läsnäolon ja vaikutuksen ja että pystymme hyväksymään arvioinnin subjektiivisuuden tästä lähtökohdasta käsin.

4.2.2 Arvion antaminen päätöksentekoprosessina

Psykologisesta näkökulmasta päätöksenteko on adaptiivinen prosessi, joka jakaantuu erilaisiin alaprosesseihin tai vaiheisiin. Psykologit eivät ole kuitenkaan esittäneet mitään yhtä yhteisesti hyväksyttyä mallia vaan päätöksentekoa kuvaavia malleja on monenlaisia ja ne jakaantuvat erilaisiin alaprosesseihin eri teorioiden mukaan (Huber 1989). Nojaan tässä käsitykseen, jonka mukaan päätöksentekoprosessi nähdään yhtenä ongelmanratkaisuprosesseista. Ongelmaa ratkaistessaan päätöksentekijällä on mahdollisuus valita useamman vaihtoehdon joukosta kaikkein järkevin, optimaalisin tai mieluisin vaihtoehto (Huber 1989; Montgomery 1989). Päätös syntyy, kun kyseinen vaihtoehto todetaan kaikkein dominoivimmaksi suhteessa muihin vaihtoehtoihin.

Huberin (1989) mukaan ongelmanratkaisija ei tiedä etukäteen, miten hän siirtyy alkuvaiheesta tavoitetilään eli päätökseen. Päätökseen ei myöskään päädytä vain yhden vaiheen kautta, vaan vaiheita on useita, ja niiden välillä päätöksentekijä suorittaa erilaisia toimintoja (operations). Nämä toiminnot itse asiassa johdattelevat päätöksentekijän vaiheesta toiseen. Toimintojen suorittamiseen päätöksentekijällä on useita erilaisia strategioita, joiden valitsemista edeltää strategian toimivuutta arvioiva prosessi.

Strategian valitseminen ei siis ole mekaaninen toiminto, vaan sen valintaan vaikuttavat erilaiset tekijät, esimerkiksi aikapaine. Jos päätös on tehtävä nopeasti, päätöksentekijä voi tukeutua yksinkertaisempien strategioiden käyttöön (Svenson & Edland 1989; Ariely & Zakay 2001). Strategiakäsitteen käyttöä näin yksinkertaistavassa muodossa on kuitenkin kritisoitu sen vuoksi, että päätöksentekijä itse ei ole kovin tietoinen abstraktien strategioiden käytöstä tai soveltamisesta. Karlssonin (1989) mukaan strategiat tulisikin ymmärtää enemmänkin teoreettisina konstruktina kuin päätöksentekijän käyttäminä keinoina tehdä päätös.

Karlssonin (1989) mukaan päätöksentekoon liittyy neljä vaihetta: 1) informaation hakeminen, 2) vaihtoehtojen vertailu ja arviointi, 3) vaihtoehtojen hylkääminen ja valinta sekä 4) valinnan vahvistaminen. Kaikkiin vaiheisiin liittyy erilaisia psykologisia prosesseja. Päätöksentekoprosessi liikkuu edestakaisin kahden tason välillä. Toinen tasoista viittaa päätöksentekijän toimintaan konkreettisessa tilanteessa ja toinen hänen itsereflektioonsa ja tietoisuuteensa, johon liittyvät esimerkiksi

strategioiden käyttö ja arvojen, toiveiden sekä antipatioiden mukanaolo (ks. myös Klaczynski ym. 2001).

Kirjoitelman arvioijan tavoitteena on tehdä päätös lukemansa tekstin arviosta. Tekstin mekaaninen kertalukeminen tai silmäily tuskin johtaa vielä päätökseen, vaan arvioijan läpikäymiä vaiheita on monenlaisia. Hän voi lukea tekstiä jaksoissa, koko tekstin useaan kertaan, joitakin osia uudelleen, ja jokaiseen lukemisjaksoon voi liittyä päätöstä valmistelevia, vahventavia tai heikentäviä vertailuja. Arvioija voi kiinnittää huomiota johonkin tiettyyn piirteeseen siitä huolimatta, arvioiko hän holistisella vai analyttisellä asteikolla. Hän voi esimerkiksi kiinnittää erityistä huomiota sanaston epätarkkaan käyttöön ja seurata sitä koko tekstin ajan. Arvioija ei kuitenkaan vielä ole välttämättä vakuuttunut vaan lukee tekstin uudelleen ja keskittyy tällä lukemiskerralla johonkin muuhun, esimerkiksi tekstin sidosteisuuteen ja rakenteisiin.

Toisen lukemiskerran jälkeen arvioija voi ajatella, että teksti on sidosteisuuden ja rakenteiden osalta taitotasolla 4 mutta sanaston osalta tasolla 3. Ehkä hän pitää sanastoa muita tärkeämpänä tekijänä ja antaa siksi nelosen sijasta kolmosen. Tai ehkä hänen strategiansa onkin silmäillä teksti vielä kerran läpi, koska hän on hieman epävarma päätöksestään. Mielestään hän painottaa liikaa omaa mieltymystään sanaston hallintaan. Kolmannella lukukerralla hän saattaa huomata, että teksti täyttää hyvin myös tehtävänannon, jolloin sanasto ei enää painotu niin paljon, ja hän antaa tekstistä nelosen. Toki tämä on yksinkertaistava esimerkki, mutta se ehkä valottaa prosessin yleistä kulkua ja lukemiskertojen ja -tapojen luonnetta suhteessa päätökseen.

Useita lukemiskertoja säätelee myös se, miten paljon arvioijan työmuisti muistaa yhdeltä lukemalta, ja toisaalta se, miten paljon hän pystyy käsittelemään informaatiota kerrallaan (Hogarth 1980, 93). Lukeminen ei myöskään välttämättä ole ymmärtävää vaan se on valikoivaa siinä mielessä, että arvioija ei pysty havaitsemaan kerralla esimerkiksi kaikkia kriteereitä, vaan niiden kunkin tarkastelemiseen tarvitaan kriteerikohtaisia lukemiskertoja. Hänen täytyy siis hakea päätökselleen lisää informaatiota. Niin tässä kuin muissakin asiantuntijataidoissa valikoiminen on rutinoituneella arvioijalla automaattisempaa ja nopeampaa kuin noviisiarvioijalla. Kokenut arvioija pystyy usein nopeampaan päätöksentekoon ja saattaa tuntea olonsa varmemmaksi päätöksen jälkeen. (Ks. esim. Freedman 1989; Cumming 1990; Huot 1990.)

Epävarmuus kuuluu kokeneenkin arvioijan työhön, ainakin jossakin määrin: useimmat arvioijat epäroivät joskus, painottavatko he kriteereitä tasapuolisesti, noudattavatko he arviointiasteikkoa muiden arvioijien tavoin tai antavatko he oikeudenmukaisen arvion. Epävarmuus näkyy

käytännössä esimerkiksi siinä, että arvioijat haluaisivat käyttää plussia ja miinusia, tai siinä, että arvioija ei osaa päättää, kumman tason hän valitsisi, jos suoritus sisältää piirteitä kahdelta eri tasolta. Kun henkilö on tehnyt päätöksensä, hänellä on taipumus suhtautua varmemmin valintaansa ja argumentoida sen puolesta kuin selittää, miksi hän ei valinnut jotakin toista vaihtoehtoa (Hogarth 1980, 37; Harvey 1997, 78). Kirjoittamisen arvioinnissa tämä voisi tarkoittaa sitä, että arvioija selittää perusteluja valitsemalleen taitotasolle kolme eikä vaihtoehtoiselle tasolle neljä tai kaksi. Epävarmuus ei aina kuitenkaan lopu arvion antamiseen vaan se voi tulla esiin jotakin toista suunnilleen samantasoista suoritusta arvioidessa tai se voi kalvaa vielä päivienkin päästä arvion antamisesta.

Lienee syytä tarkentaa vielä, että epävarmuus ei sinänsä liity arviointitilanteeseen vaan päättäjiin itseensä, minkä vuoksi sen työstäminen luonnollisena ja siksi välttämättömänä osana päätöksentekoprosessia olisi tärkeää (esim. Hogarth 1980, Harvey 1997). Arvioijan päätöksentekoon liittyy monenlaisia epävarmuustekijöitä, jotka voisi ryhmitellä esimerkiksi sen mukaan, millainen käsitys arvioijalla on (1) arvion mitattavuudesta, 2) omasta suhteestaan muihin arvioijiin, 3) arvioinnin eettisyydestä ja 4) omasta työstään yleensä inhimillisenä lukijana.

Epävarmuuden lisäksi päätöksentekoon liittyy epärationaalisia ja sosiaalisia tekijöitä, joihin myös Karlsson (1989) edellä viittaa itsereflektion yhteydessä. Epävarmuus voi myös olla seurausta näistä tekijöistä. Päätöksenteon tutkimisesta tekee vaikeaa juuri se, että päätöksentekoon liittyy niin monenlaisia, esimerkiksi sosio-kognitiivisia ja emotionaalisia, kompetensseja ja kyky tunnistaa vaikkapa tilanteen, tunteiden, mieltymysten ja motivaation vaikutuksia (vrt. Forest & Mehier 2001; Klaczynski ym. 2001). Päätöksen tekoa ei ohjaa vain yksilö, vaan yksilön tekemisiä säätelee kollektiivinen toiminta muun muassa normeineen ja arvoineen. Tätä kautta meille syntyy vakiintuneita tapoja, rutiineja ja sääntöjä, jotka ovat itse asiassa päätöksenteossa tärkeitä, koska ne auttavat mieltä käsittelemään odottamattomia tapahtumia sekä sietämään epävarmuutta (Forest & Mehier 2001, 594). Tähän liittyynee jo edelläkin esiin tullut kokemuksen tuoma varmuus ja nopeus päätöksenteossa.

Epävarmuudesta ja epärationaalisuudesta huolimatta arviointiin tuntuu liittyvän ainakin alitajuisesti absoluuttisuuden läsnäolo, ikään kuin arviotavalle olisi olemassa yksi ainoa, oikea arvio. Jos arviointiasteikko tai -kriteerit eivät kuvaa yksiselitteisesti suoritusta, arvioija voi kokea vaikeuden henkilökohtaisena kyvyttömyytenä määrittellä 'oikea' arvio tai sitten hän odottaa asteikolta liikaa, ikään kuin se pystyisi kuvaamaan tarpeeksi yksityiskohtaisesti kaikkia mahdollisia

suorituksia. Toinen epävarmuustekijä voi olla huoli siitä, pysyykö linjassa muiden arvioijien kanssa. Joskus se voi tuntua jopa tärkeämmältä kuin itse suoritus, eli arvioijan päätöstä voi ohjata hänen ajatuksensa siitä, mitä muut arvioijat mahdollisesti antaisivat samasta suorituksesta.

Kolmas epävarmuustekijä liittyy eettisiin kysymyksiin eli arvion oikeudenmukaisuuteen. Oikeudenmukaisuus korostuu varsinkin sellaisissa testitilanteissa, joissa arvioija ei tiedä mitään arvioitavasta. Joidenkin arvioijien mielestä arviointi olisi reilumpaa, jos he voisivat suhteuttaa suorituksen johonkin taustatekijään, kuten opiskeluaikaan tai -sisältöihin. Neljäs epävarmuustekijä koskee arvioijan tietoisuutta omista mieltymyksistä sekä tunne- ja vireystiloista. Mitä tietoisempi arvioija on esimerkiksi väsymyksen tai ärsyyntymisen vaikutuksesta, sitä paremmat mahdollisuudet hänellä on estää niiden vaikutus arvioon, vaikka toisaalta tietoisuus niiden vaikutuksesta voi myös aiheuttaa epävarmuutta ja voimattomuutta. Arvioijan oma käsitys arvion antamisesta päätöksentekoprosessina vaikuttaa myös siihen, miten neutraalin objektiivista tai subjektiivista hän katsoo toimintansa olevan. Pelko subjektiivisuudesta voi estää arvioimasta, ja taas usko objektiivisuuteen estää olemasta reflektiivinen tarkoituksenmukaisella tavalla.

4.2.3 Arviointiprosessi arviointikirjallisuudessa

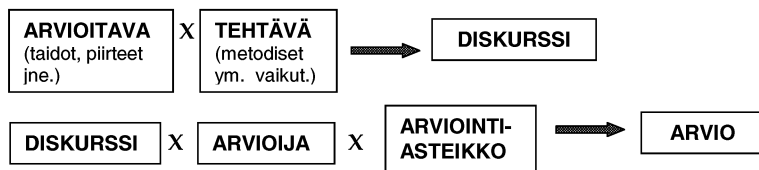
Arviointia koskevassa kirjallisuudessa käsitykset arviointiprosessin luonteesta vaihtelevat mekaanisesta pisteitystilanteesta arviointidiskurssiin (ks. esim. Cumming 1997; Upshur & Turner 1999). Kuten kuvioista 10 näkyy, yksinkertaisimmillaan todennäköisyys saada tietty arvio, esimerkiksi ennemmin 1 kuin 0 dikotomisesti arvioituna, on riippuvainen tehtävän vaikeustason ja henkilön kyvyn välisestä erosta (Upshur & Turner 1999, 83).



KUVIO 10 Yksinkertaistettu malli arviointiprosessista (Upshur & Turner 1999, 93).

Toisessa ääripäässä ovat taas viimeisimmät näkemykset, joiden mukaan arvio on osa laajempaa ja monisuuntaisempaa diskurssia. Itse asiassa erotuksena kuvioon 10 tehtävän suorittajan ja tehtävän välinen vuorovaikutus synnyttää diskurssin eikä arviota. Kuvion 11 mukaan vuorovaikutus, jonka tuloksena arvio syntyy, on kolminapainen: siinä

ovat mukana tehtävän ja arvioitavan välinen diskurssi, arvioija ja asteikko. (Upshur & Turner 1999, 85.)



KUVIO 11 Laajennettu malli arvioon vaikuttavista systemaattisista tekijöistä (Upshur & Turner 1999, 85).

Kuvioista 10 ja 11 voi huomata, että ääripäät jatkumolla, miten arvioija tai arviointiin vaikuttavat tekijät otetaan huomioon, ovat todella kaukana toisistaan. Viime vuosikymmenien aikana on julkaistu lukuisia arvioijan toimintaa koskevia tutkimuksia, mutta niissä on oltu kiinnostuneita lähinnä siitä, miten arviointikoulutusta voitaisiin kehittää tai miten mittausvirheet (measurement error) vaikuttavat arvioihin (ks. tarkemmin luvusta 3.3.2). Sen sijaan vähemmän on oltu kiinnostuneita siitä, millainen prosessi arvion antaminen on (ks. esim. Charney 1984; Engelhard 1994; White 1995; Bachman 2000). Esittelen seuraavaksi muutamia kielitaidon arviointikirjallisuudessa esitettyjä käsityksiä arviointiprosessin luonteesta ja kulusta.

Freedman ja Calfee (1983) kuvaavat arvion antamista informaation prosessointimallilla, jossa on kolme vaihetta: 1) tekstin lukeminen eli mielikuvan muodostaminen, 2) mielikuvan arviointi ja 3) arvion muotoileminen ja ulkoistaminen. Ensimmäisessä vaiheessa arvioija siis lukee tekstin muodostaakseen tekstistä mielikuvan (text image), jota arvioija sitten itse asiassa arvioi. Nämä tekstimielikuvat eivät kuitenkaan ole alkuperäisen tekstin kopioita eivätkä täysin samanlaisia eri arvioijien välillä. Ne eivät voi olla staattisia arvioija- tai tekstikohtaisestikaan, koska olosuhteet vaikuttajineen vaihtelevat eri aikoina. Arvioijien käsitykset yleensä kirjoittamisesta perustuvat maailmantietoihin, uskomuksiin ja arvoihin sekä tietoon kirjoittamisen prosessista. Myös fyysinen lukemisympäristö voi vaikuttaa siihen, millainen mielikuva tekstistä syntyy.

Tekstimielikuvaa arvioidessaan arvioija vertailee toisiinsa kirjoituksen eri osia; tekstin eri puolien keskinäinen vertailu voi tarkoittaa myös arviointikriteerien soveltamista tekstiin. Tämän arviointiprosessin eli toisen vaiheen aikana arvioija arvioi tekstiä ja päättää, miten hyvin kirjoittaja on onnistunut osoittamaan kirjoittamisen taitoaan. Prosessin

lopuksi arvio ulkoistetaan joko sanallisella tai numeerisella arviolla (Freedman ja Calfee 1983; Wolfen & Kaon 1998, 467 mukaan).

Frediksenin (1992) mallissa Wolfen & Kaon (1998, 467) mukaan arvioijan tulkitseva viitekehys (framework) ohjaa sitä, mihin piirteisiin arvioija kiinnittää huomiota ja miten hän ymmärtää ne. Tulkitsevan viitekehysten avulla arvioija sekä ymmärtää merkitykset että pystyy tunnistamaan suorituksen piirteet. Mutta koska nämä tulkitsevat viitekehukset ovat erilaisia, ovat myös toisistaan ne piirteet, joihin arvioijat kiinnittävät huomiota, ja se, millaisia tulkintoja he niille antavat.

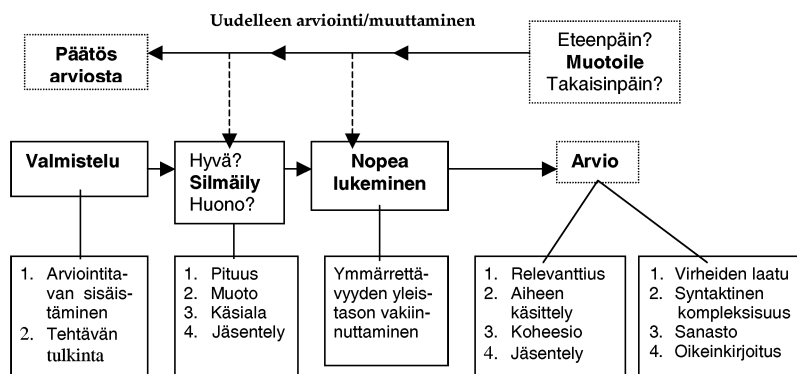
Sekä informaation prosessointimallilla että tulkitsevan rakenne-systeemin mallilla näyttäisi olevan yhtymäkohtia skeema- ja skriptiteorian ja relevanssiteorian kanssa. Näissä kaikissa teorioissa – niiden välisiin eroihin puuttumatta – yksilön skeemat ja skriptit määrittelevät hänen maailmantietoaan ja sitä kautta edelleen hänen tulkintaansa tapahtumista ja asioista sekä hänen käyttäytymistään (ks. esim. Schank & Abelson 1977; Rumelhart 1977; Sperber & Wilson 1986). Koska yksilöiden kokemukset ovat erilaisia, myös skeemat ja skriptit ovat niiltä osin erilaisia.

Informaation prosessointi- ja tulkitseva viitekehys -malli eivät ole kuitenkaan yksioikoisesti rinnastettavissa toisiinsa. Freedmanin ja Calfeen käsitys prosessin kulusta on lineaarinen, mistä käsitystä on myös kritisoitu. Frediksenin (1992) mukaan arvion antaminen on toistonomainen ja jossakin mielessä kehämäinen prosessi, jonka aikana arvioija tekee monia arvioivia päätöksiä. Esimerkiksi lukiessaan tekstiä arvioija ei siis tee vain yhtä päätöstä, vaan päätöksiä voi olla useita ja ne voivat sijaita eri kohdissa prosessia.

Lisäksi päätökset eivät ole välttämättä toisiinsa nähden kausaalisuhteessa tai ne eivät ole keskenään loogisia, mutta niille on tyypillistä, että uusi päätös muokkaa aina jollakin tavalla edellistä päätöstä. Ehkä arviointiprosessi voitaisiinkin nähdä enemmän kehämäisenä prosessina, jossa on useita erilaisia vaiheita päätöksineen (ks. myös Vaughan 1991). Vaiheet ja päätökset ovat toisiinsa nähden vuorovaikutussuhteessa mutta eivät välttämättä lineaarisesti tai kausaalisesti. Mallintamismielessä tämä voi tuntua kaootiselta tai liian monimutkaiselta, koska mallintaminen edellyttää yksinkertaistamista ja yleistämistä.

Milanovic & ym. (1996) ovat mallintaneet holistista kirjoittamisen arviointiprosessia lähinnä päätöksenteon näkökulmasta. He kehittivät mallinsa ensimmäisen version nojaten Cummingiin (1990), jonka tutkimus koski analyttistä arviointiprosessia. Sen jälkeen he ovat käyneet lukuisia keskusteluja arvioijien kanssa ja muokanneet mallia edelleen käymiensä keskustelujen pohjalta. Milanovic ym. (1996, 106) pitävät itse malliaan varsin alustavana ja näkevätkin mallinsa

eräänlaisena pohjana jatkotutkimukselle. Mallissa arvioija etenee kuvion 12 mukaisten vaiheiden kautta lopulliseen arvioon.



KUVIO 12 Päätöksentekoprosessi kirjoittamisen arvioinnissa (Milanovic ym. 1996, 95).

Arviointi alkaa siis valmisteluvaiheesta, jonka aikana arvioija tutustuu arviointikriteereihin ja -ohjeisiin sekä analysoi tehtävästä, mitä se edellyttää kirjoittajalta ja miten se suhteutuu arviointikriteereihin. Sen jälkeen arvioija silmäilee tekstin tehdäkseen siitä yleistävän arvion kuviossa 12 esitetyn neljän kriteerin (pituus, muoto, käsiäla ja jäsenty) mukaan. Sitten arvioija lukee tekstin nopeasti läpi saadakseen yleisvaikutelman tekstin ymmärrettävyydestä. Tämän jälkeen arvioija varsinaisesti arvioi tekstin siten, että hän päätyy arvioon holistisesti tarkkailemalla laatikon Arvio alla olevia kriteereitä.

Mikäli arvioija on tässä vaiheessa varma arviostaan, hän siirtynee suoraan lopullisen arvion antamiseen. Jos arvioija taas on epävarma, hän voi nostaa tai laskea arviota joko suoraan tai sitten palaamalla tekstiin uudelleen. Hän voi kuvion 12 mukaisesti ennen lopullista päätöstä lukea tekstin vielä nopeasti läpi tai silmäillä sen eli hän voi palata takaisin arviointiprosessin aikaisempiin vaiheisiin sen mukaan, miltä hänestä tuntuu. Tämä osa kuviosta tuntuu tarkoituksenmukaiselta arviointiprosessin etenemisen osalta, ja siinä on yhtymäkohtia edellä esitettyyn käsitykseen arviointiprosessin kehämäisyydestä.

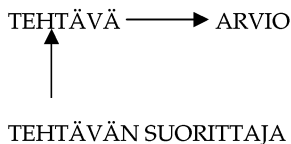
Sen sijaan kuviossa ihmetyttää se, että jos malli kuvaa nimenomaan holistista arviointia, kiinnittäisivätkö arvioijat todellisissa arviointitilanteissa huomiota kaikkiin mallissa esitettyihin piirteisiin tai kriteereihin huolimatta siitä, että ne olisi mainittu arviointiasteikossa tai -ohjeissa. Yhtä epätodennäköiseltä vaikuttaa se, että näihin piirteisiin kiinnitettäisiin huomiota juuri kuvion 12 mukaisen vaiheen kohdalla,

esimerkiksi pituuteen juuri silmäilyvaiheessa. Myös se pohdituttaa, mikä on ns. varsinaisten kriteerien painoarvo suhteessa muihin kriteereihin, esimerkiksi pituuteen ja käsialaan?

McNamaran (1996, 120–121) mukaan arviointi voidaan jakaa perinteiseen monivalinta-arviointiin (traditional fixed response assessment) ja suorituspohjaiseen arviointiin (performance-based assessment). Ero näiden kahden välillä on siinä, kuten kuviosta 13 näkyy, että perinteisessä lähestymistavassa arvio johdetaan suoraan tehtävästä. Tehtävään liittyy käsitys, jonka mukaan tehtävä itsessään pystyy ennakoimaan ja rajoittamaan mahdolliset vastaukset. Tämän vuoksi traditionaaliset tehtävät ovatkin yleensä sellaisia, että niissä voi olla vain yksi 'oikea' vastaus, eli ne ovat monivalintatehtäviä tai hyvin lyhyttä vastausta vaativia avoimia tehtäviä. Tehtävän suorittajan vastauksia käytetään hänen taitonsa mittareina ja tehtävän laadun indikaattoreina. Toisin sanoen tehtävänanalyysit viestivät tehtävän erottelukyvystä ja sen vaikeus-
tasosta.

Sen sijaan suorituspohjaisessa arvioinnissa tehtävä tuo esiin suorituksesta tai oppijan toiminnasta muutakin kuin vain sen, mitä – kärkeänsä sanoen – vastauksen valinta yksinkertaisesti indikoi. Suoritus arvioidaan käyttäen arviointiasteikkoa ja -ohjeita, joten arvion antaminen perustuu arvioijan, suorituksen ja asteikon väliseen vuorovaikutukseen. Tehtävän arviointi on näin ollen kaksivaiheinen vuorovaikutusprosessi: ensin tehtävän suorittajan ja tehtävän välinen ja sitten arvioijan ja arviointiasteikon välinen, kuten kuviosta 13 näkyy. Tosin kuvion nuolet ovat yksisuuntaisia, mikä ei oikein tue vaikutelmaa vuorovaikutuksellisuudesta.

Perinteinen monivalinta-arviointi



Suoritusarviointi



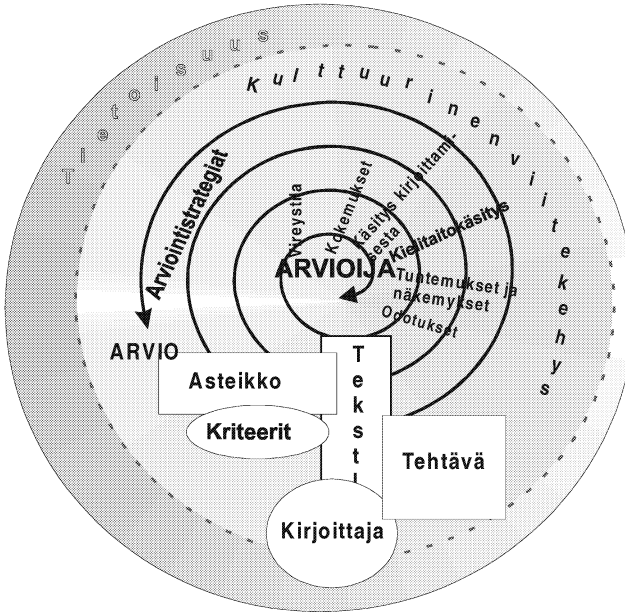
KUVIO 13 McNamaran (1996, 120) kuvaamat suorituksen arviointimallit.

McNaraman malli on hieman yksinkertainen ja yleinen, mutta se havainnollistaa hyvin paradigman muutosta vuorovaikutukselliseen suuntaan. Samaa asiaa kuvaavat aikaisemmin esiin tulleet Uspherin ja Tunerin (1999) kuviot 10–11. Edellä esitetyt mallit ovat puutteineen ja vahvuuksineen vähän erilaisia, mutta kaikille niille tuntuu olevan yhteistä se, että arvio suorituksesta ei ole yksinkertaisen putkiprosessin tulos. Itse asiassa arviointiprosessi tuntuu olevan luonteeltaan sitä monimutkaisempi ja vaikeammin kuvattava, mitä inhimillisempänä toimintana se halutaan nähdä tai suostutaan näkemään.

Toisaalta mallien idea ei olekaan pystyä kuvaamaan 'koko totuutta'. Niistä on hyötyä ennen kaikkea tarkoituksenmukaisella tavalla yksinkertaistettuina toimintakarttoina. Tarkoitan tällä sitä, että parempi tieto arviointiprosessista ja sen selkeä mallintaminen edesauttaisivat myös arvioijakoulutusta ja ehkä myös itse arviointiprosessia. Selkeä malli voisi auttaa oman toiminnan tiedostamisessa, ja parhaimmillaan se toimisi toimintaohjeena ja reflektoinnin työkaluna. Toimintamallin tulisi kuitenkin olla sen verran joustava, että sitä voisi hioa yksilöllisemmäksi sen mukaan, minkä tyyppinen kukin arvioija on vahvuuksineen ja heikkouksineen.

4.2.4 Arviointiprosessi tässä tutkimuksessa

Tässä tutkimuksessa arviointiprosessi nähdään kuvion 14 mukaisesti arvioijan, tekstin ja arviointiasteikon välisenä vuorovaikutuksena, jonka tavoitteena on tehdä päätelmä luettavasta tekstistä eli antaa arvio siitä. Kirjoittaja on mukana arviointiprosessissa lähinnä tekstin välityksellä, koska tässä yhteydessä kirjoittajat olivat tuntemattomia ja nimettömiä arvioijille. Arvioijat tiesivät kirjoittajista vain sen, että he olivat ympäri Suomea eri oppilaitoksissa opiskelevia aikuisia suomen kielen opiskelijoita.



KUVIO 14 Kirjoittamisen arviointiprosessi.

Arviointiprosessin keskiössä on arvioija, koska hänen tekemänsä tulkinnat niin tekstistä kuin arviointikriteereistäkin määrittävät sen, millainen prosessi on luonteeltaan, miten se etenee ja millainen tekstistä tehtävä päätelmä on. Ruth ja Murphy (1988) näkevät arvioijan roolin kolmijakoisena: arvioijan täytyy ymmärtää kirjoitustehtävä ja sen pohjalta tuotettu teksti sekä pystyä arvioimaan kyseinen teksti. Määritelmään tulisi lisätä myös se, että arvioijan tulee sisäistää arviointitapa, mikä on vielä eri asia kuin arviointiin pystyminen.

Arvioijan roolin kuvauksista – myös kuviosta 14 – voi helposti tulla sellainen vaikutelma, että arvioijan toiminta olisi yksisuuntaista: ikään kuin arvioija toimisi samalla tavalla kaikissa arviointitilanteissa ja lukisi samoin olettauksin kaikki tekstit ja vain arvioija vaikuttaisi siihen, miten prosessi etenee. Arviointiprosessi on kuitenkin kaksisuuntainen vuorovaikutussuhde, ensisijaisesti tekstin ja arvioijan välillä mutta myös kirjoittajan, tekstin, asteikon ja arvioijan välillä. Toisin sanoen myös kukin teksti erikseen vaikuttaa siihen, millainen prosessi on luonteeltaan ja miten arviointikriteerit tulevat sovelletuksi (ks. myös Cushing Weigle 1999).

Tässä mielessä arviointiprosessit ovat yksilöllisiä ja ainutkertaisia eli ne ovat omanlaisiaan jokaisen arvioijan ja jokaisen tekstin kohdalla. Toisaalta taas arvioijan toiminnassa sekä arvioijakohtaisesti että arvioijien välillä on varmasti paljon systemaattisuutta, mitä tavoitellaan esimerkiksi päätöksentekoa kuvaavilla malleilla. Systemaattisuuden oletus kuuluu myös reliaaбелиuden lähtöoletukseen. Mielenkiintoista tässä on se, missä määrin arvioijan toiminta noudattelee systemaattisuutta ja millaiset tekijät ja tilanteet saavat arvioijan toimimaan odotuksenvastaisesti. Missä määrin yhdenmukaisuuden ja johdonmukaisuuden tavoittelu on oikeastaan mahdollista?

Arviointi olisi ilman johdonmukaisuuden ja yhdenmukaisuuden tavoittelemista kaoottista ja epäeettistäkin, mutta arvioija ei ole kuitenkaan lineaarisesti toimiva kone. Koulutettu arvioija edustaa ammattilaisuutta ja asiantuntijuutta, joiden varassa voidaan olettaa, että arvioijat jakavat yhteisen käsityksen niin arvioitavasta taidosta kuin arviointikriteereistäkin sekä myös niiden soveltamisesta. Neuvoteltuihin käsityksiin perustuva asiantuntijuus tarjoaa myös tehokkaampia työkaluja tekstin analyttiseen tarkastelemiseen.

Ammattilaisuuskaan ei ole yksisisältöinen kokonaisuus vaan siihenkin liittyy neuvotteluista huolimatta erilaisia käsityksiä esimerkiksi kirjoittamisesta ja hyvästä tekstistä. Myös arvioijien kyky tulkita tehtävää ja arvioitavaa tekstiä voivat vaihdella arvioijittain, samoin heidän suoritukseen kohdistamansa odotukset. Lisäksi arvioijat ovat yksilöitä ihmisinä, minkä vuoksi heidän toimintaansa arviointiprosessin aikana vaikuttavat heidän tietonsa ja käsityksensä maailmasta, heidän uskomuksensa ja arvonsa, heidän kokemuksensa ja tuntemuksensa. Myös fyysiset tekijät ovat erilaisia niin arvioijien välillä kuin samalla arvioijallakin eri tekstien kohdalla, esimerkiksi väsymys tai ympäröivä tila taustaanineen voivat vaikuttaa prosessin aikana.

Miten paljon kaikki nämä seikat vaikuttavat arviointiin, on paljolti kiinni siitä, miten tietoinen arvioija on toiminnastaan. Yltiöpäisen subjektiivisuuden korostamisessa on siinäkin vaaransa. Oman kokemukseni mukaan spontaanistikin annettujen arvioiden valossa arvioijat/opettajat ovat melko yhdenmukaisia. Mikä on sitten tarpeeksi yhdenmukaista ja johdonmukaista? Mielenkiintoisempia ja ehkä myös tärkeimpiä esimerkiksi arvioinnin ja opetuksen kehittämisen kannalta ovat arvioijien käsitykset ja näkemykset. Arviointiprosessin tuntemus on tie, jonka kautta arvioijia voitaisiin ohjata tehokkaammin ja tarkoituksenmukaisemmin itse ohjaamaan omaa arviointiprosessiaan.

4.3 Tutkimuksen kertomaa: mitä arvioija arvostaa toisen/vieraan kielen kirjoittamisessa?

4.3.1 Arvioinnissa käytettyjä kriteereitä ja lukustrategioita

Millaisia ovat sitten tutkimusten mukaan arvioijien väliset tyyppierot ja mihin arvioijat kiinnittävät huomiota arviointiprosessin aikana? Useimpia testaustutkimuksia on kritisoitu siitä, että ne keskittyvät liiaksi reliaabeliuden varmistamiseen eivätkä ota tarpeeksi huomioon arvioijaa; tutkimuksissa on siis enemmän kiinnostuttu arvioiden systemaattisista vaikutuksista kuin diskurssista numeroiden takana (Upshur & Tuner 1999, 86). Toisin sanoen arviointia on tarkasteltu ja yhä tarkastellaan enemmän arvion eli pistemäärän kuin prosessin näkökulmasta (Vaughan 1991, 111; DeRemer 1998, 7; Smith 2000, 159).

Arviointikirjallisuuden perusteella näyttää siltä, että arvioijan käyttäytymistä koskevat tutkimukset on tehty pääasiassa 1990-luvulla, koskivat ne sitten kirjoitetun tai puhutun suorituksen arviointia (ks. esim. Meiron 1998; Upshur & Tuner 1999; Smith 2000). Taulukkoon 8 on koottu kirjoittamisen arviointia koskevia kvalitatiivisia tutkimuksia. Olen jättänyt kuitenkin pois sellaiset kvalitatiiviset tutkimukset, jotka tarkastelevat vain arvioijakoulutuksen merkitystä tai arvioijan taustatekijöitä ja niiden vaikutusta (ks. lisää esim. Shohamy ym. 1992; Cushing Weigle 1994; Kobayashi & Rinnert 1996). Kaikki taulukon 8 tutkimukset koskevat englantia toisena kielenä, mikä on tavallista alan kirjallisuudessa. Arvioitavien tekstien kirjoittajat ovat aikuisia kielenoppijoita, joiden suorituksia on arvioitu holistisesti viisi- tai kuusiportaisella asteikolla paitsi Smithin tutkimuksessa, jossa käytettiin competence-based-arviointia.¹⁴

¹⁴ Competence-based-arvioinnissa arvioija tekee päätöksen arvioinnista vastaamalla suoritusta kuvaaviin väittämiin kyllä tai ei sen mukaan, miten hyvin tehtävän suorittaja osoittaa osaamisensa tietyssä tehtävässä tai osataidossa (Smith 2000).

TAULUKKO 8 Vieraan/toisen kielen kirjoittamisen kvalitatiivisia arviointitutkimuksia.

Tekijä(t) ja aika	Arvioijien (1) ja tekstien (2) määrä ja metodi (3)	Arvioijat kiinnittivät huomiota	Arvioijatyyppiä/-strategioita
Cumming 1990	<ol style="list-style-type: none"> 6 kokematonta ja 7 kokenutta arvioijaa 12/arvioija ääneenajattelu 	<p>Kriteerit, kielenkäyttö, jäsentely ja sisältö annettiin otsikoina; arvioijat määrittivät niiden sisällön itse.</p> <p><u>Kokemattomat arvioijat:</u></p> <ul style="list-style-type: none"> lempeämpiä sisällön ja jäsentelyn arvioimisessa virhekeskeisempiä <p><u>Kokeneet arvioijat:</u></p> <ul style="list-style-type: none"> eri kriteerien arvioiminen ei juurikaan eronnut eri arvioijilla 	<p><u>Kahdenlaisia strategioita:</u></p> <ol style="list-style-type: none"> tulkintastrategiat arviointistrategiat <p><u>Kokemattomat arvioijat:</u></p> <ol style="list-style-type: none"> vain muutamaaan kriteeriin enemmän tekstin pintatasoon <p><u>Kokeneet arvioijat tukeutuivat:</u></p> <ol style="list-style-type: none"> useisiin kriteereihin itsensä kontrollointiin tietoon ja kokemukseen
Vaughan 1991	<ol style="list-style-type: none"> 9 arvioijaa 6/arvioija ääneenajattelu 	<p><u>5 yleisintä:</u></p> <ul style="list-style-type: none"> sisältö käsiala aikamuoto-ongelmat välimerkki-/isokirjain-ongelmat aloitus-/morfologiset ongelmat 	<p><u>Lukemistyylin perustuu:</u></p> <ol style="list-style-type: none"> ensivaikutelmaan yksityiskohtaan/ yhteen asiaan kahteen kategoriaan esim. sisältö ja rakenteet kielioppiin subjektiiviseen vaikutelmaan
Milanovic, Saville ja Shuhong 1996	<ol style="list-style-type: none"> 16 arvioijaa 40/arvioija ääneenajattelu kahdesta tekstistä ja kirjallinen retrospektio 	<p><u>Yleisimmät piirteet:</u></p> <ul style="list-style-type: none"> pituus luettavuus (käsiala) kielioppi viestivyyys idiomaattisuus sanasto oikeinkirjoitus sisältö tehtävänannon täyttäminen välimerkkien käyttö 	<p><u>4 erilaista arviointitapaa:</u></p> <ol style="list-style-type: none"> kaksi silmäily- ja yksi lukukerta kaksi silmäily- ja kaksi lukukertaa yksi lukukerta pysähtelevä yksi lukukerta
DeRemer 1998	<ol style="list-style-type: none"> 3 arvioijaa 24/arvioija ääneenajattelu 	<p><u>Kriteerien tulkinta</u></p> <ul style="list-style-type: none"> arvioijat kiinnittivät enemmän huomiota kirjoittajan yksilöllisiin piirteisiin, jos he olivat heidän omia opiskelijoitaan arvioijat painottivat eri tavalla eri kriteereitä sen mukaan, millaista arviointitapaa he käyttivät tai sen mukaan, miten kriteerit vastasivat tekstin piirteitä 	<p><u>3 erilaista arviointitapaa</u></p> <ol style="list-style-type: none"> valintaprosessi (useampi kuin yksi arvio mahdollinen) yksinkertainen tunnistamisprosessi (ei tutustu edeltä kriteereihin) monimutkainen tunnistamisprosessi (tutustuu etukäteen kriteereihin) <p><u>Lukemisstrategiat</u></p> <ol style="list-style-type: none"> suurimmaksi osaksi kerrallaan ilman keskeytystä yksi informanteista keskeytti lukemisen välillä ja mietti kriteereitä <p>► yleisvaikutelmaan perustuu: tekstit arvioitiin kerta-lukemalta analysoimatta kriteereitä</p>

(jatkuu)

<p style="text-align: center;">Smith 2000</p>	<ol style="list-style-type: none"> 1. 6 arvioijaa 2. 3/arvioija 3. ääneenajattelu 	<p><u>Yleisimmät piirteet:</u></p> <ul style="list-style-type: none"> - kielioppi - tekstin jäsentely - koherenssi - lauserakenteet - välimerkkien käyttö - oikeinkirjoitus - käsiala - tekstin pituus - sanavalinta 	<p><u>Lukemisstrategiat:</u></p> <ul style="list-style-type: none"> - "lue kerran läpi ja silmäile sitten" - tuen hakeminen tietyille kriteerille: tekstiä ei edes välttämättä lueta loppuun, jos tuki löytyy aikaisemmin - ensivaikutelma: yksi lukukerta ja sen jälkeen vahvistussilmäys
--	--	---	---

Taulukosta 8 voi havaita, että arvioijat kiinnittävät hyvin samantapaisiin asioihin huomiota kaikissa tutkimuksissa riippumatta siitä, millainen asteikko ohjaa arviointia. Kielen rakenteisiin, sanaston käyttöön, tekstin luettavuuteen, sisältöön ja sen kokonaisrakenteeseen tai jäsentelyyn liittyvät kommentit ovat odotuksenmukaisia. Vieraan/toisen kielen kirjoittamisessa myös oikeinkirjoitus on usein ymmärrettävyyden ja luettavuuden kannalta olennainen tekijä.

Arvioijat näyttäisivät kiinnittävän kaikkein systemaattisimmin huomiota rakenteisiin, kielioppiin, ainakin siinä mielessä, että siihen tai johonkin sen osaan puututaan todennäköisimmin kaikissa yhteyksissä, kun taas esimerkiksi sisällön tai ymmärrettävyyden painoarvo voi vaihdella tai jäädä jopa kokonaan huomiotta. Tämäkin on jossakin määrin odotuksenmukaista siihen nähden, miten vankka formalistisen kielenopetuksen perinne on ollut ja on osin edelleenkin.

Suhtautuminen rakenteisiin tuntuu olevan kahtalainen: toisaalta kritisoidaan niiden ylivaltaa ja toisaalta ollaan huolestuneita siitä, miten esimerkiksi noviisiarvioijat eivät kiinnitä huomiota kirjoittamisen kannalta olennaisiin rakenteisiin (ks. Cumming 1990). Esimerkiksi Lekin (1990, 61) mielestä opettajilla on taipumus kiinnittää enemmän huomiota rakennevirheisiin kuin palkita kirjoittajaa esimerkiksi vahvasta ajattelusta tai diskurssin hallinnasta. Tai sitten opettajat palkitsevat kirjoittajaa esimerkiksi aiheen käsittelystä mutta jättävät huomiotta sellaiset rakennevirheet, jotka ovat olennaisia tekstin kannalta. Tällaiset opettajat saattavat pyrkiä ohittamaan rakenteelliset seikat välttääkseen varmasti olemasta liian rakennekeskeisiä.

Rakenteiden vahva osuus kirjoittamisen arvioinnissa ei kuitenkaan välttämättä johdu pelkästään näkemyksellisestä valinnasta, jonka mukaan rakenteet olisivat kirjoittamisen tärkeitä ja vahvin kivijalka. Rakenteiden hallinta tai hallinnan puute on helppo osoittaa tekstistä, eli niitä on helpompi opettaa ja arvioida kuin esimerkiksi hyvää kirjoittamista tiedon konkreettisemmän rakenteen ja käsiteltävien yksiköiden, propositioiden, selkeyden vuoksi (esim. Davis ym. 1987). Arvioijana on paljon helpompi osoittaa, millainen ero rakenteiden hallinnassa on tasoilla 1 ja 2, kuin

osoittaa, miten sidosteisuus tai tekstin rakenne eroavat toisistaan tasoilla 3 ja 4.

Rakenteiden roolia pohdiskeltaessa on syytä huomioida myös se, että rakenteet eivät ole mikään yksi yhtenäinen yksikkö. Rakenteista saatetaan helposti tarkastella vain jotakin tiettyä osa-aluetta. Arvioijat saattavat olla erityisen kiinnostuneita esimerkiksi verbien ja aikamuotojen käytöstä mutta eivät kiinnitä juurikaan huomiota syntaksiin tai nominien taivutukseen. Johonkin tiettyyn rakenteen osa-alueeseen puuttuminen tuli tulokseksi esimerkiksi Vaughanin (1991) tutkimuksesta (ks. taulukko 8).

Toinen rakenteisiin liittyvä seikka on se, että ne ovat kielitaidon eri vaiheissa tai eri oppijoilla löyhemmässä tai tiiviimmässä yhteydessä muihin kirjoittamisen tekijöihin, mikä näkyy käytännössä eri arviointikriteerien erottelemisen ongelmana. Esimerkiksi rakenteiden yhteys oikeinkirjoitukseen voi olla hyvinkin tiivis eikä ero näiden välillä olekaan kovin selvä. Tämä seikka näkyy suomen kielessä esimerkiksi morfologian ja oikeinkirjoituksen välillä: missä määrin seuraavan esimerkin ongelmat johtuvat ehkä oikeinkirjoituksesta ja missä määrin morfologian hallinnasta: "Tämä työtö hänällä on ongilma. jälka on laitta hän kaatua pöyruilta?"

Suhtautuminen oikeinkirjoituksen ja rakenteiden väliseen suhteeseen näyttäisikin olevan kahtalainen: ne nähdään joko erillisinä taitoina tai sitten niiden katsotaan kuuluvan tiiviisti yhteen. Rakennelähtöisen näkökulman mukaan oikeinkirjoitus kulkee kiinteästi käsi kädessä rakenteiden kanssa. Oikeinkirjoitus mielletään itse asiassa osana "prosodiaa", minkä vuoksi se on osa puhuttua kieltä ja siksi olennainen osa myös kirjoitettua kielimuotoa. Tällä tarkoitetaan sitä, että kirjoitettu kieli perustuu itse asiassa mielikuvaan äänneistä; samoin esimerkiksi välimerkit toimivat vastaavassa tehtävässä kuin puheen tauot, tempo ja intonaatio (Chafe 1987a, 2, 1987b, 2).

Näistä kahdesta näkökulmasta voi erottaa vielä kolmannen näkökulman, jonka mukaan oikeinkirjoitus on mekaaninen taito, jota ei tulisi edes arvioida osana kirjoittamisen taitoa (ks. esim. Charney 1984; Hamp-Lyons 1991c; Polio 2001). Kirjoittamista koskevien tutkimusten mukaan näyttää kuitenkin tavallisemmalta, että oikeinkirjoitus erotetaan omaksi kriteerikseen, kuten myös taulukosta 8 näkyy. Oikeinkirjoitus ei kuitenkaan ole painoarvoltaan välttämättä yhtä tärkeä kriteeri kuin esimerkiksi sanasto, jäsentely ja/tai rakenteet. Joidenkin arviointikäytäntöjen mukaan oikeinkirjoitusta ei painoteta kielitaidon alemmilla tasoilla, jos se ei vaikuta tekstin ymmärrettävyyteen.

Käsialan ja tekstin pituuden saama huomio sen sijaan on hieman yllätyksellistä, koska niistä kumpikaan ei varsinaisesti ole kirjoittamisen taitoa (ks. esim. Charney 1984). Käsiala on tietysti silloin varsin

merkittävä tekijä, jos tekstistä ei saa yksinkertaisesti selvää käsialan epäselvyyden vuoksi. Muuten käsiala voi vaikuttaa arvioijiin emotionaalisesti: se voi ärsyttää tai ihastuttaa riippumatta siitä, miten luettavaa se on (vrt. Rafter & Rubin 1984; Cole & Moss 1993; Milanovic ym. 1996). Huonolla käsialalla saattaa puolestaan olla negatiivinen vaikutus arvioon (Brown 2000). Käsialan perusteella muodostetaan myös käsityksiä kirjoittajan kulttuuri- ja kielitaustasta sekä koulutuksesta ja huolellisuudesta. Käsiala yhtenä kirjoittamisen kriteerinä onkin herättänyt keskustelua pitkään: 1960-luvulla Diedrerichin tutkimuksessa arvioijat pyysivät lisäämään käsialan yhdeksi kriteeriksi analyttiseen asteikkoon (Faigley ym. 1989).

Sittemmin käydyn keskustelun perusteella yleinen näkökanta näyttää olevan, että käsialalla ei tulisi olla merkitystä arvioinnissa, koska se on vain mekaaninen taito (esim. Charney 1984; Vaughan 1991). Maalaisjärkeä käyttäen käsialalla on kaksi ulottuvuutta: se on joko luettavaa tai se ei ole luettavaa. Käsialalla on oikeastaan merkitystä vain silloin, kun sitä ei voi lukea. Toki käsitys käsialan selvydestä on jossakin määrin lukijakohtaista, ja sen nostaminen kriteeriksi voi itse asiassa vähentää sen muuten piiloon jäävää vaikutusta arvioihin. On kuitenkin vaikeaa kuvitella, miten taitoton 5 käsiala eroaisi taitotason 3 käsialasta.

Käsialan merkitystä voi vähentää tulevaisuudessa se, että tekstit kirjoitetaan yhä useammin koneella, jolloin kirjoittajat ovat tältä osin tasavertaisessa asemassa. Tasavertaisuus edellyttää tietysti, että kaikki kirjoittajat ovat tottuneita käyttämään tietokoneita. Toisaalta tietokoneella kirjoitetuissa teksteissä puutteet korostuvat selvemmin ja tekstien odotetaan olevan jollakin tavalla viimeisteltyjä, joten käsinkirjoitettujen tekstien puutteisiin saatetaan suhtautua suvaitsevaisesti ja tekstit saatetaan arvioida paremmiksi (esim. Powers ym. 1994; Brown 2000).

Tekstin pituus on verrattavissa käsialaan siinä, että pituuskaan ei ole analosoitavissa opittavana taitona, kuten tekstin jäsentely tai sanasto. Tekstin pituus on myös jossakin määrin subjektiivinen käsite. Arvioinnin näkökulmasta sillä on eniten merkitystä silloin, kun teksti on liian niukka arvioitavaksi tai jos teksti on lyhyt mutta muuten kaikella tavalla tehtävänannon vaatimukset täyttävä. Pituuden nostaminen kriteeriksi voi kuitenkin viedä huomion pois tekstin sisällön käsittelyltä ja jäsentelyltä. Hyvä teksti ei ole aina pitkä tai toisin päin.

Rakenteiden ohella sanasto on kirjoittamisen peruskriteereitä, vaikka siihen on kiinnitettykin vähemmän huomiota (Polio 2001, 98). Sanasto, kuten rakenteetkaan, ei ole yksi yksikkö vaan se sisältää useita eri alakategorioita. Tavallisesti testausarvioinnissa kiinnitetään erityistä huomiota sanaston laajuuteen eli siihen, miten hyvin se riittää jostakin tehtävästä suoriutumiseen. Jos oppija käyttää perussanastoa, se voi tarkoittaa esimerkiksi 500–2000 sanan joukkoa, joka sisältää frekven-

teimpiä, konkreettisia ja jokapäiväiseen elämään liittyviä ilmauksia (vrt. Nation 1990). Sanaston käyttöä tarkastellaan myös esimerkiksi sen varioivuuden, käytön tarkkuuden, idiomaattisuuden ja sopivuuden näkökulmasta (esim. Takala 1989; Nation & Newton 1997; Polio 2001). Sanan hallinta ei ole myöskään dikotomista, vaan siinä on erilaisia tasoja (ks. Paribakht & Wesche 1997).

Monista tutkimuksista ei käy kuitenkaan ilmi, mitä sanaston osaluettua arvioijat teksteistä kommentoivat, mikä voi johtua siitä, että tutkimustuloksissa kommentit kategorisoidaan yläotsikoiden alle. Toisaalta rakenteista näytettäisiin tuovan esiin alakategorioita, kuten aikamuotojen käyttö tai morfologiset ongelmat (ks. taulukosta 8). *Sanastoa* ehkä käytetään viittaamaan löyhemmin ja epäsystemaattisemmin eri alakategorioihin kuin *rakenteita*. Toisekseen kirjoittamisessa rakenteiden ja sanaston välinen raja, varsinkin sanaston ja syntaksin välillä, on usein hämärä ja keinotekoinen. Ehkä osa sanastollisista huomioista tulee kommentoiduksi rakenteina.

Taulukon 8 mukaan arvioijat kommentoivat teksteistä myös sisältöä, jäsentelyä ja koheesiota. Sisältöä kommentoitiin lähinnä sen mukaan, miten heikko/vahva, tylsä/elävä, hyvä/huono tai epäselkeä se on (Vaughan 1991, 119; Milanovic ym. 1996, 102, 104). Kommentit vaikuttavat sekä pintapuolisilta että subjektiivisilta sisällön moniulotteisuuteen (esim. käsittelytapa, tieto asiasta, argumentointi) nähden. Toisaalta sisältöä pidetään ongelmallisena kriteerinä varsinkin edistyneiden oppijoiden tekstien arvioinnissa, koska arvioijilla esimerkiksi akateemisissa konteksteissa ei ole useinkaan sen arviointiin tarvittavaa asiantuntemusta (esim. Hamp-Lyons 1991b, 138–139). Joidenkin mielestä taas sisällöllä ei ole juuri vaikutusta arvioon (Milanovic ym. 1996, 105). Sisällön arvioinnin subjektiivisuutta taas korostaa ehkä se, että arvioijat kommentoivat sitä sen mukaan, pitävätkö he itse siitä tai viihtyvätkö he lukijoina.

Samoin tekstin jäsentelyä ja koheesiota ei eritellä taulukon 8 kommenteissa sen tarkemmin. Huonoa jäsentelyä esimerkiksi kommentoidaan siksi, että sen vuoksi teksti tulee hylätyksi, tai siksi, että hyvä jäsentely on syy hyväksymiseen (Vaughan 1991, 117). Kirjoittamisen näkökulmasta molemmat ovat olennaisia kriteereitä, mutta taidon kehittymistä niiden osalta on ehkä vaikeaa analysoida ja osoittaa konkreettisesti. Tähän voi olla syynä, että tällaiset kriteerit ovat kyllä luontevasti sisällytettävissä osaksi opetusta, mutta niihin puuttuminen ja niiden korjaaminen teksteistä ei ole yhtä helppoa kuin esimerkiksi rakenteiden arvioiminen (Polio 2001, 104–105).

Arvioijat siis nostivat ainakin näissä tutkimuksissa esiin varsin monenlaisia erilaisia kirjoittamisen piirteitä. Toinen asia on tietysti se, miten tasapuolisesti ne ovat esillä, miten arvioijat tulkitsevat niitä ja miten

hyvin nämä piirteet edustavat kirjoittamista taitona. Tutkimusten mukaan näyttää siltä, että arvioijat tulkitsevat piirteitä eri tavalla, kuin esimerkiksi arviointikriteerien määritelmät edellyttäisivät. Tämä voi johtua siitä, että osaa piirteistä on vaikea soveltaa esimerkiksi niiden abstraktisuuden vuoksi tai että piirteiden kuvaukset jättävät liikaa tulkinnan vapautta (vrt. DeRemer 1998; Smith 2000).

Edellä käsitellyt kriteerit ovat yhteydessä arvioijien käyttämiin lukustrategioihin siten, että arvioija saattaa tukeutua vain yhteen tai kahteen kriteeriin ja hakea vahvistusta vain tietyille kriteereille (esim. Vaughan 1991; Smith 2000). Tällöin tekstin lukeminen jää varsin yksipuoliseksi. Samoin yhden silmäilykerran strategia tuntuu riittämättömältä, vaikka kyseessä olisikin holistinen arviointi. Strategioissa korostuu myös subjektiivisuus, ainakin sellaisissa tapauksissa, joissa tekstin arvo tulee punnituksi sen mukaan, pitääkö arvioija tekstistä vai ei.

Oman toiminnan reflektointi ja säätely tulevat oikeastaan esille vain Cummingin (1990) tutkimuksessa. Näyttää siltä, että niitä korostaisivat tutkimuksen kokematon-kokenut -asetelma. Joka tapauksessa nämä tutkimukset jossakin määrin vahvistavat käsitystä siitä, että arvioijat kiinnostavat huomiota melko yhdenmukaisesti samoihin piirteisiin. Toisaalta ne taas panevat ihmettelemään, miten arvioijat voivat yleensäkin päätyä samoihin arvioihin samasta tekstistä, jos he tukeutuvat kriteereihin sattumanvaraisesti tekstin, lukustrategian tai oman mieltymyksen mukaan.

4.3.2 Lyhyesti arviointiprosessin ja arvion suhteesta

Olen käsitellyt edellä arviota sekä mittauksen että prosessin näkökulmasta. Ennen kuin siirryn tuloksiin, on paikallaan pohtia hieman kolmannen ja neljännen luvun keskinäistä suhdetta. Voiko mittausvirhetä lähestyä prosessista käsin tai toisaalta voisivatko esimerkiksi arvioijien lukustrategiat olla muokattavissa identtisiksi arvioijakoulutuksen avulla?

Ensitutuntumalta vaikutelma on sellainen, että reliabiliuden näkökulmasta prosessien tarkastelu ei ole mielekäästä, koska prosessiin ei katsota päästävän käsiksi; arvio itsessään on riittävä informaation lähde. Jos arvio ei ole odotuksenmukainen suhteessa asetettuun normiin, mittausvirheen paikantamisella voidaan selittää, miksi niin ei ole. Prosessista käsin katsottuna taas arviokeskeisyys tai arvioijan tarkasteleminen mittausvirheen lähteenä jättää täysin huomiotta sen, miten ja miksi arvioija on päätenyt arvioonsa. Arviointiprosessin ymmärtäminen edellyttää, että arvioijan toiminnan inhimillisuus ja siihen kuuluva tietynasteinen epäjohtonmukaisuus hyväksytään.

Toisin sanoen arvioinnissa esiintyvien poikkeamien luokittelu viestii siitä, millaisena arvioijan rooli arviointiprosessissa nähdään: tulkitaanko poikkeamat virheiksi vai yleensäkin arviointiin vaikuttaviksi tekijöiksi. Virhelähtöinen ajattelu antaa olettaa, että virhe arvioijan käyttäytymisessä on korjattavissa ainakin jossakin määrin 'mekaanisesti', esimerkiksi vaihtamalla arvioijaa tai lisäämällä arvioijakoulutuksen määrää. Prosessitai vuorovaikutusnäkökulman mukaan kyse on monimuotoisemmasta ilmiöstä: arvioijan toimintaan voidaan kyllä vaikuttaa, mutta se edellyttää toiminnan reflektointia ja omien toimintamallien muuttamista, mikä voi tapahtua hitaasti ja asteittaisesti. Lisäksi toiminnassa on aina varmasti jotakin, joka ei ole odotuksenmukaista, esimerkiksi kunkin uuden tekstin tuoma ainutkertaisuus. Arvioijan toiminta on aina jossakin määrin yksilöllistä ja subjektiivista. Tämän ottaa huomioon myös psykometrinen näkökulma, tosin se tarkastelee arvioijaa tavallaan irrallaan eikä osana arviointidiskurssia.

Molemmat näkökulmat tarjoavat objektiivin, jolla voi tarkentaa näkymää arvioijan toimintaan. Pelkkä yksilön ymmärtäminen ja kuvaaminen jättäisivät kertomatta sen, miten arvioijat viime kädessä arvottavat suorituksia arvioilla ja missä suhteessa nämä arviot ovat toisiinsa. Pelkät numerot tai arviot eivät puolestaan kerro sitä, mitä ne arvioivat ja miten arvioija on niihin päätenyt. Tässä mielessä arvioiden laadullinen tarkastelu on myös validiuden tarkastelua eli pohdintaa siitä, arvioidaanko todella sitä, mitä on tarkoitus arvioida. Molempia näkökulmia siis tarvitaan, sillä kumpikaan yksinään ei pysty kertomaan, selittämään ja kuvaamaan tutkittavaa ilmiötä yhtä kokonaisvaltaisesti kuin molemmat yhdessä, vaikka niiden kertomat olisivat jossakin määrin ristiriitaisia.

5 TULOKSET: ARVIOIJAT VIESTIVÄT NUMEROIN JA SANOIN

Esittelen ensin kvantitatiiviset tulokset eli tarkastelen numeroiden valossa arvioijien sisäistä johdonmukaisuutta ja arvioijien välistä yhdenmukaisuutta. Yhdenmukaisuuden yhteydessä selvitän lyhyesti myös arviointilinjoja eli arvioijien lempeyttä/ankaruutta suhteessa toisiinsa sekä asteikon muuttumisen vaikutusta arvioihin. Sitten siirryn haastatteluihin, joita käsittelen haastattelu kerrallaan: ensin ensimmäisen arviointikierroksen jälkeistä haastattelua ja sitten toisen arviointikierroksen jälkeistä haastattelua. Ennen kuin siirryn kolmanteen haastatteluun, pohdiskelen lyhyesti kahden ensimmäisen haastattelun pohjalta syntyvää arvioijan lukijakuvaa. Tulosten lopuksi raportoin kolmannen arviointikierroksen jälkeisen valintaverkkohaastattelun (repertory grid) omana kokonaisuutenaan.

Kertaan tulosten aluksi lyhyesti tutkimuksen koeasetelman, koska tulosten, varsinkin kvantitatiivisten, raportointi ja tarkastelu etenee sen mukaisesti. Kuten taulukosta 9 näkyy, tehtäväsarjoja on kaikkiaan neljä (TS) ja ne kukin sisältää kolme tehtävää (TA1–TX), jotka ovat tehtäväsarjoittain eri taitotasoin suunnattuja. Kutakin tehtäväsarjaa pyrittiin kirjoituttamaan keskimäärin sitä vastaavan taitotasoin oppijoilla, esimerkiksi vaativinta tehtäväsarjaa 4 kirjoittivat mahdollisuuksien mukaan edistyneimmät oppijat. Tehtäväsarjojen tehtäväkuvaukset on esitetty tarkemmin luvussa 2.1.1.

TAULUKKO 9 Arviointiprosessin asetelma.

Tehtäväsarjat/kirjoitustehtävät										Arvioijan tunnistenumero									
TS	TA1	TA2	TB1	TB2	TB3	TC1	TC2	TC3	TX	I	II	III	IV	V	VI	VII	VIII	IX	X
TS1	50		50						50	3	4	10	16	1	5	7	11	13	12
TS2		57					57		57	3	4	10	16	1	6	9	11	14	15
TS3					101	101			101	3	4	10	16	2	8	7	9	18	17
TS4				50				50	50	3	4	10	16	5	6	15	14	13	12

Taulukosta 9 huomataan, että kaikkiaan 17 arvioijasta 4 arvioijaa (arvioijien tunnistenumerot 3, 4, 10 ja 16) arvioi kaikkien kirjoittajien tekstit ja muut arvioijat vähintään 100 kirjoittajan tekstit eli 1–2 tehtäväsarjan kirjoittaneiden tekstit. Taulukon mukaisesti tehtäväsarjojen 1 ja 4 kirjoittajat saivat arvion yhdeksältä arvioijalta ja tehtäväsarjojen 2 ja 3 kirjoittajat kymmeneltä arvioijalta. Eri tehtävätyypeistä ainoastaan tehtävä TX eli kuvasta kertominen oli sisällytetty kaikkiin tehtäväsarjoihin; muut tehtävätyypit tulivat arvioijalle tutuksi sen mukaan, mitä tehtäväsarjaa tai -sarjoja hän arvioi.

Käytän alaluvussa 5.1 arvioijista koodeja, esimerkiksi A140204, joissa koodin ensimmäinen kapiteelikirjain viittaa arviointikierrokseen: A ensimmäiseen, B toiseen ja C kolmanteen arviointikierrokseen. Koodin kaksi seuraavaa numeroa (01–18) viittaavat arvioijan tunnistenumeroon ja neljä viimeistä numeroa (1, 2, 3, 4) tehtäväsarjaan sen mukaan, mitä tehtäväsarjaa tai -sarjoja arvioija on arvioinut. Tässä yhteydessä 0 jonkin neljän viimeisen numeron paikalla tarkoittaa, että arvioija ei ole arvioinut kyseistä tehtäväsarjaa. Toisin sanoen koodissa B140204 on kyse toisesta arviointikierroksesta ja arvioijasta numero 14, joka on arvioinut toisen ja neljännen tehtäväsarjan.

Haastatteluja koskevissa tuloksissa käytän arvioijista pseudonyymejä, jotka ovat naisten/miehen nimiä. Arvioijakohtaiset tunnistenumerot ja pseudonyymit vastaavat toisiaan seuraavasti:

011200 <i>Kaisu</i>	020030 <i>Teija</i>	031234 <i>Tarja</i>
041234 <i>Eeva</i>	060204 <i>Tiina</i>	071030 <i>Inkeri</i>
080030 <i>Raija</i>	090230 <i>Heidi</i>	101234 <i>Pia</i>
111200 <i>Tuula</i>	121004 <i>Päivikki</i>	131004 <i>Pirkko</i>
140204 <i>Martti</i>	150204 <i>Merja</i>	161234 <i>tutkija</i>
170030 <i>Sanna</i>	180030 <i>Pirjo</i>	

Käytän alaluvussa 5.1 ensimmäisestä arviointikierroksesta nimitystä alkumittaus, jossa arvioijat arvioivat suoritukset itse määrittelemillään 6-portaisilla asteikoilla. Toisesta arviointikierroksesta käytän nimitystä 2. mittaus, jossa arvioijat käyttivät valmiiksi määriteltyä 6-portaista asteikkoa, ja kolmannelle arviointikierroksesta 3. mittaus, joka oli toisen mittauksen toisto.

5.1 Mitä tasoarviot kertovat johdonmukaisuudesta ja yhdenmukaisuudesta?

5.1.1 Arvioijien sisäinen johdonmukaisuus

Koska alkumittaus pohjautui arvioijan itsensä määrittelemään 6-portaiseen arviointiasteikkoon ja koska toisessa ja kolmannessa mittauksessa arvioijat käyttivät samaa, määriteltyä asteikkoa, odotuksenmukaista olisi, että arvioijan sisäinen johdonmukaisuus olisi suurempi toisen ja kolmannen mittauksen välillä kuin ensimmäisen ja toisen tai ensimmäisen ja kolmannen mittauksen välillä. Jos katsotaan arvioijakohtaisia keskimääriäisiin arvioihin perustuvia tendenssejä, tilastanalyysien perusteella näyttää siltä, että arvioijat voidaan jakaa neljään eri tyyppiin sen mukaan, millaisia kolmessa mittauksessa saatujen arvioiden erot keskimääräisesti olivat.

Tyypittely on tehty sen perusteella, miten arvioijien antamien kaikkien tasoarvioiden keskiarvot erosivat toisistaan kussakin mittauksessa. Toisin sanoen kunkin mittauskerran keskiarvoja on verrattu toisiinsa graafisesti, mikä on nähtävissä kuvioista 15–18.¹⁵

Tyyppi 1: Arvioinnit kaikilla mittauskerroilla eroavat keskimäärin toisistaan (N=7).

Tyyppi 2: Alkumittaus eroaa joko toisesta mittauksesta tai kolmannelta mittauksesta (N= 5).

Tyyppi 3: Alkumittaus eroaa sekä toisesta että kolmannelta mittauksesta (N=4).

Tyyppi 4: Eri mittauskerrat eivät eroa toisistaan tilastollisesti merkitsevästi (N=1).

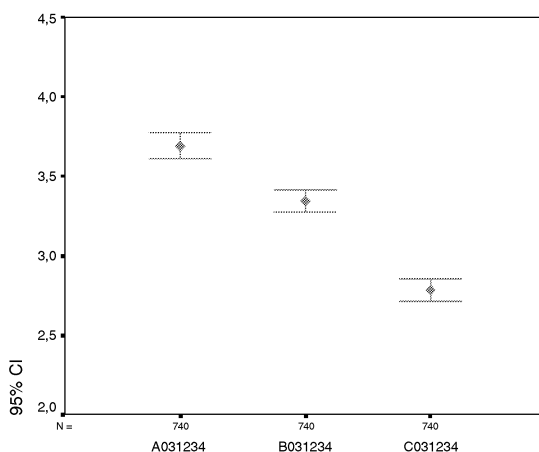
Sisäisen johdonmukaisuuden näkökulmasta katsottuna ensimmäisen tyyppin arvioijat eivät toimineet odotuksenmukaisesti, koska määritelty asteikko ei johdonmukaistanut arviointia, kuten kuvioista 15 näkyy. Kuvion 15, kuten myös kuvioista 16–18, x-akselin N tarkoittaa arvioitujen tekstien määrää ja y-akselin numerot viittaavat annettujen arvioiden tasokeskiarvoihin. Merkintä 95 % CI tarkoittaa luottamusväliä eli keskiarvo (salmiakkiruutu) on 95 prosentin luotettavuudella luottamusvälin ala- ja ylärajan välissä. Arvioija A031234 on siis antanut ensimmäisessä mittauksessa keskimäärin korkeampia arvioita kuin toisessa ja kolmannessa mittauksessa.

Kuvion 15 mukaan näyttää siltä, että arvioija on toiminut epäjohdonmukaisesti jokaisella kolmella mittauskerralla eli todennäköisesti hän on antanut arvionsa eri periaattein eri kierroksilla. Hän ei

¹⁵ Niiden arvioijien profiilit, jotka eivät näy kuvioissa 15–18, ovat nähtävissä liitteessä 6.

ole todennäköisesti tukeutunut annettuun asteikkoon joko toisella tai kolmannella mittauskerralla tai sitten ei kummallakaan, vaan hän on antanut arviot johonkin muuhun kriteeriin tai kriteereihin tukeutuen. Hän on kyllä saattanut tukeutua annettuun asteikkoon mutta on tulkinnut ja soveltanut sitä eri tavalla eri kierroksilla.

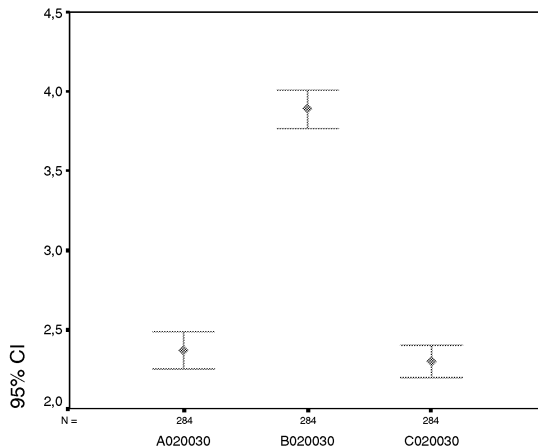
Useimmissa ensimmäisen tyyppin tapauksissa arvioijan skaala kiristyi kierros kierrokselta eli kaikkein matalampia arvioita arvioija antoi keskimääräisesti kolmannella kierroksella. Tosin erot eivät aina ole yhtä selviä kuin kuviossa 15, sillä joidenkin arvioijien kohdalla johdonmukaisuus oli toisen ja kolmannen kierroksen välillä silmämääräisesti katsottuna johdonmukaista, vaikka kierrosten välillä olikin toistettujen mittausten Anovan mukaan tilastollisesti merkitsevä ero. Tällaisten arvioijien kuvio muistuttaa paljon kolmannen tyyppin kuviota (ks. kuvio 17). Toisaalta on otettava huomioon myös se, että pienikin satunnaisvaihtelu tuo esiin tilastollisesti merkitsevän eron ja eroihin vaikuttavat myös havaintoarvojen määrät, jotka vaihtelevat arvioijittain.



KUVIO 15 Esimerkki tyyppistä 1.

Toisen tyyppin arvioijille on tyypillistä, että heidän itse määrittelemänsä asteikon mukaan annetut arviot olivat keskimääräisesti johdonmukaisempia määritellyn asteikon mukaisten arvioiden kanssa joko toisessa tai kolmannessa mittauksessa. Toisin sanoen tällainen tendenssi antaa ymmärtää, että arvioijat olisivat tukeutuneet joko toisessa tai kolmannessa mittauksessa määriteltyyn asteikkoon ja toisella näistä joko tukeutuneet itse määrittelemäänsä asteikkoon määritellystä asteikosta

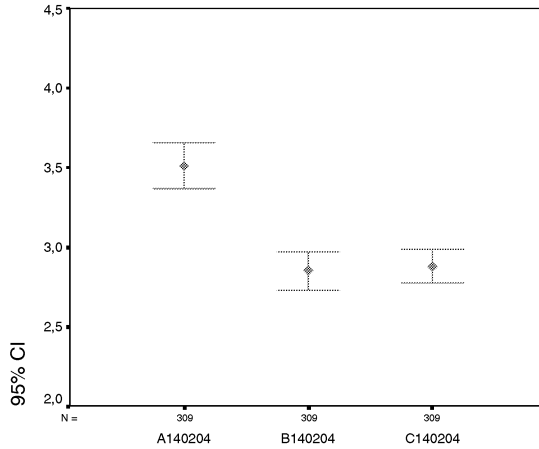
huolimatta tai sitten tukeutuneet määriteltyyn asteikkoon molemmissa mittauksissa mutta soveltaneet sitä eri tavalla eri mittauskerroilla.



KUVIO 16 Esimerkki tyypistä 2.

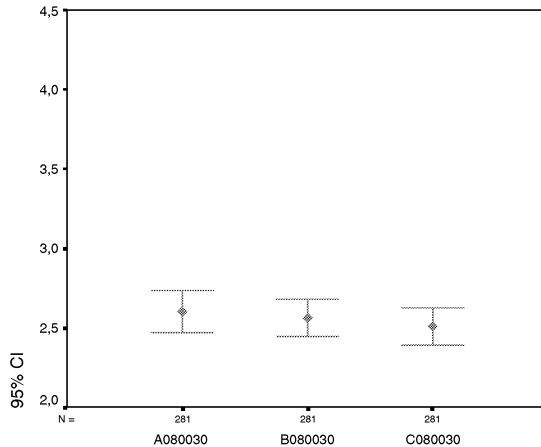
Arvioijatyyppiä 2 havainnollistavan kuvion 16 mukaan näyttäisi siltä, että arvioija on käyttänyt samoja kriteereitä ensimmäisessä ja kolmannessa mittauksessa ja toisessa mittauksessa taas joitakin muita kriteereitä. On kuitenkin mahdotonta sanoa, mihin toisen mittauksen huomattava ero perustuu ja kummalla toisesta tai kolmannelle mittaukselle arvioija on itse asiassa käyttänyt määriteltyä asteikkoa. Yksioikoinen tulkinta johtaisi päättelemään, että arvioija on voinut soveltaa määriteltyä asteikkoa toisessa mittauksessa ja palata takaisin itse määrittelemäänsä asteikkoon kolmannessa mittauksessa, mutta yhtä hyvin hän on voinut tulkita määriteltyä asteikkoa tiukemmin tai tarkemmin kolmannessa mittauksessa.

Kolmannen tyypin arvioijien (N=4) toiminta on kaikkein odotuksenmukaisinta, koska heidän kohdallaan toisen ja kolmannen mittauksen välillä ei ole tilastollisesti merkitsevää eroa ja ensimmäinen mittaus taas eroaa toisesta ja kolmannelle, kuten kuviosta 17 näkyy. Tämän perusteella voi ainakin varovaisesti olettaa, että määritelty asteikko on johdonmukaistanut arviointia sillä perusteella, että ensimmäisessä mittauksessa käytetty asteikko eroaa sisällöltään määritellystä asteikosta ja ero toisiinsa nähden johdonmukaisiin toiseen ja kolmanteen mittaukseen on selvä.



KUVIO 17 Esimerkki tyypistä 3.

Kuvion 17 arvioija on siinä mielessä hyvä esimerkki, että tyypin 3 arvioijat antoivat keskimäärin korkeampia arvioita ensimmäisessä mittauksessa kuin toisessa ja kolmannessa mittauksessa. Yhtenä syynä eroon voi olla se, että määritelty asteikko oli kriteereiltään tiukempi verrattuna arvioijien itsensä luomaan asteikkoon. Tosin arvioijien itsensä määrittelemiä asteikkoja voi olla turhan rohkeaa verrata suoraan toisiinsa, vaikka tendenssin mukaan ensimmäisen kierroksen arviot näyttäisivät olleen keskimääräisesti korkeampia.



KUVIO 18 Esimerkki tyypistä 4.

Viimeisen eli neljännen tyyppin ainoa arvioija voitaisiin liittää osaksi kolmatta tyyppiä sikäli, että johdonmukaisuudessa ei ole eroa toisessa ja kolmannessa mittauksessa. Tosin nämä kaksi mittausta eivät myöskään eroa alkumittauksesta, eli määrittelyn asteikon käyttäminen ei ole muuttanut arvioita ainakaan keskimääräisesti. Tätä voi selittää muun muassa se, että arvioijan itsensä määrittelemän asteikon sisältö oli samantyyppinen määrittelyn asteikon sisällön kanssa tai sitten arvioija tukeutui omaan asteikkoonsa tai systemaattisesti johonkin tiettyyn kriteeriin tai yleensäkin omaan kielikäsitteeseensä myös toisessa ja kolmannessa mittauksessa. Arvioijien omien asteikkojen sisältöä ja heidän kommenttejaan määrittelystä asteikosta ja sen soveltamisesta käsitellään tarkemmin luvuissa 5.2.1 ja 5.2.2.

Kun arvioijien sisäistä johdonmukaisuutta tarkastellaan korrelaatioiden valossa, odotuksenmukaista olisi, että korrelaatio olisi korkeampi toisen ja kolmannen mittauksen kuin muiden mittauskertojen välillä. Taulukossa 10 näkyvät Pearsonin korrelaatiot on laskettu arvioija-kohtaisesti kunkin mittauskerran välillä eli alkumittauksen ja toisen mittauksen arvioiden välillä (A–B), toisen ja kolmannen mittauksen (B–C) sekä alkumittauksen ja kolmannen mittauksen (A–C) välillä. Arvioitavien tekstien määrät vaihtelivat arvioijittain 282:n ja 740:n välillä.

Kuten taulukosta 10 näkyy, korrelaatiot ovat kauttaaltaan korkeahkoja. Tiivistetysti sanoen ne vaihtelivat alku- ja toisen mittauksen välillä arvioijittain siten, että matalin korrelaatio oli .62 ja korkein .83 ja kaikkien korrelaatioiden keskiarvo oli .76. Alkumittauksen ja kolmannen mittauksen välillä matalin korrelaatio oli .66, korkein .85 ja kaikkien korrelaatio keskiarvo .76. Toisen ja kolmannen mittauksen välisistä korrelaatioista matalin oli .69 ja korkein .85, ja kaikkien korrelaatioiden keskiarvo oli .79. Kaikki korrelaatiot olivat tilastollisesti merkitseviä tasolla 0.01.

TAULUKKO 10 Arvioijakohtaiset Pearsonin korrelaatiot eri mittauskertojen välillä.

	A011200	B011200	C011200
A011200	1,000	,825**	,819**
B011200	,825**	1,000	,813**
C011200	,819**	,813**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 309

	A020030	B020030	C020030
A020030	1,000	,685**	,690**
B020030	,685**	1,000	,698**
C020030	,690**	,698**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 284

	A031234	B031234	C031234
A031234	1,000	,655**	,709**
B031234	,655**	1,000	,816**
C031234	,709**	,816**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N=740

	A041234	B041234	C041234
A041234	1,000	,823**	,827**
B041234	,823**	1,000	,822**
C041234	,827**	,822**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 740

TAULUKKO 10 (jatkuu)

	A060204	B060204	C060204
A060204	1,000	,779**	,763**
B060204	,779**	1,000	,805**
C060204	,763**	,805**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 308

	A080030	B080030	C080030
A080030	1,000	,802**	,728**
B080030	,802**	1,000	,739**
C080030	,728**	,739**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 284

	A101234	B101234	C101234
A101234	1,000	,755**	,727**
B101234	,755**	1,000	,770**
C101234	,727**	,770**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 737 (AB), N = 727 (AC), N = 734 (BC)

	A121004	B121004	C121004
A121004	1,000	,773**	,762**
B121004	,773**	1,000	,788**
C121004	,762**	,788**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 294

	A140204	B140204	C140204
A140204	1,000	,818**	,797**
B140204	,818**	1,000	,807**
C140204	,797**	,807**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N=309

	A161234	B161234	C161234
A161234	1,000	,832**	,804**
B161234	,832**	1,000	,853**
C161234	,804**	,853**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 740

	A180030	B180030	C180030
A180030	1,000	,715**	,694**
B180030	,715**	1,000	,778**
C180030	,694**	,778**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 284

	A071030	B071030	C071030
A071030	1,000	,738**	,771**
B071030	,738**	1,000	,748**
C071030	,771**	,748**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 431

	A090230	B090230	C090230
A090230	1,000	,811**	,846**
B090230	,811**	1,000	,835**
C090230	,846**	,835**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 446

	A111200	B111200	C111200
A111200	1,000	,727**	,707**
B111200	,727**	1,000	,797**
C111200	,707**	,797**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 309

	A131004	B131004	C131004
A131004	1,000	,858**	,804**
B131004	,858**	1,000	,847**
C131004	,804**	,847**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 294

	A150204	B150204	C150204
A150204	1,000	,721**	,772**
B150204	,721**	1,000	,814**
C150204	,772**	,814**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 309

	A170030	B170030	C170030
A170030	1,000	,616**	,661**
B170030	,616**	1,000	,764**
C170030	,661**	,764**	1,000

** Korrelaatio on merkitsevä tasolla 0.01.
N = 284

Odotuksenmukaisesti korrelaatiot ovat suurimmalla osalla arvioijista korkeimpia toisen ja kolmannen mittauksen välillä, joskaan erot eivät ole kovin suuria muiden mittauskertojen välisiin korrelaatioihin verrattuna.

Toisin sanoen korrelaatioiden väliset erot eri mittauskertojen välillä ovat kauttaaltaan pieniä. Jos katsotaan korrelaatioerojen tilastollista merkitsevyyttä, 11 arvioijasta, joilla toisen ja kolmannen mittauksen välinen korrelaatio oli korkeampi, ero on tilastollisesti merkitsevä 8 arvioijalla eli arvioijilla

031234 ($p=.00$) 060204 ($p=.02$) 101234 ($p=.003$) 111200 ($p=.0001$)
150204 ($p=.03$) 161234 ($p=.00$) 170030 ($p=.0006$) 180030 ($p=.001$)

(ks. myös liite 7). Toisin sanoen näillä kahdeksalla arvioijalla määritelty asteikko oli korrelaatioiden valossa johdonmukaistanut arviointia. Sen sijaan kolmen muun arvioijan eli arvioijien 020030, 080030 ja 121004 toisen ja kolmannen mittauksen välisen korrelaation ero muihin korrelaatioihin ei osoittautunut tilastollisesti merkitseväksi.

Vaikka arvioijien 011200, 041234, 071030 ja 140204 korrelaatiot eivät ole odotuksenmukaisesti suurimpia toisen ja kolmannen mittauksen välillä, eri mittauskertojen välisillä korrelaatioilla ei ole kuitenkaan tilastollisesti merkitsevää eroa. Myöskään jo edellä esiin tulleiden arvioijien 020030, 080030 ja 121004 millään eri mittauskertojen välisillä korrelaatioilla ei ole tilastollisesti merkitsevää eroa. Toisin sanoen näiden arvioijien kohdalla näyttää siltä, että arvioissa ei ole tapahtunut kovin suurta muutosta eri mittauskertojen välillä, vaan he ovat antaneet samansuuntaisia tasoarvioita mittauskerrasta toiseen.

Arvioijien 090230 ja 131004 toisen ja kolmannen mittauksen korrelaatio on odotuksenvastaisesti pienempi, ja sen ero muiden mittauskertojen välisiin korrelaatioihin on myös tilastollisesti merkitsevä. Arvioijat eivät ehkä ole soveltaneet määriteltyä asteikkoa samalla tavalla toisessa ja kolmannessa mittauksessa, tai he ovat voineet tukeutua esimerkiksi toisessa mittauksessa enemmän omaan kirjoittamiskäsitykseensä kuin määriteltyyn asteikkoon.

Missä määrin arvioijat sitten antoivat samoja arvioita eri kierroksilla? Taulukon 11 sarakkeissa 1, 2 ja 3 näkyy arvioijakohtainen samanmielisyys (exact agreement) prosentteina eri mittauskertojen välillä. Samanmielisyysprosentit ovat yhteen laskettuja prosentteja ristiintaulukoinnista eri mittauskertojen välillä, eli ne osoittavat sen, missä määrin arvioijat ovat arvioineet suoritukset täsmälleen samalla tavalla eri mittauskerroilla. Mitä suurempi prosenttiluku on, sitä enemmän arvioijat ovat antaneet samoja arvioita eri kierroksilla. Ihanteellinen tulos toisen ja kolmannen mittauksen välillä olisi 100 prosenttia, jolloin arvioija olisi arvioinut suoritukset täsmälleen samalla tavalla molemmissa mittauksissa.

Lähtöoletuksen mukaan arvioijien samanmielisyyden pitäisi olla suurempaa toisen ja kolmannen mittauksen välillä kuin muiden mittaussten välillä, mikäli arvioijan itsensä määrittelemän asteikon sisältö erosi annetusta asteikosta ja mikäli arvioija nojasi arvioinnissaan määriteltyyn asteikkoon. Kuten taulukosta 11 näkyy, suurimmalla osalla arvioijista

toisen ja kolmannen mittauksen välinen samanmielisyyssprosentti on suurempi verrattuna ensimmäisen ja toisen sekä ensimmäisen ja kolmannen mittauksen välisiin samanmielisyyssprosentteihin.

Tosin prosenttilukuja verratessa on syytä ottaa huomioon, ovatko erot niiden välillä tilastollisesti merkitseviä. Taulukon 11 kolme oikeanpuoleista saraketta osoittavat, missä määrin ero prosenttilukujen välillä on merkitsevä. Tilastollinen merkitsevyys on laskettu McNemarin testillä (ks. myös liite 8). Arvioijien 060204, 090230, 111200, 140204, 150204, 161234 ja 180030 samanmielisyyss on tilastollisesti merkitsevästi suurempi toisen ja kolmannen mittauksen kuin muiden mittausten välillä. Näiden arvioijien kohdalla määritelty asteikko näyttäisi siis johdonmukaistaneen arviointia. Tosin näidenkin arvioijien kohdalla on eroa samanmielisyyssprosentteissa: alimmillaan se on 61,8 prosenttia ja ylimmillään 74,4, jota voi jo pitää varsin hyvänä (vrt. Dunn 1989; Abedi ym. 1993).

Myös arvioijien 011200, 060204, 080030, 121004 ja 170030 samanmielisyyss on suurempi prosenttien valossa toisen ja kolmannen mittauksen välillä mutta erot prosentteissa ovat juuri ja juuri tilastollisesti merkitseviä tai ne eivät ole lainkaan merkitseviä. Näiden viiden arvioijan kohdalla asteikon muuttumisella ja muuttumattomuudella ei näyttäisi olleen merkitystä samanmielisyyteen.

TAULUKKO 11 Arvioijakohtainen samanmielisyyss prosentteina.

Arvioija	A	B	C	Arvioiden määrä (N)	Eron merkitsevyys		
	1. ja 2. mittausten välillä (%)	1. ja 3. mittausten välillä (%)	2. ja 3. mittausten välillä (%)		A vs. B	B vs. C	A vs. C
011200	64.0	62.7	71.1	309		*	
020030	7.0	55.2	7.3	284	***	***	
031234	48.7	30.9	42.9	740	***	***	*
041234	61.8	62.5	68.2	740		*	**
060204	51.2	48.7	61.8	308		***	**
071030	67.0	71.4	64.9	431		*	
080030	59.7	56.7	62.9	283			
090230	52.9	59.4	72.4	446	**	***	***
101234	53.0	41.5	53.6	737	***	***	
111200	25.8	22.6	68.6	309		***	***
121004	50.3	48.2	54.0	294			
131004	42.8	21.0	39.4	294	***	***	
140204	42.0	39.0	65.6	309		***	***
150204	53.3	53.0	74.4	309		***	***
161234	60.9	53.1	72.8	740	***	***	***
170030	56.3	55.9	63.3	284			
180030	42.9	50.1	65.0	284	*	**	***

* p < 0.05, ** p < 0.01, *** p < 0.001

Neljällä arvioijalla samanmielisyys on ollut suurempaa ensimmäisen ja toisen mittauksen (031234 ja 131004) tai ensimmäisen ja kolmannen (020030 ja 071030) välillä kuin toisen ja kolmannen mittauksen välillä. Tosin erot arvioijalla 071030 eivät ole kovin merkitseviä, vaan hän näyttäisi olleen melko johdonmukainen jokaisessa kolmessa mittauksessa. Sen sijaan kolmella muulla arvioijalla erot näyttäivät selvemmilta. Varsinkin arvioija 020030 herättää huomiota erikoisella profiilillaan; arviot ovat olleet kaikkein lähimpänä toisiaan alkumittauksessa ja kolmannessa mittauksessa. Tuntuu siltä, että hän on tukeutunut toisessa mittauksessa aivan eri kriteereihin kuin alkumittauksessa ja kolmannessa mittauksessa.

Arvioijien 031234 ja 131004 korkeampi samanmielisyys ensimmäisessä ja toisessa mittauksessa saa kysymään, miksi heidän samanmielisyytensä vähenee määritellyn asteikon myötä, kun sen pitäisi vaikuttaa aivan päinvastoin. Ovatko he kenties tukeutuneet toisessa mittauksessa enemmän omaan asteikkoonsa vai osiin siitä ja vasta kolmannella kierroksella määriteltyyn asteikkoon? Vai ovatko he käyttäneet aivan eri kriteereitä tai osin eri kriteereitä kussakin mittauksessa, koska samanmielisyysprosentit eivät ole kovin korkeita? Heidän kohdallaan jokaisen mittauksen välinen samanmielisyysprosentti on alle 50, joten he ovat joka mittauksella arvioineet alle puolet suorituksista täsmälleen samalla tavalla.

Yleensä ottaen samanmielisyysprosenttien perusteella näyttää siltä, että yli puolet arvioijista on toiminut odotuksenmukaisesti. Toisen ryhmän muodostavat arvioijat, joiden samanmielisyysprosentti ei ole eroa eri mittauksien välillä. Näille arvioijille näyttäisi olevan tyypillistä, että he ovat arvioineet vähintään puolet suorituksista täsmälleen samalla tavalla eri mittauskerroilla. Mielenkiintoista heidän kohdallaan on se, miten lähellä määritelty asteikko on ollut heidän omaa asteikkoaan ja miten he ovat tulkinneet ja soveltaneet määriteltyä asteikkoa. Näihin kahteen profiilityyppiin kuuluu suurin osa arvioijista.

Kolmanteen tyyppiin voisi luokitella ne kolme arvioijaa, joiden samanmielisyysprosentti osuu joko ensimmäisen ja toisen tai ensimmäisen ja kolmannen mittauksen välille. Näiden arvioijien arviointikäyttäytyminen on vaikeammin selitettävissä ja arvattavissa kuin toisen tyyppin, koska se muuttuu päinvastoin toiseen päinvastaisesti odotukseen nähden. Syytä tällaiseen profiiliin voi vain arvailla, kuten jo edellä olen hiukan tehnytkin. Arvioijien samanmielisyysprofiilit ja -prosenttien suuruudet eivät ole kiinni ainoastaan arvioijasta, vaan myös asteikon sovelletavuudella ja sisällöllisellä yksiselitteisyydellä on merkityksensä, kuten on myös muilla arviointiin vaikuttavilla tekijöillä.

Kaiken kaikkiaan arvioijien sisäisen johdonmukaisuuden tarkastelu osoittaa, että arvioijien johdonmukaisuudessa on arvioijakohtaisia eroja. Vaikka eri mittauskeinot tarkastella johdonmukaisuutta perustuvat erilaisiin mittaus- ja tulkintaperiaatteisiin, on niiden osoittamissa tuloksissa myös yhteisiä piirteitä. Tällaisia olivat esimerkiksi arvioijat 080030 ja 170030, joiden kohdalla kaikki tunnusluvut osoittivat, että eri mittauskertojen välillä ei ollut eroa, ja arvioija 031234, jonka kohdalla kaikkien mittauskertojen välillä oli eroa. Arvioijan 140204 kohdalla taas kaikki kolme mittausastapaa osoittivat, että määritely asteikko yhdenmukaisti hänen arviointiaan.

Karkeasti tiivistäen arvioijat voitaisiin jakaa johdonmukaisuutensa perusteella kahteen ryhmään eli arvioijiin, jotka toimivat hypoteesin mukaisesti, ja arvioijiin, jotka eivät toimineet. Toisin sanoen noin puolella arvioijista määritely asteikko johdonmukaisti heidän arviointiaan ja noin toisella puolella määritellyllä asteikolla ei ollut vaikutusta heidän johdonmukaisuuteensa.

5.1.2 Arvioijien välinen yhdenmukaisuus eri tehtäväsarjojen arvioinnissa

Tarkastelen arvioijien välistä yhdenmukaisuutta ja arviointityyliä tehtäväsarjoittain, koska kussakin neljässä tehtäväsarjassa arvioijat arvioivat samojen suorittajien kirjoitukset samoista tehtävistä. Tehtäväsarjoja oli kaikkiaan neljä, ja kussakin niissä kirjoitukset samoista tehtävistä. Kussakin tehtäväsarjassa oli kolme tehtävää, joista yksi tehtävä oli sama kaikissa tehtäväsarjoissa (ks. taulukko 9). Tehtäväsarjojen vaikeustaso vaihteli siten, että kaikille yhteisen tehtävän lisäksi ensimmäinen tehtäväsarja sisälsi A- ja B-tason tehtävän, toinen tehtäväsarja A- ja C-tason tehtävän, kolmas tehtäväsarja kaksi B-tason tehtävää ja neljäs tehtäväsarja B- ja C-tason tehtävän. Ensimmäinen tehtäväsarja oli siis oletettavasti kaikkein helpoin ja neljäs tehtäväsarja kaikkein vaativin. Eri tehtäväsarjojen tehtävien vaikeustasoihin ja tehtävätyypppeihin voi tutustua tarkemmin luvussa 2.1.1.

Jokaisessa neljässä tehtäväsarjassa arvioijat osoittivat faktori-analyysin mukaan arvioivansa yhtä taitoa tai piirrettä, koska muuttujat latautuivat selvästi yhdelle faktorille ominaisarvokriteerin mukaan. Toisin sanoen analyysin mukaan löytyi vain yksi yli ykkösen suuruinen ominaisarvo (Eigenvalue), joka vaihteli mittauskerroittain ja tehtäväsarjoittain siten, että ensimmäisessä mittauksessa tehtäväsarjan yksi ominaisarvo oli 6,98 ja faktorin selitysosuus kokonaisvarianssista 78 %, toisessa tehtäväsarjassa 7,52 (75 %), kolmannessa tehtäväsarjassa 7,07 (71 %) ja neljännessä tehtäväsarjassa 6,58 (73 %).

Vastaavat arvot olivat toisessa mittauksessa ensimmäisessä tehtäväsarjassa 6,26 (70 %), toisessa 7,60 (76 %), kolmannessa 7,36 (74 %) ja neljännessä 6,10 (68 %). Kolmannessa mittauksessa linja pysyi samantyyppisenä eli ominaisarvo oli ensimmäisessä tehtäväsarjassa 6,40 (71 %), toisessa 7,60 (76 %), kolmannessa 7,17 (72 %) ja neljännessä 6,41 (71 %).

Taulukkoon 12 on koottu pääkomponenttia selittävien kommunaliteettien ja latausmatriisien arvot. Taulukon 12 merkinnöistä T1–T4 viittaavat tehtäväsarjojen numeroihin, K kommunaliteetteihin ja L latausmatriisien arvoihin.¹⁶ Kuten taulukosta 12 huomaa, faktoria selittävien kommunaliteettien arvot olivat melko korkeita. Muista arvioista poikkeavalta näyttää oikeastaan vain arvioija 031234, jonka kommunaliteetti oli toisen mittauksen ensimmäisessä tehtäväsarjassa .45 ja neljännessä .43. Hän ei kuitenkaan eroa muista arvioijista muissa tehtäväsarjoissa eikä yleensääkään ensimmäisessä mittauksessa, joten kyseinen ero ei houkuttele yksityiskohtaisempaan pohdiskeluun.

TAULUKKO 12 Mittauskertakohtaiset kommunaliteetit (K) ja latausmatriisit (L).

Arvi- oija	1. MITTAUS								2. MITTAUS								3. MITTAUS							
	T1		T2		T3		T4		T1		T2		T3		T4		T1		T2		T3		T4	
	K	L	K	L	K	L	K	L	K	L	K	L	K	L	K	L	K	L	K	L	K	L	K	L
011200	.74	.85	.78	.88					.75	.89	.74	.86					.71	.86	.73	.85				
020030					.67	.82							.66	.82							.53	.72		
031234	.76	.87	.70	.83	.76	.87	.72	.86	.45	.48	.65	.80	.60	.78	.43	.65	.51	.62	.71	.83	.69	.84	.64	.81
041234	.74	.87	.80	.90	.72	.87	.73	.87	.71	.81	.75	.87	.73	.87	.68	.84	.77	.86	.76	.88	.74	.88	.74	.88
060204			.79	.90			.64	.80			.78	.89			.57	.78			.75	.86			.67	.84
071030	.75	.86			.71	.86			.71	.83			.66	.82			.72	.86			.62	.80		
080030					.71	.86							.71	.86							.68	.84		
090230			.79	.90	.75	.88					.76	.88	.75	.88					.75	.87	.78	.90		
101234	.73	.85	.75	.87	.57	.74	.70	.84	.64	.81	.71	.84	.69	.84	.61	.79	.71	.84	.69	.83	.59	.78	.66	.82
111200	.75	.88	.51	.68					.71	.84	.75	.87					.75	.86	.73	.86				
121004	.75	.86					.71	.85	.71	.86					.66	.83	.66	.80					.73	.78
131004	.77	.88					.77	.89	.75	.84					.75	.89	.73	.84					.54	.87
140204			.67	.82			.62	.78			.75	.87			.61	.79			.77	.88			.66	.74
150204			.70	.82			.59	.77			.65	.80			.65	.82			.72	.84			.69	.83
161234	.75	.87	.77	.88	.64	.79	.74	.86	.78	.89	.75	.87	.71	.86	.63	.80	.71	.84	.76	.87	.67	.83	.68	.84
170030					.56	.73									.70	.85					.70	.86		
180030					.62	.79							.67	.82							.64	.81		

Faktoripistemäärien väliset Pearsonin korrelaatiot olivat eri mittausten välillä jokaisessa tehtäväsarjassa vähintään .94 ja merkitseviä tasolla 0,01, joten tältäkin osin arvioijat näyttäsivät arvioineen samaa taitoa eri kierroksilla.

¹⁶ Kommunaliiteetti kertoo muuttujan – tässä yhteydessä arvioijan – kyvystä mitata faktoria eli se on muuttujan selitysosuus faktorirakenteessa. Kommunaliiteetti on rajoitettu välillä 0–1. Latauksen suuruus kertoo siitä, kuinka paljon muuttujalla on tekemistä faktorin kanssa.

Miten yhdenmukaisesti arvioijat sitten arvioivat suorituksia? Jokaisessa neljässä tehtäväsarjassa jokaisessa kolmessa mittauksessa arvioijien väliset erot olivat pieniä, sillä osiotestikorrelaatiot vaihtelivat yhtä arvioijaa lukuun ottamatta .71:n ja .88:n välillä. Tämä muista eronnut arvioija oli sama (031234) kuin edellä, ja ero muihin oli suurin tässäkin yhteydessä toisessa mittauksessa ja ensimmäisessä tehtäväsarjassa. Koska matalinkin kerroin oli niinkin korkea kuin .46, ei tapaukseen ole syytä kiinnittää huomiota sen enempää.

Sen sijaan näissä reliabiliteetikertoimissa on mielenkiintoista se, että erot arvioijien välillä olivat pieniä myös ensimmäisessä mittauksessa, jolloin arvioijat tukeutuivat itse määrittelemiinsä tasokuvauksiin. Tämän perusteella määritelty kuvausasteikko ei lisännyt arvioijien välistä yhdenmukaisuutta, vaan he näyttivät arvioineen hyvin yhdenmukaisesti riippumatta siitä, kumpaaako asteikkoa he käyttivät.

Taulukosta 13 nähdään jokaiselta kolmelta mittauskerralta tehtäväsarjakohtaisesti laskettu kaikille arvioijille yhteinen Cronbachin alpha ja mittauskertojen välinen (interclass) korrelaatio. Taulukon merkinnät TS1–TS4 viittaavat tehtäväsarjoihin. Kuten taulukosta 13 huomataan, mittauskertojen väliset korrelaatioarvot ovat systemaattisesti pienempiä kuin alphan arvot, mikä johtuu siitä, että mittauskertojen välisiä korrelaatiota mitattaessa virhekomponenttiin on sisällytetty arvioijien välinen vaihtelu (Craemer 1998). Cronbachin alpha osoittaa sitä, miten johdonmukaisesti arvioijien antamat arviot mittaavat samaa taitoa, ja mittauskertojen välinen korrelaatio taas osoittaa arvioijien antamien arvioiden välistä yhdenmukaisuutta.

TAULUKKO 13 Kolmen mittauskerran alpha- ja mittauskertojen väliset korrelaatiot tehtäväsarjoittain.

	Cronbachin alpha				Mittauskertojen välinen			
	TS1	TS2	TS3	TS4	TS1	TS2	TS3	TS4
1. kierros	.96	.96	.95	.95	.59	.64	.53	.60
2. kierros	.94	.96	.96	.94	.52	.67	.52	.48
3. kierros	.95	.96	.95	.95	.52	.68	.59	.53

Kaikki korrelaatiot ovat merkitseviä *** $p < 0,001$.

Taulukon 13 korrelaatiokertoimet osoittavat, että erot arvioijien välisessä yhdenmukaisuudessa eri mittauskertojen välillä ovat minimaalisia. Tehtäväsarjojen 1 ja 4 arviointi näyttäisi olleen jopa yhdenmukaisempaa ensimmäisellä kierroksella kuin toisella ja kolmannella. Ero on tosin niin pieni, että sillä ei ole juurikaan merkitystä. Kaiken kaikkiaan määritelty asteikko ei taulukon 13 korrelaatioarvojen mukaan näytä yhdenmukais-taneen arviointia.

Miksi arviointi on sitten näinkin yhdenmukaista? Osoittihan arvioijien sisäisen johdonmukaisuuden tarkastelu kuitenkin, että arvioijien välillä oli yksilöllisiä eroja. Tosin ne eivät olleet suurimmalla osalla arvioijia kovinkaan suuria. Implikoiko yhdenmukaisuutta osoittavat arvot sitten sitä, että arvioijien itsensä määrittelemien asteikkojen sisällöt olivat keskenään niin samanlaisia, että se johti edellä esitetyn kaltaiseen yhdenmukaisuuteen? Vai ohjasiko arvioijien työtä jokin kaikille yhteinen konstrukti, joka oli niin vahva, että määritelty asteikko ei siihen vaikuttanut? Vai ohjasiko kyseinen konstrukti ehkä määritellyn asteikon tulkintaa ja soveltamista teksteihin näin voimakkaasti ja samansuuntaisesti?

Ehkä kyse oli siitä, että arvioijien omat käsitykset olivat keskenään samanlaisia ja myös yhdenmukaisia annetun asteikon kanssa? Vai osoittaako tämä vain sitä, että toisaalta näiden opettajien käsitykset kirjoittamisen arvioinnista olivat samanlaisia, mikä näkyy ensimmäisessä mittauksessa, ja toisaalta sitä, että he kaikki tulkitsivat ja sovelsivat määriteltyä asteikkoa samansuuntaisesti, mikä näkyy toisessa ja kolmannessa mittauksessa? Numeeriset tulokset sinänsä eivät anna vastausta näihin kysymyksiin, joten kysymyksiä täytyy tarkastella uudelleen haastatteluaineiston tarjoamien tulosten perusteella.

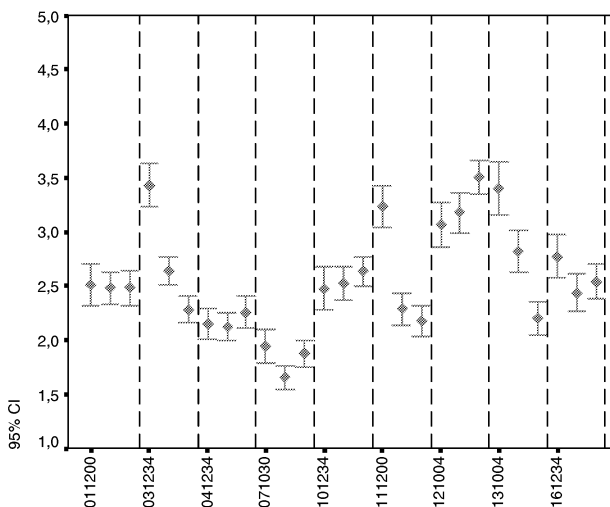
Toisaalta kysymysten esittäminen voi olla turhaakin, koska arviointi täyttää hyväksyttävälle reliabeliudelle asetetun rajan (>.80). Arviointi vaikuttaisikin riittävän yhdenmukaiselta, jos tutkimuksen koeasetelmassa ei olisi ensimmäistä mittauskertaa, joka itse asiassa herättää edellä esitetyt kysymykset ja ajaa tarkastelemaan arvioijien tulkintoja määritellystä asteikosta ja heidän omia kirjoittamiskäsityksiään.

5.1.3 Arviointilinjan tarkastelua tehtäväsarjoittain

Miltä kunkin arvioijan linja näyttää suhteessa toisiin arvioijiin? Missä määrin esimerkiksi arvioijan lempeys, ankaruus tai asteikon keskivälin käyttö nousee esiin, vaikka arvioijat näyttävät arvioineenkin suoritukset melko yhdenmukaisesti? Tässä yhteydessä arviointilinjaa voi tarkastella vain tehtäväsarjoittain, koska arvioijat arvioivat samojen kirjoittajien vastauksia samoihin tehtäviin nimenomaan tehtäväsarjoittain. Kuvioissa 19–22 on nähtävissä tehtäväsarjoittain kunkin arvioijan antamien arvioiden keskiarvo kultakin kolmelta mittaukselta. Yhden pystysuuntaisen katkoviivan sisällä on aina kolme keskiarvoarviota, joista ensimmäinen viittaa ensimmäiseen mittaukseen, toinen toiseen ja kolmas kolmanteen. N kuvioiden 19–22 otsikoissa viittaa arvioitujen suoritusten määrän.

Arvioijalinjaa tarkasteltaessa on tarkoituksenmukaisinta seurata toista ja kolmatta mittausta, joissa arvioijat käyttivät samaa asteikkoa. Ensimmäisen mittauksen osalta arvioijien keskiarvojen toisiinsa vertaaminen antaa lähinnä yleisluonteisia viitteitä siitä, missä määrin arvioijien itsensä määrittelemät asteikot erosivat toisistaan tiukkuudessaan. Toisen ja kolmannen mittauksen osalta voi taas tehdä hyvin varovaisia päätelmiä siitä, miten eri tavalla tai samalla tavalla arvioijat tulkitsivat asteikkoa sekä millaisia eroja arvioijien arviointilinjassa oli.

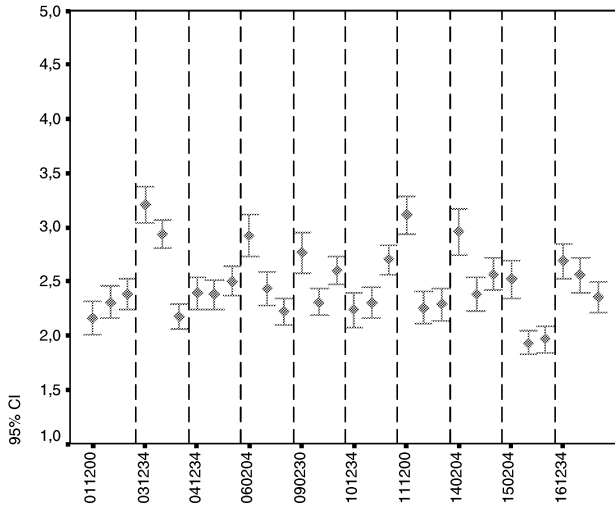
Kuvion 19 mukaan ensimmäisen tehtäväsarjan keskiarvot kasaantuvat pääasiassa tasoarvion kaksi ja hieman alle kolmen tietämille. Tästä linjasta arvioijat poikkeavat sekä tiukempaan että lempeämpään suuntaan. Eroa aiheuttaa hieman myös se, mitä tarkasteltiin jo luvussa 5.1.1, että joidenkin arvioijien sisäinen johdonmukaisuus vaihtelee toisia enemmän, esimerkiksi arvioijien 031234 ja 130204. Näiden molempien arvioijien arviointityyli lähenee kuitenkin joka mittauksesta muita arvioijia ja on kolmannessa mittauksessa melko samansuuntainen yleisen linjan kanssa.



KUVIO 19 Tehtäväsarja 1 (N=146).

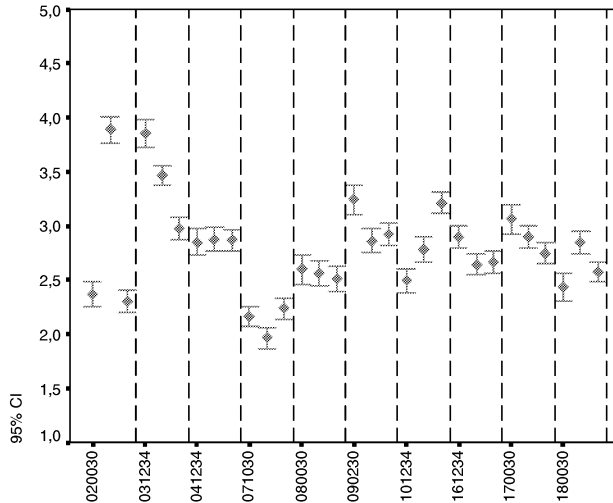
Jos katsotaan varsinkin toista ja kolmatta mittausta, ensimmäisen tehtäväsarjan arvioijista muita arvioijia tiukempaa linjaa edustaa 071030 ja muita löyhempää taas 121004, jonka arviointilinja näyttäisi höllentyvän

kierros kierrokselta. Nämä molemmat arvioijat poikkeavat systemaattisimmin yleisestä tendenssistä jokaisessa kolmessa mittauksessa, mikä viittaa siihen, että annettu asteikko ei näyttäisi juuri vaikuttaneen heidän omaan arviointityyliinsä. Keskiarvojen mukaan se näyttäisi pysyneen muita arvioijia tiukempänä tai lempeämpänä asteikosta riippumatta.



KUVIO 20 Tehtäväsarja 2 (N=162).

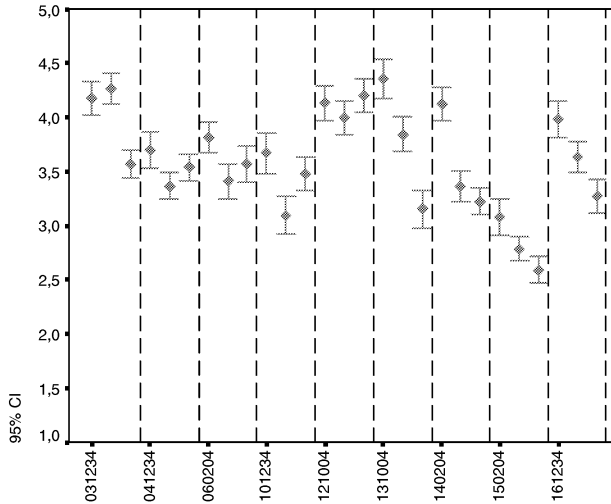
Toisen tehtäväsarjan keskiarvot kasaantuvat ensimmäistä tehtäväsarjaa tiiviimmin kahden ja hieman alle kolmen paikalle, varsinkin toisessa ja kolmannessa mittauksessa lukuun ottamatta arvioijan 031234 toista mittausta (ks. kuvio 20). Toisessa tehtäväsarjassa eri arvioijien linjat näyttävät siis varsin yhdenmukaisilta toisella ja kolmannella mittauksella. Samantyyppinen tendenssi näkyy myös taulukossa 13, jonka mukaan arvioijat olivat kaikkein yhdenmukaisimpia toisen tehtäväsarjan arvioinnissa, vaikka erot olivatkin pieniä muihin tehtäväsarjoihin verrattuna.



KUVIO 21 Tehtäväsarja 3 (N=284).

Kuten kuviosta 21 näkyy, kolmannen tehtäväsarjan keskimääräiset arviot näyttävät kasaantuvan lähinnä kahden puolen ja kolmen välille. Toisessa mittauksessa arvioijat 020030 ja 031234 näyttäisivät olleen muita arvioijia löyhempiä. Sisäisen johdonmukaisuuden vaihtelua havainnollistaa hyvin se, että toisessa mittauksessa ero lempeimmän (020030) ja tiukimman arvioijan (071030) välillä näyttäisi olleen keskimäärin jopa kaksi taitotasoa mutta sitten taas kolmannessa mittauksessa heidän välillään ei ole lainkaan eroa.

Arvioijan 071030 tiukka linja on systemaattinen ja sisäisesti johdonmukaisen oloinen sekä jokaisessa kolmessa mittauksessa että molemmissa hänen arvioimissaan tehtäväsarjoissa 1 ja 3. Muuten tämän tehtäväsarjan arvioijien keskimääräiset tasoarviot toisessa ja kolmannessa mittauksessa liikkuvat kahden puolen ja kolmen välillä eli erot näiden arvioijien välillä näyttävät olevan pieniä.



KUVIO 22 Tehtäväsarja 4 (N=147).

Kuviosta 22 huomataan, että neljännen tehtäväsarjan keskimääräiset arviot ovat korkeampia kuin muiden tehtäväsarjojen; ne sijoittuvat pääasiassa tasojen kolme ja neljä väliin. Tässä tehtäväsarjassa tehtävät koostuivat kuvatehtävän lisäksi B-tason kirjeestä ja C-tason työpaikkahakemuksesta, joten tehtäväsarjan vaativamman vaikeustason lisäksi sen suorittajatkin olivat kirjoittajajoukon edistyneimpiä oppijoita siinä määrin, kun tämän tehtäväsarjan tehtäviä onnistuttiin kirjoittamaan edistyneemmillä oppijoilla. Muita tehtäväsarjoja korkeampiin tasoarvioihin voi vaikuttaa sekin, että näissä tehtävissä oli mahdollisuus osoittaa edistynyttä kielitaitoa paremmin kuin esimerkiksi ensimmäisessä tehtäväsarjassa.

Ensimmäisen tehtäväsarjan tavoin tässäkin erottuu muita lempeämpänä arvioijana 121004, jonka keskiarvot ylittävät taitotason neljä. Muita tiukempi arvioija on taas 150204, jonka keskiarvoarvo jää alle kolmen toisessa ja kolmannessa mittauksessa. Arvioijan 150204 linja oli myös toisessa tehtäväsarjassa tiukan puoleinen, joskaan hänen linjansa ei eronnut näin selvästi muista arvioijista. Muilla arvioijilla (lähinnä 031234, 041234, 060204, 101234, 140204 ja 161234) keskiarvot näyttävät olevan yhdenmukaisempia kolmannessa kuin toisessa mittauksessa.

Kaiken kaikkiaan edellä esitettyjen kuvioiden perusteella arvioijien (N=17) vaihtelu keskiarvojen perusteella asettuu suunnilleen yhteen tasoon. Arvioijakohtainen tyyli ei näytä aiheuttavan isoja eroja, mikä

näky myös yhdenmukaisuutta osoittavissa korrelaatiokertoimissa. Muita hiukan tiukempaa linjaa keskiarvojen perusteella edustavat kuitenkin arvioijat 071030 ja 150204 ja lempeämpää taas vastaavasti 121004. Muut poikkeukset ovat lähinnä mittauskohtaisia siten, että arvioija on ollut ensimmäisellä tai toisella mittauskerralla muita lempeämpi mutta ei enää kolmannella, kuten arvioijat 020030, 031234, 101234 ja 131234. Näissä tapauksissa on siis kyse ehkä enemmänkin eri syistä johtuvan arvioijan sisäisen johdonmukaisuuden vaihtelusta, joka näkyy myös kuvioissa 15–18, kuin systemaattisesti lempeämmästä arviointityylistä.

Tarkastelen seuraavaksi lyhyesti, miten eri tehtäväsarjat eroavat toisistaan annettujen tasoarvioiden perusteella ja millaisia arvioija-kohtaiset erot olivat kaikki tehtäväsarjat arvioineiden arvioijien välillä. On oletettavissa, että eri tehtäväsarjojen arvioijakohtaiset keskiarvot poikkeavat toisistaan niiden tehtävien vaikeustason mukaan ja tietysti myös tehtäviä suorittaneiden kirjoittajien taitotason mukaan. Yleisestihän näyttää siltä, että kolmessa ensimmäisessä tehtäväsarjassa keskiarvot kasaantuvat kahden ja kolmen väliin ja neljännessä tehtäväsarjassa kolmen ja neljän väliin kuitenkin niin, että toisessa tehtäväsarjassa suurin osa keskiarvoista on ryppääntynyt kaikkein tiiveimmin 2 ja 2½:n väliin.

Tendenssi on samansuuntainen kaikissa kolmessa mittauksessa, koska muutamaa poikkeusta lukuun ottamatta keskiarvot ovat samansuuntaisia tai vähän korkeampia myös ensimmäisessä mittauksessa. Toki tasoarvioissa on myös hajontaa: arvioijien antamien tasoarvioiden vaihteluväli on ensimmäisessä mittauksessa pääosin 1–5 tai 1–6 ja myös muissa mittauksissa, varsinkin neljännessä tehtäväsarjassa. Edellä esiin tullut tasoarvioiden keskivälisyys näkyi myös tasoarvioiden vaihteluvälin kiristymisessä varsinkin toisen mittauksen toisessa tehtäväsarjassa ja kolmannen mittauksen ensimmäisessä ja toisessa tehtäväsarjassa, joissa 3 arvioijan tasoarvioiden vaihteluväli on 1–4.

Myös tasoarvioiden frekvenssit osoittavat mittauskohtaista vaihtelua (ks. liite 9). Taitotasojen 6 määrä vaihteli ensimmäisessä mittauksessa arvioijittain 0–42:n välillä, mihin tietysti vaikutti sekin, mitä tehtäväsarjaa ja kuinka monta suoritusta arvioija arvioi. Taitotasojen 5 määrä vaihteli ensimmäisessä mittauksessa kahdesta yli kaikki suoritukset arvioineen arvioijan 156 tasoarvioon. Sen sijaan toisessa ja kolmannessa mittauksessa taitotasojen 5–6 määrä väheni, varsinkin kolmannessa mittauksessa: jos taitotasoa 6 oli annettu, niitä oli vähimmillään 1 ja enimmillään 16.

Tasoarvioiden frekvenssien mukaan varsinkin toisessa ja kolmannessa mittauksessa näytettäisiin annetun eniten taitotasoa 2–3 eli noin 60–80 prosenttia kaikista tasoarvioista, lukuun ottamatta muutamaa painotuspoikkeusta taitotason 1 tai 4 suuntaan.

Koska keskiarvot seuraavat tällä tavalla samansuuntaista tendenssiä, ne ehkä heijastelevat enemmänkin kirjoittajien 'todellista' keskimääräistä tasoa tai arvioijien samantyyppistä kirjoittamiskäsitystä kuin arvioijasta johtuvaa oletettua mittausvirhettä, lähinnä arvioijan taipumusta käyttää asteikon keskiväliä tai taitotasoa 2–3. Toisaalta keskiarvojen kasaantumisen tasolle 2–3 on odotuksenmukaista, kun otetaan huomioon arvioitavien suoritusten jokseenkin suuri määrä ja se, että tasoarvioita tarkastellaan keskiarvoina.

Keskiarvoihin vaikuttaa tietysti myös se, mitä tehtäväsarjaa arvioija oli arvioinut. Esimerkiksi arvioijien 060204 ja 140204 antamien arvioiden keskiarvo on lähes yhtä tasoa parempi neljännessä kuin toisessa tehtäväsarjassa jokaisessa mittauksessa, kun taas ensimmäistä ja toista tai ensimmäistä ja kolmatta tehtäväsarjaa arvioineiden arvioijien 011200, 071030 ja 111200 keskiarvot ovat yhtenäisesti 2–2½:n välillä erityisesti toisessa ja kolmannessa mittauksessa. Tämä selittyy sillä, että neljännen tehtäväsarjan tehtävät olivat tasoltaan vaativampia ja niiden suorittajat myös keskimäärin hiukan edistyneempiä kuin muiden tehtäväsarjojen kirjoittajat.

17 arvioijasta kaikkiaan 4 arvioi kaikki tehtäväsarjat. Näistä arvioijista 031234 on ollut jokaisessa tehtäväsarjassa kaikkein tiukin kolmannessa mittauksessa ja antanut systemaattisesti keskimääräisesti alempia arvioita jokaisessa tehtäväsarjassa. Samantyyppisesti erisuuntainen tendenssi, mutta ei yhtä voimakkaasti, näkyy 101234:llä, joka on ollut lempeämpi kolmannessa mittauksessa. Hänen keskiarvonsa ovat kaikissa tehtäväsarjoissa 2:n ja 3 ½:n välissä, kuten on myös 041234:lläkin ja osin 161234:llä.

Kaiken kaikkiaan yli seitsemän sadan suorituksen arviointi on ollut pääasiassa asteikon puolen välin ja sen alapuolella olevien tasoarvioiden käyttämistä. Yleinen linja antaa kuitenkin olettaa, että suorittajat ovat – ainakin tässä käytettävillä asteikoilla mitattuna – suurimmaksi osaksi myös olleet tämän tason kirjoittajia (ks. myös liite 9). Tätä tukee omalta osaltaan myös kirjoittajien itsearviointit, joiden mukaan 60 % kirjoittajista arvioi kirjoittamisen taitonsa tasolle 2–3. Itsearviointinissa käytetty asteikko oli kielellistä mukauttamista lukuun ottamatta sama, mitä arvioijat käyttivät toisessa ja kolmannessa mittauksessa. Vaikka kyse ei olekaan suorista vastaavuuksista, itsearviot antavat kuitenkin epäsuoraa tukea arvioijien antamille keskiarvoille.

5.2 Teemahaastattelujen kertomaa

5.2.1 Ensimmäinen haastattelukierros: omaan taitokäsitykseen nojaaminen

5.2.1.1 Yleisvaikutelmasta liikkeelle

”No se että oli hirvittävän mukavaa. Alussa oli raskasta yhtä paperia kauan pikku hiljaa sitä niinku pääsi vauhtiin.” Tarja

Yleisvaikutelmasta keskusteltiin heti haastattelun aluksi, ja tämän keskustelunaloituksen tarkoitus oli toimia lämmittelyvaiheena koko haastattelulle. Yleisvaikutelmasta puhuminen johdattelikin aika kätevästi itse prosessista puhumiseen, koska haastateltavat saivat heti haastattelun aluksi purkaa päällimmäisen olotilansa, liittyi se sitten tuntemuksiin omasta varmuudesta, työmäärään tai annettuihin taitotasoihin. Yleisvaikutelmasta puhumiseen käytettiin vain vähän aikaa, eikä se tuntunut enempiä aikaa vaativankaan.

Ensimmäiset ja spontaanit yleisvaikutelmat jakaantuivat yleisesti ottaen siten, että 6 arvioijaa koki työn enemmän negatiivisena, 4 positiivisena ja 6 sekä negatiivisena että positiivisena. Arvioijien molemmansuuntaiset kommentit liittyivät lähes kaikissa tapauksissa itse työhön joko ajallisesta näkökulmasta ja/tai sitten varsinaiseen työmäärään ja sen laatuun. Monet kommentoivat myös kriteerien puuttumisesta ja siitä johtuvaa epävarmuutta. Myös itse tekstit saivat kommentteja siitä, millaisia tuntemuksia ne herättivät ja miten ne vaikuttivat lukutilanteeseen.

”mmm no sinänsä aa siinä mielessä palkitsevaa että aina hirveen opettavaista ja tota sellasta niinku aika hyvä kattoo joskus semmone isompi määrä kerralla kus siinä näkee jotenki sitä semmosta vaihteluväliä tai sillä tavalla mutta toisaalta niinku siinä mielessä aika tylsii piirteitä aika tylsää alko olla niinku ties tasan mitä kaikki venäläiset kirjottaa ja mitä kaikki somalit kirjottaa et niinku niillä on niin samat argumentitki et se oli ihan uskomatonta ja tota ja sitten mikä mun ja niinku tavallaan et se oli semmosta ylä ala ylämäki alamäki välillä tuntu et noh nythän tää pelittää joo kyl tää niinku ihan hyvin menee välillä tuntu et paska en jaks enää ja tota et et että mä en saa tähän niinku mitään johdonmukaisuutta ja sehän siinä tietysti on et se epäilemättä se...” Heidi

Heidin kommentti työprosessin luonteesta kuvaa hyvin työhön liittyviä erilaisia tunnetiloja ja sitä, miten molemmat jaksamisen ja innostumisen sekä väsymisen ja kyllästymisen ääripäät ovat työssä läsnä. Vastaavanlaiset tunnetilat ovat löydettävissä lähes kaikkien arvioijien haastatteluista, mutta eivät välttämättä yhtä suorasti ilmaistuna. Myös ne arvioijat, jotka kokivat työn enemmän negatiivisena esimerkiksi tiukan

aikataulun vuoksi, saattoivat pitää itse tekstien lukemista mielenkiintoisena. Myös arvioija, joka suhteutti työmäärän rahaan, piti sinällään hyvänä ideana, että arvostelijat lukisivatkin suuremman määrän tekstejä kerrallaan, jolloin hänen mielestään on helpompi olla johdonmukaisempi.

Yleisvaikutelmasta puhuttaessa haastateltava saattoi myös tuntea tarvetta purkaa väsymyksen tunteensa heti aluksi, koska myöhemmin haastattelun edetessä hänen kommenttinsa sävy saattoi muuttua positiivisemmaksi tai hän saattoi innostua työprosessin jonkin vaiheen tai tietyn tekstityypin yksityiskohtaisemmasta kuvaamisesta.

”tylsää, onha se aika tylsää nii (mikä on tylsää miksi) mm se on toi toisaalta aika helpoo sillä tavalla että että niiku siinä ei tuu mitään yy uutta ja yllättävää et se on sitä samaa että niinku tuntee että että tietää taikka luulee tietävänsä niinku tän ja tän oppijatyyppin ja nää ja nää virheet jaaha tässä on nyt taas tää tapaus ja tuo tapaus et se niinku se ei oikeestaa tuo mitään uutta et sitte ja sit se voi niinku ninku havaita senkin et millä tavalla tai niinku kyllästyy siihen et ihmiset niinku näkee ne tehtävät tietyllä tavalla et ei siitä nii” Päivikki

Kuten edellä käy ilmi, yleisvaikutelmat heijastelevat tuntemuksia ensimmäisestä arviointikierroksesta yleisellä tasolla: vaikutelmat ovat kuitenkin yksilöllisiä havaintoja erilaisuutensa vuoksi. Ne eroavat jonkin verran toisistaan sen mukaan, mihin kukin arvioija kiinnitti huomiota ensimmäisenä tai mikä hänellä oli mielessään päällimmäisenä. Sen sijaan yhteistä yleisvaikutelmille on se, että ne yksistään antaisivat liian yksioikoisen kuvan arvioijien kommentteista, heidän tuntemuksistaan ja käsityksistään, koska haastattelun edetessä arvioijat saattoivat syventää ja perustella monipolisemmin sanomaansa sekä esittää päinvastaisia ja keskenään ehkä ristiriitaisiakin kommentteja.

5.2.1.2 Arvioijien luomat asteikot

”Mä ajattelin et se on sit mun nii tää on just sitä että että oonks mä nyt vaan tämmönen joka päättää että asiat on näin vai onko se oikean suomalaisen mielestä sitten ihan toisin et se oli tavallaan just se se asia ett kuinka luotettavana mä voin niitä omia kriteereitani pitää.”

Ensimmäisen arviointikierroksen haastavin osuus oli se, että arvioijan täytyi luoda annetulle 6-portaiselle asteikolle sisältö, jotta kirjoitusten arviointi olisi yleensäkin mahdollista. Arvioijan saamassa ohjeessa häntä pyydettiin arvioimaan suoritukset asteikolla 1–6, jolla taso 1 on heikoin taso, 6 vahvin ja tasot 2–5 näiden väliltä. Hänelle myös kerrottiin, että tässä vaiheessa hänellä ei ole käytettävissä valmiita kriteereitä. Arvioijan

täytyi siis itse tehdä päätös siitä, millainen suoritus esimerkiksi edustaisi taitotasoa 3 ja millaisen suorituksen hän saattoi sijoittaa taitotasolle 6.

Suurin osa arvioijista aloitti tekstien arvioimisen siten, että he selailivat ensin tekstejä tai lukivat tarkemmin joitakin tekstejä pinon alusta ja sitten tekstejä vertaamalla miettivät tasojen sisältöjä. Tällä tavoin syntynyt asteikko muotoutui ”pikku hiljaa”, kuten arvioijat itse totesivat. Arviointiprosessin aikana asteikko myös tarkentui ja osin muuttui varsinkin keskitasojen kohdalla, koska mitä enemmän arvioija luki tekstejä, sitä useamman tasoisia tekstejä hän kohtasi.

”vaan alotin ihan siis vaan alotin ku mie katoin ne ohjeet et täs kyll oo mitään muuta ohjetta ku vaan että lue ja ja aloin lukea mie luin ensimmäiset pariin kertaan läpi siis mie niinku aluks vähän vertailin et ahaa et tää on minust kolmone mikäs muu siellä oli kolmone toi onks nää niinku samaa” Tiina

Kolme arvostelijaa suunnitteli asteikkonsa ennen lukemisen aloittamista ja kirjasi myös suunnitelmansa ainakin jonkinlaisina muistiinpanoina. Teija laati kriteerit tasoille 3 ja 6 ja kirjoitti ne paperille. Tekstit, jotka eivät täyttäneet näitä kriteerejä, sijoittuivat välitasoille siten, että esimerkiksi kolmostasoa parempi teksti sai taitotason neljä ja kuutostasoa huonompi teksti sijoittui taitotasolle viisi. Myös Tarja jakoi asteikon kahtia siten, että tasot 1, 2 ja 3 muodostivat toisen puolikkaan ja tasot 4, 5 ja 6 toisen puolikkaan. Hän siis määritteli aluksi vain nämä puolikkaat ja mietti sitten tekstejä lukiessaan, miten ne sijoittuivat tarkemmin puolikkaan sisäisille tasoille. Pirkko taas kirjoitti tehtävätyyppikohtaisen listan asioista, joihin hän kiinnitti huomiota.

”mä määrittelin vaan kolmosen ja kutosen ja mä panin siihen kol kolmoseksi että tota et se on niinku ymmärrettävää ja hallittua mutta ei millään lailla niinkun ansiokasta ja sitten pistin siihen kuutoseen että se on idiomaattista ja rikasta sit myöhemmin ajattelin et tähän pitäis vielä lisätä että sujuvaa mut että näitä mä niinku käytin ja sitten sinne välliin aattelin et pitää vaan jollain lailla sijoitella ne” Teija

”noh tietysti alussa piti hakee se että ja tehä vähän muistiinpanoja et mitä asioita minä katon ihan joka kysymyksen kohalla mä sitten tein semmosen listan että näitä minä tarkkailen ja ja sit se rupesi löytymään se se linja” Pirkko


Tämäntyyppistä etukäteissuunnittelua oli jossakin määrin lähes kaikilla arvostelijoilla. Aloittaessaan selailua arvioijat lähtivät siitä, että ykköstaso on lähellä nollatasoa ja kuutostasoa on syntyperäisen kielenkäyttäjän kaltainen taso. Kaikilla arvostelijoilla oli kokemusta arvioinnista, joten oli selvää, että asteikon tasojen määrä ohjasi ainakin

jonkin verran miettimään sitä, miten moneen osaan taito piti jakaa. Joidenkin arvioijien 6-portainen asteikko jakaantui lähinnä puoliksi, joidenkin kolmeen osaan siten, että tasot 1 ja 2 edustivat läheistä taitotasoa toisiinsa nähden, samoin tasot 3 ja 4 sekä 5 ja 6. Yhteistä oli myös se, että asteikko sai varsinaisen sisältönsä tai se tuntui muuttuvan konkreettiseksi ja sovellettavaksi vasta tekstejä lukiessa. Tässä mielessä arvioijien raportoiman mukaan asteikot saivat konkreettisen sisältönsä pääasiassa tekstilähtöisesti eli tekstejä selaamalla ja niitä toisiinsa vertaamalla.

Millaisia sitten asteikon sisällöt olivat? Pyysin haastattelussa arvioijia kuvaamaan taso tasolta sekä luomaansa että käyttämäänsä asteikkoa. Toisin sanoen kysyin heiltä, millainen suoritus heidän mielestään oli tyypillisesti esimerkiksi taitotason 2 suoritus tai millaiselle suoritukselle he antoivat taitotason 4. Koska suurimmalla osalla arvioijista asteikko oli vain heidän omassa päässään, on selvää, että asteikon yksityiskohtaisuus ei välity samalla tavalla, kuin jos he olisivat kirjoittaneet sen paperille. Toisaalta paperille kirjoittaminenkaan ei vielä osoita sitä, miten asteikkoa tulkitaan ja sovelletaan. Olen koonnut kunkin arvioijan asteikot taulukoihin 14–18. Niiden sisältö perustuu siis siihen, miten arvioija muisteli mielessään olevaa asteikkoa tai miten hän sitä ehkä osin muotoili uudelleen haastattelun aikana.

Olen siis koonnut asteikot sen mukaan, mihin arvioija sanoi kiinnittäneensä huomiota kullakin tasolla. Olen pyrkinyt siirtämään asteikkoon arvioijan sanoman mahdollisimman alkuperäisessä muodossa, vaikkakin puheelle tyypilliset esimerkiksi epäröinnit ja toistot on jätetty pois. Joissakin kohdin jokin tasoa koskeva kriteeri on poimittu keskeltä muuta pitempää puhumisjaksoa, mikä saattaa vääristää kriteeriä, koska se on erotettu irralleen sen alkuperäisestä ja laajemmasta kontekstista. Suluissa olevat huomautukset ovat lisäämiäni selvennyksiä.

TAULUKKO 14 Arvioijan raportoimat tasokuvaukset.

TASO	SANNA	TUULA	RAIJA
1	helppo arvioida	jotain raapustettu mut siitä ei saa mitään tolkkua, suomen kielen sanoja mutta niissä ei ollu välttämättä päätteitä	täyty olla kirjoitettuna älyllistä, joku ajatus piti olla ykkösessäki
2	sellaista lapsen tekstiä ni et siinä on pelkkää preesenssiä ja suhteellisen oikein mut ei kuitenkaan mitään rikkaampaa rakennetta	jo jotakin tolkkua mut tosi semmosta hataraa ja heikkoa, ei mitään rakenteita, jo lauseen näköisiä tekeleitä	aika heikko semmone tyydyttävä, jo muutamia sanoja oikein kirjoitettuna ja jotakin jonkin verran joku ees taivutusmuoto
3	osas jonkin verran rakenteita mut ei kuitenkaan hallinnu niitä hyvin, fraaseja ja muotoja yritetty enemmän vaikkakin virheitä, perusteksti luonteeltaan, näki et se oli ollu Suomessa jo jonkin aikaa	sai olla aika paljon virheitä vielä mut kuitenkin jo järkevää	keskiverto, lisää sanoja, taivutuksia ja tavallaan jo lauseita ja jonkin näköne semmone koheesio siinä ihan pientä koheesiota tekstissä
4	riikkaampaa kieltä ja enemmän oikein, aika selkeä	saatto olla hirveen hyviä, hyvin käyettyä sanastoa mutta saatto olla ne muodot aika heikkoja tai toisin päin niin että virheetöntä ja rakenteet hyviä mut siit ei jotenki muodostonu jollain tapaa semmosta järkevää juttua	hyvä, piti lisääntyä sen koheesio ja sanojen ja kieliopin ja siinä piti olla jo jonkin verran pituutta
5	aika selkeä	tosi hyvä mut kuitenkin joitain juttuja mitkä oli vähän menny pieleen	sellasia virheitä joita suomalainen ei enää tekis, nii monta kuitenkin, kuin suomee mut kuitenkin joku astevaihteluvirhe tai rektiovirheitä ja sellasia muuten hirveen hyvä, niin kun melkein täydellinen, tavallaan kuin äidinieltä mutta siinä oli selkeitä sanastovirheitä, kielioppivirheitä, pitää olla tyyliä, oikein voimakkaasti puhekielen omasia sanoja ei sais olla
6	lähellä suomenkielisen tuotosta	tosi hyvä, lähes täysin kaikki oikein, ei ollu mitään (vrt. edelliset puutteet)	kirjottaa niin kuin suomalainen kirjoittaa, tavallaan sellane äidinkielen taso, täydellinen
	 <p>Kriteerit</p> <ul style="list-style-type: none"> • rakenteet • sanasto • ymmärrettävyys/luettavaus • idiomaattisuus 	<ul style="list-style-type: none"> • ymmärrettävyys • sanasto • rakenteet • sisältö 	<ul style="list-style-type: none"> • sisältö • pituus • sanasto • oikeinkirjoitus • rakenteet • koheesio • idiomaattisuus • tyyli

TAULUKKO 15 Arvioijan raportoimat tasokuvaukset.

TASO	HEIDI	KAISU	TEIJA
1	ei oikein pystynyt arvioimaan-kaan tai siitä ei saanu tolkkua	ykslauseiset on ykkösiä tai jotain kuhan on, toistetaan aika paljon sitä mitä siinä on ja kerrotaan että <i>sairaam lääkäri</i> ja näitä juttuja	täytyy olla jollakin tavalla puheena-aihe ja hahmotettavissa, mut voi olla semmostakin mikä jää täysin epäselväks
2	sai jo kohtalaisen tolkun ja sitä oli jo pikkusen niinku enemmän (tekstiiä)	ois sitä asiaa kerrottu enemmän ja vähän tarinanomaisemmin (kuvateht.) mut ei vielä esimes kaikki kieliooppimuodot ihan täysin hallussa, mä saan selvän mitä se tarkoittaa, preesensissä menneestä ajasta	vähän vähemmän epäselväks jäävää (vrt. taso 1), sai olla semmosta et jotakin semmosta sivuseikkaa ei tajua välttämättä
3	piti olla jo selkeesti tekstiä, sai olla virheitä ja runsaasti, saatto olla semmosta pientä vaikeaselkisuutta, ei luettelaviivoi kolmoseen, sit jo semmosia kirjeen tunnusmerkkejä vaikei oiskaan tarkasti tilanteeseen sopivia alotuksia tai muita	aika selkeesti koko se tehtävä hanskassa, kaikki niinku ne asiat sieltä, loogisesti etenevä kuitenkin sallitaan muutama lapsus	ei oo mitään erikoista, siin on virheitä mut sen kuitenkin ymmärtää, ymmärrettävää ja hallittua mutta ei millään lailla ansiokasta, ei mitään kovin osuvia suomenkielisiä sanontoja tai sanoja, sisältö tavanomainen
4	ero kolmoseen virheitte mää- rässä, tuottamisen helppou- dessa ja semmosessa tekstin koheesiossa, toisen tilanteeseen astettumista niinku tunteen vä- littämisen (suruvalitteluviesti)	tosii hyvää	joitakin hyviä osuvia sanoja mitkä osottaa et seuraa suomenkielistä keskustelua mut niitä ei käytetä mitenkään johdonmukaisesti, yritystä johonkin semmosiin rakenteisiin joita mä piijn vaikeina tai joita vasta edistyneemmät osaa
5	kohtuullisen virheetön, teksti- nä aika hyvä siis tämmösenä koherenttina tekstikokonaisuus- tena, idiomaattisuus sidosta- misessa, sanastollisissa ja käsit- teellisissä valinnoissa	jo hirveen hyvä, sä osaat jo tosi hyvin sitä kieltä, siihen tehtävän- antoon ja tuotukseen nähden kaikki oli kohdallaan, edellisen lisäksi kielio- pillinen kriteeri, tuotos oli niinku sujuvaa ja toimivaa ja ym- rrettävää ja hyvää suomee ja vaihte- levia lauserakenteita ja sanotaan vaikka sivulauseita paljon ja ehkä jo joitain upotuksiaki	esimerkiks jotaki idiomeja, oli yritystä mut ne ei välttämättä ihan onnistunu, määrittelin vitosen ku- tosen avulla
6	idiomaattisuus sidostamisessa, sanastollisissa ja käsitteellisissä valinnoissa, konventionaalinen suomalaistyypinen reagointitapa	asettu kauheen korkeelle, voi olla muutama virhe tai ihan kielio- ppi- virhe mut ois sanomaltaan hirveen hyvin ymmärrettävä ja sitten kau- heen luonteva kauheen suomalaisen olonen, kielio- pillinen kriteeri, ne rakenteet ylöspäin tultaessa	idiomaattista, rikasta ja sujuvaa, virheitä sai olla

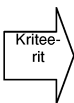
Kriteerit

- ymmärrettävyys
- pituus
- tekstimäisyys/koheesio
- konventioiden hallinta
- rakenteet (virheettömyys)
- sujuvuus
- sanasto

- pituus
- sisältö
- tehtävänannon täytyminen
- rakenteet
- ymmärrettävyys
- tekstin rakenne
- sujuvuus
- idiomaattisuus

- ymmärrettävyys
- sisältö
- sanasto
- rakenteet
- sujuvuus
- idiomaattisuus

TAULUKKO 16 Arvioijan raportoimat tasokuvaukset.

TASO	INKERI	TARJA	PIRJO	PIA
1	ei oikein saanu tolkun tolkkua, ne oli käsittämättömiä jollakin tapaa tai jos niistä käsitti jotain niissä oli esimes taitutus täysin pois tai taivutettu erittäin vähän	keskenjäänyt, ei ymmärtänyt tehtävää, aika heikko, ei ollut paljon mitään	preesens, perusmuodossa jotain sanoja, viesti ei välity, ykkönen voi olla myös pelkkä kaunis lause hän menee kauppaan, missä-mistä-mihin-muo-dot	kuhan on pantu sanoja peräkkäin, jonkinnäköne järki mut melken nollataso
2	jos muuten kolmonen mut tehtävänanto ei täyttynyt	oli jotain mut sit se oli tosi niukasti, sanastoo ei ollut oikeastaan	imperfekti, jos oli yritystäkin sinne päin, joitakin rakenteita oikein, esim. täytyy-lause	meni sekasin ykkösen kanssa
3		rupes olemaan sanastoo jonkun verran, lyhyttä tyyliä	perfekti	kolmonen kultainen keskite että ne perusasiat on suunnilleen, osaa käyttää esimerkiksi verbimuotoja mut aikamuodot voi mennä sekasin tai meneekin sekasin niissä on virheitä
4		tyyli aika rönsyilevää, tosi paljon niinku virheitä, mutta siinä oli jo tuotettu	passiivi	aika lähellä kutosta, näyttää aika täydelliseltä mut pieniä vipauksii, luku suomea kolmisen tai neljä vuotta
5	(viestissä) niillä on joku alkupuhuttelu ja mahdollisesti jotain muuta siinä ja sit vielä melko oikein tai on vain joitaki kielivirheitä	samantyyppinen kuin kutonen	lauseenvastikkeita jonkun verran	aika lähellä kutosta, plus miinus niinku tämmösiä näitä virheitä mukana
6	ois vastannu tehtävänantoa, melko lähellä natiivin tasoa	aika hyvä, onko tämä edes nyt kielenoppijan kirjottamaa	tosi hyvä, voi olla vielä joitakin virheitä, esim. objektivirheitä	vastaa niin sanottua natiivin tasoa, semmonen täydellinen suoritus, jos semmosta voi olla
	 <ul style="list-style-type: none"> • ymmärrettävyys • rakenteet • tehtävänannon täytyminen • konventioiden hallinta • idiomaattisuus 	<ul style="list-style-type: none"> • tehtävänannon täytyminen • pituus • sisältö • sanasto • rakenteet • idiomaattisuus 	<ul style="list-style-type: none"> • rakenteet • ymmärrettävyys 	<ul style="list-style-type: none"> • ymmärrettävyys • rakenteet • idiomaattisuus

TAULUKKO 17 Arvioijan raportoidut tasokuvaukset.

TASO	TIINA	MERJA	MARTTI
1	surkea ja huono, kirjottaa tekstin jonka kirjoittanut jo kymmeniä kertaa aikaisemminkin, oli saanut paperille vähän jotakin	kolme neljä sanaa ihan tämmöne vajaa kymmenen sanaa, vaikka olisi aika virheetöntä	ei ollut mitään tolkkua, ei ollut lauseita oikein kunnolla ja sanoja peräkkäin, esimerkiksi <i>tytön ja huone</i> (kuvatehtävä) ja sit siin olis ollut vähän jotain muutaki
2	jotakin on jotakin siltä (ykkösen ja kolmosen) väliltä, kirjottaa tekstin jonka kirjoittanut jo kymmeniä kertaa	virheitä perusasioissa, ei tajua astevaihtelua, puuttu esimerkiksi verbit luki et <i>minä onnellinen</i>	yritystä oli mut saatto olla kauheen yksinkertainen mut siin oli joku kokonainen lause vaikka oli ollut väärinki niinku <i>tytön makaa huone</i> (vrt. taso 1)
3	saa selvän ja missä on kuitenkin virheitä varsinkin verbeissä, kohtia joita ei ymmärrä, mokia, minkä voi ymmärtää väärin, imperfekti (kuvatehtävässä), pystyy kirjoittamaan omasta päästään jotakin vaikka virheellisesti	keskitaso, oli perusasiat niinku just astevaihtelu, tietysti nyt oli niitä virheitä saatto olla	ei enää niin riisuttu, esim. <i>tytön makaa huone</i> ja sit siinä olis ollut vähän jotain muutaki
4	enemmän virheitä ku vitosessa, kaikissa muissa voi olla virheitä mut verbit täytyy osata nelosen kirjoituksessa, teksti täytyy ymmärtää kerta lukemalla	kaikki niinku tämmöset (perusasiat) kunnossa ja oli vielä niinku tekstikin silleen hyvä	oli ihan hyvä tarina vaik se oli just kieliopillisesti päin helvettiä (vrt. tasot 1 ja 2, kuvatehtävä), tekstuaalisesti hyvä vaik kielellisesti heikko
5	ei saa olla kovin paljon virheitä, ihan kaikki oikein mut joku pieni juttu, rektioissaki täytyy olla suurin osa oikein varsinkin yksinkertaisimmissa	vähän vaille huippu, oikein hyvä lähes niinku virheetöntä ja tekstiä oli enempi	jäi vähän epämääräseks, ei karkeit kielioppivirheitä, aika hyvä, vitosta mä en antanu vaik olis ollut oikein mut jos se oli kauheen suppeesti et mä halusin et on vähän yritystä
6	olis niinku suomalainen, sataprosenttisesti oikein	lauseenvastikkeita ja oikein, hyvin hyvä	jäi vähän epämääräseks, ei karkeit kielioppivirheitä, aika hyvä, ei kovin paljon korjauksista, kutosta mä en antanu vaik olis ollut oikein mut jos se oli kauheen suppeesti et mä halusin et on vähän yritystä
	<div style="border: 1px solid black; padding: 2px; display: inline-block;">Kriteerit</div> <ul style="list-style-type: none"> • sisältö • pituus • rakenteet • ymmärrettävyys • idiomaattisuus 	<ul style="list-style-type: none"> • pituus • rakenteet • tekstimäisyys 	<ul style="list-style-type: none"> • ymmärrettävyys • rakenteet • sisältö • pituus • tekstimäisyys • yritys/paneutuminen

TAULUKKO 18 Arvioijan raportoimat tasokuvaukset.

TASO	PÄIVIKKI	PIRKKO	EEVA
1	sanoja peräkkäin että siitä ei välttämättä viesti välity, ei osaa esimes tehdä lauseita sil ei oo virkerakennetta, esimerkiksi niinku subjekti ja joku verbi sit jotain muuta, yleensä aika lyhyt	pystyy ymmärtämään mistä siinä on kysymys, pystyy viestimään ainakin ihan semmosen perusasian sieltä tehtävänannosta, ei tarvi kovin paljon osata	saanu jotakin kirjoitettua ja jonku sillä tavalla, että siitä ymmärsi, muoto ei tärkeä
2	lausemaisuuatta jonkun verran, joitakin semanttisesti oikeita verbejä mut sit voi olla vielä paljon rajummin vielä rikkinäistä, ei välttämättä kaikkee ymmärrä mitä se on tarkottanu, vielä aika lyhyt mut siinä voi olla pitempiäki tekstejä mut siinä on niitä ykköksen piirteitä	osas enemmän ku ykkönen, oli jonkilainen lähtö	jos käytetty esimes paritiiviä (ostoslistatehtävä), muoto ei tärkeä, pystyy kertomaan vähän laajemman jonkun tai vähän enemmän kertoo
3	melkein kaiken ymmärtää mut ei välttämättä ihan kaikkee, yrittään esimes jo taivuttaa verbejä ja jotaki sijamuotoja laittaa, ne voi olla vielä päin vitelikkö mut siinä on ihan selvästi taju, jos kolmonen on vaativa teksti siin on tosi isoja tyyhillisiiä mokia	asetteli ei oo nii kovin suomalaisstandardin mukainen, ajatus voi katketa kesken kaiken ja muuttua joksiki muuksi, rakenteissa ja kielessä virheitä, semmonen pääasia ei se niinku yhdeltä lukemalta ihan selviä eikä vielä kahdeltakaan että sitä joutuu sitte sepiittämään lukija vielä lisää	jos vaikka kiitos tai semmone niinku viestinnällisesti (ostoslistatehtävä) rupesin enemmän kattoomaan mitä muotoja siellä on, helppo lukea ei tarvi arvata
4	välittyy se (viesti) jo aika hyvin, voi olla semmosia vähän hauskoja sanoja tai vähän niinku semanttisesti väärin valittuja verbejä tai jotenki alatyylisiä tai niin liian yleviä tai no vaan sopimattomia sanoka mutta muuten se on hyvin tekstimäinen, aikamuotoja havaittavissa ja paikallissijoja tämmösissä peruspaikallissijoissa on jo voi olla silleen hajanaista	pitäs pystyä sanomaan et mitä haakee, semmonen hyvä keskitaso	sisällöt, et miten järkevää se on, onks se yleensä kirje, paljonko ihan niitä rakenne ja sanostovirheitä, kirjoituksen muoto
5	kunnon teksti et siin ei mitään mikä esimes lukijan pysähdyttää, yritetty mennytty aikamuotoo rakentaa, perusrakenteet niinku partitiivi ihan hyvin paikalla, haettu semanttisesti jo niinku semmosta hienosäättö sanoihin, yritetään jo niinku sävyjä tai sanontoja tai fraaseja	jos ei täyttäny kuutosta sit se oli vitonen, standardin mukasesta saa vähän olla poikkeusta (työpaikkahakemus), mahtuu vielä kieliooppivirheitä	sisällöltään täysin ok, kaikesta sai selvän, täytti kokonaan sen mistä tehtävä oliki, kirjoitettu muodot kaikki ok, esimerkiksi se työhakemus oli liian tavallinen tai ei ollu sitten oikeita asioita tai oli jotain ihan epäoleellista. Tai sitte saatto olla hirveen hyvä ne asiat itessään mut sitten oli jonkun verran muodollisia virheitä
6	sisältö ei saa ärsyttää, hyvin pieniä rikkaita esimerkiksi sijamuodoissa, hyvin yksittäisiä esimerkiksi verbin käyttövirheet aika-	asiasisältö, vastaukset kaikkiin tehtävänannon kysymyksiin, monipuolisuus lauserakenteissa kirjoen aloitus ja lopetus	melkein ihan täydellinen, huomaa kuitenkin että ei oo ihan suomalaisen kirjoittama

TAULUKKO 18 (jatkuu)

	muotovirheet hyvin pieniä tai jotain semmosia niinku että ei oo geminaattaa, ä:n pisteet on vilah-tanu jonneki, sanajärjestys voi olla vähän hauska tai voi olla niitä tyylivirheitä pikkusen vielä et voi olla vähän liian turhanhienosti sanottu	semmonen täydellinen, ei kärsi kovin montaa kielioppivirhettä, kuutosen arvonen vaik joku k puuttu jostakin tai joku objek-tisija oli värin	
Kritee-rit	<ul style="list-style-type: none"> • <i>ymmärrettävyys</i> • <i>rakenteet</i> • <i>pituus</i> • <i>sanasto</i> • <i>tyyli</i> • <i>tekstimäisyys</i> • <i>sanasto</i> 	<ul style="list-style-type: none"> • <i>ymmärrettävyys</i> • <i>sisältö</i> • <i>konventioiden hallinta</i> • <i>rakenteet</i> • <i>tehtävännannon täytyminen</i> • <i>oikeinkirjoitus</i> 	<ul style="list-style-type: none"> • <i>ymmärrettävyys</i> • <i>pituus</i> • <i>rakenteet</i> • <i>sisältö</i> • <i>konventioiden hallinta</i> • <i>sanasto</i> • <i>tehtävännannon täytyminen</i> • <i>idiomaattisuus</i>

Lähes kaikissa arvioijien luomissa asteikossa esiintyy vähintään kolme kriteeriä, vähimmillään kaksi ja enimmillään kahdeksan. Tosin kriteereitä ei kuvata systemaattisesti kaikilla tasoilla, vaan niihin saatetaan viitata vain yhdellä tasolla tai niihin viitataan käyttämiäni koontitermejä epäsuoremmin, esimerkiksi "puheenaihe hahmotettavissa", "jo jotakin tolkkua" = ymmärrettävyys, "selkeesti tekstiä", "yksilauseiset on ykkö-siä" = pituus, "lähellä suomenkielisen tuotosta", "kauheen suomalaisen olonen" = idiomaattisuus. Vaikka asteikoissa viitataan joihinkin kriteerei-hin epäsystemaattisesti ja jossakin määrin myös yksipuolisesti, esimer-kiksi rakenteista vain verbien käyttöön, asteikot kokonaisuudessaan näyttävät kuitenkin kattavan suurimman osan kirjoittamiselle tyypillisistä kriteereistä (vrt. esim. Faigley ym. 1985; Connor 1991; Carlisle & McKenne 1991).

Asteikkojen lähempi tarkastelu kuitenkin osoittaa, että niistä ei voi tehdä kovinkaan perusteellisia johtopäätöksiä kriteerien laadun tai määrän perusteella. Itse asiassa asteikot voisi jakaa lähtökohtansa puolesta kahteen ryhmään. Ne näyttävät olevan joko 1) rakenne- ja/tai sanastolähtöisiä, joita edustavat Sannan, Tuulan, Raijan, Inkerin, Tarjan, Pirjon, Pian, Tiinan, Merjan ja Martin asteikot, tai 2) ymmärrettävyys- ja/tai sisältölähtöisiä, jollaisia ovat Teijan, Heidin, Kaisun, Eevan, Päivikin ja Pirkon asteikot. Rakennelähtöisille asteikoille on tyypillistä, että niissä korostetaan yleensä kielen rakenteiden ja jonkin rakenteellisen osa-alueen, esimerkiksi taivutuksen hallintaa. Asteikon perusvire on siis lähinnä rakennekeskeinen ja muutamassa tapauksessa sanastokeskeinen. Tälle rakennelähtöiselle näkökulmalle on tyypillistä myös kielen oikeellisuuden tai virheiden määrän korostaminen.

Tämä ei kuitenkaan tarkoita sitä, että rakennelähtöisissä asteikoissa ei kiinnitettäisi huomiota muihin kriteereihin; rakenteet ovat enemmän-

kin kivijalka, jolle kirjoittamisen taidon määrittely näyttäisi systemaattisimmin perustuvan. Lisäksi rakennelähtöiset asteikot ovat keskenään erilaisia siinä, miten niukasti tai laajasti tasoja kuvataan yleensä tai mitä asioita rakenteista nostetaan esille. Esimerkiksi Inkerin asteikossa välitasojen kuvausta ei juuri ole, mikä ei tosin tarkoita sitä, ettei niitä olisi ollut käytössä itse arviointiprosessin aikana. Myös Pirjon asteikko vaikuttaa melko niukalta, ja se perustuukin lähes yksinomaan rakenteiden hallintaan. Itse haastattelussa Pirjo kertoikin kiinnittävänsä erityisesti huomiota rakenteiden hallintaan. Samaa totesi haastattelussa itsestään myös Tiina, jonka asteikko nojaa varsinkin verbien hallintaan.

”mie painotan varmaan kie siis tarkoitat kiel kielioppii kielioppii mä painotan ihan ihan aika tai siis paljon suorastaan mie tiedän it itseki siitä et painotan sitä ja nimenomaan kielioppivirheitä että noh miun mielestä kyllä tietyt perusasiat täytyy täytyy osata et verbintaivutus täytyy osata jos haluaa saada nyt vai nelostason numeron ja ja sitte niinku rektioissaki täy täytyy olla suurin osa oikein varsinkin yksinkertasemmissa et jos haluaa saada vitosen” Tiina

Ymmärrettävyys- ja/tai sisältökeskeisille asteikoille on taas tyypillistä se, että niissä korostetaan tekstin luettavuutta ja viestin välittymistä ja/tai sisällön järjestyttä tai sen suhdetta tekstityyppiin tai tehtävänantoon. Toki näissäkin asteikoissa on kaikissa mukana myös rakenteiden ja/tai sanaston kriteeri, mutta erotuksena on se, että ymmärrettävyyden ja/tai sisällön progressio kulkee rakenteita systemaattisemmin tasolta toiselle. Osassa näistä asteikoista korostetaan myös sitä, että virheitä voi olla mutta ne eivät mene sisällön ohii tai että niitä voi olla myös asteikon ylimmällä tasolla, kuten tehdään Teijan ja Pirkon asteikossa. Lähtökohtana ei ole siis samalla tavalla kielen oikeellisuus kuin useimmissa rakennelähtöisissä asteikoissa.

Kriteerejä tarkasteltaessa on kuitenkin varottava tekemästä liian yksiselitteisiä tulkintoja näin suppean kontekstin perusteella, koska arvioijat voivat tarkoittaa samaa asiaa erilaisin ilmauksin tai eri asioita samalla ilmauksella. Esimerkiksi ”hyvä kieli” voi tarkoittaa arvioijasta riippuen rakenteellisesti hyvää kieltä tai se voi myös viitata konventioiden hallintaan tai tekstimäisyyteen. Samoin ”ymmärrettävyys” voi viitata sekä konkreettiseen ymmärtämiseen että tekstin helppoon luettavuuteen ja sen sujuvuuteen. Myös ”idiomaattisuus” voidaan tulkita eri asteikkojen kohdalla vähän eri tavalla. Itse olen viitannut sillä kriteerien kategorisoinnissa myös kaikkiin sellaisiin kommentteihin, joissa korostetaan ilmauksen samankaltaisuutta syntyperäisen ilmauksen kanssa, esimerkiksi ”täydellinen” ja ”lähellä natiivin tasoa”.

Kokoamani asteikot ovat otteita haastatteluista, joten koko puhumisen konteksti ei ole niissä läsnä. Jo editoitua versiota täytyy siis

tarkastella varauksellisesti, ettei asteikossa esitettyjen kriteerien perusteella tee ylitulkintoja suuntaan tai toiseen. Asteikkojen tasokuvaukset eivät siis tee täysin oikeutta sille, miten arvioijat kuvasivat yleisesti ottaen kirjoittamista taitona, koska useissa haastatteluissa arvioijat kommentoivat kirjoittamista asteikkoja analyttisemmin ja moniulotteisemmin. Esimerkiksi arviointiin liittyviä ongelmia kuvatesaan arvioijat puhuivat tehtävänannon merkityksestä ja sanaston käytöstä, vaikka se ei välttämättä näy heidän asteikossaan.

Edellä esitetyistä varauksista huolimatta tai ne huomioon ottaen asteikot kuitenkin peilaavat jossakin määrin sitä, mitä arvioijat ovat arvioineet kirjoituksista ja millaisena taitona he näkevät kirjoittamisen. Asteikkojen herättämän käsityksen mukaanhan kirjoittaminen näyttäisi olevan kriteerien runsaasta lukumäärästä huolimatta selkeimmin rakennekeskeinen taito ja siinäkin joissakin tapauksissa lähinnä jonkin tietyn yksittäisen rakenteen osa-alueen taito, esimerkiksi taivutuksen hallintaa, verbien käyttöä tai lauseiden muodostusta koskeva. Rakenteiden painava rooli näkyy myös siinä, minkä tyyppisiä asioita pidetään rakennevirheet kuin tekstimäisyydestä, tekstin sisällöstä tai vaikkapa konventioiden hallinnasta johtuvat puutteet.

Vaikka yleensä ottaen kirjoittaminen näytti haastattelujen perusteella olevan monen arvioijan mielestä rakenteita moniulotteisempi taito, sen arviointi tai kirjoittamistaidon jakaminen eri kriteereihin ja edelleen taitotasoihin ei kuitenkaan peilannut tuota moniulotteista käsitystä. Kirjoittamisesta monisyisenä taitona voi olla helpompaa puhua yleisellä tasolla kuin kuvata sitä, mitä esimerkiksi tekstin sisältö tai koheesio tarkoittaa kuudella eri taitotasolla. Ehkä rakenteet ovat tässä mielessä konkreettisemmin jaettavissa osataidoiksi kategorisen ja hierarkkisen luonteensa vuoksi. Vai nojaammeko taitoajattelussamme sittenkin ensi sijassa – ehkä opetuksenkin lähtökohtana olevaan – formalistiseen kielikäsitteeseen? Davisin ym. (1987) mukaan kyse on siitä, että rakenteita on helpompi opettaa ja arvioida kuin hyvää kirjoittamista, koska rakenteisiin liittyvä tieto on konkreettista ja rakenteiden yksiköt, propositionit, ovat selkeitä.

”noh oli ne periaattees mut ku ei sit se käytäntö menny sen mukaan et se soveltamine oli mut kyllä mä se mä sen ihan tietosesti niinku tein et mä sit mietin et mikä on mikäki kyl se silleen oli mut kyl se ei se auttanu siinä käytännössä et siin oli sit paljon semmosii arviointi vaikeuksii ja oli niinku hankala sitä sitten miettiä et paljon painottaa sitä niinkun että e semmosii tekstuaalisia asioita ja paljo kielipillisiä siis et minkäarvone o semmone et siin oli ihan kiva tarina mut niinku sitten oli niinku ihan päin helvettiä ja minkäarvone on sitte semmonen” Martti

”niinku hain ensin sitä että onko se keskeinen asiasisältö onko ymmärretty oikein se kysymys ja ja saako se sanottua sen viestin mitä on tuotava ja sitten näihin yksityiskohtiin että onko määrät oikein nii paljon ku tai siinä sanottiinkii et piti olla määrät ja sitten oliko kieli niinkun suunnilleen kunnossa ja sitten oliko siinä mukana semmoset ää tervehdysfraasit jos aatellaan että se ystävä aamulla heräs ja luki sen paperin puhuteltiin sitä jotenki ja sitten samoten oliko siinä allekirjotukset ja sitten joku muu semmone et siellähän saatto olla semmosia niinku kielivalinnalta tai tyyliä va oli semmosia hyvin hauskojaki juttuja et oli ihan niinku käytetty jotaki muuta ilmasutapaa kun ihan semmosta asiallista esimerkiksi tämmöne siinä sitten vaikutti ja sitten noh kaiken kaikkiaan nin tämmöne yleislinja...mutta semmosia kriteerejä mulla ei oo erikseen että tämmöne tarkoittaa ykköstä ja tämmöne kuutosta että se vaan niinkun jotenki tuntuu että mille tasolle se sattuu” Pirkko

Yhteistä kaikille asteikoille on pääosin se, että taitotasolla 1 teksti on hyvin lyhyt ja siitä ymmärtää jotakin, taitotasolla 2 tekstiä on vähän enemmän ja kirjoittaja pystyy osoittamaan jonkin verran jonkin kielen perusrakenteen hallintaa, taitotasolla 3 suoritus on ”teksti” ja siinä on viestitty jokin sisältö suhteellisen yhtenäisesti, ja kirjoittaja osoittaa hallitsevansa verbintaivutusta, taitotasolla 4 teksti on suhteellisen yhtenäinen ja ymmärrettävä, vaikka siinä on joko kielellisiä tai tyyllisiä virheitä, taitotasolla 5 teksti on melko virheetöntä, tyylin ja kielen rakenteiden käyttö monipuolisempaa ja taitotasolla 6 kirjoittaja taito on lähellä syntyperäisen taitoa.

Mistä arvioijat ovat ammentaneet asteikon sisällön? Kahdeksan arvioijaa kertoi saaneensa vaikutteita pääasiassa omasta opetuksestaan. Toisin sanoen omassa opetuksessa tai oppilaitoksessa käytettävää asteikkoa vähän muuttamalla sen sai sopimaan tässä yhteydessä käytettävään asteikkoon. Esimerkiksi jos arvioija arvioi opetuksessaan 5-portaisella asteikolla, hänen tarvitsi lisätä siihen vain yksi taso, niin asteikosta tuli 6-portainen. Jos opetuksessa taas käytettiin 3-portaista asteikkoa, se oli arvioijan mukaan lähes sama, koska opetuksessa käytettävässä asteikossa ovat mukana myös puolikkaat, plussat ja miinukset.

Tämä ei kuitenkaan tarkoita sitä, että arvioijat olisivat suoraan lainanneet opetuksessa käytettävää asteikkoa sisältöineen ja tasoineen. Kyse lienee enemmänkin epäsuorasta vaikutuksesta. Eräällä tavalla nämä opetukseen pohjautuvat asteikot tuntuivat olevan yhdistelmä opetuksessa käytettävää asteikkoa ja opettajan omia näkemyksiä ja painotuksia. Haastatteluissa tulikin esiin myös se, mitä arvioijat pitävät tärkeänä omassa opetuksessaan ja millaisiin asioihin he kiinnittävät huomiota. Tässä käytettävässä asteikossa ehkä painottuukin enemmän opettajan oma näkemys ja vähemmälle jäi taas se, miten arviointi liittyy

opetuksessa asetettujen tavoitteiden saavuttamiseen, koska kirjoittajat taustoinen olivat arvioijille tuntemattomia.

”mut tää pohjautu hirveen paljon siihen mikä on meillä tää systeemi täällä et siinä oli toi kutonen oli niinku ylimäärästä et sen takia niinku et siin oli niinku yks yks enemmän ni meillä periaatteessa on täällä pyöriny niitä nel neljän tason kurseja et sitä viidettä ei oo ollukkaan” Pirjo

”mun tietysti on helppo tavallaan arvostella sitä ku meillä on neliportainen asteikko ollu tähänki asti että niinkun niin no tavallaan niinku et tota mä niinkun suurinpiirtein tuntumalla jo tiesin et mitkä mä laitan sinne alkupäähän ja mitkä sitten on niitä huippuja” Sanna

”mistä mä oon ne repiny ni tota ni se tulee siitä siitä kyllä että tota kun noita niinku äidinkielenä suomea niitä ylioppilaskokeita työkseni arvioi nin tota siinä on se selvä heija heijastus et esimerkiks tuo kolmone et mikä se kolmone on se tulee kyllä sieltä selvästi” Teija

Puolella arvioijista oli kokemusta Yleisten kielitutkintojen testien arvioimisesta, mikä tuli esiin myös haastatteluissa. Nämä arvioijat viittasivatkin 9-portaiseen YKI-asteikkoon, ja osa heistä sanoi saaneensa vaikutteita siitä omaan asteikkoonsa, varsinkin alemmille tasoille. YKI-asteikon vaikutusta ei kuitenkaan koettu pelkästään positiivisena vaan myös häiritsevä. Osa raportoikin, että he tietoisesti yrittivät pitää sen erillään tästä asteikosta, mikä tuntui onnistuneen haastattelijoiden omien kommenttien mukaan vaihtelevasti.

”mua sitte sotki tää yki että sillan ensimmäisiä ku katto ni ensin aateli aina sen että miten tää ois siinä ykin sitte rupes miettimää että no hetkinen kun näitä on nyt kuus että mitenkäs se menee mutta sitte ei siinä menny kauan em sitte aatellu enää” Eeva

”noh mua häiritsee se että mä mietin niitä ykin kriteereitä et mä niinku nappaan tavallaan sieltä niitä perusteita mä en onneksi muista niin tark ja enkä ja olin fiksu etten kaivannu niitä et et että oisin niinku suoraan ruvennu niinku ajattelemaan että ykkönen vois olla joo ykkönen ja kuutonen vois sitte olla tuolla niinku tuolla kohalla että en onneks sitä ruvennu tekemään mutta mutta niinku ihan selvästi niihi kriteereihin tuli tukee sieltä et niitä että mitä mä katon tästä nimenomaan niitä havaintoja tukemaan että et mä hain sieltä tekstimäisyyttä mä hain sieltä tätä ja tätä ja tätä vaikken suoraan niinku niihin numeroihin mitenkää korreloi” Päivikki

Pari arviolijaa raportoi myös tuntevansa Euroopan neuvoston asteikon, mutta sillä ei tuntunut juurikaan olevan arvioijien mielestä vaikutusta. Näin oli ehkä siksi, että arvioijilla ei ollut siitä arviointikoekemusta samalla tavoin kuin joillakin oli kokemusta Yleisten kielitutkintojen testien arvioimisesta. Haastatteluista kävi ilmi myös se, että vaikka asteikko nojautuikin jossakin määrin johonkin muuhun arviointijärjestelmään ja/tai arvioijan omiin kokemuksiin ja käsityksiin, arvioinnissa oli mukana myös tekijöitä, jotka vaikuttivat asteikon soveltamiseen. Tavallaan tämä arviointi oli suhteessa Yleisten kielitutkintojen tasoarviointiin ja opetukseen verrattuna avoimempaa, kuten yksi arvioija totesikin: ”tässä niinku mm koska tässä oli niinku aina jätetty se skaala auki niin tota niin sitä sitten teki niinku aika lailla semmosena tajunnanvirtana tai sillä tavalla että.”

5.2.1.3 Millaista oman asteikon soveltaminen oli – tuntumaa arviointiprosessiin

”mulla kuitenkin loppupuolella muuttu ehkä vähän että et sit mä tuntu et mä annoin sit jolleki paremman mitä mä edellisell entinen teksti oliki mä olin antanut kakkosen nyt mä annoin kolme” Pirjo

Yleisesti ottaen arvioijat pitivät arviointia hankalana ilman valmiiksi määriteltyjä kriteereitä. Arviointia tuntui vaikeuttavan myös se, että useimpien arvioijien mielestä vain omassa mielessä olevan asteikon soveltaminen oli vaikeaa. Asteikko oli kaikilla muilla paitsi kolmella arvioijalla koko arviointiprosessin aikana vain mielessä. Asteikon soveltamista tekstien arviointiin pidettiin vaikeana monista syistä. Vaikeuksia aiheuttivat 1) *tasojen erottaminen toisistaan*, 2) *taukojen aiheuttamat katkokset*, 3) *arvioitavien tekstien samantasoisuus* ja toisaalta taas niiden *eritasoisuus* sekä 4) *eri kriteerien painoarvosta päättäminen*.

1) *Tasojen erottaminen toisistaan*

Kaikkein helpoimpana arvioijat pitivät tasojen 1 ja 6 soveltamista. Varsinkin taso 1 tuntui toimivan lähes kaikkien arvioijien mielestä hyvin, koska se oli helposti määriteltävissä muutaman sanan tai lauseen pituiseksi tekstiksi, kuten arvioijien omista asteikoista näkyy (ks. taulukot 14–18). Tason 6 soveltaminen oli ongelmallisempaa, koska tekstien yleinen taitotaso painottui enemmän asteikon alemmille tasoille, joten vain harvat tekstit ylsivät heidän mielestään asteikon ylimmälle taitotasolle. Tämä näkyy myös luvussa 5.1.3, jossa tarkastelen eri mittauskertojen keskiarvoja tehtäväsarjoittain, sekä liitteen 9 arvioija- ja mittauskertakohtaisia taitotasofrekvenssejä kuvaavissa taulukoissa. Ääri-

päiden kokeminen helpommaksi arvioida ei ole yllätyksellistä sikäli, että tekstien joukosta on tavallisesti helpointa poimia niin sanotusti heikot ja vahvat tekstit, koska ne ovat yleensä homogeenisempiä puutteineen ja vahvuuksineen kuin välitason tekstit.

”mutta tuo keskiosa oli vaikeempi ja se että niistä mulla oli nyt joku näkemys näistä ykkösistä vitosista ja kutosista mutta niitten kakkosen kolmosen nelosen keskinäisistä rajoista ne oli vaikeita että ei mulla semmosta ei se mun mielestä muodostunu tarkaks niissä niissä myöhemminkään” Inkeri

Ongelmia siis tuottivat varsinkin tasot 3 ja 4 mutta myös muutamalla arvioijalla tasot 2 ja 5. Monien mielestä välitasojen välillä oli vaikea tehdä eroa: ”että tuota aika lailla veteen piirrettyjä viivoja kyllä ja semmosta hyvin häilyvää niinku sekin että millon on esimerkiksi kakkone ja millon kolmone.” Keskiarvojen vaikeusvaikutelmaa voi korostaa myös se, että arvioijat itse raportoivat käyttäneensä eniten taitotasoja 2 ja/tai 3, jolloin saman tason toistaminen suorituksesta toiseen voi tuntua esimerkiksi uskon puutteena tai epävarmuutena siitä, miten laajaksi tason rajat voivat venyä eli miten erilaiset ja määrällisesti monet suoritukset voivat olla samaa tasoa: ”siinä oli paljon sellasia jotka mä oisin laittanu sinne keskelle [mumisee] ku piti päättää ni sit mä päätin.”

”no joo no ei ei et tietysti ois voinu kaikki ne laittaa siihe kakkoseen mutta mut jos mä huomasin et siin oli jotain semmosii et fraaseja tai jotain sellasii muotoja mitä mitä ne niinku periaatteessa kakkos kakkosen ei niinku osais käyttää kyllä mä sen oon varmaan laittanu sit siihen seuraavaan sitte mut voi hyvin olla että se ois niinku todellinen taso siel paljon ylempänä et tota noh sen ne on ongelmallisia tekstejä en mä ni em mä tiedä oliko mulla niissä mitään hirveen johdonmukasta tapaa mä vaan laitoin ne johonki” Sanna

”se toisaalta sitte se kolmone ja nelosen raja ni oli kyllä vähän häilyvä et em mä niinku pystyny sitä ihan itselleni semmoseks kauheen tarkaks laittamaan et se oli kyllä vähän semmone mutujuttu” Tuula

2) Taukojen aiheuttamat katkokset

Arvioitavia tekstejä oli yhdellä arvioijalla vähimmillään sadalta kirjoittajalta eli noin 300 tekstiä ja enimmillään 247 kirjoittajalta eli 740 tekstiä.¹⁷ On selvää, että tällaisen, massa-arvioinniksikin kutsutun, arvioinnin tekeminen vaatii useampiakin taukoja ja varsinkin tällaisessa

¹⁷ Kuten luvusta 5.1 näkyy, arvioitavien tekstien määrät vaihtelevat jonkin verran sen mukaan, oliko kukin kirjoittaja kirjoittanut kaikki kolme tehtävää tai oliko arvioija antanut arvion kaikille suorituksille. Arvioija saattoi jättää antamatta arvion esimerkiksi sellaisille suorituksille, jotka eivät hänen mielestään täyttäneet taitotason 1 vaatimuksia.

tapauksessa, jossa arviointia tehtiin oman työn ohessa iltaisin, öisin ja viikonloppuisin. Muutamalla arvioijalla tauot saattoivat olla useamman päivänkin mittaisia; joku taas teki arvioinnit melkein kertaistumalta pitämällä vain lyhyen tauon.

Sekä pitkät että lyhyet tauot työn lomassa koettiin siten, että niiden jälkeen oli hankala palata siihen, mihin ajatukset, lähinnä käytettyjen kriteerien sisällöt, jäivät tauolle lähdeittäessä. Otteet Eevan ja Pian kommentteista havainnollistavat hyvin tauon jälkeistä tilannetta, jossa on vaikeaa palauttaa mieleen tarkalleen sitä asteikkoa tai niitä kriteereitä, joita on soveltanut kirjoituksiin ennen taukoa.

”minusta itestäni tuntu et se muuttu koska mä tein tätä niin katkonaisesti että varsinkin aluks sillä tavalla että puoli tuntia yhtenä päivän seuraavana päivänä tunti niin sillon tuntu että se muuttuu koko ajan et just noitten nuitten en hel tavallaan sellasten helpompien tehtävien osalta just ne et siinä pysykö se miten mää niinku aattelinkaan tään koska em mä sitä mitenkään iteleni kirjannu ylös niin että varmaan varmaan muuttu kyllä” Eeva

”ja sit huomas jos niinku piti paljon taukoo esimerkiksi joka kertaku alotti ni tuntu niinku että ois niinku alottanu alusta jos sitä teki niinku yhtäsoitto ni siihen tuli mut sit jos tuli jotain taukoja väliin yhtään pitempiä ni se oli tavallaan uelleen käytävä se koko prosessi läpi niinku mielessä et sitte pysty taas jatkamaan sitä tai sitte täytyy alkaakin tarkastella missä meni missä aikasemmi” Pia

Eevan ja Pian kommenttien perusteella tuntuu vähän oudolta se, että arvioijat eivät kirjoittaneet kriteereitään ylös, vaikka lähes kaikki arvioijat kommentoivat kriteerien puutetta ja siihen liittyvää epävarmuutta ja mielivaltaisuutta, sekä sitä, että he olivat tottuneet käyttämään kirjallisia kriteereitä. Kriteerien karkeakin merkitseminen paperille olisi voinut antaa arvioijille tukea ja lisätä varmuutta varsinkin taukojen jälkeen: ”se oli mulla vaan päässä sit jos sen ois pannu paperille sit ois voinu tarkistaa et ne ois kaikille samat” tai ”se oli mun päässä mut ehkä mun on nyt syytä sanoo täss et se ei välttämättä ollu sen paussin jälkeen.”

Toisaalta kriteerien kirjaaminen ylös saadettiin kokea liian aikaa vievänä, sillä arviointiin haluttiin ehkä päästä nopeammin ja suoremmin käsiksi. Yleensäkin tekstien arvioinnin koettiin vievän enemmän aikaa kuin mitä odotettiin: ”noh aikaa vievää se oli siihe verrattuna et mä ajattelin et mä ehkä lukisin tunnissa tai kahessa.” Vaikka tällaiset kommentit liittyvät enemmänkin arvioinnin jälkeisiin tunnelmiin, kriteerien kirjaaminen paperille on voinut tuntua arviointia aloitettaessa kohtuuttomalta lisätyöltä. Joissakin tapauksissa vaikutti siltä, että arvioija ei ehkä tullut ajatelleeksi kirjaamisen tarpeellisuutta ennen kuin vasta urakan jälkeen. Syyinä tähän on voinut olla myös se, että arviointia ei ehkä

otettu kovin vakavasti, koska arvioijat tiesivät olevansa mukana tutkimuksessa eivätkä esimerkiksi tekemässä testimäistä taitotaso-arviointia todistuksineen.

Joka tapauksessa väsymys ja tauot ovat myös tutkimusten mukaan (esim. Harwell 1999) merkittävästi sisäiseen johdonmukaisuuteen vaikuttavia tekijöitä. Väsymys tulee esiin esimerkiksi silloin, kun arvioitavaa on paljon ja siihen on vähän aikaa, kuten tässä oli: ”siin lopussa et et mä luin ne aika niinku mä taisin saattaa parhaimmillaan jopa pari teht tekstii kattoo minuutissa tai niin sillai ku tiiätsä oli väsyny ja oli kiire.” Toinen sisäiseen johdonmukaisuuteen heikentävästi vaikuttava tekijä on se, että arviointityötä tehdään pitkällä aikavälillä, mikä näkyikin edellä olevasta Pian esimerkistä. Tällaisten tekijöiden vaikutus korostuu varsinkin silloin, kun asteikkoa ei ole kirjattu mitenkään ylös.

3) Arvioitavien tekstien samantasoisuus ja eritasoisuus

Asteikon soveltamisessa oli vaikeaa myös se, että kirjoitukset olivat arvioijan mielestä hyvin samantasoisia. Arvioija saattoi joutua antamaan samoja tasoarvioita – kuten jo edellä on käynyt ilmi tavallisesti kakkosta ja kolmosta – jopa useita kymmeniä peräkkäin: ”ja sit rupes tuntuu et mä oon antanu puol sivuu nelosii ni sit oli pakko vaik väkisin antaa muita.” Martin kommentista käy kokoavasti ilmi, millaisia tuntemuksia ja mistä syystä samojen tasoarvioiden antaminen herättää arvioijassa sekä millaista epävarmuutta siihen liittyy.

”sitte niinku et sitte rupes tuntuu et on vähän liian ankara ankara et sitte ku oli niinkun semmone syvälle piintyny tarve et pitäs tulla kaikkii arvosanoi silleen niinku ainaki vähä et vähä semmonen gaussin käyrä kummitteli mielessä sit välil tuntu et kyl ne vaikka ne oli olevinaan selvät ne kriteerit ni ne silleen niinku että välillä jumittu johki arvosanaan ja sit rupes tuntuu et mä oon antanu puol sivuu nelosii ni sit oli pakko niinku vaik väkisin antaa muita et sellane rupee tommosessa isos vaikuttaa ku ne vetää ruksin siihen kaavakkeeseen peräkkäin ni semmosee kiinnittää varmaan paljon huomiota ku jos ei olis merkannu niit ylös jos ois vain pistäny et se vaikuttaa mun mielestä et sitä rupee vähä kattoo että mä oon nyt varmaan huono arvioija ku mulla on noin mont nelosta” Martti

Jotkut haastateltavat kiinnittivät huomiota myös siihen, missä vaiheessa eli kuinka monen suorituksen jälkeen he olivat antaneet ensimmäisen viitos- tai kuutostason arvion. Jotkut myös laskivat, kuinka monta viitos- tai kuutostason suoritusta heidän arvioimiensa tekstien joukossa on. Yhtä lailla arvioijat raportoivat siitä, jos he eivät olleet antaneet yhtään kuutosta. Kuutostason arvioiden puuttuminen tuntui aiheuttavan epävarmuutta, lähinnä epäilyksiä siitä, tuliko tekstit arvioi-

duiksi oikein, oikeudenmukaisesti tai samassa linjassa muiden arvioijien kanssa.

Tällaisen ajattelun taustaoletuksena näyttäisi olevan se, että koko asteikkoa pitäisi voida käyttää, jotta arviointi tulisi tehtyä 'kunnolla'. Kommentit kertovat ainakin siitä, että ylätasoja täytyi jollakin tavalla perustella: "pitää tätä koko skaalaa käyttää mut sitten se et kuitenkin mulla oli jotkut kriteerit et ehkä päässä nitte suhteen etten mä voi antaa kutosta jos siel ei kerran mun mielest oo sellasii."

Asteikon muuttumiseen kesken arviointiprosessin vaikutti myös se, missä määrin teksteistä syntyvä vaikutelma aiheutui tehtävyyypistä. Osa tehtävistä oli tarkoitettu alkeistason tehtäviksi (A-taso), esimerkiksi kauppalista ja tuttavallinen esittelykirje, ja osa taas edistyneemmän tason tehtäviksi (C-taso), esimerkiksi suruvalitteluviesti ja työpaikkahakemus. Jos tehtävät vaihtuivat vaikeammaksi kesken arvioinnin, muuttuivat myös tekstit paremman tasoiseksi, koska oli todennäköistä, että niitä olivat kirjoittaneet kieltä paremmin hallitsevat oppijat. Tehtävien muuttuminen saattoi vaikuttaa siihen, että arvioijan linja tiukentui, koska tekstitkin muuttuivat paremmiksi.

"et sit se tuntu siinä jossain puolen välin jälkeen et hei nähän paranee ihan huimasti et et sopiikohan tää nyt enää ja sitte mulle tuli just se se epävarmuus siitä että mie aloin et mi en alkaa vertailee mut mie aloin ajatella et no ni et nyt varmasti siellä alkupäässä on kolmosia joille mie nyt en antais ku kakkosen ku taas nyt mie annan sellasille kolmosen jos se sama teksti ois ollu alkupäässä ni mie oisin antanu sille nelosen et si et ei se ollu oikein helppo eikä kiva koska se ei ollu kuitenkaan ni semmonen ja sit siinä tuli niinku tohonki mie kirjoitin siin tuli niinku vertailtuu niitä toisiinsa mikä taas ei sit välttämättä oo" Tiina

Kuten Tiinan esimerkistä näkyy, hänen asteikkonsa sisältö muuttui tehtävyyppien muuttuessa varsinkin keskitasojen osalta. Tällainen muutos ei ole kovin yllättävä, jos arvioija oli rakentanut asteikkonsa tekstipinkan alkutekstien tason mukaan ja jos hän sattui arvioimaan myös neljännen tehtäväsarjan tehtäviä. Jos tekstit muuttuvat parempitasoisiksi kesken arviointiprosessin, on kovin todennäköistä, että asteikon keskiväli kasvaa eli sille sijoittuvien tekstien kirjo laajenee entisestään. Tämä edellyttää tietysti sitä, että tässä vaiheessa tulevat tekstit eivät ole arvioijan mielestä niin hyviä, että niitä voisi sijoittaa asteikon kahdelle ylimmälle tasolle. Osalla tämän tutkimuksen arvioijista varmasti kävikin näin, koska arvioijien itsensä määrittelemien asteikkojen perusteella varsinkin taitotaso 6 edellytti erittäin hyvää taitoa, ainakin kielen hallinnan osalta (ks. taulukot 14–18).

Näin ollen asteikon välitasoista saattoi tulla kovin laajoja, mikä saattoi tuntua arvioijista epäreilulta sekä heikompi- että parempitasoisia

tekstejä kohtaan. Ongelmallista tässä oli erityisesti se, että asteikon muuttaminen johdonmukaisesti, myös alkutekstien osalta, ei onnistunut, koska se olisi vaatinut liian isoa työmäärää. Arvioijien mielestä olisi ollut liian ylivoimaista käydä suurta määrää tekstejä uudelleen läpi: "se arviointiskaala vaihteli siin matkan varrella koska se niinku se oma prosessi meni sitten ja sen ja ku ei voinu nin isossa määrässä palata takasin päin." Toisaalta arvioijien asteikot saattoivat muuttua muutenkin tekstikohtaisesti mutta eri syistä, koska asteikot olivat vain heidän mielessään. Jos asteikko olisi kirjattu ylös, voi olla, että tekstien tason muuttuminen ei olisi vaikuttanut asteikkoon samalla tavalla.

4) Eri kriteerien painottaminen

Oman asteikon soveltamisen systemaattisuus ja toisaalta asteikon sisällön pysyvyys epäilyttivät ainakin jossakin määrin kaikkia arvioijia. Suurin osa arvioijista pohdiskeli sitä, millaista heidän itsensä määrittelemää asteikkoa he oikeastaan sovelsivat ja oliko asteikko sama arviointiprosessin eri vaiheissa. Arvioijia arvelutti muun muassa se, mikä eri kriteerien painoarvo oli ja tuliko niitä painottaneeksi suunnilleen saman verran eri tekstien kohdalla. Varsinkin rakenteiden painoarvoa kommentoitiin: "must tuntu et kielioppipohja vei pitemmän korren mulla kuitenkin loppupuolella muuttu ehkä vähän." Kriteerien välinen painoarvo korostui myös silloin, kun suoritus oli kahden tason rajalla.

"mä sit mietin et mikä on mikäki kyl se silleen oli mut kyl se ei se auttanu siinä käytännössä et siin oli sit paljon semmosii arviointi vaikeuksii ja oli niinku hankala sitä sitten miettii et paljon painottaa sitä niinkun että e semmosii tekstuaalisia asioita ja paljo kielioppillisii siis et minkäärvone o semmone et siin oli ihan kiva tarina mut niinku sitten oli niinku ihan päin helvettii ja minkäärvone on sitte semmone että kielioppi oli mut sit oli vaan että pari lausetta et niitten rinnastaminen oli hankalaa" Martti

Toisaalta arvioija saattoi olla hyvinkin varma siitä, mikä kriteeri tai mitkä kriteerit laskivat tai nostivat suorituksen tasolta toiselle: "usein mä kakkosen ja kolmosen erotin sen mukaan että minkä verran siellä on oikeita muotoja että niitä oli helppo lukea." Toisaalta taas päätökseen tasosta ei tuntunut liittyvän kovinkaan paljon varmuutta tai selkeitä perusteluita: suoraan sanottuna tuntu just joskus kakkosen ja kolmosen välilläkin nimenomaan tämmösissä pitemmissä tekstissä että ei tavallaa ollu semmosta selkeätä juttua millä pysty erottamaan."

Monissa rajatapausten tai yleensäkin tasojen yksityiskohtaisemmissa kuvauksissa arvioijat puhuivat rakennepainotteisemmin kuin puhuessaan vaikkapa kirjoittamisesta yleensä. Tällöin arvioijat kommentoivat esimerkiksi partitiivin käyttöä, verbien taivutusta, aikamuotojen käyttöä tai rektioiden hallintaa. Toisaalta taas arvioijat vetosivat siihen, että

useimpien tekstien kohdalla kokonaisuus ratkaisi tasoarvion tai että jokin yksi rakenteellinen seikka ei ohittanut muita tekstin ansioita. Kuitenkin silloin kun teksti oli kahden tasoarvion välissä, kielen rakenteelliset seikat saattoivat painaa enemmän kuin esimerkiksi tekstin looginen rakenne.

”toki partitiivi nyt muuten mut just syödä ja juoda nimenomaan niiden verbien yhteydessä mut tota nin eikä suinkaan lauseenvastikkeissa mutta mut just se esim se astevaihtelu ja verbin taivutukset oliko oliko niitä ja sit semmoset ihan oikein kirjoitus perus tämmösissä tavallisissa sanoissa ja sitte verbit että joiltaki puuttu ne...itse asiassa loppujen lopuksi tuossa koko jutussa kyll se oli kuitenkin kokonaisuus mikä sen ratkas et em mä nyt suinkaan ni takerru ahaa nyt sillä oliki että tyttön et tää on kakkone et saatohan se olla muuten hirveen ja sit se oliki kolmonen ei ne nyt ollu ettei suinkaan voinu saada kolmosta ja oli erehtyny kirjottamaan ni ni ei suinkaan vaan joitakin et jos oli siinä jos siitä niinku katso koska tuleehan virheitä vaikka on kuin muuten hyvä kirjottamaan ne oli ratkasevia joo et...sit esimerkiks ku mä mietin onks toi kakkonen vai kolmonen mä saatoin katsoa ai niin tossa on toi astevaihteluki noi et ei totakaan oo ymmärtäny et tää on sit kakkone ne oli semmosissa rajatapauksissa varmaan” Merja

Rakenteiden lisäksi muita rajatapausten kriteereitä olivat muun muassa oikeinkirjoitus, tekstin hauskuus, pituus ja siihen liittyen riskinotto kyky. Esimerkiksi Päivikin arvioinnissa suoritusta alensivat rakenteellisten puutteiden lisäksi välimerkittömyys yleisesti, kirjaimien epäselkeys ykkös-kakkostason rajoilla, isot ja pienet kirjaimet kolmosnelostason rajoilla; suoritusta taas nostivat esimerkiksi luovuus ja riskinotto kyky, joiden määrittelemisestä ja soveltamisesta tasokohtaisesti ei tosin keskusteltu sen yksityiskohtaisemmin.

Muutama arvioija käytti myös sellaisia kriteereitä, joita he eivät maininneet taitotasoa kuvattaessaan tai jotka eivät olleet varsinaisia rajatapauskriteereitäkään. Näitä kriteereitä voisi kutsua yleiskriteereiksi, joiden hallintaa arvioija ei jaotellut kuudelle taitotasolle. Tässä mielessä ne muistuttavat yllä kuvattuja rajatapauskriteereitä, joskin näitä kriteereitä ei sovellettu tiettyjen tasojen rajoilla, vaan ne toimivat yleisemmin kaiken tasoissa teksteissä joko niiden tasoarviota korottavasti tai alentavasti. Toisin sanoen jos kirjoitus täytti muuten tason edellyttämät kriteerit, mutta ei tätä yleisluonteista kriteeriä, se sai yhtä tasoa alemman tasoarvion.

Tällaisia kriteerejä olivat esimerkiksi tekstin niukkuus, tehtävänannon täyttäminen ja kommunikoi vuus eli tekstin luettavuus. Yleiskriteerit eivät kuitenkaan aina toimineet näin kategorisesti ja systemaattisesti, tai ne eivät olleet yhtä olennaisia kaikissa tehtävätyypeissä.

”sehän muuten oli yksi yksi ongelma tuos koko hommassa että mitä tehdään semmosille joissa selvästi on niinkun rikottu sitä tehtävän käskyä et on ymmärretty jotenki väärin se mä tein sitte niin että vetä vejin sinne ykkösen helposti vaikka se ois ollu muuten muuten aika hyvinki kirjoitettu (miten se vaikutti kuvassa) noh sanotaan että joku oli kirjottanu siihen asti kun sen ois pitäny tulla ni tota ihan sujuvasti ja tota ymmärrettävästi ja sillä lailla hallitusti nin niin tota et mie oisin ollu antamassa sitä kolmosta ni mä vejin sen sitte siihen kakkoseen sitte et sillä lailla mie annoin sen vaikuttaa kyllä (suunnilleen yhdellä tasolla vai) mm mm” Teija

”ensinnäki semmosii periaatteita mulla tuossa oli aika paljon sellasii jossa ei ollu tota se tehtävänanto täytetty jostain syystä esimerkiksi yleisönosastokirjotuksen oli oikeestaan harvoja jotka oli yleisönosastokirjotuksia hyvin harvat nim mää katoin sit silleen et mä pu pudotin yhdellä yhdellä tuota nin nin aa arvosanalla sitä sitä tuota esimes jos muuten ois saanu esimes kolmosen nim mä pudotin kakkoseen tai kakkosesta ykköseen aa jos se ei ollu se tehtävänanto täytetty” Inkeri

”joku oli selvästi kirjottanu paljon mutta oli tehtävän ymmärtäny väärin nin mä vähensin siitä yhden pisteen ja ja sitte et jos oli melkeen kokonaan väärin ymmärtäny tehtävän ni mä vähensin jopa kaks pistettä vaikka se oli hir että pitkä tarina hyvä muuten ihan hyvä kieliki mutta jos kerta kaikkiaan mun mielestä siis se ei kommunikoinu niinkun tavallaan se tilanne sillai et siinä ei ollu sitä kommunikaatiota” Raija

Niukkojen tekstien kohdalla arvioijat saattoivat päätyä antamaan omasta mielestään hyvänkin tasoarvion, jos teksti oli kielen rakenteiltaan ansiokas. Yksi arvioija totesi, että ei hän ollut ihan varma, noudattiko hän mitenkään johdonmukaisesti systemaattista linjaa niukkuuden kohdalla. Hänen mielestään jokin teksteistä olisi voinut ansaita paremmankin tasoarvion, kuin hän antoi. Yleisesti ottaen tekstin niukkuus ohjasi arvioijat kahtalaiseen tulkintaan: 1) kielellisesti hyvä teksti lupasi, että kirjoittaja pystyy kirjoittamaan hyvää tekstiä yleensä; 2) teksti ei ollut niukkuudessaan riittävä näyttö siitä, että siitä voisi antaa esimerkiksi taitotasoa kahta enemmän.

”mie yritin kattoo sen niin ainaki yritin vaikka siin ois ollu vaan puoltoist rivii tai siis pari hassuu rivii jos se olis sen näköstä sanotaan vaik mistä sen sitte niinku näki mut jotenki jos se oli sen näköstä että et siin on ihmine joka kyllä pystyy kirjottamaan ja on kirjottanu ihan ihan ihan hyvin nin nin se on voinu saada lyhyestäki niukastakin tekstistä ihan jonkun vaikka nyt kolmosen tai jopa nelosen” Tiina

Tasojen välisenä kriteerinä oli joillakin arvioijilla myös sanasto tai tarkemmin sanan käyttö tarkassa merkityksessä. Myös sanaston käyttö rajatapauskriteerinä saattoi olla rajautunutta, jos se toimi vain yhdellä taitotasolla. Alla olevassa Kaisun esimerkissä esiin tuleva sana *riikki* liittyi

kaikille kirjoittajille yhteisestä tehtävätyypistä kirjoitettuun tekstiin. Yhden sanan saamaa – ehkä voimakkaaltakin tuntuva – painoarvoa voi selittää se, että *rikki*-sanana käyttö oli varsin yleistä *poikki*-sanana merkityksessä ja se oli keskeinen sana tehtävän kannalta (ks. tehtävä X liite 2). Toisaalta Tuula pohti samaa tehtävää samasta näkökulmasta, mutta hän ei mielestään pystynyt tekemään yhtä selkeää rajanvetoa kuin Kaisu.

”että kuitenkin kolmosessa mun mielestä vielä sallitaan muutama la lapsus mut että esimes mulla oli yks sana jota mä pidin niinku se et tää et ku jalka moni oli kirjottanu et jalka menee poikki tai jalka menee rikki ni se rikki oli mun mielestä se kakkonen monessa mut joillaki oli kyllä sit niin hyvin muuten et sit joutu sen tiettyyn tai tai niinku valitsin sen kolmostason tää on positiivista tietysti mutta mut se rikki oli yks semmone sana joka joka jäi mieleen niistä rajatapauksis e et ei se mun mielestä ni niinku semmone häilyvä raja mutta joku tuntuma siihen tuli kuitenkin” Kaisu

”siinä tavallaan niinku täytyi osata myös tietyt sanat ja sit ku niitä ei tienny ni hirveen monet just sitte sano jotain et tyttö putosi ja ja sitte mulla heti niinku jotenki aina voi ei täähän nyt ei ainakaan sit kolmosta enempää saa sit mä rupesin miettii no en mä nyt voi rokottaa yhestä sanasta niinku niin kauheest mutku toisaalta sitku joku kaatua sana ei niinku tuntunu olevan tuttu juuri kenellekään et jalka on rikki ni sit siinäki jo heti alko ensin putoo ja sit jalka on rikki ni kyl aah että että niinku voi ei et nyt se meni niinku tavallaan ehkä ei ansaitse näin huonoa numeroa näiden sanojen takia siinä mulla oli kyllä vähän ongelmia kanssa ku kamppailin itseni kanssa et voiko antaa sit vaikka käyttää silleen sanoja ni sit hyviä arvioita” Tuula

Tasokohtaisten, rajatapaus- ja yleiskriteerien lisäksi asteikon saattoi jakaa kahteen osaan: ala- ja yläpäähän, jolloin alemmilla tasoilla saattoi käyttää omia yleiskriteerejä ja ylätasolla omiaan. Esimerkiksi Pian mukaan kirjoitukset saattoi ensin jakaa kahteen luokkaan tiettyjen kriteerien mukaan ja vasta sen jälkeen hän päätti siitä, onko esimerkiksi asteikon alapäähän sijoittuva teksti ykkönen, kakkonen vai kolmonen. Myös Eeva jakoi tekstit ykkösestä kolmoseen ja nelosesta vitoseen, mutta siten, että alatasoilla hän kiinnitti enemmän huomiota tekstin ymmärrettävyyteen ja sen järkevyyteen. Ylätasolla hän taas kiinnitti huomiota enemmän rakenne- ja sanastovirheisiin sekä tekstin muotoon ja tyyliin, eli näillä tasoilla hän katsoi sitä ”että onko ne nyt niinku semmosia että ne ihan pomppaa silmille vai ei.”

”no se riippu niin paljon siitä tehtävästä tehtävänannosta ja eri tavallahan sen nyt arvostelee sen suru tai just et kuinka se esitti ja sieltä se tarve ni se vaikutti arvosteluun aika paljon mutta tota nin ylempällä tasolla oli ne muodolliset seikat ja sitte alemmilla tasoilla ni lähinnä mun mielestä niinku nää objektimuodot aika hyviä mun mielestä ja sitte aikamuotojen

käyttö on mun mielestä yks aika hyvä kriteeri ja sitte yleensä se että osaa käyttää niinku osaa käyttää oikeita persoonamuotoja niinku verbit ja objektit ne on ne semmosia semmosia perusjuttuja ku taas ylempillä tasoilla se muoto ja sitte sanaston laajuus ne on ne ne on en pysty sanoo niinku tarkemmin niinku jotain ihmeellistä jotain tehtävää tai...perusjutut kuitenkin oli nää" Pia

Vaikka arvioijat saattoivat raportoida aika yksityiskohtaisestikin kriteerien sisältöä, osa heistä oli huolissaan siitä, käyttivätkö he sittenkään raportoimiaan kriteereitä vai tulivatko he ehkä käyttäneeksi myös joitakin muita kriteerejä: "mut kuitenkin niin ku tekee voi huomata yhtä äkki et ei enää tiääkään tekeeks niitä kriteereillä vai jollain muulla." Tällaisena muuna kriteerinä nousi esiin esimerkiksi käsiala, jota voisi kutsua myös haamukriteeriksi sen epäsuoran luonteen vuoksi. Arvioijat eivät halunneet nimittää tai kategorisoida sitä varsinaiseksi kriteeriksi, vaikka tunnustivatkin, että se vaikutti tekstistä syntyvään vaikutelmaan ja muutamalla jopa tasoarvioon: "joillaki käsiala oli sellane ettei siitä saanu oikein selvää et se on saattanu sitten hiukan heikentää sitä arviointia."

Käsialan haamuvaikutusta osoittaa sekin, että vaikka arvioijat nostivat sen haastatteluissa taajaan esiin, sitä ei ole kuitenkaan kirjattu yhteenkään taulukoissa 14–18 kuvatuista asteikoista. Tosin käsiala ei ollut aina tulkittavissa kriteeriksi, vaan siihen viitattiin enemmänkin sitä kautta, mitä se kertoi esimerkiksi kirjoittajan taustaan liittyvistä tekijöistä, kuten kirjoittajan äidinkielestä tai koulutustaustasta. Onkin vaikea sanoa, miten tällaiset havainnot vaikuttivat tasoarvioihin.

"jos siel tulee joku pieni virhe sen ehkä antaa helpommin anteeks hyvässä käsialassa pieniä virheitä kirjoitusvirheitä ku se että jos sun täytyy koko ajan muutenkin takeltaa sitä käsialan lukemista niin sitte siellä jos tulee vielä joku kielioppivirhe nin mä pelkäsin sitä että että mä en niinku liian ankara sen käsialan takia et tää oli m niinku mun mun semmonen tunne koska siellä oli niin muutama todella surkee käsiala mut luettava sitte taas joitaki niinkun niinkun melkeen ku koneella kirjoitettua tekstiä joka oli tosi ihana lukee semmosta selkeetä hyvää isot kirjaimet ja pienet kirjaimet hyvin paikallaan ja pisteet ja pilkut niin sitten heti tulee semmonen tunnelma että ahaa tähän on hyvä et mä koin sen siinä kohtaa mulle tuli semmonen tunne että et py pysyks mä niinku tavallaan samoissa kriteereissä" Raija

Kaiken kaikkiaan oman asteikon ja sen kriteerien soveltaminen aiheutti melko paljon epävarmuutta omasta arviointilinjasta. Yksi syy epävarmuuteen arviointilinjasta oli se, että arvioijat eivät kirjoittaneet asteikkoja ylös: "no mun mielestä tää oli paljon sattumanvarai no just sen takia että että kysyn miten mää nyt sitten aattelenkin no just silloin kun mä satuin ottamaan nää paperit esille ja sitte mikä paperi sattuu ensimmäisenä käteen jos mä otin semmosen joka oli tosi hyvä ja ja ehkä

en sillä hetkellä nyt miettinyt että miten mää olin aatellu toissa päivänä ni ehkä mä annoinkin sille vähemmän ja siitä seurauksena mä annoin seuraavallekin vastaavasti vähemmän ja koska mulla oli vain oma pää mihin mä vertaan ja oma muisti että miten mä niinku oon eilen antanu tai jotain mutta ne arvioinnit mitä mä oon tehny muuten ni niissähän on ollu sanalliset kuvaukset joissa on voinu aina joka ikisen kirjoituksen kohalla kattoo että mihin kohtaan tää menee.”

Melko varma arviointilinjastaan ja sen säilymisestä oli oikeastaan vain yksi arvioija, Pirkko. Hän sanoi kiinnittäneensä tietoisesti huomiota linjansa säilymiseen. Hänen mielestään, päinvastoin kuin monista muista arvioijista, tämä arviointi oli helpompaa kuin arviointi esimerkiksi Yleisissä kielitutkinnoissa annettujen kriteerien mukaan. Pirkko olikin tehnyt arvioitavista asioista tehtävätyyppikohtaisen listan, jota hän sanoi seuranneensa arvioinnin aikana. Varmuuden, johdonmukaisuuden tunnetta, edesauttoi se, että hän oli kirjannut ylös ”tarkkailtavat asiat”.

”minusta se oli ihan ihan semmonen työtapana että ei musta siinä ollu mitään ongelmia en nyt ainakaan muista että olisin nyt erikoisesti miettinyt ja sitten jos tuli joskus semmonen ni sitten piti niinku palata siihen alkuun ja ja kerrata se mitä mä oon katonu muilta että jos tuli semmonen niinku löp lepsahduksen tunne välillä mut että varmastiki tuota semmoset kirjoitetut kriteerit auttaa mutta ehkä silloin siinä alussa kun se nipun saa mä en tiijä miten tämä jatkuu tämä juttu mutta että vois olla voi olla niin että siinä on niinkun sanottu jotakin mutta musta se ei pahaa tee jos sitä joutuu itekki niinkun miettimään ensiksi että että mitä minä tässä arvioin ja mitkä asiat vaikuttaa ja mikäpä estää tekemästä niin vaikka tuliski kriteerit että ensin mietti niinku itelleen selväksi ja sitten että saa sitä tuntumaa että mitä on toiset ajatelleet tavallaan siinä käy niinku kyllä semmosta keskustelua niitten arviointitehtävääntajien kanssa silloin kun mietti ite ensiksi minä en nyt sillä tavalla osannu oottakaan niitä” Pirkko

Muilla arvioijilla vaihtelivat epävarmuuden ja varmuuden tunteet. Varmuus ja epävarmuus oman asteikon soveltamisen johdonmukaisuudesta ei siis ollut kategorisesti jompaa kumpaa, vaan molempia ja jotakin siltä väliltä. Yhdessä kohtaa haastattelua heistä saattoi tuntua, että linja piti melko hyvin, ja toisessa kohtaa taas omien kriteerien uskottavuus ja oikeellisuus sekä niiden keskinäinen painoarvo epäilyttivät. Esimerkiksi Martti oli sitä mieltä, että asteikko vakiintui kyllä päässä mutta ei käytännössä. Ehkä epävarmuus ja varmuus liittyivät osin ristiriitaan mielen ja toiminnan välillä: ”mut asteikko oli helppo palauttaa mieleen”, ”pystyn luomaan kriteerit ja et se kohtuullisesti toimii”, ”sit ku niitä oli tullu sata ni rupes mietti et kyllä tää onki aika hyvä” ja toisaalta taas: ”oma linja epäilyttää”, ”linja muuttui puolessa välissä”, ”välillä ei tuntunut luotettavalta”, ”skaala vaihteli matkan varrella.”

Kaiken kaikkiaan kymmenen arvioijan mielestä asteikkoa oli vaikea soveltaa ja kuuden mielestä se sujui ihan hyvin, vaikkakin mukana kulki epävarmuus omasta linjasta. Asteikon soveltamisen vaikeuteen tuntui ennen kaikkea vaikuttaneen kriteerittömyys ja yleinen epävarmuus siitä, tekeekö arviointia oikein perustein. Yksi arvioija totesikin, että toivottavasti hänen ei enää tarvitse arvioida ilman kriteereitä, koska hän tunsii olevansa ”niin tuuliajolla”.

Arvioijien kommenttien perusteella herää kysymys siitä, johtuiko epävarmuus tehtävien ja asteikon laadusta vai arvioijasta itsestään. Toisaalta epävarmuutta pidettiin asteikon avoimuuden, kriteerittömyyden, huonojen tehtävien ja samantasoisten kirjoittajien aiheuttamana, ja toisaalta Hogarthin (1980) mukaisesti epävarmuuden katsottiin johtuvan päättäjistä itsestään. Lienee selvää, että päättäjä on se, joka tuntee itsensä varmaksi tai epävarmaksi, mutta varmuuden määrään ja laatuun vaikuttanee kuitenkin myös sekä kaikkien muidenkin arviointivuorovaikutuksessa mukana olevien tekijöiden laatu että niiden keskinäinen vuorovaikutus.

Arvioijien tulkinnat epävarmuuden aiheuttajista vaihtelivat sen mukaan, tarkastelivatko he itseään arvioijana, tehtävien toimivuutta, asteikon laatua vai tekstiä itseään: ”aika monissa kohdissa oli semmonen olo että et et niinku että aa mikä tuomari mää oon ja et se semmonen arvioinin ikän ikuisuuskyymyksiä varsinkin sen takia ku ei ollu mitään muuta ku ne numerot” tai ”ehkä tarvinnu yhen lukemisen että mä oisin ollu ite jotenki niinku vakuuttuneempi niistä varma en koskaan mutta vakuuttuneempi.” Myös se, että epävarmuus ei liittynyt vain tähän arviointiin vaan arviointiin yleensä, puhuu puolestaan: ”ja sit jäi sellane niinku oikeistaa vähän sama ku niistä muistaki arviointihommista että pieni epävarmuus mistään ei voi olla varma ja sitten kuitenkin joku semmonen tuntuma.”

Epävarmuuden taustalla tuntuisi olevan ajatus tai oletus siitä, että ikään kuin suorituksille olisi löydettävissä yksi oikea, absoluuttinen tasoarvio. Toisaalta arvioijan oletetaan pyrkivän siihen, vaikka sitä ei ole olemassakaan. Ehkä vaatimus tulee ylitulkituksi, mikä voi stressata arvioijaa entisestään. Paine siitä, että on tulkinut suoritusta ’oikein’, soveltanut kriteereitään ideaalitason mukaisesti ja antanut muiden kanssa linjassa olevan tasoarvion, heijastuu jossakin määrin arvioijien kommentteissa. Paine on voinut voimistua entisestään, jos arvioija ei ole ollut varma omista eikä muiden arvioijien kriteereistä.

Näiltä osin on mielenkiintoista nähdä, miten toinen arviointikierros eroaa ensimmäisestä. Auttaako arvioijia yhtään se, että heillä on käytettävissään yhteiset tasokuvaukset? Kiinnittävätkö he silloin

enemmän vai vähemmän huomiota omaan toimintaansa? Suuntaako valmis asteikko arvioijien huomiota johonkin muuhun?

5.2.1.4 Millaiset asiat nousevat pintaan arviointiprosessissa?

”Karsee duuni, kusist hommaa”

VAI

”No se et oli hirvittävän mukavaa”?

Olen käsitellyt tähän mennessä haastatteluaineistoa lähinnä asteikon muodostamisen, sen laatimisen ja sen soveltamisen näkökulmasta. Tarkastelen seuraavaksi arvioijien arviointiprosessia luvussa 4.2.4 esittämäni kuvion mukaan (ks. kuvio 14), koska haastatteluissa esiin nousseet arviointiprosessia kuvaavat kommentit voidaan ryhmitellä melko luontevasti sen mukaisesti, kuten kuvioista 23 näkyy. Kuvioiden 14 ja 23 mukaisesti arvioijien kommentit liittyivät joko arvioijaan itseensä, asteikkoon, tekstiin, tehtävään tai kirjoittajaan. Vaikka erottelenkin kommentit tällä tavalla, en kuitenkaan tarkoita, että ne liittyisivät vain kirjoittajaan tai vain asteikkoon, vaan kyse on enemmänkin vuorovaikutuksesta arviointiprosessissa mukana olevien osapuolten (arvioija, teksti, tehtävät, asteikko) välillä.

Koska olen jo käsitellyt yksityiskohtaisemmin arvioijien määrittelemiä asteikkoja, nostan tässä yhteydessä esiin vain sellaisia tekijöitä, jotka liittyvät asteikon perusrunkoon eli esimerkiksi tasojen määrään, jota arvioija ei voinut muuttaa, vaikka olisi halunnutkin. Tarkastelin edellä myös arvioijien epävarmuutta ja arviointiin käytettävää aikaa taukoineen, joten pyrin välttämään toistoa. Sen sijaan käsialaan palaan vielä, mutta tässä yhteydessä en puutu siihen kriteerinä vaan tekijänä, joka arvioijien mukaan välitti erilaista tietoa kirjoittajan taustasta.

ongelman siten, että he kirjoittivat arviointilomakkeeseen suorituksen kohdalle: *ei voi arvioida*. Tosin, kuten alla olevista esimerkeistäkin näkyy, tällaisten suoritusten määrä näytti olevan vähäinen (ks. myös liite 9).

”no en siinä itte asiassa tuli just just semmone ongelma et mä kirjoitinkin sinne niihin lomakkeisiin että et tota mä toivoin semmosta nollatasoa tai se hyvin äkkiä tuli tarpeelliseksi että et koska joistakin ei saanu niinkun mitään selvää niitä kyllä oli harvoja mutta muutamia että ei ollu tolkkua niin aattelin että et tota se nolla olis semmonen ja sitten ykkösessä täytyy täytyy kuitenkin olla jollaki tavalla puheenaihe ja se että et mitä se siitä on niinkun sanomassa nin hahmotettavissa” Teija

”tuossa ois oikeastaan pitäny olla nollaki sen takia että siellähän oli semmosii vaa että oli raapustettu jotaki kirjaimia sinne tai puoli sanaa ni se on mun mielestä oikeasta nolla eikä ykkönenkään” Inkeri

Asteikko ei herättänyt pelkästään negatiivisia tunteita. Parin arvioijan mielestä sen avoimuus oli itse asiassa työtä helpottavaa: ”tää on periaatteessa mun mielestä aika silleen niinku selkee” ja ”ja sen paljon yksinkertaisempaa mä en tiedä kuin kehittäisi.” Yhden arvioijan mielestä asteikko tarjosi arvioijalle vapauden toimia katsomallaan tavalla, mikä oli todella vastakkaista sen kanssa, mitä on aikaisemmin tullut esille: ”ja sit ku tässä ei ollu ohjattu oli et oman tuntemuksen mukaan se tästä teki helpon koska ei ollu olemassa oikeeta ja väärää vastausta.” Arvioijan suhde asteikkoon saattoi myös muuttua prosessin aikana myönteisemmäksi: ”ensin vaikee mut sit se kääntyi mielenkiintoseks vaikee ja mielenkiintone ne ei must sulje pois toisiaan.”

Teksti suorituksena/tekstityyppi/kirjoittaja

Tehtävänannon ohjaamaan tekstityyppiin, tekstiin suorituksena ja kirjoittajaan liittyviä tekijöitä on oikeastaan aika keinotekoisia irrottaa toisistaan – vaikka olen tehnytkin niin kuviossa 23, koska lukija tekee päätelmiä kirjoittajasta tekstin perusteella ja tekstistä tehtävänannon perusteella sekä toisin päin että ristikkäin. Mielenkiintoista tässä on se, mitä asioita ja miten arvioijat nostivat teksteistä esille ja miten ne liittyivät kirjoittajaan, ehkä arvioijaan itseensäkin, ja miten tämä kaikki sitten edelleen vaikutti tekstistä syntyvään vaikutelmaan. Kyse oli siis monitahoisesta vuorovaikutussuhteesta, josta viime kädessä on vaikea sanoa, mikä oli seurausta mistäkin. Ehkä kausaalisuus tai kausaalittomuus ei olekaan tärkeätä, vaan tärkeätä on tunnistaa tekijöitä, jotka vaikuttivat arvioijien käsityksiin kirjoittamisesta taitona. Vaikka haastattelujen perusteella ei pysty suoranaisesti osoittamaan, miten nämä tekijät vaikuttivat arvioihin, ne kuitenkin osoittivat, millaisia erilaisia käsityksiä, odotuksia ja tuntemuksia arviointiprosessissa on läsnä.

Tekstien herättämiä kirjoittajalähtöisiä huomioita olivat muun muassa kirjoittajan tapa jäsentää tekstiä ja käsitellä tiettyjä sisältöjä, joita arvioijat kommentoivat sellaisenaan tai jotka edelleen johdattelivat arvioijia tekemään päätelmiä kirjoittajan kulttuurista ja äidinkielestä. Eri tehtävätyypeissä kulttuuriseikat tulivat esille eri tavalla: joissakin kulttuuriset vihjeet tuntuivat jäävän pääosin rivien väliin, kuten tarinatehtävässä, ja joissakin ne taas nousivat pintaan ärsytystä herättäen, kuten esittely- ja suruvalittelutehtävässä. Sinänsä neutraalilta tuntuvassa esittelytehtävässä, jossa pyydettiin tekemään lyhyt oman kulttuurin esittely, arvioijia ärsytti venäläisten kirjoittajien tapa käsitellä aihetta esimerkiksi siten, että monet venäläiset kirjoittivat luettelon oman kulttuurinsa tunnetuimmista kirjailijoista ja kulttuurirakennuksista.

”nääh kulttuuriset kulttuuriset seikat on kans hirmu niinku ongelmallisia tän arvioinnin kannalta ja sit se on hirveetä ku semmosista venäläisistä näkyy kaikki niitten koulujärjestelmät lä niin hirveesti läpi ja se et miten tämmöseen on tapana vastata ja mitä asioita omasta kulttuurista on tapana esitellä siis se oli kyl niinku se on hirveän ristiriitoja herättävä ja ja siin kyllä helposti kokee olevansa vähän epäoikeudenmukainen siinä” Heidi

”sit oli jotain semmosia että mitä ne nyt kirjotti et oli et oli vaan luetteltu niinku jotain et mitä se oli se tehtävä et mitä omasta kotimaastasi haluat no venäjällä on ää pushkin ja tolstoi ja dostojevski ja ni ei se niinku kerro oikeestaa mitään” Pirjo

Suruvalittelun kohdalla arvioijat nostivat esiin siihen liittyviä konventioita: mitä suomalaisessa kulttuurissa sanotaan ja miten suoraan (vrt. Kalliokoski 1996). Osa arvioijista oli sitä mieltä, että tässä tehtävässä ei ole sopivaa leikkiä huumorilla. Kritiikki koski sellaisia suruvalitteluviestejä, joissa esimerkiksi otettiin ilolla vastaan tuleva perintö tai katsottiin parhaaksi, että inhottavasta tyypestä päästiin eroon. Siitä huolimatta, että arvioijat itsekin mielsivät humoristiset aiheen käsittelytavat kulttuurieroista johtuviksi, tällaiset konventioiden vastaiset tekstit vaikuttivat emotionaalisella tasolla ja ärsyttivät epäsovpuudessaan.

Vastaavanlaisia kielellisiä tekijöitä, jotka liittyivät kirjoittajien kulttuuritaustaan, olivat rakenteiden hallinta ja puhekielen käyttö. Tämä näkyi esimerkiksi sillä tavalla, että yksi arvioija sanoi antaneensa vähän enemmän anteeksi kiinalaisille, joiden tekstissä r- ja l-kirjaimet menivät sekaisin. Jonkun toisen mielestä taas venäläiselle, joka ei hallinnut lainkaan menneitä aikamuotoja, ei voinut antaa missään nimessä taitotasoa 4, koska tällainen kirjoittaja ei ollut oivaltanut kielen perusasioita: ”eräänlaisen kontrastiivisen idean tajuamista.” Puhekielisyyys taas pohditutti paria arvioijaa siksi, että he eivät olleet oikein varmoja siitä, kuinka oletettavasti inkerinsuomalaisten kirjoittamiin puhekielisiin

teksteihin tuli suhtautua. Heidän mielestään nämä kirjoittajat varmasti puhuivat sujuvasti, vaikka eivät pystyneetkään tuottamaan yleiskielistä tekstiä, mikä vahvistaa Kalliokoskenkin (1996) esiin tuomaa kirjoitetun kielimuodon itsenäistä luonnetta.

Kaiken kaikkiaan arvioijat näyttivät toisaalta olevan hyvin tietoisia siitä, miten erilaiset kulttuuriseikat saattoivat vaikuttaa heidän arviointiinsa, ja toisaalta ne taas tuntuivat vaikuttavan ainakin jonkin verran tiedostamisesta huolimatta: "noh siinä täytyy myöntää et siinä kyllä varmaan vähän et joskus ku tuntu että ku niitä venäläisiä on niin hirveesti etku niitä tuli siinä peräkanaa ni sitä varmaan niinku nyppiinty välillä ja välillä sakotti ja välillä ei." Mielestäni kirjoittajan läsnäolo osoittaa hyvin sitä, miten teksti ei elä irrallista elämää kirjoittajastaan ja toisaalta myös sitä, miten tekstejä arvioidaan sosiaalisen yhteisön tuotoksina ja miten niitä peilataan sosiaalisiin ja kulttuurisiin normeihin (vrt. esim. Rubin 1988; Barton & Hamilton 1998).

Tekstin niukkuus sai arvioijat pohtimaan sitä, missä määrin tekstin pituus tai laajuus kertoi kirjoittajan rohkeudesta ottaa riskejä, kielitaidosta johtuvista puutteista, strategiasta selviytyä tekstin tuottamisesta, asennoitumisongelmasta, tehtävän motivoimattomuudesta vai ajan puutteesta. Nämä kaikki vaihtoehdot tulkinnot nousivat esiin pohdiskelussa, eikä niihin tuntunut olevan yhtä ainoata ratkaisua. Yleensä ottaen tekstin niukkuus aiheutti epävarmuutta arviosta, ja arviolle haettiin tukea mahdollisesta niukkuuden syystä. Joku palkitsi "riskinotosta" ja joku toinen taas arvosti "varman päälle kirjoittamista". Yksi tapa suhtautua tekstin pituuteen oli tällainen: "mitta mitta musta se riskinotto kyky joskus sitä mittaakin et jos sä pystyt paljon tuottamaan nopeesti niih silloin sulla on paremmat edellytykset tavallaan niinku osata et et lyhyt teksti on se joka mua yleensä mua kyllä ärsyttää että et se on semmonen joka saattaa niinku kiepauttaa sen alaspäin."

Tekstiin liittyvistä ominaisuuksista käsiala vaikutti ainakin kolmella tavalla: 1) se saattoi toimia kriteerinä, mitä olen käsitellyt jo aikaisemmin, 2) se liittyi tekstin luettavuuteen mutta vaikutti vain epäselvissä tapauksissa, 3) se viesti erilaisia taustatietoja kirjoittajasta, esimerkiksi äidinkielestä, koulutuksesta, huolellisuudesta ja motivoitumisesta tehtävään. Käsialan yhteydessä arvioijat pohdiskelivat myös sitä, miten käsiala korreloi tekstin taitotason kanssa, varsinkin jos käsiala oli epäselkeä ja tekstin taso heikohko: "näytti siltä että ei ei se auttais vaikka se käsiala olis selvempääki että et ne oli sitä ykköstä taikka sitä nollaa tasoa."

Vaikka arvioijat tekivät päätelmiä kirjoittajien taustoista käsialan perusteella, "tää on ihan selkee venäläisen käsiala", he tuntuivat kuitenkin olevan hyvin tietoisia käsialan vaikutuksesta ja siitä, että siihen ei pitäisi kiinnittää huomiota. Arvioija saattoikin kokea mielihyvää siitä,

että luki tekstin useampaan kertaan läpi eikä antanut käsialan vaikuttaa: "silloin mulle tuli tietysti semmone tyytyväine olo että vitsi että et olipa hyvä että luin ton huolellisesti koska se oli kuitenkin nii hyvä." Yleensä ottaen arvioijat itse totesivat, että he kyllä kiinnittivät käsialaan huomiota, mutta se ei vaikuttanut heidän arviointiinsa.

Käsialan vaikutus tuntuu olevan, niin tässä kuten muissakin tutkimuksissa, kahtalainen. Toisaalta se on – ainakin epäsuorasti – yksi kirjoittamisen arviointikriteeri (esim. Vaughan 1991; Milanovich ym. 1996) ja toisaalta sillä ei tunnu olevan vaikutusta arvioon. Esimerkiksi tutkimuksessa, jossa arvioijat arvioivat sekä koneella että käsin kirjoitettuja tekstejä, käsialalla ei ollut tilastollisesti merkitsevää vaikutusta kirjoittamisen arvioihin (Powers ym. 1994). On myös osoitettu, että arvioijat saattavat antaa alempia arvioita koneella kirjoitetuille teksteille, koska niissä korostuvat helpommin virheet ja koska niiltä myös odotetaan enemmän 'huolitellulta' näyttävän ulkonäkönsä vuoksi (Powers ym. 1994).

Edellisen lisäksi muutamat arvioijat kokivat arvioinnin vaikeaksi siksi, että he eivät tienneet tarpeeksi kirjoittajien taustoista, lähinnä siitä, miten kauan ja intensiivisesti kirjoittajat olivat opiskelleet suomea. On luonnollista, että ilman tätä tietoa arvioijat eivät mielestään tienneet, mitä kirjoittajien saattoi olettaa osaavan. Tällainen taitotasoarviointi, jossa kirjoittajat olivat tuntemattomia ja nimettömiä ja jossa asteikkokaan ei tarjonnut tukea, poikkesi varmasti opetukseen liittyvästä saavutus-tyyppisestä arvioinnista, jossa sekä arvioinnin tavoitteet että opiskelijat ovat tavallisesti arvioijalle tuttuja.

"et sit tuntu pahalta niinku silleen et ajatteli että jos mä tuntisin ton tyypin ni mä voisin antaa sille niinku yrityksestä ettei pystyny arvioimaan sitä yritystä koska ei tienny sitä oppilasta et mä koin sen arvostelun siis semmosena et voishan sitä tietysti ajatella niin että arvioi vaan abstraktii suoritusta ja silloha sil ei oo merkitystä mut mää se tuntu musta häijyltä et se tuntu ongelmalliselta" Martti

"no on tässä mun mielestä silleen vähän vaikeempi arvioida ku ei tiedä mitä voi olettaa ja odottaa se musta on se vaikeus ku tietää että on opettanu sen astevaihtelun ni sit voi sit voi niinku olettaa et he osaa se se niinku oli must se vaikeus ja sit tietysti se taso et miten niitä ku sit suhteuttaa" Merja

Arvioijat siis pahoittelivat sitä, että he eivät voineet palkita kirjoittajaa yrittämisestä eivätkä osaamisesta, oli kyse sitten asenteesta tai jonkin kielipiillisen seikan hallinnasta. Vastaavasti taas joidenkin arvioijien mielestä oli hyvä, että oppijat jäivät ikään kuin kasvottomiksi, koska arviointi oli tässä mielessä eettisempää, koska siihen eivät vaikuttaneet persoonalliset tekijät. Jonkun mielestä kasvottomuus oli kyllä oppijan

kannalta haitta, mutta taas eräänlaisen vertailtavuuden kannalta ”ihan hyvä asia”. Arvioijien päättelyt osuvat osin yksiin testauskirjallisuuden kanssa, jonka mukaan yksi yleisin tapa estää sädekehävaikutus on peittää kirjoittajien henkilöllisyys (esim. Linn & Gronlund 1995).

Arvioija

”tuntu et paska en jaks enää” – Miten tästä vireystilasta eteenpäin? Kuten edellä on jo käynyt ilmi, arvioijat kokivat arvioinnin raskaana, joskin he myös näkivät sen mielenkiintoisena prosessina. Raskaus liittyi nimenomaan prosessin aikana kumpuilevaan väsymykseen ja keskittymisvaikeuksiin. Väsyneenä arvioijan oli vaikeampi pitää kiinni asettamistaan kriteereistä tai olla tietoinen arviointinsa vaikuttavista subjektiivisista tekijöistä, kuten kirjoittajan taustasta tai joistakin tekstin ärsyttävistä piirteistä. Väsymys onkin yksi syy siihen, että arviointi ei ole yhdenmukaista (Davis ym. 1987). Monelle arvioijalle tämä oli myös ensimmäinen kerta, kun he arvioivat kerralla näin suurta määrää tekstejä. Suuri määrä toi arviointiin myös oman lisänsä, mikä näkyy muun muassa Tuulan kommentista.

”sitte rupes jo itseään epäilemään että onko onko nyt pystyny oleen kauheen johdonmukainen ku niitä teki niin paljon ja sitä tavallaan teki niitä aamulla ja illalla et väsyneenä ja virkeenä ja sit rupes niinku epäilemään et sit vähän senki mukaan niinku heittää sitte se arviointi et sitte olin itselleni iha vihane et kauheeta et miten voikin tälleen vaikuttaa että pitäis olla kauheen johdonmukainen ja tarkka nyt sitten ja mut kai se vaan on aina vähän subjektiivista tää homma et se se niinku toi tää suuri määrä tämmösen huomion mitä ehkä ei niin pienissä määrissä oo huomannu koska sen tekee niinku aina samassa mielentilassa” Tuula

”just näitä ongelmia mistä mä sanoin keskittymiskyky ja just tää epävarmuus siitä että onko ne omat sitte kuitenkin kohallaan tai pysyykö niissä omista asettamisissaan... tota niin se oli se että sitte ku tuntu että sekoaa ite siinä hommassa et ei enää tee vaan pitää niinku tauon siinä sen jälkeen miettii uuestaan katoen niitä mitä aikaisemmin arvioin jos ne olis semmosia vähän niinku niitä vertaamalla sitte onko nää nyt sisäkkäisiä” Pia

”kyllä mulle tuli pikkusen semmonen olo et et kyllä se et on se kyllä eettisesti aika paha paikka että että vähän paha mennä lupaamaan sellasta sellasta loppuun asti sellasta niinku objektiivisuutta ja sellasta aina edes hyvää tahtoa koska siis ne omat fiiliksetki vaikuttaa hirve se on muuuten kans yks asia joka on siinä tehtävien ulkopuolella siis semmonen niinku oma uupumusprosessi siis sillai et ku välillä niinku kahelta yöllä korjaat ja välillä johonkin muuhun aikaan ni se niinku se oma vireystila kyllä vaikuttaa ihan selvästi siihen miten niitä kokee” Heidi

Juuri tekstien suuren määrän vuoksi arvioija ei voinut jäädä yhden tekstin kohdalle kovin pitkäksi aikaa. Kuten arvioijat itsekin totesivat, oli pidettävä yllä tietynlaista vauhtia. Myöskään taaksepäin palaaminen ei ollut kovin mielekästä, koska se olisi vienyt liikaa aikaa ja koska tekstejä ei olisi voinut kuitenkaan vertailla kovin yksityiskohtaisesti niiden suuren määrän vuoksi. Tekstien suuri määrä toi mukanaan myös rutinoitumisen ja tylsyyden tunteet. Jossakin vaiheessa työ alkoi uuvuttaa, ja tekstit alkoivat näyttää toistensa kopioilta. Tämän seurauksena, kuten arvioijat itsekin totesivat, ote alkoi herpaantua ja oman työn kontrolli pettää.

Päätöksentekoprosessissa aikapaine ja työn määrä vaikuttavat osaltaan siihen, millaisia strategioita henkilö valitsee päätöksenteon eri vaiheissa (vrt. Huber 1989). Tässä tutkimuksessa tekstien määrä ja käytettävissä oleva aika näyttivät vaikuttavan arvioijien kommenttien mukaan siihen, miten tekstien lukemisprosessi eteni. Käytännössä tämä tarkoitti sitä, oliko arvioijalla esimerkiksi aikaa ensin silmäillä tekstit ja lukea ne vasta sitten huolellisesti läpi tai lukea ne tarvittaessa toiseen tai kolmanteen kertaan vai perustuiko päätös yhteen lukemiskertaan tai jopa vain tekstin silmäilemiseen. Monet arvioijat totesivat myös lukeneensa tekstejä liukuhinnamaisesti, jos he tekivät työtä väsyneenä tai jos tekstit olivat tylsiä tai samantyyliisiä.

Toisaalta taas esimerkiksi epäselkeän käsialan tai joidenkin rajatapausten kohdalla samat arvioijat sanoivat lukeneensa tekstit tai jonkin tekstin kohdan useampaan kertaan läpi. Tämän perusteella näyttäisi siltä, että arvioijat eivät kommenttiensa mukaan lukeneet kaikkia tekstejä systemaattisesti samalla tavalla, vaan vireystila ja aika vaikuttivat osaltaan lukemisstrategioihin. Tekstien suuri määrä ei ole kuitenkaan pelkästään haitta, vaan se voi myös rutinoida lukijaa positiivisesti (vrt. Cooper 1977). Rutinoitumisesta saattoi tässä yhteydessä seurata sekä tietynlaista varmuutta että mahdollisuutta tarkastella tekstejä lintu-perspektiivistä: ”siinä mielessä palkitsevaa että aina hirveen opettavaista ja tota sellasta niinku aika hyvä kattoo joskus semmone isompi määrä kerralla kus siinä näkee jotenki sitä semmosta vaihteluväliä.”

Kuten aikaisemmin on jo käynyt ilmi, lähes kaikki arvioijat viittasivat kommentteissaan siihen, miten he tunnistivat arviointiinsa vaikuttavia tekijöitä ja miten he yrittivät välttää negatiivisia tekijöitä. Tällaisia välttämiskeinoja olivat esimerkiksi tauon pitäminen vireystilan laskiessa tai epäselkeällä käsialalla kirjoitetun tekstin lukeminen useampaan kertaan. Arvioijien mukaan eroja oli siinä, miten hyvin he sitten onnistuivat itsekontrollissa. Joku suhtautui hyvinkin epäluuloisesti koko arviointiin, ja joku taas oli sitä mieltä, että kontrollointi onnistuu. Tosin näkemykset vaikutusmahdollisuuksista epätoivoon saattoivat vaihdella yhdellä ja samalla arvioijalla haastattelun eri kohdissa: ”se

vaihtelee onnesta epätoivoon kaikki siltä väliltä koko tunteiden kirjo käydään läpi.”

”siis tämmönen käytännön tekijä että ku torstaina oli kiire moneen muuhun asiaan yhtä aikaa sit mä aattelin et nyt mä vaan arvioin näitä nopeesti ni sit ensinnäki niinku kerroin eilen se koko si se yks sivu täh mitä mä olin niinku väentäny tällai ja sitte sit tavallaan siihen ei semmosella semmosella nopeudella ni ni mun mielestä siinä ei oo semmos siihen ei tuu semmost ei tuu semmosest otetta niihin teksteihin kun niitten tekstien pitää antaa puhua ja siinä mielessä myös mie jouduin ne monta kertaa lukee koska ja ehkä osittain se että että ei oo kuitenkaan niin rutinoitu niihin” Kaisu

”noh oli ne periaattees mut ku ei sit se käytäntö menny sen mukaan et se soveltamine oli mut kyllä mä se mä sen ihan tietosesti niinku tein et mä sit mietin et mikä on mikäki kyl se silleen oli mut kyl se ei se auttanu siinä käytännösä et siin oli sit paljon semmosii arviointi vaikeuksii” Martti

Väsymiseen ja negatiivisiin tunnetiloihin liittyvät myös ärtyamisen tunteet, joita erilaiset teksteihin tai niiden kirjoittajien taustoihin liittyvät tekijät saattoivat aiheuttaa. Kuten seuraavista esimerkeistä näkyy, arvioijien johdonmukaisuuteen ja yhdenmukaisuuteen vaikuttavat tekijät eivät ole kategorisoitavissa erillisiksi luokiksi, vaan ne ovat tavallaan sisäkkäisiä tai jopa keskenään vastavuoroisessa kausaalisuhteessa. Tarkoitin tällä sitä, että esimerkiksi kirjoittajan tapa käsitellä aihetta voi saada arvioijan ärsyntymään, koska hän on väsynyt, minkä takia arvioija taas saattaa antaa alemman tasoarvion (vrt. Davis ym. 1987; Linn & Gronlund 1995).

”se että mä nään et tää on venäläinen tää on arabipoika tää on somalipoika tää on tää on noh ruottalainen ihan selvästi ni et mä yritin niinku sitä ensin niinku etten ota sitä huomioon tämmösiä omia noh sympatioita antipatioita” Päivikki

”noh siinä täytyy myöntää et siinä kyllä varmaan vähän et joskus ku tuntu että ku niitä venäläisiä niin hirveesti etku niitä tuli siinä peräkanaa ni sitä varmaan niinku nyppiinty välillä ja välillä sakotti ja välillä ei että mutta oikeestaan mulle ei kyllä oo siihen mitään hirveen hyvää niinku semmosta ratkasua et se ois kyllä oikeestaa mun mielestä jonkun kaa jutella että mitä muut tekee mut että että minusta tuntu et mä siinä saaton kyllä käydä jonkun verran tuulella” Heidi

Arvioijat pohtivat omaa rooliaan myös ymmärtäjinä siinä mielessä, miten paljon tekstien arviointi oli itse asiassa kiinni siitä, että he opettajina, alan ammattilaisina, ymmärsivät tekstejä paremmin kuin tavalliset suomalaiset lukijat, koska he pystyivät kielitietoutensa ja kokemuksensa perusteella paremmin päättelemään merkityksiä. Toisin

sanoen arvioija voi itse tietoisesti tai tiedostamattaan yliymmärtää tai aliyymmärtää muun muassa kirjoittajan kulttuuritaustan, väsymyksen, virheiden tai tekstin sisällön aiheuttamien syiden vuoksi. Seuraavat esimerkit heijastelevat sitä, miten arvioija itse kohtaa tämän ongelman.

”että haluanko mä ymmärtää enkö mä halua ymmärtää ymmärtääkö mun äiti vai oonks mä nyt mä vaan tiijän mitä täs on nyt ihan selvästi tää heikko vartalo tai mut sen takia se näyttää nyt vähän toiselta verbiltä että et et tää kuinka naiiviksi itsensä laittaa nis se on musta yks keskeinen ongelma siinä arvioinnissa et haluanko mä ymmärtää enkö mä halua” Päivikki

”sitten mä rupesin aina miettimään että että onko tää nyt niin sitä että tätä ei ymmärtäisi normaali ihminen että mä nyt vaan satun ymmärtämään ku mä luen tätä niinku silleen että mä täytän sen suoraan sen tekstin sinne pisteet ja näin että kai sitä lukee jotenki niin eri tavalla että usein lukee sillä tavalla että huomaa että joku on kirjoitettu väärin se on niinku ihan pikku virhe vaikka yks kirjain nin sen lukee ihan oikein ja sitä rupes aina vaan miettimään että tekeekö n niinku sitä että korjaa niinku suoraan ne virheet sinne niin että ymmärtää paremmin” Eeva

Päivikin esimerkki on mielenkiintoinen siinä mielessä, että se nostaa esiin sen, miten paljon arvioijalla on valtaa tässä suhteessa ja miten sen käyttö riippuu siitä, miten tietoinen arvioija on toiminnastaan ja miten kykenevä ja/tai motivoitunut hän on kontrolloimaan omaa toimintaansa. Tähän sitaattiin ehkä tiivistyy kaikkein olennaisin asia arvioijan työssä. Järjestäytyneessä arviointitoiminnassa arvioijaa tuetaan muun muassa arviointikriteereillä, maamerkeillä ja koulutuksella. Tässä yhteydessä arvioijalla ei ollut mitään näistä käytettävissään. Tosin en usko, että edellä esiin tulleet piirteet ja inhimilliset tekijät katoavat tai lakkaavat olemasta tukikeinojen myötä, joskin niiden vaikutus voi kylläkin heiketä. Ehkä nämä haastattelut osoittavat sen, että arviointiprosessista keskusteleminen olisi hyödyllistä, jotta arvioinnin kaikkinaiset tuntemukset saisivat kohdata päivänvalon.

Arviointia koskeissa arkikeskusteluissa on paljonkin puhuttu esimerkiksi pәрstäkerroniarvioinnista eli lähinnä sädekehävaikutuksesta, joka liittyy varsinkin kouluarviointiin. Entäpä institutionaalistuneessa testi-arvioinnissa, esimerkiksi ylioppilastutkinnoissa, millainen avoimuus kuuluu siihen liittyvään keskusteluun? Ja mitä avoimuudella voidaan saavuttaa? Jos ei muuta, niin ehkä ainakin eettisempiä mielikuvia arvioinnin luonteesta.

”et ku siellä on muutamissa teksteissä esimerkiks semmosia hirveen hyvin lauseita et se se on niinku ihan selvästi niinku se taiv et se siel on just tätä luovuutta niinku esimerkiks sanojen käytössä tai ilmasussa että että se sais niinku liian paljon painoarvoa koska sitte siinä ku ite turtuu ni sitä tykkää sit semmosta vitsit et täällä on joku joka on vähä vähän yrittäny revitellä et

siis ettei se sais niinku sen takia paljon parempaa et tämmösiä ongelmia”
Heidi

5.2.1.5 Ensimmäisen haastattelun satoa

Arvioijien itsensä laatimat kriteerit, sellaisena kuin ne kullakin arvioijilla olivat, toimivat ikään kuin normeina siinä mielessä, että jos ja kun arvioijat poikkesivat niistä, poikkeaminen aiheutti epätasapainoa. Tämä ei kuitenkaan tarkoita, että kriteerit olisivat olleet yksiselitteisen selkeitä ja toiminnallisia ja että niitä olisi sovellettu systemaattisesti. Kyse on lähinnä siitä, että mielikuvamaisilla, holistisesta kirjoittamiskäsityksestä johdetuilla kriteereillä on myös normimainen luonne. Arvioijat pitivät kriteereitään eräänlaisina toiminnan maamerkkeinä, joita he pyrkivät seuraamaan. Arvioijat olivat itse varsin tietoisia siitä, että he eivät aina pystyneet toimimaan mielessään olevien kriteerien mukaisesti ja että kriteeritkin saattoivat vaihdella jonkin verran eri aikoina tai eri kriteerit olivat etualalla eri tekstien kohdalla.



Ristiriitaa ajattelun, mielikuvien ja toiminnan välillä.

Se, että ajattelun tasolla olevien kriteerien ja konkreettisten arvioiden suhde ei ollut kovin suora eikä myöskään läpinäkyvä, sai arvioijat pohtimaan sitä, mihin heidän arviopäätöksensä perustuivat. Haastattelujen perusteella ei voi sanoa, mikä vaikutti suoranaisesti yksittäiseen arvioon. Sen sijaan arvioijien kommenttien perusteella voi tehdä päätelmiä tekijöistä, jotka vaikuttivat teksteistä syntyviin vaikutelmiin arviointiprosessin aikana. Tällaisia tekijöitä olivat muun muassa arviointiin käytettävissä oleva aika, tekstien taso (tässä lähinnä samantasoisuus) ja tekstityyppi (vrt. esim. ostoslista ja suruvalittelu) ja kriteerien painottaminen.

Arviointiprosessia tuntui hankaloittavan päätöksenteon kulku; se saattoi nimittäin edetä vähän eri tavalla kuin arvioija ehkä odotti. Esimerkiksi käytettävissä oleva aika pakotti tekemään kompromisseja tai arviosta päättäminen painottamaan sellaista kriteeriä tai kriteereitä, jotka eivät johtaneet arvioijaa tyydyttävään tasoarvioon. Arvioijien itsensä määrittelemien kriteerien käyttö aiheutti myös epävarmuutta muun muassa siitä, miten vertailtavia omat arviot olivat suhteessa muiden arvioijien antamiin arvioihin, miten oikeudenmukaisia omat arviot olivat yleensä ja miten yleistettävää tai neutraalia arvioijan oma ymmärtäminen oli lukijana.



Arviointiprosessiin kuuluu monenlaista epävarmuutta.

Arvioijien epävarmuuteen, epäilykseen omien kriteerien oikeellisuudesta ja omasta kyvystä toimia lukijana, tuntui liittyvän käsitys siitä, että arvioinnin pitäisi olla totta. Totuus tuntui liittyvän siihen, että arvioija arvioi oikein, mikä puolestaan tuntui edellyttävän, että kullekin tekstille oli olemassa yksi oikea arvio. Yksi ja oikea arvio liittyy taas absoluuttisuuteen eli siihen, että jokaisen kirjoittajan kohdalla voidaan määritellä hänen absoluuttinen taitotonsa. Oikeellisuuden ja totuuden läsnäolo ei ole kuitenkaan kovin suoraa tai se ei ole koko ajan etualalla, vaan heijastuu ennemminkin epävarmuuteen liittyvissä kommentteissa. En kuitenkaan puhu tässä nyt totuudesta ja oikeellisuudesta valheellisuuden ja mielivaltaisuuden vastakohtana, vaan päätöksentekoon liittyvänä inhimillisenä epävarmuutena.



Miten totta ja oikein arviointi voi olla tai sen pitäisi olla?

Arvioijien kaikkalainen pohdinta osoitti, että arvioijat liikkuivat päätöksenteossaan molemmilla Karlssonin (1989) esittämällä tasoilla. Toinen näistä tasoista viittaa päätöksentekijän toimintaan konkreettisessa tilanteessa ja toinen hänen itsereflektioonsa, johon liittyvät erilaisten strategioiden käyttö ja muun muassa arvojen, uskomusten, toiveiden ja antipatioiden mukanaolo. Toiminnan taso näyttää osoittavan numeroiden valossa, että arvioijat arvioivat omilla asteikoillaankin suorituksia yllättävän yhdenmukaisesti (ks. luku 5.1). Haastattelujen perusteella yksittäisistä kriteereistä arvioijat näyttivät nojaavan kaikkein yhdenmukaisesti kielen rakenteiden hallintaan, tai ainakin se oli kriteeri, joka tuntui olevan monella arvioijalla etualalla sekä asteikkokuvausten että esimerkiksi rajatpauskriteerien kohdalla.



Onko arvioijien yhdenmukaisuuden takana yksi muita voimakkaampi, systemaattisimmin etualalla oleva kriteeri eli kielen rakenteet?

Päätöksenteon toinen taso eli itsereflektointi – ainakin jälkikäteen – osoitti, että arvioijat olivat tietoisia siitä, millaiset tekijät heihin vaikuttivat arviointiprosessin aikana. He myös sanoivat toisaalta tunnistaneensa subjektiivisten tekijöiden läsnäolon ja toisaalta pohdiskelleensa sitä, missä määrin he pystyivät kontrolloimaan niiden vaikutusta. Kaiken kaikkiaan asioiden arvottamisessa tuntuivat olleen läsnä järki, tahto ja tunteet (Niiniluoto 1994). Se, miten voimakkaana esimerkiksi tunteet tulivat

esiin, tuntui riippuvan ennen kaikkea arvioijan vireystilasta, joka herpaantuessaan vaikutti muun muassa tekstin laadusta tai kirjoittajasta syntyneeseen vaikutelmaan.



Arvioinnissa ovat läsnä järki, tahto ja tunteet.

5.2.2 Toinen haastattelu: määritellyn asteikon soveltaminen

5.2.2.1 Yleisvaikutelmia arvioinnista

”arvioinnista niinkun niin helposti puhutaan vaan siitä että mi et mitkä numerosarja se on ja se just must siinä ei oo mitään merkistystä onko se ykkösestä yheksään vai ykkösestä kolmoseen kun ne rajat on selkeet se on musta niinku se isompi ongelma aina see” Raija

Toinen haastattelu alkoi, kuten ensimmäinenkin, yleisvaikutelmasta, päällimmäisistä tuntemuksista puhuen. Tälläkin kertaa ensimmäisen kysymyksen tarkoituksena oli toimia tilanteen lämmittäjänä. Haastattelut lähtivät kuitenkin saman tien täydellä vauhdilla käyntiin, koska arvioijat tuntuivat paneutuvan itse aiheeseen oma-aloitteisesti ilman lisäkysymyksiä. Haastattelun aloituksista heräsi sellainen vaikutelma, että arvioijat olivat miettineet asioita ennen haastattelua ja he todellakin tuntuivat ilmaisevan heti sen, mitä heillä oli mielessä kaikkein päällimmäisenä.

”saanko nyt heti verrata (joo) joo tekee mieli koska tota oikestaan kun mä sain sitten tämän arviointiasteikon ni mä olin vähän järkyttynyt siitä koska se oli mun mielestä todella erilainen kuin mitä sitä oli sillon ekalla kerralla sitten arvioinut et mulla niinku yleisesti ottaen ne arviot kaikki ni varmaan laski tasot hyvin runsaasti ja niinku tää oli aika tiukka” Tuula

Yleisvaikutelmat toisesta arviointikierrroksesta voisi jakaa melko tasaisesti kolmenlaisiin kokemuksiin: 1) nyt helpompaa, 2) helpompaa ja odotuksenvastaista 3) odotuksenvastaista. Lähtökohtana jaottelussa on ensimmäisen ja toisen arviointikierroksen vertailu ja se, millaisin odotuksin ja toivein arvioijat suhtautuivat toiseen arviointikierrokseen ensimmäisen jälkeen. Suurin erohan kierrosten välillä oli siinä, että tällä toisella kierroksella arvioijilla oli käytettävissään asteikko tasokuvauksineen.

1) Nyt helpompaa

Osa arvioijista koki tämän arviointikierroksen helpompana siinä mielessä, että arvioija tiesi tällä kertaa, mistä on kysymys: ”mä en tuntenu itseäni niin aseettomaks tämän tehtävän eessä.” Tuttuuden tunne liittyi koko prosessiin, arvioitavaan materiaaliin ja ajankäyttöön. Tuttuuden lisäksi näihin – jossakin mielessä positiivisiin ensivaikutelmiin – viittasivat myös helpottuneet kommentit siitä, että arviointi tuntui ”helpommalta”, ”ohjattummalta”, ”luontevammalta” ja ”järkevämmältä” valmiiden arviointikriteerien vuoksi. Yleisesti ottaen arvioija tunsu olevansa Kaisua lainatakseen ”varmemmilla vesillä”.

”no tää oli vähän helpompi sen takia että mun mielestä oli vähän niinku ehkä oppinu että ties mitä on niinku tulossa ni oli helpompi tehdä” Inkeri

”noh siis niinku siitä arvostelusta yleensä (joo) noh oli vähän helpompi ku mut mut niinku mä tohon kirjoitin et se on niinku et tuntu helpommalta mut et mä en tiedä jo kuinka paljon se johtu niistä et oli noi kriteerit olemassa ja kuinka paljon sit siitä tavallaan ku jo ties et mitä se on ja oli tavallaan niinku kokonaiskuva siitä aineistosta et niinku osas vähän silleen ajatella et mikä on hyvä ja mikä huono mut kyl se niinku meni vähän sujavammin oli niinku sikäli parempi” Martti

”asteikko oli musta selkeesti hyödyllinen ja helpottava asia” Heidi

2) Helpompaa ja odotuksenvastaista

Toinen ryhmä arvioijista suhtautui tähän arviointikierrokseen lähinnä kahtalaisesti: ”ehkä sillä tavalla jos oli hitaampi niin niin helpompiki arvioida tämän paperin kanssa.” Toisaalta arviointi oli heidän mielestään ”hitaampaa”, ”vaikeampaa” tai ”tylsemppää”, mutta toisaalta käytössä ollut arviointiasteikko ohjasi heitä esimerkiksi kiinnittämään huomiota sellaisiin asioihin, joihin he eivät omien kommenttiansa mukaan tulleet kiinnittäneeksi huomiota ensimmäisen arviointikierroksen aikana. Lisäksi esimerkiksi ajanvieyvyyden vastaparina tuotiin esille asteikon apu yleisemmin tai tuttuuden tunne siitä, mitä oltiin tekemässä.

”hyvin epämääräinen mä en tykänny yhtään tästä arvio arviointiasteikosta koska mun mielestä se ei erotellu melkein kaikki tuli sit siihen keskiryhmään mut tota toisaalta siinä tuli semmosiaki ajatuksia että ettei tekis sitä arvioo arvion vuoksi ettei kiinnittä huomioo liikaa semmosiin yksittäisiin asioihin toisaalta on tässäki puolensa” Sanna

”em mie osaa sanoo ei ei mitään niinku kovin erityistä ku ties nyt mitä mistä on kysymys ties mistä on kysymys ties suurinpiirtein kui kauan siihen menee aikaa ja ja sit mie odotin mielenkiin kiinnolla näit näit arviointiasteikkoo mutta tota mut pettymys koska siis mie petyin tähän” Tiina

3) *Odotuksenvastaista*

Kolmas ensivaikutelmien kategoria kuvastaa arvioijien suurimmaksi osaksi negatiivista suhtautumista arviointikierrrokseen: "no ei tää ollu mun mielestä yhtään sen helpompaa mitä oli oikestaa edelliselläki kerralla" Nämä arvioijat kuvasivat yleisiä tuntemuksiaan adjektiiveilla "vaikeaa", "raskasta" tai "epäselvää". Yleisvaikutelmista käy ilmi, että tämä arviointikierrros oli jollakin tavalla pettymys, joka johtui pääasiassa asteikosta. Odotukset siitä, että asteikko tekisi arvioinnista helpompaa, sujuvampaa ja selkeämpää, eivät täyttyneet.

"että hohhoijaa [nauria] tää oli tää toinen kierros vähän raskaampi et jäi vähän semmonen olo kun tää oli loppu se oli niinku ihan siis tälleen näin mutta muuten mulla oli yhtä su suurinpiirtein ihan samanlainen olo kun edelliselläki kerralla että että se mikä oli epäselvää silloinki oli mulle vieläki vähän epäselvää kaikesta huolimatta" Pia

"tota tuota tuota öö mä odotin että tää ois ollu paljon helpompaa kun sillon ekalla kerralla jollonka oli niinku vaan omassa päässä se se tota mmm ne tasot tai se että mitä ne pitää sisällään mutta ei se ollukkaan paljon helpompaa" Eeva

Jos tämän kierroksen ensivaikutelmia vertaa ensimmäisen kierroksen vaikutelmiin, näyttävät kokemukset samantyyppisiltä: arvioijat ovat tyytyväisiä, tyytymättömiä ja jotakin siltä väliltä. Myös kommentteissa, jotka koskevat jaksamista, arvioinnin vaikeutta, epävarmuuden tunnetta, ajankäyttöä, työmäärää ja asteikon toimivuutta, on paljon yhteistä. Näkökulma on vain vähän erilainen, koska arviointi-asteikko oli erilainen. Myös arvioijien suhde asteikkoon oli vähemmän henkilökohtainen toisella arviointikierröksellä.

Kommentit eroavat ensimmäisestä haastattelusta siinä, että ne vaikuttavat jonkin verran edelliskertaa avoimemmilta. Tämä voi johtua siitä, että arvioijat tiesivät, minkä luonteinen haastattelutilanne on ja mitä odottaa siltä. He saattoivat myös tuntea olonsa muuten vain varmemmaksi, tai ehkä heidän suhteensa määriteltyyn asteikkoon ei ollut yhtä henkilökohtainen, minkä vuoksi asteikkovaikutelmista oli helppo lähteä liikkeelle. Pelkkien ensivaikutelmien perusteella syiden syvempi pohdiskeleminen ei liene kovin tarkoituksenmukaista, joten siirryn arvioijien tämänkertaisiin arviointikokemuksiin sen kautta, miten he kommentoivat ja analysoivat määriteltyä asteikkoa.

5.2.2.2 Asteikosta kuultua

Tällä arviointikierroksella arvioinnin aloittaminen kävi yleensä ottaen vähän helpommin, koska asteikossa olivat tasokuvaukset mukana. Arvioijat aloittivatkin työn lukemalla ensin asteikon ja ryhtyivät vasta sitten soveltamaan sitä teksteihin. Arvioijien välillä oli kuitenkin eroa siinä, miten tarkkaan he tutustuivat asteikkoon. Suurin osa kertoi lukeneensa asteikon läpi. Muutama arvioija etsi asteikosta etukäteen, esimerkiksi alleviivaamalla, ydinsanat tai ydinkohdat kultakin tasolta. Muutama arvioija taas teki asteikon lisäksi muistiinpanoja, omia pikkumerkintöjä tai sivuhuomautuksia ”tärkeemmistä kohista mihin niinku kiinnittää huomiota”.

Aika pian alun jälkeen asteikko herätti monenlaisia tuntemuksia siitä, miten hyvin tai vaivattomasti sitä saattoi soveltaa teksteihin. Kuten jo yleisvaikutelmista kävi ilmi, asteikon soveltaminen ei sujunut kaikilta vaivatta. Kymmenen arvioijan mielestä asteikon soveltaminen oli vaikeaa ja yhden mielestä lähes mahdotonta. Hänen mukaansa toisella arviointikierroksella oli vain yksi todellinen ongelma: ”tää oli se tää oli se ongelma tää as asteikko siinä.”

Kolmen arvioijan mielestä asteikon soveltaminen ei ollut suoranaisesti helppoa mutta se ei ollut myöskään vaikeaa; se oli aikaa vievää. Parin arvioijan mielestä asteikko taas helpotti arviointia, nopeutti sitä ja teki siitä tässä mielessä sujuvampaa edelliseen kierrokseen verrattuna: ”asteikko kuitenkin antaa semmosta niinkun tai antaa paljonkin apua em mä sitä niinkun huonoksi sanoo missään muodossa en ollenkaan että siihen pysty sitten tukeutumaan tosiaan kun alko tuntua niinku että että annanko mä nyt pelkkiä kolmosia.” Yhteistä kaikille arvioijille oli se, että he esittivät kritiikkiä sekä asteikon sisällöstä että sen soveltamisesta.

Asteikkoa kritisoitiin liian ”ympäripyöreäksi”, ”ylimalkaiseksi” tai ”summittaiseksi”. Puolet arvioijista piti asteikkoa kokonaan summittaisena, ja kuuden arvioijan mielestä se oli osin summittainen. Vain yksi arvioija, Teija, oli asteikkoon pääosin tyytyväinen. Hänen mielestään asteikko oli ”aika helppokäyttöinen ja aika toimiva ja melko yksiselitteinen ja aika tarkkakii loppujen lopuksi”. Asteikkoon tyytyväinen arvioija oli kokenut äidinkielisten lukioikäisten kirjoittajien tekstien arvioija, ja hän näkikin monia yhtäläisyyksiä tämän asteikon ja opetuksessa ja ylioppilastutkinnossa käytetyn asteikon välillä. Myös Tiinalle tämä asteikko toi mieleen ylioppilastutkinnon kriteerit ja äidinkielisten kirjoittamisen asteikon, kuten seuraavasta esimerkistä näkyy.

”tämmöstä kriteeristöä vois käyttää niinku suomalaisen kirjottamassa kirjoittamien te tekstien arviointiin et oikeestaan tää ei niinku ei täs ei mittaa sit kielioppipainotteisuus ku mie olisin halunut jotain sellasta sellasta ni nih minkälaista sanastoa osaa käyttää ja minkälaisia rakenteita osaa käyttää eihän täs oo mitään sitä täs vaan kerrotaan et minkälaisia minkätyyppisiä tekstiä osaa kirjottaa sillähän minuu niinku sit rupes raivostuttaa oikeestaa” Tiina

Asteikon osin summittaiseksi kokeneet arvioijat tunsivat, että he pystyivät arvioimaan suoritukset jokseenkin hyvin tällä asteikolla. Heidän kritiikkinsä oli hyvin samansuuntaista niiden arvioijien kanssa, jotka kokivat asteikon pääasiassa ylimalkaiseksi tai epäselkeäksi. Erona oli tosin se, että asteikon osittain summittaiseksi kokeneiden mielestä toinen arviointikierros tuntui helpommalta nimenomaan asteikon tarjoaman avun vuoksi tai siksi, että asteikko toimi joiltakin osin hyvin. Tätä näkökulmaa luonnehtii tietynlainen suhteellisuus siinä, miten valmiita tai yksityiskohtaisia arviointiasteikot voivat yleensäkkään olla: ”jotenkin sitä asteikkoo tai jotenkin jos ne ois niinku ihan enkä mä nyt tiijä pystyskö niit sitten kuitenkin ni niin tarkasti sitten.”

Asteikkoa summittaisena ja epäselkeänä pitävien arvioijien mielestä asteikkoa oli todella vaikea soveltaa sellaisenaan. Heidän mielestään asteikkoa tuli ainakin jossakin määrin muuttaa. Myös asteikkoon tyytyväinen arvioija ehdotti asteikkoon joitakin lisäyksiä. Muutokset ja lisäykset nähtiin siis toisaalta välttämättöminä ja toisaalta asteikon sovellettavuutta parantavina. Välttämättömyyden puolesta puhuu se, että pystyäkseen soveltamaan asteikkoa arvioija joutui tekemään omia tulkintojaan ympäriryöistä kuvauksista, tekemään asteikkoon lisäyksiä tai luomaan jopa omia kriteereitään – siitäkin huolimatta, että hän mielestään nojasi arviointiprosessin aikana ennen kaikkea asteikon kuvauksiin.

”noh kyl mä pyrin aina nojaamaan asteikkoon mutta ei ne kyllä ihan suoraan löytyny asteikosta yleensä et tota et kyllä tänne esimes niinku koska tänne asteikkoon joutu tavallaan ite niinku omassa mielessään kirjottamaan lisää määritelmiä” Heidi

5.2.2.3 Asteikon soveltaminen: Millaisia tulkintoja asteikon tasokuvaukset herättivät?

Tasot 1 ja 2: ”must ne oli hyvät nää ykkönen ja kakkonen”

Seuraavassa raportoin siitä, miten arvioijat tulkitsivat ja kommentoivat asteikon eri tasoja ja niiden sovellettavuutta. Käyn arvioijien tulkintoja läpi taso tasolta siten, että esittelen aina aluksi määritellyn asteikon kuvauksen kyseisestä tasosta ja sitten siihen liittyviä arvioijien kommentteja ja tulkintoja. Käsittelen kaksi tasoa kerrallaan siten, että yhdistän tasot 1 ja 2, 3 ja 4, 5 ja 6. Myös Euroopan neuvoston asteikon kuusi tasoa jaetaan yleensä kahden tason taitoryhmiin siten, että kaksi ensimmäistä tasoa muodostavat A-tason, välitasot 3 ja 4 B-tason ja ylimmät tasot 5 ja 6 C-tason. Samantyyppistä tasojen ryhmittämistä esiintyi myös tämän työn arvioijien kommentteissa.

TASO 1	Osaa kirjoittaa hyvin lyhyitä ja yksinkertaisia tekstejä, esimerkiksi postikortin ja/tai täyttää lomakkeeseen itseään koskevat perustiedot. Teksti koostuu toisistaan irrallisista yksinkertaisista, ehkä ulkoa opituista sanaryhmistä, lauseista ja tavallisimmista fraaseista.
---------------	--

Puolet arvioijista koki ykköstason soveltamisen sujuvana tai jopa helppona: ”on helppo huomata et mikä on ykköne.” Usein ykköstason suorituksen laatua kommentointiin ilmauksin, että ”jos nyt oli jotain tai että jos oli jotakin vähänkin enemmän tai mitään yritystä tai yhtään mitään tekstiä” se oli kakkonen. Ykköstasoinen suoritus oli siis arvioijien mielestä todella niukka tekstinä. Ehkä juuri tämän vuoksi ykköstä pidettiin tasona, jonkatasoiset suoritukset erottuivat helposti muiden suoritusten joukosta. Arvioijien kokemukset olivat samansuuntaisia myös ensimmäiseltä arviointikierrokselta, jolloin taitotaso 1 oli melkein kaikkien mielestä helpoin taso arvioida.

Myös asteikon tasokuvaukseen oltiin yleensä tyytyväisiä. Varsinkin asteikon ilmaus *teksti koostuu toisistaan irrallisista yksinkertaisista, ehkä ulkoa opituista sanaryhmistä, lauseista ja tavallisimmista fraaseista* – tai sen jotkin osat – kuvasi arvioijien mielestä selkeästi ykköstason tekstiä. *Irrallisuudella* arvioijat viittasivat hyvin lyhyisiin teksteihin, joissa ei ollut kokonaisia lauseita ja/ tai joiden lauseet olivat lähinnä fraaseina opittuja ilmaisuja.

”et parhaiten mun mielestä tässä oli tää ykkönen ja tää tää toisistaan irrallisista yksinkertaisista ehkä ulkoaopituista sanaryhmistä” Kaisu

”oikestaan ykköset oli kauheen helppo sieltä raakata heti koska ne oli sit semmosii että ne oli kerta kaikkiaan heikkoja ja just silleen täällä oli mun mielestä tää oli hirveen hyvin et se koostuu tämmösisistä irrallisista yksinkertaisista ehkä ulkoaopituista sanaryhmistä tai fraaseista eli se oli silleen selkeä” Tuula

Ykköstason kuvaus sai kritiikkiäkin muutamalta arvioijalta. Esimerkiksi Teijan mielestä koko ykköstason kuvaus oli huono, koska se ei kuvannut tarpeeksi tekstiä itseään. Hän kritisoi myös kakkostasoa samasta syystä mutta nosti kuitenkin siitä myös positiivisia seikkoja esille, mitä hän ei tehnyt ykköstasosta. Teijan kanssa samantyyppisiä kommentteja ykköstasosta ja sen suhteesta kakkostasoon esitti myös Merja, jonka mielestä ykköstasoon mahtui liian paljon erilaisia suorituksia eli sen sisällä oli hänen mielestään paljon ”hajontaa”.

”sitten täällä ykkösessä ja kakkosessakin niitä tuli mulla yleensäkin vähän sinne et et tuota näissä jotenki näitä ei ollu niin helppo käyttää näissä ei sanota sitä niin tarkasti sitä että minkälainen se semmonen teksti on joka ansaitsee ykkösen taikka kakkosen” Teija

Vaikka ykköstason kuvauksen kohta *teksti koostuu toisistaan irrallisista yksinkertaisista, ehkä ulkoa opituista sanaryhmistä, lauseista ja tavallisimmista fraaseista* pidettiin selkeänä ja hyvin ykköstasoa kuvaavana, se sai myös osakseen kritiikkiä, lähinnä ilmauksen *ulkoa opituista sanaryhmistä* osalta. Esimerkiksi Heidin mielestä oli mahdotonta tietää, mikä on ulkoa opittua ja mikä ei. Hänen mielestään ilmaus oli hyvä lähinnä teoreettisesti.

”sitten tää ulkooapitut sanaryhmät nin se kuulosti musta alkuun niinku hyvältä mutta sitä onkii hirveen vaikee kyllä lopulta niinku sitten taas semmosena auten tekstissä niinku miettii on tämä nyt ylkea ulkoa opittu sanaryhmä [hekotusta] et tota että se on vähän niin sellanen omituinen” Heidi

TASO 2	Osaa kirjoittaa lyhyitä ja yksinkertaisia tekstejä, esim. kirjelappuja ja viestejä jokapäiväisistä asioista sekä hyvin yksinkertaisen henkilökohtaisen kirjeen, esim. kiitoskirjeen. Teksti koostuu ainakin osin yksinkertaisista lauseiden ja fraasien jaksoista, joita yhdistävät tavallisimmat konjunktiot: <i>ja, mutta, koska</i> , joskin väärinymmärrykset ja lukurytmin katkeaminen ovat tavallisia.
---------------	--

Kakkostasoa kommentointiin usein suhteessa ykköseen, lähinnä siitä näkökulmasta, miten hyvin se auttaa erottelemaan ykköset ja kakkoset toisistaan, mikä näkyi myös ykköstasoa kuvaavissa esimerkeissä. Mieli-piteet kakkostasoon toimivuudesta hajaantuivat kuitenkin ykköstasoa enemmän. Muutama arvioija piti kakkostasoon kuvausta yleensä hyvänä: ”*must tuo kakkonen on aika hyvä*”. Yhden arvioijan mielestä kakkostasoon kuvaus kokonaisuudessaan auttoi erottelemaan kolmos- ja kakkostasoon tekstit toisistaan.

Muuten yksimielisempää hyväksyntää sai oikeastaan vain ilmaus *teksti koostuu ainakin osin yksinkertaisista lauseiden ja fraasien jaksoista, joita*

yhdistävät tavallisimmat konjunktiot: ja, mutta, koska. Varsinkin konjunktiotesimerkkejä pidettiin arviointia helpottavana: ”enemmän tämmöstä niinku tukea.” Näiden arvioijien mielestä konjunktiot auttoivat myös erottelemaan ykköstason kakkosesta. Konjunktioihin viitattiin ja ne mainittiin usein esimerkkinä myös siitä, että ne olivat ainoa rakenteita koskeva yksityiskohta koko asteikossa.

Kakkostason kuvausta taas kritisoitiin siitä, että se oli liian tekstityyppilähtöinen ja siksi liian kapea määrittelyltään. Tekstityyppi-lähtöistä määrittelyä kritisoitiin myös liian ylimalkaiseksi, koska henkilökohtaisen kirjeenkin voi kirjoittaa niin monella tavalla, että tekstityyppiluokitteluna *kirje* ei anna tarpeeksi yksityiskohtaista kuvaa mahdollisesta variaatiosta. Tekstityyppien osalta muutamaa arvioijaa häiritsi myös se, että esittelytyyppinen teksti puuttui kokonaan kakkostason kuvauksesta, vaikka sellainen oli suoritettavien kirjoitus-tehtävien joukossa. Tekstityyppi-lähtöisyyttä kommentoineen Heidin mielestä asteikon kuvaus toimisi paremmin, jos se olisi funktiolähtöinen.

Tekstityyppi-lähtöisyyden lisäksi ilmaisu *väärinymmärrykset ja lukurytmin katkeaminen ovat tavallisia* sai kritiikkiä osakseen. Alla olevien esimerkkien mukaan molempien arvioijien mielestä ilmauksen osoittamia väärinymmärryksiä ja lukurytmin katkeamisia ei juuri tule vastaan. Arvioijat ovat mielestään tottuneita lukemaan hyvinkin erilaisten ja eritasoisten oppijoiden kirjoittamia tekstejä ja saamaan selvää erilaisista käsitteistä. Näin ollen he ovat lukijoina tottuneet täydentämään tekstiä, jos se on kielellisesti tai sisällöllisesti epäselkeää. Toisaalta arvioijat kokivat täydentämisen myös ongelmallisena, koska he valittelivat taipumustaan ymmärtää enemmän kuin he itse asiassa ymmärtävät. Arvioijat epäilivät myös ymmärtävänsä oppijoiden tekstejä paremmin kuin tavallinen suomalainen.

”siinä skaalassa sanotaan että väärinymmärrykset ja lukurytmin katkeaminen tavallisia itse asiassa väärinymmärryksiä tulee hirveen vähän siis koska nää on niin jotenkin sim semmosii aika selkeitä nää tehtävätyypit että em mä must itse asiassa hirveen vähän oli sellasta jossa olis niinku jossa mä oisin kokenu et mä oisin ymmärtäny jonkun tai en ymmärrä ku mitä joku sano tarkoittaa ja mm sit sitä lukurytmin katkeamistaki on vähän vaikee silleen arvioida kus sitä on niin harjaantanu lukemaan näitä mä huomaan et mä usein luen tiiätsä mielessäni ne oikein siis silleen että ro korjaan ne niinku mielessäni rakennevirheetki suoraan oikein” Heidi

”sitte tääki niinku väärinymmärrykset ja lukurytmin katkeaminen ovat tavallisia niin tää tässä kyllä autto mut sitte niin se että että mikä nyt sitten lasketaan väärinymmärryksiks ja lukurytmin katkeamiseksi että nyt niitä ei tullu paljon” Teija

Tasot 3 ja 4: "että ehkä se vaikein oli justiinsa tuolla niinku keskitasolla"

TASO 3	Osaa kirjoittaa yksinkertaisia mutta sidosteisia tekstejä tutuista tai henkilökohtaisesti kiinnostavista aiheista. Osaa kirjoittaa mm. henkilökohtaisen kirjeen omista koke- muksistaan ja tuntemuksistaan. Kielitaito riittää kuvailemaan yllättäviä tilanteita, selittämään pääidean tai ongelman abstraktisesta tai kulttuuriaiheesta, esim. musiikista tai elokuvista sekä kertomaan itseään läheisistä aiheista, esim. perheestä, harrastuksista tai työstä, joskin esim. sanastolliset puutteet aiheuttavat toistoa.
---------------	--

Suurin osa arvioijista oli sitä mieltä, että arvioitavat tekstit sijoittuivat pääasiassa tasolle kolme. Myös tason neljä suorituksia oli arvioijien mielestä paljon, jos he olivat arvioineet lähinnä neljättä tehtäväsarjaa. Runsautensa lisäksi tai juuri sen vuoksi välitasojen tekstit tuntuivat aiheuttaneen eniten ongelmia. Yli puolet arvioijista oli sitä mieltä, että joko asteikon taso kolme tai sitten tasot 3–4 olivat kaikkein ongelmallisimmat soveltaa: "et tässä niinku tarvittas eniten minusta täs täsmennystä kolmosen ja nelosen välillä." Sama tendenssi sekä tämäntasoisten suoritusten määrän että tasojen soveltamisvaikeuden osalta tuli esille myös ensimmäisellä arviointikierroksella.

Kolmostason ongelmallisuutta lisäsi osaltaan se, että siitä tuli tasona ehkä kaikkein laajin ja vaihtelevin. Kolmostason tekstit saattoivat siis olla keskenään hyvin erilaisia.

"se kolmonen ehkä muodostu niin hirveen laajaks koska tavallaan niinku hyvin virhe jos oli kirjotti hyvin huo niinku sillai rakenteellisesti puutteellista mutta se oli kuitenkin sidosteista koska mä pidin sidosteisuutta aika niinku semmosena keskeisenä kolmosen kriteerinä nih tota mm nin sillä sidosteisuudella saatto saada kolmosen vaikka se oiski sitten ollu niinku rakenteellisesti aika niinku sillai alkeellista vielä tai tai ihan peru ihan perusasioissaki ongelmia" Heidi

Kritiikkiä kolmostason kuvauksesta sai ilmaus *tekstejä tutuista tai henkilökohtaisesti kiinnostavista aiheista*, koska arvioijien mielestä tuttuuden tai henkilökohtaisen kiinnostavuuden asteen voi tietää vain kirjoittaja itse. Toisaalta taas esimerkiksi Kaisun mielestä kaikki, mitä kirjoittaja kirjoittaa, on henkilökohtaista, koska hän kirjoittaa sen kuitenkin viime kädessä omasta näkökulmastaan. Näin ollen hänen mielestään "nää on tälläsiä ilmalauseita joilla ei oikestaan oo sit niinku ei nää hirveesti pysty auttaan." Joitakin arvioijia mietitytti myös ilmaisu *sidosteisuus ja toisto*. Heidän mielestään oli vaikea tietää ja/tai ymmärtää, mitä kolmostason *sidosteisuus* oikein tarkoittaa tai mitä *toistolla* haetaan.

"se on musta aika ongelmallinen tapa määritellä että et kirjottaa jonkinlaisia tekstejä tutuista tai henkilökohtasista kiinnostavista aiheista mistä pirusta minä tiijän mikä niille on henkilökohtasesti kiinnostavaa siis

et jos ne kirjottaa omasta kulttuurista nih onko se niille kuitenkaan henkilökohtasest kiinnostavaa että tai sitten jos on kysymys jonkun mielipiteen kirjottamisesta tai muusta ni emmä voi tietää onko se heistä kiinnostavaa tai onko se tuttuakaan” Heidi

”toi sidosteinen se on kans hauska sana et mitä sekin ny sitten tarkoittaa mut siinä mä sit ajattelin sitä et siin täytyy olla niinku sit se tekstissä niinku tekstinä niinku järki” Martti

”just tää osaa kirjottaa yksinkertaisia mutta sidosteisia tekstejä ni mitä sillä tarkotetaan tai ilman huomattavia mitä esimerkiksi on huomattavat sanastolliset ja rakenteelliset puutteet että tietysti se niinku niitten tarkempi määrittely ni se jää koko lailla sen arvioijan vastuulle tässä tapauksessa” Inkeri

Muutaman arvioijan mielestä kolmostason soveltaminen teksteihin ei kuitenkaan ollut vaikeinta tai edes vaikeatakaan. Esimerkiksi Sannan mielestä kolmostason kuvaus toimi ja sen mukaan oli helppo arvioida, koska hänestä se edusti ”tyypillistä tämän arviointipinkan tekstiä”. Samoin Pirkon mielestä kolmosen kuvaus oli osuva: ”tämä oli minusta aika hyvä hyvä tämä kuvaus.” Yhteistä Sannalle ja Pirkolle oli se, että he puhuivat ensisijaisesti siitä, mitä kolmostasolle yltävä kirjoittaja jo osaa, mikä saattoi osaltaan vaikuttaa kolmostason sovellettavuuteen.

”no sittenhän yksinkertaisia mutta sidosteisia tekstejä tutut ja aiheet tutut ja kiinnostavat aiheet ni se oli jotenkin tavallaan semmonen helppo ryhmä et kaikki sit jotka pysty jo vähän enemmän käyttämään erilaisia muotoja ja pikkasen eri rakenteita ja pysty selittämään jotakin jostakin mutta kuitenkin siinä sai olla vielä jonkin jonkin verran virheitä tutut aiheet ei tarvinnu olla vielä mitään yksityiskohtaisia tekstejä nin siihen ne siirty sitten suurin osa” Sanna

TASO 4	Osaa kirjoittaa selkeitä ja yksityiskohtaisia tekstejä monenlaisista henkilökohtaisesti kiinnostavista aiheista. Osaa kirjoittaa mm. informoivan tai argumentoivan esseen tai raportin. Osaa kirjoittaa tekstejä, esim. kirjeen itselleen merkittävistä tapahtumista ja kokemuksista. Pystyy ilmaisemaan itseään ilman huomattavia, esim. sanastollisia ja rakenteellisia, puutteita.
---------------	---

Nelostasosta kommentoitiin eniten ilmausta *ilman huomattavia sanastollisia ja rakenteellisia puutteita*. Tähän tasokuvauksen kohtaan kiinnittivät huomiota kaikki arvioijat. Huomiot vaihtelivat pohdiskelevista havainnoista ärsyntyneisiin kommentteihin ilmauksen mitään-sanomattomuudesta ja ympärilyöreydestä. Eniten tunteita herätti ilmauksen moniselitteisyys ja subjektiivisuus: arvioija saattoi nimittäin tulkita ilmauksen haluamallaan tavalla: ”joku voi olla sitä mieltä että hirvee puute jos ei osaa monikon genetiivin toisen mielestä se ei oo puute ollenkaan että niinku menee sit tähän niin paljon siihen sen arvioitsijan

semmoseen henkilökohtaseen näkemykseen et mikä on tärkeätä ja mikä ei oo tärkeätä.”

Monet arvioijista olivat sitä mieltä, että ilmaus *ilman sanastollisia ja rakenteellisia puutteita* sai nelostason kuulostamaan vaativalta, varsinkin suhteessa kolmoseen. Esimerkiksi Päivikin mielestä se jakoi koko asteikon kahteen osaan siten, että tasot 1, 2 ja 3 muodostavat asteikon alemmat tasot ja tasot 4, 5, ja 6 ylempät tasot. Hänen mielestään tämä ilmaus aiheutti sen, että asteikkoon tuli ”ihan selvästi semmonen sentin rako tähän kohtaan.” Samansuuntaisia kommentteja esitti myös Sanna, jonka mukaan turhan monet tekstit jäivät kolmostason suorituksiksi nelostason vaativuuden vuoksi.

Kolmos- ja nelostasojen välistä eroa pidettiin muutenkin suurena. Näiden arvioijien mielestä kynnys antaa nelostasoja nousi, koska hyppäys kolmoselta neloselle tuntui yleensäkin liian suurelta, kuten Sannan repliikistä näkyy. Jos nelostaso tuntui liian vaativalta, arvioijasta oli helpompaa ja tavallaan myös oikeampaa antaa kolmonen kuin nelonen. Antaessaan kolmosen arvioijan ei tarvinnut tuntea olevansa liian avokätinen, lempeä tai vastuuton. Vastaavasti taas ne arvioijat, jotka kokivat nelosen matalampana, pitivät hyppäystä neloselta viitoselle korkeana.

”no joo tää tämmönen lause et pystyy ilmaisemaan itseään ilman huomattavia esimerkiks sanastollisia ja rakenteellisia puutteita kuinkaha moni niistä ei varmaa pari kolme hassua pääsisi ehkä yli niissä on niinku ei oo ollenkaan huomattavia jompia kumpia sanastollisia tai rakenteellisia puutteita plus sitten se että jos ei vastannu kysymykseen ni sekin pudotti”
Sanna

Mitä arvioijat sitten ymmärsivät huomattavilla sanastollisilla ja rakenteellisilla puutteilla? Suurin osa arvioijista mielsi ilmauksen kielen rakenteita koskevaksi, joskin arvioijien näkemykset *huomattavista rakenteellisista puutteista* saattoivat vaihdella arvioijittain. Kaksi arvioijista piti mahdollisena myös sitä, että *rakenteelliset puutteet* voisivat viitata tekstin rakenteeseen eli enemmänkin tekstuaalisiin taitoihin. Tosin, kuten arvioijatkin itsekin päättelivät, *sanastollisen* parina *rakenteellinen* tuntuisi viittaavan enemmän kielen rakenteiden hallintaan.

Millaisten rakenteiden hallinnan arvioijat sitten näkivät puutteellisena tai millainen sanaston hallinta olisi taas puutteetonta? Lähes kaikki arvioijat ottivat kantaa vain ilmauksen toiseen puolikkaaseen eli rakenteiden hallintaan eivätkä kommentoineet lainkaan sanaston hallintaa. He joko puhuivat yleensä rakenteista tai yksityiskohtaisemmin tietyn rakenneseikan hallitsemisesta. Kaksi arvioijaa nosti esille myös sanaston. Toinen heistä kommentoi sanastoa varsinkin sen

käytön tarkkuuden näkökulmasta, ja toinen puhui aika pitkälti yleensäkin siitä, että asteikko painottaa liikaa sanaston hallintaa.

Osa arvioijista painotti kommentoissaan ilmauksen sanaa *huomattavia*. Se yleensä tulkittiin joko, kuten edellä jo mainitsin, hyvin vaativaksi, erittäin hyväksi rakenteiden hallitsemisen tasoksi tai kyvyksi kirjoittaa ymmärrettävää tekstiä. Jälkimmäisessä tapauksessa arvioijat katsoivat ilmauksen viittaavan lähinnä pikkuvirheisiin, jotka eivät enää vaikuta tekstin ymmärrettävyyteen tai luettavuuteen. Tässä mielessä ero huomattavien puutteiden välillä saattoi muodostua huomattavaksi, jos tarkastelun toisessa ääripäässä oli ymmärrettävyys ja toisessa varsin systemaattinen perusrakenteiden hallinta.

”joo sitä mää yritin siinä matkan varrella vähän määritellä itelleni ja tota aaa noh semmoset selkeet virheet jotka johtaa metsään et mä en ymmärtäny sit enää mikä oli ni ne oli semmosii huomattavia et sitten pikkuvirheistä mä en nyt tässä vaiheessa et ne mä sitten jos oli pikkuvirheitä jonkin verran ni ne mä laitoin varmaan tohon nelosiin” Sanna

”vaikka tässä sanotaan että ilman huomattavia esim sanastollisia rakenteellisia puutteita mut mä otin taas tän huomattavia niinkun sillä tavalla että että ne oli semmoset että niistä kuitenkin ymmärsi mistä on kysymys” Eeva

”se saatto olla vähän fiiliksen mukana et sit välillä huomattavaa oli niinku vähemmän ku jokus mut tota ni em mä nyt sitä tietenkään et jos oli tonnosii et siellä nyt joku vaikka kestoasteet heitti nih ei se ollu huomattava tai jos nyt joku rektio oli väärin et lähinnä kai mä sitä yritin niinku sen ymmärrettävyyden kannalta niinku miettii et sit se on huomattava puute et niinku siit ei saa et niinku ei saa selvää tai joutuu arvailee” Martti

”esimerkiks jos sä nyt vaikka käytit väärää sanaa em mä nyt muista olik siellä siellä joku ainaki et sun äitisi on oli huono nainen ni just tai joku tämmönen nin nin et se oli nyt tämmönen selvä tai joku väärä verbi tai ihan joo koska ei niissä sit mihin mä nelosen annoin et ne kyl oli mun mielestä jo aika hyviä et ei siellä enää niinku semmosia” Merja

Arvioijien näkemykset siitä, minkä rakenteiden hallintaa he pitivät tärkeänä nelostasolla, vaihtelivat jonkin verran. Jonkun mielestä aikamuotojen epäsystemaattinen hallinta oli huomattava puute, ja jonkun toisen mielestä taas rektiovirheitä ei voinut enää sallia nelostasolla. Yleisimmin rakenteellisena puutteena nähtiin verbimuotojen hallinta joko aikamuotojen käytössä tai aikamuotojen negaatioissa tai yleisemmin verbien taivutuksessa. Eroja oli siinä, miten tiukasti arvioijat pitivät kiinni rakennekriteereistään. Joku sanoi olleensa ehdoton, ja joku toinen taas sanoi tehneensä myönnytyksiä. Esimerkiksi Heidi sanoi ensin olleensa

ehdoton negaatioiden kohdalla mutta tehneensä sittemmin kompromisseja muiden kriteerien hyväksi.

”mä ajattelin että osaa peruskieli opit siis sanotaanko näin et suurinpiirtein kaiken muun paitsi lauseenvastikkeet et niinku silleen kielitaito on sillä tasolla et tietää kaikki nää perusjutut mutta niinkun just tää hienosäättö ja lauseenvastikkeet ja tämmöset niinku puuttuu ja erikoisrektiot voi olla vielä vähän epäselvänä ja niinku silleen mä ajattelin sen mielessäni että et siihen menis niinku jos kaikki muu suurin piirtein hallussa tai ees hajulla mutta kyllähän mä luulen et käytännössä se meni niin että et tonne neloseen kyllä lipsahti to tämän asteikon mukaan niinku huomattaviaki rakenteellisia puutteita” Pia

”et mun mun mielestä niinku aikamuodot ja tavalliset rektiot sit sit tää on kyl sellane asia mikä mikä tota voi olla että tää on vaan mun niuhottamista” Tiina

”noh varmaan niitä verbejä mä aina niihin verbeihin sitten kiinnitin huomiota että että se verbioppi jotenkin niinku korostuu ja et jotain et se vaikka muuten sitä mut jos ei sit jotain negatiivista imperfektiä käytä ollenkaan ei osaa niissä hirveesti tulee virheitä tietysti” Pirjo

”nelonen noh just rektio virheitä semmosia niinku että lähinnä ehkä sillä tavalla että että sanastollisesti se on niinku ookoo että siellä niinku tai mistä sitä tietää mitä se on halunnu sanoa siellä se ihminen mutta kuitenkin niinku että että se vaikuttaa siltä että se on löytänyt oikeet sanat mutta se ei välttämättä tiijä niitten ihan ihan hienosäättö niinku muodollisesti että joo varmaan just sillä tavalla että ei sanastollisia semmosia semmosia oo ongelmia että se viesti ei välittyys mutta sitten just rektioita mun mielestä” Eeva

”noh m mun mielestä siinä on se että jos jos ver jos tota noin ni subjekti ja predikaatti ei kongruoi keskenään ni mun mielestä se on niinkun huomattava ja ja sitten niinku rektiovirheet on huomattavia että et mä mun mielestä sitten see että ne et tota vokaaliharmonia ei oo huomattava eikä se ei eikä tota noin ni niinku kirjainvirheet että jos niinku sanotaan kirjoitusvirheet ei oo mun mielestä huomattavia mut että siellä oli joitakin semmosia ihmisiä jotka teki esimerkiksi niinku sitten kongruenssi ja rektiovirheitä ja se oli kuitenkin musta tyylillisesti niinku nelosen teksti mutta et niitä oli sen vertaa monta siellä että et sitten rupee miettiin sitä että jos niinku joka lauseessa on yks sellanen virhe ja objektivirheitä et objektivirheet mun mielestä kans se että että tota noin ni et niist objektivirheistäki osa on selvästi mun mielestä huomattavia et ei ei kaikki et et joskus se ei ei sillai niin häiritse sitä kuvaa” Rajja

”nelostasolla mä sit aloin vaatii jo sitä että et aikamuodot piti olla niinku suht hallinnassa ja et esimes ei et ei käyny enää että jotenkin ei voi olla enää sen tasosta että et ei ajoi proo ei ajanut että niinku mun mielestä sit et siellä piti niinku siellä mä olin sit jo aika tarkka” Heidi

Kaiken kaikkiaan kommentit huomattavista rakenteellista puutteista vaihtelivat holistisesta perusrakenteiden hallinnasta jonkin yksittäisen, lähinnä morfologisen, rakenneseikan hallintaan. Kommenteissa oli havaittavissa myös arvioijien omia henkilökohtaisia mieltymyksiä jonkin tietyn rakenneseikan painottamiseen. Siihen, miksi tällaiset mieltymykset nousivat esiin, on vaikea antaa vastausta. Ehkä testityyppilähtöisellä asteikolla ja arvioijien käsityksillä ei ollut kosketuspintaa, tai ehkä asteikko ei tarjonnut tuttuja, kieleen liittyviä työkaluja tekstien arvioimiseen. Olisi mielenkiintoista nähdä, olisiko tilanne sama, jos arvioijat arvioisivat rakennepainotteisella asteikolla, jonka rakennepainotukset eroaisivat arvioijan omista painotuksista.

Vaikka nelostasolla huomiot kiinnittyivät pääasiassa ilmaukseen *ilman huomattavia sanastollisia ja rakenteellisia puutteita*, arvioijat kiinnittivät huomiota myös muutamaa muuhun seikkaan. Nelostason kuvausta kritisoitiin ylimalkaisuudesta ja myös siitä, että se ei antanut tarpeeksi tukea arviointiin. Arvioijat olisivat kaivanneet yksityiskohtaisempaa kuvausta, varsinkin rakenteiden osalta.

Nelostason kuvauksesta pidettiin taas hyvänä ilmausta *selkeitä ja yksityiskohtaisia tekstejä*, koska ”siinä on jo semmosta selkeyttä enemmän et et tota ja ja niinku enem niinkun että enemmän sellasta niinku tarkempaa selittämistä ja mm vähän semmosta enemmän semmosta niinku et kykenee niinkun raportoimaan.” Nelostason kuvauksista konkreettisesti ohjaavana pidettiin myös ilmausta *osaa kirjoittaa argumentoivan esseen*, minkä vuoksi esimerkiksi Pirjo edellytti nelostason kirjoitukselta argumentteja.

Tasot 5 ja 6: ”mie en huomannu mitää eroo on näil kahel ylimmäl tasol”

Tasot viisi ja kuusi nähtiin hyvin korkeina tasoina. Arvioijat myös spekuloiivat sillä, miten moni äidinkielenkään yltäisi näille tasoille; joidenkin arvioijien mielestä jopa heidän itsensä olisi vaikeaa kirjoittaa kuutostason tekstiä. Pari arvioijaa oli sitä mieltä, että harva maahanmuuttaja saavuttaa näitä tasojakaan tai ainakaan kuutostasoa: ”niinku ei monetkaa koskaan pääse niinku sinne kuutoseen.” Yhden arvioijan mielestä jo taso viisi ”on ihan enkelitaso” ja yhden arvioijan mielestä ”ihan natiivinki kirjottama niin helposti jäis tohon vitoselle ainakin tän kuvauksen mukaan.”

Parin arvioijan mielestä ylemmät tasot ovat täysin akateemisen kirjoittamisen tasoja, joten ne eivät luonnollisestikaan koske läheskään kaikkia maahanmuuttajia. Akateemisuuteen liittyen yksi arvioijista kritisoi sitä, että tason kuusi kriteerit ovat kovemmat kuin yliopiston maturiteetissa. Toisaalta arvioijien itsensäkin luomissa asteikoissa (ks. taulukot 14–18) taitotasossa 6 viitattiin syntyperäisen tasoiseen kieli-taitoon.

Arvioijat käyttivät tasoja 5 ja 6 kaikkein vähiten. Kaikki arvioijat eivät "edes päässeet antamaan" yhtään kuutosta, jos he arvioivat kolmea ensimmäistä tehtäväsarjaa ja tulkitsivat tason 6 syntyperäisen tasoiseksi. Tosin niin sanotusti vahvempienkaan suoritusten joukossa joidenkin arvioijien mielestä "ei ollu kyllä niin niin täydellistä suoritusta." Myös viitostasoa useimmat arvioijat käyttivät "hyvin vähän" tai niitä tuli "vain muutamia" tai "ehkä jokunen". Pari arvioijista sanoi antaneensa muuttaman kuutosen mutta oli kuitenkin varma, että kukaan näistä kirjoittajista ei voisi oikeasti olla tällä tasolla, tai epäili muuten sitä, että kirjoittajat olisivat olleet niin hyviä.

TASO 5	Osaa ilmaista itseään selkeästi ja rakenteellisesti hyvin. Osaa kirjoittaa erilaisia tyyllillisesti ja muodollisesti hallittuja tekstejä, jotka on osoitettu tietyille lukijalle. Osaa kirjoittaa mm. yksityiskohtaisesti kuvailevan kirjeen monimutkaisesta aiheesta sekä omia näkökulmia korostavan esseen tai raportin. Kirjoittaja pystyy ilmaisemaan sanottavansa ilman rajoituksia.
---------------	---

Arvioijat poimivat viitostason kuvauksesta esimerkiksi ilmaisun *kirjoittaja pystyy ilmaisemaan sanottavansa ilman rajoituksia*, mikä tuntui monen mielestä osoittavan hyvin "hyvää" kielitaitoa, "korkeaa" tai "huimaa" taidon tasoa. Mitä korkeammalle kielitaidon taso nousi, sitä enemmän arvioijat pohdiskelivat sitä, millainen teksti yleensä edustaisi kyseistä tasoa ja missä määrin heidän arvioimansa tekstit voisivat olla tämäntasoisia. Eräässä mielessä pohdiskelussa haettiin rajaa kielenoppijan ja syntyperäisen kirjoittajan välillä: voiko sujuva, ilman rajoituksia kirjoittava kirjoittaja olla kielenoppija, ja jos on, miten se on todistettavissa kirjoitustehtävillä.

"että siitä tätä mä ihmettelin että vitostasolla on kuvaus että kirjottaja pystyy ilmaisemaan sanottavansa ilman rajoituksia niih mä olin ihan että mitä se on että että voinko mää niinku tavallaan onko mulla semmone valmis teksti täällä päässä mitä mä aattelen et se pitää olla just se ja ja sitten sitä mukaa mä niinku arvioin että osaako se sanoo sen" Tarja

"edellisellä kierroksella oon arvioinu tasolle kuus nii sellaset mä laitoin sitten tohon vitos ryhmään ku siel oli ne pystyi ihan samaan ilman rajoituksia ni se oli jo aika huimaa että et siinä sitten saatto olla jonkun verran pieniä virheitä vielä mutta kuitenkin et mä saatoin niinku kuvitella et tää tyyppi pystyy kirjottamaan mistä tahansa mitä tahansa niin sen mä laitoin" Sanna

"mut sitten mitäs sitten hmm voiskohan tommonen olla kriteeri tuo mikä on tuossa viitosessa että et pystyykö se kirjottaja ilmasemaan sanottavansa ilman rajoituksia mitä se sitten ois noh ei se voi olla oikein hmm noh sitä on niin vaikea toisaalta sitten todistaa että millon millon jollakin on ollu rajoituksia taikka millon ei et se riippuu niin paljon siitä tehtävästä" Teija

Kuten esimerkeistä näkyy, *ilman rajoituksia* voidaan tulkita eri tavoin. Se voi tarkoittaa sitä, että teksti on täysin virheetön tai että kirjoittaja pystyy kirjoittamaan mistä tahansa aiheesta tai että hän yleensäkin pystyy sanomaan sen, mitä haluaa. Myös tasokuvauksen ensimmäinen lause *rakenteellisesti hyvin* jättää varaa tulkinnoille siitä, mistä ilmauksessa on kyse: ”ni mun mielestä kyllä se et jos sanotaan rakenteellisesti hyvin se tarkoittaa kyllä silloin että siellä ei oo siis semmosia rektiovirheitä eikä tämmösiä että silloin musta se niinku tota noin ni että tää oli niinku aika tiukka toi vitonen.”

Epävarmuutta herätti myös se, missä määrin kyseiset tehtävyyt pit mahdollistivat tämätasoisien kielitaidon osoittamisen. Miten kirjoittaja kirjoittaisi jonkin muun tekstin kuin arvioitavana olevan tekstin? Toisin sanoen arvioijia mietitytti, pystyisikö kirjoittaja tuottamaan arvioitavaa tekstiä vastaavia tai jopa vaativampia tekstejä todellisessa elämässä. Yleensä ottaen joitakin arvioijia tuntui epäilyttävän, oliko arvioitava teksti tarpeeksi vahva näyttö, jotta sen perusteella voisi antaa ylempiä tasoarvioita.

Joidenkin arvioijien mielestä se, että asteikossa mainitut tekstityypit, kuten raportti ja essee, puuttuivat arvioitavista tehtävistä, vaikeutti ylempien tasojen soveltamista. Pari arvioijaa piti esseetä ja raporttia erittäin vaativina tekstityypeinä, osin myös akateemisina, varsinkin esseetä. Yksi arvioija pitikin etenkin ylintä tasoa epäsopivana arviointiin: ”sit tää monimutkaisia kirjeitä raportteja artikkeleita lukijaa helpottava looginen tekstin pääasioita korostava rakenne ää tää on must hirveen akateemista paskaa jollon sitä ei voi niinku soveltaa ollenkaan.” Joidenkin arvioijien mielestä taas työpaikkahakemuksesta ja mielipidetekstistä voi kirjoittaa kuutostasonkin tekstin, jos kirjoittajan taito riittää siihen.

”näissä kriteereissä oli se hankala puoli et tota ni kun näissä on tämmösiä et o osaa kirjottaa erilaisia tyylillisesti ja muodollisesti hallittuja tekstejä ja niinku vitosessa ja sama nelosessa osaa kirjottaa selkeitä ja yksityiskohtasia tekstejä monenlaisista henkilökohtaisista kiinnostavista aiheista ni mistä hitosta mä tiedän et se osaa kirjottaa monenlaisista ku sillä on vaan se yks teksti eli se on niinku teki hankalaks soveltaa tota yläpäätä” Martti

”sit taas toi vitonen oli siinä mielessä mun mielestä hankala että et siinä puhutaan niinku monimutkaisista aiheista ja omia näkökulmia korostavia esseitä ja tämmösiä et niinku taas tämmöset näin lyhyet tehtävät ni ei oikeastaan pysty tuomaan sellasta kykyä esille et niinku hirveän vaikea antaa sitte jotain vitosta” Tuula

Kaiken kaikkiaan viitostason kuvauksessa tuntui aiheuttavan ongelmia se, että lukijakeskeisyys, tekstityypin tai tietyn muodon ja tyylin hallinta yleistettiin koskemaan muitakin kirjoituskonteksteja. Arvioijien

kommenttien mukaan he ”ehkä” halusivat arvioida vain sellaista taitoa, joka oli pääteltävissä arvioitavasta tekstistä: *”osaa kirjoittaa erilaisia tyyllillisesti ja muodollisesti hallittuja tekstejä jotka on osoitettu tietyille lukijalle noh aa erilaisia tyyllillisesti hallittuja tekstejä mistä mä voin tietää.* Myös asteikon ja tehtävänannon vastaamattomuutta suhteessa kirjoitustehtävään kritisoitiin.

”minusta tämä vitonen on hyvä siis se oli helppo käyttää tässä ei oo semmosta yks asia vaan mikä minkä mä oon sit laittanu niinko kysymysmerkiksi ja huomautukseksi etko tässä sanotaan näin että et osaa kirjoittaa tekstejä jotka on osoitettu tietyille lukijalle ni tota nin nin tässähän se riippuu siitä että onko se minkäläinen se on ollu tehtäväksanto että onko siinä o annettu se lukija esimerkiks siitä tarinatehtävästä nin oli vaikea ajatella tältä kannalta koska siinä ei ollu sanottu mittään että kelle se on” Teija

TASO 6	Osaa kirjoittaa selkeitä ja sujuvia tyylliltään tilanteeseen sopivia tekstejä. Osaa kirjoittaa monimutkaisia kirjoituksia, raportteja tai artikkeleita, joissa on lukijaa helpottava looginen, tekstin pääasia korostava rakenne. Osaa kirjoittaa tiivistelmiä ja katsauksia ammatillisista aiheista. Tekstissä ajatukset on muotoiltu tarkasti ja yksiselitteisesti. Tekstissä ei ilmene lainkaan merkkejä siitä, etteikö kirjoittaja pystyisi ilmaisemaan sitä, mitä hän haluaa ilmaista.
---------------	---

Kuutostason teksti konkretisoitui usein tekstinä, joka on samanlainen kuin äidinkielen kirjoittama teksti tai lähes sitä tasoa tai jopa keskivertosuomalaisen kirjoittamaa tekstiä parempi. Toisaalta arvioijissa oli jonkin verran kahtalaista suhtautumista siinä, pitäisikö kuutostason tekstin olla äidinkielen tasoinen vai ei. Olisiko reilua tai tarkoituksenmukaista vaatia tekstiltä niin korkeaa tasoa? Toisaalta syntyperäisen tasoa ei sitäkään pidetty yksiselitteisenä: millaisen syntyperäisen kirjoittajan taso olisi sopiva vertailukohta?

Arvioijien kommenttien perusteella äidinkielen taso rinnastui joko yleensä suomalaisiin, heidän tuttavapiiriinsä tai heihin itseensä. He viittasivat esimerkiksi joissakin kommentteissa itseensä kirjoittajana tai aviomieheensä kysyen, pystyisivätkö he itsekään tai heidän miehensä kirjoittamaan kuutostason tekstiä. Yhden arvioijan mukaan hänen oma tekstinsä olisi siinä tapauksessa kuutostason teksti, jos se olisi ”ihan pilkun päälle”. Suomalaisiin rinnastuva äidinkielisyys saattoi olla myös sitä, että kirjoittaja hallitsi samat rakenteet ja saman sanaston kuin äidinkielenkin kirjoittaja. Arvioijien kommentteille on kuitenkin yhteistä se, että asteikon taitotaso 6 nähdään jossakin määrin syntyperäiseen rinnastettavana taitona.

”tuossa kuutostasolla näkee näkyy ehkä parhaiten se että on jotakii tämmöstä erikoissanastoo et se sanavarasto on jo niin hyvä että se rakenteissa se ei enää jo vielä niinku ei oo niinku enää mitään oikeestaan

uutta et se rupee niinku olemaan semmosta tavallaan jo äidinkielistä vaikka sitä mä tietysti pyrinki varoa etten sitten ajattele niin” Tarja

”kuutonen eihän suomalainen välttämättä pääse tommoseen kuutoseen että se ois mun mielestä kuus ja puoli kyllä tuo tasokuvaus mikä tossa on että en mä välttämättä pysty aina kirjottamaan lukijaa helpottavaa loogista rakennetta ja tota nin pääasioita korostavaa ja ja ehkä se johtu niistä että mun mun tehtävätki että niissä ei ollu oikeastaan sellast joka ois ollu kauheen vaativa et jos mun ois pitäny tehdä joku” Päivikki

”mut et et se on tota ilmeisesti ollu sit kyllä niinku hyvin semmonen niinku et siin on ollu välistä jonkunlaista tyyliä jua kyllä semmosta et se ois voinu joo nyt mä muistan et se oli mun mielestä ihan sellanen et se ois voinu olla niinku suomalaisen niinku voi kuvitella jonkun suomalaisen kirjottama et se oli niinku semmonen et siinä oli vähän semmosta puhekielisyyttä ja tämmöstä et se oli kuitenkin niinku et kuka tahansa nuori ihminen vois ihan hyvin kirjottaa semmosen ettei siitä tietäis et se on enää ulkomaalainen” Tuula

Kuutostason vaativuudesta tai sen suhteesta syntyperäisen kirjoittamaan tekstiin kertoo sekin, että muutamat arvioijat epäilivät, että tekstipinon oli sotkettu muutamia syntyperäisten kirjoittamia tekstejä, koska nämä tekstit olivat niin suomalaisen oloisia. Kyseiset arvioijat arvelivat – joskin leikilliseen sävyyn – että tekstit olisi laitettu tarkoituk-sellisesti pinon, jotta nähtäisiin, miten he suhtautuisivat niihin. Arvioitavien tekstien joukossa ei kuitenkaan ollut, ainakaan tarkoituk-sellisesti, ainoatakaan syntyperäisen kirjoittamaa tekstiä.

”joo mä en muista laitoinko mä siihen sitten yhen ku saatto siel et sit mä yhen pistin armosta ku yks oli mä rupesin nyt kattomaan et et perkele et onko tää suomalaisen kirjottama joo et siellä oli muutama oli niin hyvii et mä aattelin et nää voi olla kyl suomalaisen kirjottamii et nyt mua yritetään täs vähän tota noin ni no joo mut siit mä sit laitoin kutosen” Martti

Virheisiin kuutostasolla suhtauduttiin kahtalaisesti: osa arvioijista niitä ei sallinut ja osa taas sanoi joitakin pikkuvirheitä vielä hyväksy-vänsä. Virheettömyys liittyy osin äidinkielimäisen tekstin vaatimukseen ja osin taas mielikuvaan kuutostason absoluuttisuudesta. Joidenkin arvioijien mielestä suomalaisen tekstissä ei ole virheitä ja joidenkin mielestä taas suomalaisetkin tekevät virheitä, mutta koska kuutostaso on niin korkea, vain harvat suomalaisetkaan saavuttavat sitä.

”et tos on vitosen ja kutosen välissä mun mielestä ei ollu niin kauheen tän tekstin tuosta niin suurta eroo tosin mä tein sen selkeen eron sillä tavalla että kutonen oli mun mielestä niinkun tavallaan täydellinen ja vitonen oli semmonen että jos ei tavallaan tyylist et se oli pikkusen kankee ni mä

ajattelin et sitten se niinku tavallaan et virheitä ei hyväksytä kumpaakaan”
Raija

Virheitä sallivat arvioijat taas kertoivat hyväksyvänsä joitakin, muutamia pieniä virheitä. Pikkuvirheinä arvioijat pitivät pääasiassa oikeinkirjoitusvirheitä, kuten pitkien/lyhyiden vokaalien ja konsonanttien merkitsemistä ja/tai ä:n ja ö:n pisteiden puuttumista tai joitakin selkeästi huolimattomuudesta johtuvia puutteita. Ero arvioijien suhtautumisessa virheiden olemassaoloon, laatuun ja merkitsevyyteen tuli esiin jo viitostason kohdalla. Kahtalainen suhtautuminen virheisiin voi osittain johtua asteikon monitulkintaisuudesta mutta myös arvioijien omista näkemyksistä, koska samantyyppisiä eroja oli nähtävissä myös arvioijien omassa asteikoissa.

”varmasti virheitäki saa olla vielä vähän näin minä ajattelin saa jotakin vikaa olla (minkälaisia vikoja sää sallisit sinne) kyllä minä jonkun tämmösen ää astevaihtelu siis tämmösen kaksi ään äänteen merkkiä ni semmoset kaks koota tai yks koo ja tuota ku aattelee et kyllähän suomalainen aika paljon rikkoo näitä omia oman kielen sääntöjä vastaan että ne et jos tulisi joku joku objektivirhe ni en minä niin pahana sitä pitäs” Pirkko

Mielenkiintoista virheiden laadussa on se, että ne katsotaan pääasiassa kielellisiksi tai oikeastaan rakenteellisiksi, vaikka huomiota tosin kiinnitettiin myös siihen, miten tehtävät ja niiden edellyttämät sisällöt pystyvät osoittamaan kuutostason mukaista kirjoittamista. Myös kommentit kielen helppoudesta ja vaivattomuudesta, esimerkiksi ”ni semmonen niinku rennosti kirjoitettu jotenkin et siis siinä oli niinku esimes kieli ja jotenkin niin si sit välitty mu mulle sellanen vaivattomuus”, viittasivat pääasiassa kieleen ja sen käyttöön. Painopiste näyttäisi olleen enemmän kielessä kuin esimerkiksi kirjoittamisen konventioissa, tekstin rakenteessa tai sisällön monipuolisessa tarkastelussa. Tai sitten arvioijat viittasivat *kielellä* laajempiin kokonaisuuksiin, esimerkiksi myös sanastoon ja tekstin rakenteeseen tai kirjoittamiseen yleensä.

Vaikka kuutonen koettiin hyvin korkeaksi tasoarvioksi ja osin myös vaikeaksi arvioida, pari arvioijaa oli kuitenkin asteikon tasokuvaukseen yleensä tai sen johonkin tiettyyn yksittäiseen ilmaisuun tyytyväisiä. Esimerkiksi tekstin *yksiselitteisyyttä* pidettiin hyvänä ja sitä, että *tekstissä ei ilmene lainkaan merkkejä siitä, etteikö kirjoittaja pystyisi ilmaisemaan sitä, mitä hän haluaa ilmaista*. Molempien ilmaisujen toimivuutta perusteltiin sillä, että teksti on selkeää ja täysin ymmärrettävää lukijan näkökulmasta.

”kuutonenkin kyllä että nämä oli tarkimmat” Teija

”mutta tää yksiselitteisyys oli musta aika hyvä niinku kuvain tälle kutoselle et siinä ei niinku et siinä ei niinku tavallaan tarvii sit mitään enää arvailta vaan se niinku on ja tosiaan niinku se että et et niinku se on kyllä hirveen musta semmosta niinku aika hyvinkin lähellä nat semmosta natiivitasoo niin mä sitä ajattelin ja tosiaan tää ty tilanteeseen sopivuus tietenkkin kauheen oleellinen” Heidi

”tos kutosessa on musta se on nyt et ei ilmene lainkaan merkkejä etteikö kirjottaja pystyisi ilmasemaan sitä mitä haluaa ilmasta eli se on aika selkeästi kyllä sanottu” Tuula

Rajatapausten ratkaiseminen

Tason päättäminen niin sanotuissa rajatapauksissa oli odotuksenmukaisesti ongelmallista. Joillekin arvioijille se tuntui kuitenkin olevan suurempi ongelma kuin toisille. Esimerkiksi osa arvioijista olisi halunnut käyttää plussia ja puolikkaita riippumatta siitä, miten moniportainen asteikosta olisi niiden myötä tullut. Sama arvioija saattoi samassa haastattelussa sekä toivoa 6-portaista pienempää skaalaa että mahdollisuutta käyttää plussia ja miinuksia. Toisen arviointikierroksen haastatteluissa arvioijat kuitenkin nostivat vähemmän esille rajatapausongelmia ja/tai erityisesti rajoilla vaikuttavia kriteereita, koska pääpaino tuntui olevan enemmän asteikon tulkinnassa ja sen sovellettavuudessa.

Haastatteluissa esiin nostetuissa rajatapauksissa arvioijien ratkaisut vaihtelivat systemaattisista käytänteistä tiettyjen tasojen rajoilla erityisesti vaikuttaviin kriteereihin. Systemaattisena käytänteenä voi pitää muun muassa Tarjan toimintatapaa, jonka mukaan hän yleensä antoi rajatapauksissa korkeamman tasoarvion, koska puolikkaiden käyttö ei ollut mahdollista. Pian vastaava ratkaisu oli taas verrata tekstejä toisiinsa ja etsiä tukea aikaisemmin arvioiduista teksteistä. Eevan esimerkki havainnollistaa puolestaan sitä, miten kriteerien vaikuttavuus voi vaihdella eri tasoilla, ja myös sitä, miten arvioija pyrki jättämään suorituksen tasojen rajoille, vaikka se ei ollutkaan mahdollista. Eevan tapauksessa päätöksen tekeminen rajatapauksissa näyttää vaikealta, lähes mahdottomalta.

”noh sit mä vertasin johonkin aikasempaan niinku minkä mä olin arvioinut niinku et tää on nyt kakkonen ja tää on nyt nelonen ni tai kolmonen ni sit mä niinku vähän vertasin sitä tekstiä niihin että kumpaa se niinku lähe ois lähempänä se” Pia

”tai tai tota nin nin joku oli ykkösen ja kakkosen välillä että laitoin sitten että että tota nin nin oli paljon tekstiä että ei niinku sinänsä ykkönen ei kuvannu sitä että ois ollu tosi erittäin lyhyt ja tämmönen mutta sitten mä

en yksinkertaisesti saanu ymmärtäny siitä joihinki saato in kirjottaa et käsiala tai joihinki saato in kirjottaa että en ymmärrä viestiä tai jotain tämmösiä no just aina sillon jos niinkun niinku se jotenkin poikkes tästä jo joko niin että se oli ihan selkeesti mä olin laittanu ku ykkösen että en laittanu rajoille ja se poikkes tästä kuvauksesta hyvin paljon tai sitten että että oli niinku kahen eri numeron kuvauksesta ni sillon usein laitoin sitten rajalle” Eeva

Rajatapauksissa yleisin yksittäinen kriteeri oli rakenteet. Varsinkin ylemmillä taitotasoilla rakenteiden merkitys näytti korostuvan. Rakenteiden rooli rajatapauksissa näytti yleensä olleen sellainen, että jos kirjoittaja osasi käyttää jotakin tiettyä rakennetta oikein, hän sai korkeamman taitotason, vaikka jonkin muun kriteerin osalta teksti ei olisikaan täyttänyt tasovaatimuksia. *Rakenteilla* ei kuitenkaan viitattu aina rakenteisiin yleensä, vaan jonkin tietyn osa-alueen hallintaan, esimerkiksi aika-muotojen käyttöön tai nomien taivutukseen. Nämä olivat pitkälti samoja rakenteita, joita arvioijat arvostivat kommentoidessaan nelostason kuvausta *ilman huomattavia sanastollisia ja rakenteellisia puutteita*. Arvioija saattoi rajatapauksia koskevissa kommenteissaan viitata myös virheiden määrään: jos virheitä oli ”liikaa”, suoritus sai alemman taitotason.

”on nelonen mut sitten see että onks se nelonen vai vitonen ni mä ajattelin että siinä tietysti jos tästä antaa vitosen ni se ehkä sitten kuitenkin putoo sinne neloseen niitten kahden muun ansiosta mut jotenkin niinku tuntu siltä että siinä oli joitakin semmosia niin kerta kaikkiaan siis kun ne oli käyttäny niinku lauseenvastikkeita oikein ni mun mielestä se on sitten jo rupee oleen niinku se lauseenvastikeitten käyttö on semmonen asia että jos niitä osaa käyttää oikein nin ni se menee sitten jo tonne vitosen puolelle et vaikka siellä ihan joku semmonen ihan jotain pientä saatto ollakki” Raija

”ni musta tuntu et eniten mä varmaan kiinnitin huomiota verbeihin taas kerran ni sitten mä laskin sitä arviota alaspäin ja ainakin niinku jotenkin tänään musta tuntu jotenki et mä oisin ollu tiukempi tänään tai ei sais olla” Pirjo

”joo taivutus oli kyllä yks joka selvästi niinku erotti sen et tuota jos oli semmonen missä oli vaan sanoja taivuttamattomia peräkkäin ni ni vaikka se oiski ollu jotenkin ymmärrettävä se ni kyllä se sitten tiputti sen varmaankin sitten ykköseen vai oisko se sitten ollu kakkonen joski vähän riippuu siitä kui paljon sitä siellä on mut se varmasti niinku alens” Inkeri

”yks oli se musta se semmonen noh yks oli pituuskin yks yks tekijä kuitenkin et jos se oli kovin lyhyt siis semmonen ni ei se sitte ollu musta kolmosen vaikka nyt siitä sairaala jutusta jos nyt oli sitten ihan lyhyt nin nin et se oli yk yks semmonen ratkaseva ja sitten tuo et jos se oli semmonen kovin niinku hajanainen et ne oli vaan niinku semmonen ja si sit se oli niinku kakkonen ja sit siellä oli virheitä selvästi ihan väärin” Merja

Tasorajoilla vaikuttivat kolmantena ryhmänä subjektiiviset tekijät. Arvioija saattoi tehdä vaikeassa tilanteessa, väsyneenä tai suorituksesta innostuneena, päätöksen, joka perustui pääasiassa tunnepohjaisesti hänen omaan mieltymykseensä. Esimerkiksi tekstin hauskuus tai tekstin taso suhteessa sen ympärillä oleviin teksteihin saattoi olla tasoa korottava tekijä. Myös arvioijan oma mieliala tai vireystila saattoi vaikuttaa sen mukaan, millä tuulella tai miten väsynyt arvioija oli kunkin arviointisession aikana.

”jos siitä tuli hyvin sellasia niinku kaavamaisia ne niinku vastaukset niihin aika helposti sitten niinku sokeutu ja sitten niinku rutiinilla sitte laitto sen heti suoraan sitten jos sieltä oli ihan joku poikkeava juttu tuli ni sitten niinku että se oli ihan niinku pelastava piristys että taas siitä niinku heräs ja sitte rupes miettimään että no kumpaam mä nyt tämän laitan” Tarja

”et sit viel mä niinku huomasin et se oli ihan kauhee ku sit ne ympärillä olevat tekstit jollain tavalla kuitenkin vaikutta jos on ollu niinku pitkä aika semmosia aivan huonoja ja sit tulee joku hyvä ni sitä niinku tosi mielellään anto sille heti jonkun hyvän ettei nyt kaikki oo näin huonoja et sitä väkisin menee siihen semmoseen vaik sit niinku heti soimaa itseään et ei näin pidä tehdä mm” Tuula

”mä vähän veikkaan että tässä arviossa mä pudotin sen sitten tohon alempaan etku nää oli kuitenkin aika kovat vaatimukset tossa nelosessa vitosessa kuutosessa nin että melko varmaan mä oisin sitten ankaralla kädellä rankaisu et siinä saattaa olla kyllä jonkinlaista heitto kun mä luin kolmena eri iltana niitä ja sitten eri mielialoissa ainakin aluks olin varmaan ankara” Sanna

Arviosta päättäminen rajatapauksissa havainnollistaa hyvin arviointiprosessin monikerroksista ja subjektiivista luonnetta. Arvioija päättäjänä toimii järkipäisesti ja tunnepohjaisesti sekä myös tilannekohtaisesti (vrt. Niiniluoto 1994). Kognitiivisesta päätöksentekoprosessin näkökulmasta yhteen tai muutamaan kriteerin tukeutuminen näyttää myös perustellulta, koska lukiessaan arvioija ei pysty tukeutumaan useaan kriteeriin kerrallaan (Hogarth 1980; Vaughan 1991). Toisaalta tässä oli kysekin holistisesta arvioinnista, jolloin arvioijan ei oletettukaan lukevan tekstejä kriteeri kriteeriltä. Kaiken kaikkiaan arvioijan näkökulmasta hänen valikoivuutensa tuntuu tarkoituksenmukaiselta, koska se lienee keino selviytyä työstä, välttää kognitiivista ylikuormittumista. Miten oikeudenmukaista tällainen toiminta on arvioitavaa kohtaan ja miten luotettavana näin annettua arviota voidaan pitää, jää tietysti pohdittavaksi?

Pitkin matkaa esiin tullut arvioinnin rakennepainotteisuus ja jossakin määrin vaikuttava subjektiivisuus eivät ole kumpikaan kovin yllättäviä tuloksia kirjoittamisen arviointia koskevien tutkimusten mukaankaan

(vrt. taulukko 8). Tämän tutkimuksen arvioijat arvostivat erityisesti verbien oikeaa käyttöä ja morfologian hallintaa, mikä myötäilee koko lailla Vaughanin (1991) tutkimustuloksia, vaikka nämä koskevatkin englannin kirjoittamista toisena kielenä.

Vaughanin (1991) tutkimuksessa nousi esiin myös subjektiivisuus yhtenä lukemistyylinä. Itse en ehkä kategorioisi kielioppikeskeisyyttä, yhteen asiaan fokusoitumista tai subjektiivisuutta itsenäiseksi lukemistyyliksi, koska ne kuuluvat olennaisena osana kaikkiin lukemistyyliihin ja lukustrategioihin. Vaikka en tutkinutkaan yksittäisen suorituksen arviointia, uskaltaisin siitä huolimatta olettaa, että arvioija ei ole vain yhtä tyyppiä, vaan hän voi painottaa eri asioita eri suoritusten kohdalla ja jopa saman suorituksen kohdalla. Arvioijan suhtautuminen ei ole kiinni vain arvioijasta itsestään, vaan myös arvioitavasta tekstistä ja tilannetekijöistä, jotka vaikuttavat siihen, millaiset tekijät ovat päätöksessä kulloinkin etualalla. Arvioijina emme ole joko rationaalisia tai emotionaalisia, vaan molempia ja kaikkea. Arvioinnin onnistumisen kannalta kyse on lähinnä siitä, miten tietoisia olemme arviointiin vaikuttavista tekijöistä ja kuinka hyvin pystymme kontrolloimaan ja ohjaamaan toimintaamme.

Cummingin (1990) tutkimuksen mukaan kokeneet arvioijat pystyvät kontrolloimaan toimintaansa, mikä on nähtävissä myös tämän työn arvioijien toiminnassa. Ainakin kommenttiansa perusteella arvioijat olivat myös tietoisia siitä, miten erilaiset emotionaaliset ja vireystilaan liittyvät tekijät vaikuttivat. On tietysti mahdotonta sanoa tämän aineiston perusteella, mistä kaikista tekijöistä arvioijat olivat kulloinkin tietoisia ja missä määrin he pystyivät minimoimaan niiden vaikutusta konkreettisesti. Toisaalta arvioijien yhdenmukaisuutta osoittavien korrelaatioiden mukaan he toimivat varsin yhdenmukaisesti, mikä puhuu puolestaan – elleivät sitten muun muassa subjektiiviset tekijät ole olleet samanlaisia ja samalla tavoin vaikuttavia eri arvioijilla.

Asteikkoon ehdotettuja muutoksia

Koska arvioijat kritisoivat asteikkoa ja sen lähtökohtia kokonaisuutena sekä yksittäisten tasokuvausten toimivuutta, he ehdottivat asteikkoon monia konkreettisia muutoksia, jotka liittyivät pääasiassa koko asteikon painotuksen muuttamiseen, suoritusta kuvaavien ilmausten tarkentamiseen ja/tai yksityiskohtaisemmin kielellisten, lähinnä rakenteellisten, kuvausten tai joidenkin yksittäisten kriteerien lisäämiseen. Olen luokitellut muutosta koskevat ehdotukset kolmeen kategoriaan: 1) asteikon painopisteen muuttaminen, 2) tasokohtaiset tarkennukset ja 3) asteikon konkretisointi.

1) Asteikon painopisteen muuttaminen

Asteikkoa arvosteltiin ennen kaikkea siitä, että se oli liian tekstityyppi-

painotteinen, minkä vuoksi se ei arvioijien mielestä tarjonnut riittävästi tukea arviointiin. Kriitikistään huolimatta arvioijat eivät kuitenkaan kysyttäessäkään pystyneet ehdottamaan vaihtoehtoa olemassa olevalle asteikolle. Arvioijien ehdotukset liittyivät lähinnä yksityiskohtiin, kuten jonkin yksittäisen tasokuvauksen ilmaisun tarkentamiseen tai muuttamiseen tai joidenkin kriteerien, lähinnä kielellisten, lisäämiseen asteikkoon. Tämantyyppiset ehdotukset eivät kuitenkaan olisi ratkaisseet asteikon painopisteeseen kohdistunutta kritiikkiä. Yksi arvioijista kuitenkin ehdotti koko asteikon lähtökohdan muuttamista funktio-keskeiseksi.

”koska tässä ei oo niitä funkt nyt mä vasta tajuun itse asiassa tässä ei oo niiku funk jollain tavalla tässä ei oo funktioita niinkään et mitä mitä kielen funktioita pystyy ilmasemaan vaan tää on hirveen sillai niinku tekstityypikeskeinen tää skaala noh sen takia niitäki tuntu noh niinku asteikon tasolla noh em mä nyt tiijä olik se nyt muuten itse asiassa iso ongelma niitä sitten sinne rimmata mutta mut tavallaan että musta se vois olla monipuolisempi tai vähän erilainenkin lähtökohdiltaan toi skaalan kuvaus” Heidi

Koko asteikkoa koskevia ehdotuksia olivat tavallaan myös sellaiset muutosehdotukset, jotka koskivat jonkin tietyn kriteerin tai jonkin tietyn asian tai ilmauksen poistamista kaikilta tasoilta tai sen lisäämistä kaikille tasoille. Kommenttien sisällöt vaihtelivat Eevan maininnan yleisluonteisuudesta Heidin ja Raijan esittämään kriteerilistaan, johon tulisi ottaa kantaa yleensäkin asteikossa. Kaisun mielestä taas konkreettisia esimerkkejä pitäisi joko olla kaikilla tasoilla tai sitten ei ollenkaan.

”mä haluaisin tähän lisää niinku kuvausta tästä it itse siitä tekstistä tai siitä tuotoksesta että minkälainen se nyt tyypillisesti olis just esimerkiksi että mitä noh tähän tyyliin niinku on ykkönen kakkonen ja kolmonen tässä että se ei oo pelkästään että osaa kirjoittaa semmosta ja tämmöstä vaan että siellä on myöski että että teksti koos minkälainen se teksti on mistä se koostuu minkälaisia puutteita siinä ehkä on” Eeva

”ja sit se tekstin pituus ei pelkästään pituus vaan tietenk se että mi minkälainen se on tekstinä miten koherentti se on ja ylipäättään siis ylipäättään se että on pystynyt tuottamaan semmosen määrän tekstiä ja tota ja no sitten joissakin tapauksissa tosiaan tää tämmönen niinku tilanteeseen sopivuus niinku jossain suruvalittelussa ja kirjeessä ja tämmösissä ja se sidosteisuus” Heidi

”mä ajattelin että kirjottaminen on niinku sitä että ton pituuden ja lyhyden mun mielestä että et täytyy jos kerta kaikkiaan teksti on kauheen lyhyt vaikka se on kuinka hyvä ni miten se tehdään tai sitte taas contra se että jos se on kauheen pitkä ja se on sujuva erittäin sujuvakin mulla oli muutama todella hyvä sujuva pitkä teksti mut että siellä on aivan maho

taikka no sillai sanotaan että joka lauseessa muutamia kielioppivirheitä ni sen arvostelu on aika vaikeeta sitten see että et millä tavalla me otetaan huomioon puhekielisyys koska tuossa arv näissähän ei puhuttu siitä mitään et millä tavalla siihen niinku suhtaudutaan yleisesti sitten tota noin ni välimerkithän on semmonen jotka arabi arabi maailmasta tulevalta ihmiseltä puuttuu kokonaan ja mun mielestä sekin täytyy sopia toi että joko jos ne puuttuu kokonaan tai sitten esimerkiksi niinku joillakin venäläisillä on se että ne käyttää tietysti niinku venäläistä välimerkkisysteemiä jollon ne pilkku ja pilkkuja tulee paljon enemmän kuin suomessa tai sitten englantilaiset joilta puuttuu pilkut kokonaan ni kyllä musta siinäkin täytyy sitten miettiä että että tota noin ni et miten niihin siihen välimerkki-systeemiin suhtaudutaan” Rajja

”mun mielestä jo joko täs kakkosessa pitäis niinku poistaa nää esimerkit tai sit kaikkiin lisätä jotain esimerkkejä semmosia että se et et tavallaan ne kriteerit ois niinku samalla tasolla että ois niinku ylemmällä tasolla et ois sellaset hier hierarkia kriteerit jotka sitten jatkus ne asiat alusta loppuun” Kaisu

2) Tasokohtaiset tarkennukset

Hyvin samantyyppisiä muutosehdotuksia edellisten kanssa olivat tasokohtaiset tarkennukset. Tosin ne kohdistuivat vain tietylle tasolle. Arvioijille oli tyyppisempää esittää tasokohtaisia kuin koko asteikkaa koskevia muutoksia. Tähän tietysti osaltaan vaikutti se, että arvioijat olivat tyytyväisiä joihinkin yksittäisiin ilmauksiin tai sitten jonkin tietyn tason kuvaukseen. Tasokohtaisissa tarkennuksissa toivottiin lisää kuvausta tietyn tason suorituksista. Sen lisäksi niissä puututtiin joidenkin yksittäisten ilmausten sanavalintaan, esimerkiksi vaadittiin *toiston* korvaamista jollakin muulla ilmauksella. Jotkut arvioijat halusivat muuttaa nelostason kuvauksen kokonaan sisällölliseksi poistamalla siitä rakenteisiin ja sanastoon liittyvät viittaukset, ja toiset puolestaan halusivat vaihtaa *esimerkiksi*-ilmauksen viitostasolta kuutostasolle.

”näistä ykkösestä ja kakkosesta mä niitten kans niin vähän siinä touhusin että muuta että ku toivosin niihin lissää kuvausta että et niinku tarkemmin se ero että mikä ero on esimerkiksi ykkösen ja kakkosen välillä” Teija

”noh olis ehkä pitäny jättää pois joku tällänen just tää lause että ei mitään sanastollisia tai rakenne rakenteellisia puutteita vaan et se ois ollu niinku puhtaasti sisällöllinen tai tietenkin tyyli tai aihepiirisidonnainen tai jotenkin et et kus siinä tässä oli niinku vähän sieltä täältä joku kommentti nin sitten toinen esti toisen kun mä oisin muuten ajatellu että se on vaikka kyllähän tää nyt kirjottaa tosi hyvin ja tää on varmasti vitonen nin sit siellä onki jotain katastrofaalisia virheitä ni emmä voinu sitä nostaa sinne vaikka varmasti sen henkilön kielitaito olis ollu kuitenkin ihan kohtuullisen lähellä” Sanna

”no ainakin ton toistoa sieltä kolmosesta ehkä siihen pitäis keksiä joku toinen sana sitte mitä se on et mitä mitä ne sanastolliset puutteet aiheuttavat kyllä ne aiheuttaa jotakin mut ei se kyl toistoa ole... toi abstrakti mua kyllä kans vähän kaiveli että oisko se sitten kuuluuko se tähän kolmoseen että kuitenkin joku essee ja raportti mun mielestä ne on niitä abstrakteja et se ois niinku pikemminkin nostettava ehkä tonne ylemmälle täs lähtee toi kolmonen mun mielestä vielä niinku ehkä oo semmonen”
Tuula

3) *Asteikon konkretisointi*

Suurin osa arvioijista piti asteikkoa liian ylimalkaisena, joten siihen liittyvät lisäykset koskivat ennen kaikkea asteikon konkretisointia. Omaksi yhtenäisimmäksi ja myös yleisimmäksi ryhmäkseen voi erottaa rakenteelliset muutosehdotukset ja/tai lisäykset. Niille kaivattiin näkyvämpää roolia asteikossa. Rakenteelliset lisäykset liittyivät joko yleensä kielioppikriteerien lisäämiseen tai yksittäisten rakenteiden, kuten taivutusten ja aikamuotojen, sisällyttämiseen asteikkoon.

”mä niinku kaipaisin tähän rinnalle jotain jotain ehkä jonkunnäkösiä kielioppikriteereitä...sitten justinsa miksei sanaston kannaltakii että et minkäläinen se sanasto sitten pitäis olla mitä sanoja hallita että kyllä sen huomaa et jos jotkut osaa hyvin käyttävät kuvailevia adjektiiveja tai jotain abstrakteja sanoja tai muuta ni ni kyllähän niinku sillä tavalla niihin kiinnitti kyllä huomiota että sitten saatto niinku neloseen nostaa sitä” Pirjo

”mut ehkä et enempi vois olla semmosta phh et olis jotain semmosta vaikka että et niinku pitäis hallita vaik jo taivutukselliset perusasiat tai just noh täs on nyt jotain täs oli näitä tää oli must ihan kiva että täs kuitenkin oli tätä syntaksii oli vähän käsitelty...vaik niinkun että ymmärtääkö niinku tapaileeko verbin taivutusta vai jättääkö verbit kylmästi aina taivuttamatta ja noh tossa on ykköses on see et ehkä u ulkoo opittui nih jotain semmosta semmost niinku morfologista tähän vähän ois kaivannu sitä ei täs oo oikestaa yhtään”
Martti

”no jos sä nyt laittaisit vaikka tänne taikka laitettas että täällä oli tavallisimmat konjunktiot ja mutta koska ni sitten jos täällä olis nyt sitten lueteltu lisää konjunktiota vaikka kolmosessa et olis niinku et oli osattu käyttää sitten vaikkapa jotta nii tää nyt ois ihan hullua varmaan mutta kuitenkin kuitenkin siis ideana jos olis esimen enemmän niinkun esimerkiks esimerkkejä et mitä on osattu käyttää tai ei oo käytetty” Merja

”vois olla yksityiskohtasempikii jotenkin se kriteeristö semmonen yksityiskohta että esimerkiks niinku jo mm mun mielestä suomen aikamuodotki tai kaikki maailman aikamuodotki vois olla jo jossain kohti hallussa tietyllä tasolla” Kaisu

Kaiken kaikkiaan asteikkoa kritisoitiin siitä, että se oli liian tekstityyppikeskeinen, mikä edelleen vaikutti siihen, että asteikkoon kaivattiin suorituksen laatua paremmin kuvaavia, lähinnä kielellisiä, kriteereitä tai kuvauksia. Arvioijien esittämä kritiikki on odotuksenmukaista suhteessa asteikon käyttäjakeskeiseen tarkoitukseen (vrt. Alderson 1991). Toisaalta asteikkojen yleinen ongelma on niiden abstraktius. Esimerkiksi DeRemerin (1998) tutkimuksessa (ks. taulukko 8) arvioiden ei katsottu heijastelleen validisti tekstin piirteitä, koska asteikko ei kuvannut tarpeeksi konkreettisesti tekstin piirteitä.

Abstraktiuden ja ylimalkaisuuden lisäksi asteikkoa pidettiin myös liian osaamiskeskeisenä, mikä voi vaikuttaa osaltaan siihen, että arvioijat tulksivat ylimmät tasot todella vaativiksi tasoiksi. On hieman eri asia ajatella, mitä kaikkea kirjoittajan täytyy osata, kuin pohtia, mitä kaikkea kirjoittajan ei tarvitse vielä osata tai millaisia puutteita hänen kieli-taidossaan sallitaan, vaikka hänen suorituksensa sijoittuisikin taitotasolle 6. Voi olla, että jälkimmäinen tapa hahmottaa ylimmät tasot olisi ollut oppija- ja ehkä arvioijaystävällisempi.

Asteikon soveltuvuus eri tehtävätyyppien arviointiin

Koska arviointiasteikkoa pidettiin tekstityyppilähtöisenä, sen voisi olettaa sopineen hyvin ainakin tältä osin erilaisten tehtävätyyppien arviointiin. Sopiko se silti kaikkien tässä arvioitujen yhdeksän erilaisen ja eritasoisen tehtävätyypin arviointiin? Edellähän kävi jo ilmi, että joidenkin arvioijien mielestä kaikilla tehtävätyypeillä ei voinut osoittaa ylimpien taitotasojen, varsinkaan taitotason 6, tasoista kirjoittamista. Näiden arvioijien mielestä esimerkiksi ostoslistatehtävä oli tällainen tehtävätyyppi. Myös vaativimpien tehtävien, työpaikkahakemuksen ja mielipidetekstin, kykyä erotella viitos- ja kuutostason kirjoittajia epäiltiin.

Suruvalittelua sen sijaan kritisoitiin siitä, että se ei joidenkin arvioijien mielestä sopinut kolmostasoa alemmille tasoille. Sitä pidettiin muun muassa sanastollisesti vaikeana. On selvää, että ostoslistaa ei voi arvioida tasokuvaustenkaan perusteella kuutostason tekstiksi, mutta suruvalittelun voi arvioida ykkösen suoritukseksi, vaikka se onkin määritelty C-tason (taso 5 ja 6) tehtäväksi. Ostoslista ei tekstityyppinä sisällä näkökulmaa korostavaa esseemäistä tekstiä, eikä siinä ole kyse monimutkaisesta aiheesta. Suruvalittelussa voi kuitenkin olla jo ykköstason kirjoittajalla joitakin yksittäisiä aiheeseen liittyviä fraaseja, esimerkiksi *olen pahoillani* tai *minä surullinen*, tai esimerkiksi viestin aloitus tai lopetus.

Arvioijat kritisoivat erityisesti sitä, että arvioitavat tehtävät eivät sisältäneet raportin ja esseen kirjoittamista, kuten asteikon tasokuvaukset edellyttivät. Toisaalta he eivät myöskään maininneet, että asteikossa viitattiin taitotasolla 5–6 myös yksityiskohtaisesti kuvailevaan kirjeeseen,

lukijaa helpottavaan loogiseen, tekstin pääasioita korostavaan rakentamiseen, tiivistelmään ammatillisesta aiheesta, oman näkökulman korostamiseen (essee), joita kaikkia C-tason tehtävistä työpaikkahakemus ja mielipideteksti edellyttävät tavalla tai toisella. Toisaalta tehtävätyyppien vaativuuteen vaikuttaa ennen kaikkea se, mitä niiltä odotetaan. Mielipideteksti voi olla kolme riviä ja koskea liiallista makkaran syöntiä ja oluen juontia, mutta toisaalta esseekin voi vaihdella koevastauksen tyyppisestä tekstistä pitkähköön lähteiden pohjalta kirjoitettuun akateemiseen pohdiskeluun. Toisin sanoen epäselvyys voi johtua pitkälti siitä, miten arvioija määrittelee mielessään teksti- ja tehtävätyypit.

”taustalla näkyy sellanen juttu että tavallaan verrataan niinku semmosiin ylempiin tehtävänantoihin että pystyisköhän tämä ihminen johonkin noh mistä me tiedetään ku mejän pitää arvioida se tuotos eikä sitä että pystysköse kirjottaa laajahkon esseeseen jossa se rakentelis sitten niinku vaille jotain nin nin tota se ehkä jotenkin vaikuttaa siellä taustalla sit kuitenkin että jos se ei suoriudu tästä yksi yksinkertaisesta tehtävänannosta ni se ei varmaan suoriudu siitäkää” Inkeri

*”sit tuol nelosessa just ku aattelee että toivotan informoivia ja argumentoivia esseitä ja raportteja ja sitku miettii joku raporttiki et niinku miten semmonen toimii että niinku kuka osaa voisko tän kirjottaa ja ajatella et sitten tavallaan laajentamaan sitä ajatusta jos se on kirjottanu nyt tämmösen teksti ni voiskohan tää sama kirjottaa vaikka raportin jollain tapaa sitte aina siinä niinku mietti et ei tää kyl varmaan yltäis siihen ni sit se on kolmonen esimerkiksi et mä ajattelin aika paljon ihan täl tavalla”
Tuula*

Arvioijista suurimman osan mielestä asteikko sopi hyvin johonkin tiettyyn tehtävätyyppiin ja johonkin toiseen taas ei. Muutaman mielestä asteikko toimi ihan hyvin kaikissa tehtävätyypeissä, koska ”em mä sillä lailla niinku arvioinnin kannalta ni en em mä osaa sanoo et ois ollu mitään eroo.” Muutaman arvioijan mielestä tehtävätyyppien välillä ei ollut oikeastaan mitään eroa, joskin jotakin tehtävätyyppiä saattoi olla vähän helpompi arvioida tällä asteikolla.

Yksittäisistä tehtävätyypeistä asteikko tuntui sopivan parhaiten kirjeen arvioimiseen: ”musta se henkilökohtanen kirje oli oikeastaan aika hyvä se hajotti tän asteikon aika hyvin” tai ”jotenkin ne kirjeet oli sillä tavalla helpompia että ei tätä tarvinnu sentäs joka paperin jälkeen kattoo että miten et siinä jotenkin oli helpompi hahmottaa se semmonen jatkumo.” Asteikko ja tehtävätyyppi tuntuivat sopivan toisiinsa sitä paremmin, mitä selvemmin kuvaukset sopivat suorituksiin ja mitä laajemmin suoritukset sijoittuivat asteikolla, mikä näkyy seuraavasta Tuulan lausumasta. Se myös havainnollistaa sitä, että toisella arviointikierroksella suruvalittelua pidettiin melko hyvin erottelevana tehtävänä.

”mä aattelin et kauhee mikä tehtävä oli se suruvalittelu juttu ja tuota et sitä niinku et se jotenki oli menny kauheen huonosti mut nyt oikestaan nää taas nää sisällöt autto mun mielestä sen arvioinnissa jollain kummallisella tavalla paljon enemmän nyt se skaala niinku tavallaan tuli paljo kirjavammaksi että tavallaan siinä nää jopa autto...tää oli mun mielestä aika kiva tää kirje ystävälle et se niinku myös oli hyvin semmonen et siin niinku todella sai ihan laajasti kaikenlaisia numeroita et sillä tavalla vaik tääl oli sanottuki se että niinku osaa kertoa itsestään ja henkilökohtasista asioista mut silti niinku jos siinä onnistu hyvin ni se oli sit kyl tosi hyvä ja sit taas tai siit näki ehkä se on niinku vois sanoa et se on erotteleva mun mielestä” Tuula

Ongelmallisista tehtävätyypeistä puhutti eniten työpaikkahakemus, lähinnä siitä näkökulmasta, miten tärkeänä hakemuksen muodon hallintaa tulisi pitää. Osa arvioijista pohdiskeli, oliko muodollisten seikkojen huomioon ottaminen yleensäkin tärkeää ja missä määrin. Osa taas kritisoi sitä, että asteikko ei tarjonnut tältä osin mitään tukea päätöksentekoon. Tukea päätökselleen arvioijat hakivat siitä, miten suomalainen kirjoittaisi vastaavan tekstin, tai siitä, miten puutteellinen työpaikkahakemus otettaisiin vastaan todellisessa elämässä. Pohdiske-luissa mietittiin myös, millainen kuutostason kirjoittaja olisi ja tulisiko hänen hallita myös hakemuksen muoto.

Arvioijat olivat yleensä ratkaisseet ongelman niin, että he edellyttivät perusasioiden hallintaa, esimerkiksi päiväystä, sopivaa aloitusta ja lopetusta: ”mä sitten aattelin että vitosella niinku vois vaatia että ois se muodollisesti sitten niinku oikein aseteltuna” tai sitten niin, että he eivät kiinnittäneet huomiota muotoon: ”ajattelin et mie luen niinku luen sen tekstin ja ja sitten et tekstinä yritän sitä arvioida enkä sitä et onko se osannu nyt laittaa päiväyksen ja puhuttelun ja muistanu laittaa oman nimensä.”

Muita arvioinnin kannalta ongelmalliseksi katsottuja tehtävä-tyyppejä oli muun muassa tarina kuvasta, jota ensimmäisellä arviointi-kierröksellä pidettiin helppona tehtävätyyppinä. Tällä arviointikierrök-sellä arvioijat eivät enää olleet kovin varmoja siitä, miten hyvin määrittely asteikko sopi tämän tehtävän arviointiin. Esimerkiksi Martin mielestä tarinan arviointi oli vaikeaa, koska tehtävän instruktio oli liian vapaa, vaikka muuten hänen mielestään tarinatekstejä oli mukavinta lukea. Teijan mielestä tarinatekstejä oli taas vaikea arvioida sen vuoksi, että tarinoita kirjoitettiin kirjoittamisen vuoksi: ”tuo tarina oli just ongelmal-linen siinä että et se oli miun mielestä tällä kierroksella vaikein tähän asteikkoon taikka niin tähän suhteuttaa...että et se ei ollu semmonen luonteva niinku oikea kielenkäyttötilanne.”

”itse asiassa tulin siihen tulokseen että et hemmetti et sehän on ihan yhtä niinku yhtä n mun mielestä se vois olla vaikka ja i ihan niinku joku joku tietynlainen niinku diskurssitehtävä enempi mut ei niinku tota ei sellanen sellasenaan ainakaan näillä kriteereillä siis sehän menee niinku aivan ohi se koko tehtävistä tosta ja sillon siihen ei voi soveltaa mitään mitään argumentoivaa informoivaa loogisesti etenevää näitä tälläsii sääntöjä mut että sen aa aa niitten” Kaisu

Ostoslistatehtävä herätti kommentteja lähinnä siksi, että se ei yksinkertaisuudessaan sopinut kuin asteikon alemmille tasoille, joille se tosin oli suunnattukin: ”sitten tässä nelosessa tää osaa kirjottaa muunmuassa informoivan tai argumentoivan esseen tai raportin ja tää on sit taas niin kaukaa haettu mun mielestä tää koko lause siihen mitä minä arvioin ruokalista sanastoo.” Vaikka tehtävä oli viestipohjainen eli tehtävänantoon kuului, että ostoslista oli osa lyhyttä viestiä, arvioinnissa tunnuttiin kuitenkin kiinnittävän huomiota vain itse listaan ja siihen, missä määrin kirjoittajilta vaadittiin partitiivin käyttöä.

Arvioijien mielipiteet vaihtelivat siinä, edellyttivätkö he ostoslistatehtävältä viestin tapaan aloitusta ja lopetusta. Joidenkin arvioijien mielestä suomalaisen ostoslista saattaisi yhtä hyvin olla myös perusmuotoisia sanoja ilman määriä (*maito, leipä* jne.) ja myös ilman puhuttelua, vaikka se olisi jätettykin viestinä pöydälle: ”ei mun tartte kirjottaa siihen ystävän nimekään jos mä tiedän kelle mä jätän jos mä asun jonkun yhden ihmisen kans jolle mä jätän niinku tän et et se on se on erittäin vaikee tehtävä kirjottamisen arviointiin.”

Kaiken kaikkiaan arvioijien kommentteista päätellen arviointiasteikon lisäksi tehtävyyppikohtainen ohjeistus olisi ollut paikallaan. Eri tehtävyyppien tulkinta ja arviointi olisivat saattaneet olla yhdenmukaisempia ja yksiselitteisempiä, jos mukana olisi ollut tarkempi kuvaus siitä, mitä kunkin tehtävänannon täyttämislle edellytetään. Pelkkä kirjoittajan saama tehtävänanto ei siis näyttäisi riittävän arvioinnin tueksi. Lisäksi maamerkkien eli esimerkkitekstien mukanaolo olisi voinut helpottaa suoritusten sijoittamista asteikolle. Toisaalta arvioijat olivat kuitenkin arvioineet suoritukset melko yhdenmukaisesti (ks. luku 5.1.3). Olisi mielenkiintoista tietää, olisiko tulos ollut vielä parempi vai jopa heikompi, jos arvioijia olisi ohjeistettu enemmän ja heillä olisi ollut maamerkkitekstejä käytettävissään.

5.2.2.4 Arvioija keskiössä

Vaikka arvioijilla oli toisella arviointikierroksella käytössä määritelty asteikko, he kokivat päätösten olevan liikaa heidän ohjaksissaan. Arviointi tuntui siis liian subjektiiviselta. Subjektiivisuuden tunnetta – ja

siihen kietoutuvaa epävarmuutta ja varmuutta – voisi eritellä seuraavasti sen mukaan, miten subjektiiisuuden tunne liittyy arvioijien kommenttien perusteella heidän toimintaansa:

1) *Arvioijan vastuu*

- Missä määrin arvioija katsoo vastuun olevan hänellä itsellään?
- Missä määrin se on taas jollakin ulkopuolisella arvioinnin osalla, lähinnä asteikolla?

2) *Subjektiiisuuden inhimillisuus*

- Missä määrin subjektiivisuus on osa arviointiprosessia, jonka olemassaoloa ja sen aiheuttamaa ahdistusta ei voi kieltää tai poistaa kokonaan?
- Missä määrin objektiivisuuden ihanne voi totetua?

3) *Arvioija geneerisenä lukijana*

- Missä määrin arvioija tekstin lukijana pystyy tekemään johtopäätöksiä kaikkien lukijoiden puolesta?
- Missä määrin arvioija tietyn tekstin lukijana pystyy tekemään yleistyksiä kirjoittajan taidosta muissa kirjoittamiskonteksteissa?

Toisella arviointikierroksella arvioijat kaipasivat asteikolta systemaattista tukea, mutta haastattelut osoittavat, että asteikko ei vastannut tältä osin arvioijien odotuksia. Vaikka kysymykset siitä, ”mikä minä olen päättämään”, eivät olleet yhtä yleisiä ja reaktioiltaan voimakkaita kuin ensimmäisellä kierroksella, arvioijat epäilivät itseään. Vaikka asteikkoa ei pidetty aivan toivottomana, se ei kuitenkaan tuntunut auttaneen pitämään omaa toimintaa tarpeeksi objektiivisena. Missä määrin vastuun täytyy olla ja se yleensäkin voi olla asteikolla? Missä määrin kyse on asteikosta ja missä määrin ehkä jostakin muusta? Seuraava Pian repliikki havainnollistaa arvioijan opetuskokemuksen vaikutusta varmuuden tunteeseen. Hänellä on kokemusta pääasiassa perustason kurssien opettamisesta.

”noh kyllä se meillä nii no enpä mä osaa oikein sanoa ku musta tää on välillä nii tulkinallista kyllä mä noista a alatasoissa niistä pysty sitten niinku lohduttaa itteensä et kyllä tän on pakko olla tämmönen että mut kuitenkin hirveen subjektiivistaha tää on että oikeestaan aika pelottavan subjektiiivista joo” Pia

Raijan lausuma osoittaa puolestaan, miten nimenomaan asteikko tuo varmuuden, koska vastuu on asteikolla ja se takaa oikeudenmukaisuuden. Jos taas asteikko on huono, se on systeemin eikä esimerkiksi arvioijan vika. Tähän näyttäisi liittyvän sellainen ajatus, että asteikon

soveltaminen ei suoranaisesti olisi arvioijakohtaista eikä subjektiivista, jos asteikko vain on hyvä ja perusteellinen; subjektiivuus, taidon yksilöllinen tulkinta, kuuluisi ikään kuin vain omien kriteereiden käyttöön.

”kun on selkeet kriteerit ni onhan se niinku sitten jotenkin tuntuu siltä että et mä oon ollu niinku oikeudenmukanen ainakin tämän systeemin mukasesti...nii ja sil on tavallaan vastuu on sillä systeemillä joo nii nimenomaan siltä että se siirtyy sinne sitten se ja ja ja oppilas voi sitten tavallaan arvos niinku tavallaan sitten sen voi kritisoida ainoastaan sitä systeemiä eikä ehkä niinkään sitä yksittäistä opettajaa joka sen on katonnu koska se on kuitenkin noudattanu niitä ohjeita mitä sille on annettu” Rajja

Arviointiprosessiin liittyvää subjektiivisuutta voidaan pitää karkeasti kategorisoiden joko arvioinnin ikuisuusongelmana tai luontevana osana arviointitoimintaa. Osa arvioijista kertoi kokeneensa omat mieltymyskensä ja ärsytyksensä ahdistavina tekijöinä, joista piti olla tietoinen ja joita piti yrittää välttää. Tällaiseen suhtautumiseen liittyi hyvin vahvana myös epävarmuus omasta toiminnasta ja sen eettisyydestä. Tällöin ei myöskään oltu kaukana syyllisyyden tunteesta. Osa subjektiivisuuden esiin nostaneista arvioijista puolestaan totesi pohdiskeltuaan asiaa, että näin se vain on ja sille ei voi mitään. Se, miten asian tiedostaneet arvioijat ohjasivat tai kontrolloivat subjektiivisten tekijöiden vaikutusta, saattoi arvioijien kertoman mukaan vaihdella tilanteittain ja hetkittäin. Aineiston perusteella ei voi kuitenkaan tehdä päätelmiä siitä, miten arvioijat todellisuudessa toimivat.

Kuten seuraavista arvioijien lausumista näkyy, tälläkin arviointikierroksella omat mieltymykset ja ärsytykset liittyivät usein arvioijan omaan taipumukseen kiinnittää huomiota johonkin tiettyyn tekstin piirteeseen, esimerkiksi käsialaan tai sisällön mielekkyyteen, tai sitten tekstin kautta tehtäviin päätelmiin kirjoittajan taustasta. Arvioijien reaktiot liittyivät lähinnä kirjoittajien lähtökulttuuriin, äidinkieleen, koulutukseen ja yleiseen kielitaitoon, joista arvioijat tekivät päätelmiä muun muassa tekstin asettelun, lauserakenteiden, käsialan ja tekstin sisällön perusteella. Tämänäyttävät reaktiot olivat tavallisesti tunnepohjaisia, ja niiden skaala tuntui vaihtelevan melko neutraaleista huomioista vähän voimakkaampiin, lähinnä ärsyyntymisen ja kyllästyksen tunteisiin. Mielenkiintoista näissä tunnereaktioissa oli se, että ne tuntuivat syntyvän melko pienenoloisista asioista, kuten siitä, mihin kohtaan kirjoitustilaa kirjoittaja oli sijoittanut tekstinsä.

”mikä mä oon näitä räpimään oikein [aurua] vähän niinku sem menee vähän liian henkilökohtasuksiin nää mun ajatukset välillä että et se ehkä siinä tekee semmosen niinkun että silleen kylmän viileesti aina jaksanu suhtautua että et nää herättää tunteita nää tekstit kaikennäkösiä että must

se kyllä voi alkaa vaikuttaa tohon arviointiin jos antaa mennä mukana itensä siinä kun jos pääsee kuitenkin kurkistamaan joissakin aika niinku mun mielestä niinku aika henkilökohtasiin asioihin monissa noista ni sit tuntuu välillä vähän niinku pahalta mennä arvioimaan tai silleen mut se on tuommosta normaalia näihän on aina” Tarja

”arviointiin vaikutti mielenkiintosesti myös se et semmonen oma samanmielisyyys että kyllä jotenkin niinku sitä taipuu pitämään niinku jotenkin niinku monipuolisempänä tai niinku tekstinä parempana jos itse jotenkin on niinku samaa mieltä et ku itse asiassa harvoin sit näkee sellasia tekstejä joista itse olis eri mieltä mut jotka ois niin hyviä joista tulis semmonen vaikutelma et tää on kuitenkin hyvin kirjoitettu” Heidi

”niinku rupes ärsyttää loppu loppu tota nin nin osasta jo ne venäläiset ja ja sitten mä vieläkin mietin niitä et onks tää kuka tää on tää on varmaan semmonen afrikkalaispoika joka on tää on taas niinku luonnonmenetelmällä oppinu et siis semmosia ihan niinku tyhmiä tyhmiä niinku miettimisiä siitä itse kirjottajasta ja ehkä jollakin tavalla niinku ku tässä näissä arviointiasteikoissa tällä kertaa oli niinkun tätä kirjottamisen intentiota et mihin se pyrkii e kirjottaja niitten omien asioiden ilmasemiseen esimerkiksi ni se varmaan vei vielä enemmän niinku siihen että et rupes miettimään sitä että onko semmosen ihmisen maailmassa tämmöset asiat tärkeitä henkilökohtasia kiinnostavia et niinku siihen tavallaan siihen sanottavaan rupes enemmän kiinnittää huomiota” Päivikki

”nytkin toi esimes toi käsiala on mun mielestä semmonen asia jota on kauheen vaikee kyllä nii tota unohtaa sitä huonoo käsialaa että siin on kyllä siitä on niin vaikee kyllä päästä yli että sitäkin vaikka täytyy noudattaa tietysti sen aikaa kun se on yhdellä kertaa selkeästi luettavaa ni pitäis unohtaa kokonaan se semmonen että se hapertaa niinku rivillä mutta mut ei sille must mahda oikein mitään et siinä tulee vähän semmonen tunne et ku sen näkee sen ensisilmäyksen et se on huono käsiala ni tulee semonen tunne että tää ei voi olla ainakaan kutonen mulle ainakin tulee semmonen tunne välillä että et tää on ilman muuta vaik sehän saattaa ihan hyvin olla että joku suomalainenkin saattaa kirjoittaa hyvin huonol käsialal et” Raija

”nii niistä joo et et et ku vois voihan sitä ihan kuvitella et se on eihä se käsiala nyt niin että et se ehkä se on niin että ne toleen noin ne ei tunne ollenkaan ketä ne on et se ei et siin ei pääse mikään pörstäkerroin vaikuttamaan kyl se kuitenkin ikään kuin vaikuttaa ehkä tässä tapauksessa se on vielä pelottavampaa sikäli että tässähan ei todellakaan näe sitä ihmistä vaan tässä näkee sen ryhmän että et vaikka vaikka se kirj kirjottaja ei yhtään sanois olen venäläinen nainen nin sen näkee siitä käsialasta et niinku mie sanoin just se että et kyl jossakin tuli mieleen et et tai niinku tää persoona etku se oli kirjoitettu silleen niin hassusti että ku tässä on tää tää tää laatikko ni se oli tungettu jotenkin tonne noin ja se rupes niinku ärsyttää se ihan vaan se se asettelu siihen paperille” Tiina

Koska toisen arviointikierroksen haastatteluissa korostuivat asteikon käsittely ja siihen kohdistuva kritiikki, arvioijat toivat ehkä määrällisesti vähemmän mutta kuitenkin emotionaalisesti voimakkaammin esille subjektiivisia olotilojaan kuin ensimmäisen arviointikierroksen haastatteluissa. Kommenttien emotionaaliseen laatuun ja määrään voi olla useita syitä. Ehkä arvioijat tunsivat haastattelutilanteen tutummaksi ja olonsa vapautuneemmaksi toisella haastattelukerralla. Ehkä samojen tekstien lukeminen toisen kerran kirvoitti esiin sellaisia tuntemuksia, joita ei voi syntyä yhden arviointikerran jälkeen. Voi olla myös, että valmiiksi määritelty asteikko ei tuntunut niin henkilökohtaiselta, mikä vapautti arvioijia jollakin tavalla. Ehkä kyse olikin vain siitä, että itse asteikko yhdessä muiden tekijöiden kanssa ohjasi vaikutelmia tähän suuntaan.

Ikään kuin subjektiivisuutta suodattaakseen arvioijat kiinnittivät huomiota omaan rooliinsa geneerisenä lukijana. Kuten ensimmäisellä arviointikierroksella, niin myös toisella kierroksella arvioijat epäilivät sitä, missä määrin heidän ymmärtämisensä edusti yleensä suomalaisten kykyä ymmärtää vastaavanlaisia tekstejä. Toisaalta on vaikea sanoa, onko ymmärtämisessä kyse geneerisen lukijan tuskasta, yleensä arviointiin liittyvästä epävarmuudesta, tasoarvioista päättämiseen liittyvästä vastuusta vai siitä, että puutteelliseksi katsottu asteikko jätti liikaa tilaa oman työn pohdiskelulle.

”että että mistä helvetistä mä voin tietää onko tää suomalaisen mielestä siis onko onks mä nyt enää mikään arvioija siis suomalainen vaan oonks mä niinku joku hemmetin kielitieteilijä joka selittää niinku että noh tämä johtuu nyt tästä ja tässä nyt on tämä on ihan varmasti ymmärrettävä suomalaisen mielestä tää niinku sitä että tekis kauheen mieli niinku verrata niitä omia ajatuksia johonki tavallisen aa sähköasentajan käsitykseen semmosesta tekstistä että et ite tuns semmosta niinku riittämättömyyttä alkupuolella siihen että että niinku et onkohan tää nyt itte asiassa ollenkaan niinku mä ajattelen tän olevan pitäns niinku vähän peilata jon jonkun tavallisen keittäjän käsitykseen” Päivikki


Mielenkiintoista on joka tapauksessa se, miten arvioija kielen ammattilaisena ei koe olevansa tavallinen suomalainen, vaan heistä irtaantunut asiantuntija, joka sanoo, mitä suomalainen ymmärtää: ”mä ajattelin et se on sit mun nii tää on just sitä et et oonks mä nyt vaan tämmönen joka päättää et asiat on näin vai onko se oikean suomalaisen mielestä sitten ihan toisin.” Yhtä kiinnostavaa on todeta, että kukaan arvioijista ei esittänyt – ainakaan eksplisiittisesti – päinvastaista näkökulmaa, jonka mukaan arvioijat eivät ymmärtäisi tai saattaisivat olla ymmärtämättä esimerkiksi kielellisten puutteiden vuoksi sellaista tekstiä, jonka ns. maallikkolukija ymmärtäisi ja hyväksyisi. Arvioijien epäilythän


näkyivät lähinnä oman toiminnan vertaamisena toisiin arvioijiin tai oletettuihin yleisesti hyväksyttäviin arviointitapoihin.

Kaiken kaikkiaan merkille pantavaa arvioijien subjektiivisuuteen, päätöksentekoon ja geneeriseen lukijuuteen liittyvissä tuntemuksissa ovat epävarmuus ja jossakin määrin vastuun aiheuttama ahdistus. Toisaalta tällaisten tuntemusten vastapainona esiintyy myös jonkin verran varmuutta, tyytyväisyyttä ja viihtymistä tekstien lukijana: ”on niin mukavia kirjoituksia et ihan haluais olla tekemisissä näitten kirjottajien kanssa mut et ihan ihan mukava seurata ja nähä että jotkut pääsee hirmu pitkälle siinä kielessä josta ne jotkut on asunu hirmu kauan maassa ja ne ei oo oppinu hyvin mut jotkut toiset on oppinu hirmu hyvin että antasin niinku semmosta rokkasua tähänki arkityöhön mitä tekkee näihen maahanmuuttajien kieltenopettaja.” Arviointiprosessista syntyvä mielikuva ei siis pysähdy tunneatkumolla neutraaliuden kohdalle, vaan liikkuu myönteisten ja kielteisten tunteiden välillä.

5.2.2.5 Toisen haastattelun satoa


Monet arvioijat odottivat toisella arviointikierroksella käytettävältä asteikolta helpotusta arviointiin, mutta sitä se ei tuonut kaikille. Osalle arvioijista toinen kierros oli helpompi mutta se oli myös osalle pettymys, koska asteikko ei vastannut odotuksia. Arvioijien mielestä asteikko ei soveltunut tarpeeksi hyvin arviointiin, se ei poistanut epävarmuutta, eikä se myöskään vähentänyt subjektiivisia vaikutelmia. Asteikko kuitenkin toi arvioijien mukaan jonkin verran varmuutta ja ohjeisti arviointia. Asteikkoon kohdistuva kritiikki myös ilmentää osaltaan, että arvioijan asema oli muuttunut ensimmäisestä arviointikierroksesta.

 *Asteikon sanalliset kuvaukset vaativat neuvottelua merkityksistä.*

 *Sanalliset kuvaukset voivat vähentää epävarmuutta, mutta eivät poista sitä.*

Kritiikissä asteikkoa pidettiin liian monitulkintaisena ja tekstityyppi-lähtöisenä, mikä on odotettavaakin asteikon käyttäjakeskeisyyden vuoksi. Arvioijien kommentteissa, muutosehdotuksissa ja hyvää kirjoittamisen taitoa kuvaavissa kriteereissä nousivat esille kaikkein yhtenäisimmän rakenteelliset seikat: arvioijat tuntuivat arvostavan kielen rakenteiden virheettömyyttä ja rakenteisiin liittyviä tasokuvauksia, vaikka yksittäisten rakenteiden painotus vaihteli arvioijittain. Väärän mielikuvan välttämiseksi on kuitenkin todettava, että arvioijat toki kiinnostavat huomiota


myös muihin kirjoittamisen kriteereihin, mutta eivät yhtä yhdenmukaisesti ja taajasti.


 *Missä määrin kirjoittaminen nähdään rakenteiden hallintana?*

Toinen yhteinen nimittäjä arvioijien kommentoissa oli se, että asteikkoon vaadittiin yksityiskohtaisia ja konkreettisia kriteereitä. Holistinen arviointi tekstityyppilähtöisellä asteikolla tuntui useimmista arvioijista liian summittaiselta. Osa arvioijista kritisoi myös sitä, että arvioitavat tekstityypit eivät vastanneet tarpeeksi hyvin asteikossa mainittuja tekstityyppejä. Lisäksi tehtävätyyppien taso ei aina vastannut asteikon tasoa, esimerkiksi ostoslistaviestin perusteella ei voinut arvioida taitotasoja 5 ja 6.

 *Tehtävä- ja tekstityyppien tulisi vastata mahdollisimman suorasti asteikon sisältöä.*

On odotettavissa, että huono asteikko aiheuttaa enemmän epäyhdenmukaisuutta ja epävarmuutta kuin hyvä asteikko. Siitä huolimatta arvioijien kritiikki saa pohtimaan myös sitä, missä määrin ongelma on asteikossa ja missä määrin vastuu on arvioijalla itsellään. Asteikon puutteellisuuden tiliin ei voitane täysin laittaa sitäkään, että arvioijat kertoivat myös subjektiivisista, tunnepohjaisista vaikutelmista, jotka liittyivät kirjoittajiin ja heidän taustoihinsa. Lisäksi arvioijia pohditutti heidän oma roolinsa lukijana, joka ymmärsi ehkä joskus liikaakin ja joskus liian vähän. Asteikon ja arvioijan suhde, arvioijan subjektiiviset havainnot ja hänen roolinsa geneerisenä lukijana nostavat esiin eettisiä kysymyksiä, jotka liittyvät sekä arvioitavien tasapuoliseen ja reiluun kohteluun että arvioijien sosiaaliseen rooliin tekstien arvottajina.

 *Arviointi on myös subjektiivista – viireystila säätelee ja kontrolli vastustaa mielivaltaisuutta.*

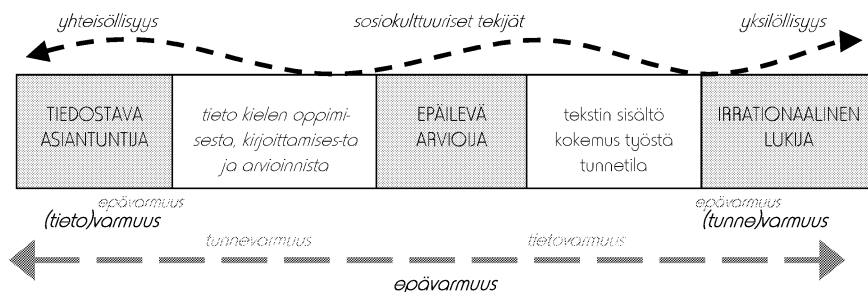
 *Pohdituttaa: mikä on arvioijan rooli tekstin lukijana ja arvottajana?*

5.2.3 Arvioija arvottavana lukijana

Haastatteluaineistojen perusteella olisi vaikea kuvitella, että arvioijan päätöksentekoprosessi etenisi lineaarisesti tai että arvioijan lukema teksti olisi autonominen yksikkö, joka elää omaa elämäänsä kirjoittajasta ja sosiaalisesta yhteisöstä irrallisena. Vaikuttaa siltä, että arviointiprosessi lukemis- ja päätöksentekoprosessina sisältää lukuisia ja taas lukuisia

huomioita ja päätöksiä, jotka eivät ole välttämättä kausaalisessa suhteessa toisiinsa vaan voivat olla hyvinkin ristiriitaisia keskenään ja ajallisesti päällekkäisiä niin, että niitä on hankala asettaa kronologiseen järjestykseen. Jos nousee yksittäisten tekijöiden, huomioiden ja havaintojen keskeltä abstraktisemmalle tasolle, arvioijan voi nähdä koko prosessin keskiössä lukijana, joka on yksi mutta jolla on kuitenkin monta lukijatyyppiä.

Arvioijan toiminta lukijana voitaisiin tiivistää kuvion 24 kaltaiselle jatkumolle, jossa on yhteisöllisyys–yksilöllisyys-, asiantuntija–irrationaalinen lukija - ja varmuus–tunne-akselit. Arvioija näyttää toimivan haastattelujen perusteella sekä sosiokulttuurisen ja sosiaalisen yhteisön jäsenenä että yksilönä (vrt. Marton 1994; Savage 2000). Sosiaalisuus ja yksilöllisyys eivät kuitenkaan ole selvästi toistensa ääripäitä tai helposti eroteltavissa toisistaan, vaan ne tuntuvat olevan jossakin määrin sisäkkäisiä ja jossakin määrin limittäisiä niin, että toinen niistä on toisinaan paremmin nähtävissä (vrt. Marton 1994). Kyse on siitä, että haastattelujen aikana arvioijalla saattoi olla eri lukijatyyppi etualalla eri aikoina sen mukaan, mitä hän kommentoi, mitä hän muisteli arviointiprosessista tai mitä häneltä kysyttiin. Arvioijan lukijatyyppi siis vaihteli tilanteittain.



KUVIO 24 Arvioijan lukijatyyppiä.

Arvioijan toiminta yksilöllisenä lukijana korostui varsinkin silloin, kun hän viittasi siihen, miten mielekkäänä hän koki tekstien lukemisen. Nimitän tällaista lukijatyyppiä irrationaaliseksi, koska arvioijan toiminta ei hänen kommenttiansa perusteella ollut järkipäristä vaan enemmänkin oman sen hetkisen mielialan tai vireystilan mukaista. En kuitenkaan kutsuisi tätä lukijatyyppiä emotionaaliseksi lukijaksi siitä syystä, että se ei oikein toimi asiantuntijan vastaparina yksinään. Lisäksi emootiot ovat

mukana myös asiantuntijan toiminnassa; kuviossa ei olekaan kyse dikotomioista, vaan siitä, millainen läsnäolo on kulloinkin etualalla, hallitsevana. Näin ollen jatkumon toisessa päässä oleva irrationaalisuuskaan ei tarkoita järjetöntä mielivaltaisuutta vaan lähinnä tunne- tai viireystilan oloa etualalla.

Irrationaalinen lukijatyyppe korostui arvioijalla esimerkiksi sellaisissa kommentteissa, jotka liittyivät väsymiseen, kirjoittajan taustaan tai tekstien yksitoikkoisuuteen. Tällaisille kommentteille oli tyypillistä, että niissä nousivat esiin sellaiset arvioon joko suoraan tai epäsuorasti vaikuttavat kriteerit, joita ei pidetä taidon kannalta tarkoituksenmukaisina ja arvioitavan kannalta oikeudenmukaisina. Irrationaaliset kriteerit viestivät kuitenkin arvioijan senhetkisestä käsityksestä kirjoittamisen taidosta, jos ne vaikuttavat hänen antamiinsa arvioihin. Tällaisiin perusteisiin annetut arviot – oli se hyväksyttävää tai ei – heijastelevat irrationaalisen lukijan käsitystä taidosta.

Irrationaalisen lukijan rooli voi liittyä hetkittäisiin tekijöihin, kuten viireystilaan, mutta myös ehkä arvioijan pysyvämpään ominaisuuteen eli hänen persoonallisuutensa ja temperamenttiinsa. Viireystilan muuttumisesta, esimerkiksi väsymisestä, nousevat kommentit olivat usein painokkaita, minkä perusteella kutsuisin niitä tunnevarmoiksi. Niiden tunnevarma sävy saattoi johtua konkreettisesti koetusta tunnetilasta, esimerkiksi kyllästymisestä tai ärsyyntymisen kirjoittajien samanlaisuuteen tai tapaan käsitellä jotakin aihetta ”mielikuvituksettomasti”. Tunnevarmat kommentit eivät olleet pelkästään negatiivissävyisiä, vaan saattoivat liittyä myös positiivisiin kommentteihin, esimerkiksi tekstin kykyyn naurattaa. Irrationaalisuus ja tunnevarmuus eivät kuitenkaan kulkeneet käsikkäin, vaan irrationaalisuudesta puhuminen saattoi yhtä lailla heikentää arvioijan varmuutta, mikä näkyi oman arviointilinjan ja sen oikeudenmukaisuuden epäilynä.

Irrationaalinen lukijatyyppe heijastelee yksilöllisyyttä, subjektiivisia kokemuksia ja käsityksiä. Arvioinnin johdonmukaisuuden ja yhdenmukaisuuden näkökulmasta nimenomaan irrationaalista lukijatyyppeä yritetään häivyttää, kun arvioijaa koulutetaan poisoppimaan subjektiivisuudestaan. Toisaalta arvioijien antamat arviot osoittavat, että vaikka irrationaalinen tyyppi lukee myös tekstejä, se ei välttämättä vaikuta arvioijien yhdenmukaisuuteen.

Yhteisöllisyys korostuu puolestaan sellaisissa kommentteissa, joissa arvioija puhuu odotuksenmukaisen ammattimaisesti, esimerkiksi analysoi tekstejä kielellisestä näkökulmasta, kritisoi arviointikriteereitä tai pohdiskelee arvioinnin luotettavuusongelmaa ja omien arviointiensu suhdetta muiden arvioijien antamiin arvioihin. Kun arvioijan lukijatyyppe on asiantuntija, hän osoittaa kuulumistaan arvioijayhteisöön esimerkiksi

arviointikokemustensa kautta, jotka voivat liittyä omiin arviointikäytänteisiin, saatuun arviointikoulutukseen, luokkahuonetilanteisiin tai kontrastiivisiin päätelmiin eri lähtökielen oppijoista.

Yhteisöllisyyttä osoittava asiantuntijalukija voi vaikuttaa myös hyvin tiedostavalta, koska hän reflektoi omaa toimintaansa: kyseenalaistaa omia toimintatapojaan ja kertoo ohjailleensa niitä. Oman toiminnan ohjailuun liittyen arvioija saattoi esimerkiksi epäselkeän käsialan kohdalla pakottaa itsensä lukemaan tekstin useampaan kertaan tai pitää tauon huomattuaan väsymyksen vaikuttavan työhönsä.

Asiantuntijalukijalle oli myös tyypillistä oman toiminnan perusteleminen ja tekemiensä valintojen pohdiskeleminen ammattikieleensä tukeutuen: "joku oli selvästi kirjottanu paljon mutta oli tehtävän ymmärtäny väärin nin mä vähensin siitä yhden pisteen ja ja sitte et jos oli melkeen kokonaan väärin ymmärtäny tehtävän ni m vähensin jopa kaks pistettä vaikka se oli hir että pitkä tarina hyvä muuten ihan hyvä kieliki mutta jos kerta kaikkiaan mun mielestä siis se ei kommunikoinu niinkun tavallaan se tilanne sillai et siinä ei ollu sitä kommunikaatiota."

Asiantuntijan argumentit tuntuivat myös melko varmoilta, koska ne perustuivat konkreettisesti tekstistä osoitettaviin faktoihin, tietoon ja tutkittuihinkin käsityksiin kirjoittamiseen liittyvistä osatekijöistä. Tämän vuoksi kutsunkin asiantuntijalukijaan liittyvää varmuutta tietovarmuudeksi. Asiantuntijalukija tai irrationaalinen lukija ei kuitenkaan ollut automaattisesti tieto- tai tunnevarma, vaan asiantuntijalukijakin saattoi esimerkiksi toisaalta vakuutella rakenteiden hallinnan tärkeyttä ja toisaalta miettiä, miten oikein ja tarkoituksenmukaista sen korostaminen on.

Olen sijoittanut epäilevän ja epävarman arvioijan jatkumon keskelle, mutta totean toistamiseen, että jatkumo ei ole sekventiaalinen vaan eräänlainen katkeamattoman liikkeen virta. Epäilevän arvioijan voi kuitenkin erottaa jatkumon keskipisteeksi epävarmuusaksellilla ja myös arvioijan lukijatyyppinä sillä perusteella, että siinä kohtaavat sekä asiantuntijuus että irrationaalisuus epävarmuudessaan. Epäilevän arvioijan epävarmuus voi kulminoitua sekä päätöksenteon, valitsemisen vaikeuteen että tietoisuuteen omasta toiminnasta ja sen suhteellisuudesta.

Viitataan tässä tiedostamisella ja epävarmuudella siihen, miten arvioija toisaalta reflektoi omaa toimintaansa ja oli tavallaan varma siihen vaikuttavista tekijöistä ja toisaalta taas juuri tästä syystä epävarma: "varmaan tavallaan ehkä ei ansaitse näin huonoa numeroa näiden sanojen takia siinä mulla oli kyllä vähän ongelmia kanssa ku kamppailin itseni kanssa et voiko antaa sit vaikka käyttää silleen sanoja ni sit hyviä arvioita." Epäilevälle arvioijalle oli tyypillistä, että vastausta ongelmaan ei ehkä löytnyt suoraan tieto- tai tunneperusteluista, vaan hänen täytyi

vain tehdä ratkaisunsa ja jäädä sitten käsittelemään omaa epävarmuuttaan. Epäilevän arvioijan lukijatyyppi siis korostui epävarmoissa valintatilanteissa, esimerkiksi silloin kun arvio oli kahden taitotason rajalla tai silloin kun arvioija epäili antaneensa arvion puutteellisin perustein.

Haastattelujen lähtöoletuksen, etnografisen ja fenomenografisen, näkökulmasta arvioija viestii toiminnallaan sekä yksilöllisiä että yhteisöllisiä käsityksiä. Kuten lukijatyyppikuvio – ehkä vähän kömpelöstikin – osoittaa, arvioija ei ole vain asiantuntija tai vain irrationaalinen lukija, vaan eri lukijatyyppit ovat etualalla eri aikoina sen mukaan, miten arvioijan käsitykset muuttuvat. Käsityksille näyttäisikin olevan Martonin (1994) mukaisesti tyyppillistä dynaamisuutta: ne voivat muuttua nopeastikin ja laajalla skaalalla.

Arvioijien käsitysten ja sitä kautta ilmenevien toimintatapojen yksilöllisyys ja yhteisöllisyys eivät heijastu vain haastatteluista vaan myös arvioijien antamista arvioista. Arvioista ei juuri keskustella tässä valossa, vaikka informaatio, jota ne kantavat mukanaan, on niin monisyistä. Itse asiassa arvioijien toiminta on ehkä sittenkin vain ja ainoastaan asiantuntijuudessaan inhimillistä. Asetan kysymyksen seuraavasti: pitäisikö meidän keskustella avoimemmin ja enemmän arvioiden inhimillisyydestä?

5.3 Valintaverkkohaastattelu

Käsittelen seuraavassa tutkimukseni kolmatta haastattelua, jonka arvioijat (N=15) kävivät läpi kolmannen arviointikierroksen jälkeen. Erotuksena ensimmäisen ja toisen arviointikierroksen jälkeisiin puolistrukturoituihin teemahaastatteluihin, tämä kolmas, valintaverkkohaastattelu (repertory grid) ei liittynyt suoraan itse arviointikierroksesta nouseviin kokemuksiin, vaan oli itsenäinen kokonaisuus. Se oli itsenäinen ja irrallinen myös sillä tavalla, että valintaverkkohaastattelu on haastattelumetodi, jolla on vahva siihen liittyvä teoreettinen pohjansa. Haastattelu linkittyi kuitenkin tutkimuksen aikaisempiin vaiheisiin siten, että valitsin haastattelussa käytettävät yhdeksän oppijan tekstiä toisen arviointikierroksen arvioiden perusteella. Valintaverkkohaastattelun tavoitteena on valottaa sitä, millaisia konstruktia eli teksteihin liittyviä piirteitä arvioijat nostavat esiin näistä yhdeksästä tekstistä sekä sitä, miten he arvottavat samoja konstruktia suhteessa esiin nostamiinsa konstruktuihin. Käytän tässä luvussa *konstruktia* samantyyppisessä merkityksessä, jossa olen aikaisemmin tekstissä käyttänyt *käsitys-* ja joissakin yhteyksissä myös *kriteeri-*termiä.

5.3.1 Millaiseen käsitykseen ja menetelmään valintaverkkohaastattelu pohjautuu?

George Kelly esitteli kliinisen psykologian tarpeisiin kehittämänsä valintaverkkometodin vuonna 1955.¹⁷ Valintaverkko on strukturoitu haastattelumetodi, joka tarjoaa haastattelijalle mahdollisuuden keskustella haastateltavan kanssa siitä, millainen hänen mentaalinen karttansa on eli miten hän tulkitsee ympäröivää maailmaa (Stewart 2000a, 2). Metodi pohjautuu konstruktiteoriaan, jonka lähtökohtana on se, miten henkilö havaitsee ja konstruoi omaa henkilökohtaista ja sosiaalista maailmaansa, kertoo myös siitä, millainen hän itse on (Fransella & Bannister 1977; Smith 1993). Henkilöt eroavat toisistaan sen mukaan, miten he konstruivat tapahtumia. Konstruktiteorian mukaan maailma on yksilöllinen tältä osin, koska tulkinta maailmasta on viime kädessä yksilön tekemä. Kellyn mukaan ihmistä voidaan ymmärtää ymmärtämällä hänen konstruktisysteemiään (Bannister & Mair 1968, 12).

Konstrukti voidaan määritellä yleisesti ottaen seuraavasti: konstruktit määrittelevät sen, miten yksilö havaitsee asioita (Bannister & Mair 1968, 12). Tämä ei ole kuitenkaan ainoa määritelmä, vaan määritelmiä tai oikeastaan luonnekuvauksia on useita, joista kukin paljastaa jotakin konstruktille tyypillistä. Esimerkiksi seuraavat ominaisuudet ovat konstruktiteille tyypillisiä. Konstruktit

- tekevät maailmasta ennustettavamman
- vaikuttavat odotuksiimme ja havaintoihimme
- voivat systeeminä muuttua ja kasvaa
- ovat eriarvoisia; jotkin konstruktit ovat toisia tärkeämpiä, vakiintuneempia ja siksi vaikeampia muuttaa, kuten esimerkiksi perusarvoihin liittyvät konstruktit
- peilaavat yksilöllistä ymmärrystä, vaikka ovat myös sosiaalisesti konstruoituja
- eivät ole aina sisäisesti johdonmukaisia.

Konstruktieihin liittyy ajallisesti sekä historiallinen että tuleva aspekti, koska ne toisaalta muotoutuvat laisikseen kokemusten perusteella eli ne ovat keino määritellä ja jäsentää tapahtumia ja toisaalta keino ennakoida tulevaa kokemusten pohjalta. Kellyn mielestä jokainen

¹⁷Valintaverkkometodia käytetään psykologiassa yhä erilaisten potilasryhmien, esim. änkyttävien ja ei-änkyttävien ryhmien, keskinäiseen vertailuun. Jotkut psykologit ovat jopa irrottaneet gridit täysin alkuperäisestä yhteydestään ja käyttävät niitä esimerkiksi etsiäkseen erilaisista ryhmistä tilastollisesti merkittäviä konstruktisuhteita. (Smith 1993.) Psykologian lisäksi valintaverkkoa on käytetty niin kirjallisuuden tutkimuksessa, oppimista ja opettamista koskeissa tutkimuksissa kuin markkinointitutkimuksissakin, esimerkiksi asiakasprofiilien selvittämisessä (Fransella & Bannister 1977; Solas 1992; Stewart 2000). Valintaverkon käyttö näyttäisikin vakiinnuttaneen asemansa yhtenä tutkimusmetodina eri tieteenaloilla (Jahoda 1986).

ihminen on eräänlainen tiedemies, joka muotoilee ja kokeilee hypoteeseja maailmasta omien kokemustensa kautta (Fransella & Bannister 1977, 4; Stewart 2000a, 1).

Konstrukteilte on erityisen tyypillistä, että ne ovat luonteeltaan dikotomisias eli "a construct is a way in which two or more things are alike and thereby different from a third or more things" (Fransella & Bannister 1977, 5). Joidenkin käsitysten mukaan juuri dikotomisuus erottaa *konstruktiin käsitteestä* (concept), jota käytetään psykologisissa teorioissa vastaavassa merkityksessä (Shaw & Gaines 1990, 5). Jokaisesta konstruktista tai sen luonnehdinnasta – esimerkiksi jos arvioimme autoa urheilulliseksi tai suhdetta kollegaan ristiriitaiseksi – voidaan siis kysyä: "Verrattuna mihin?" Näin ollen jokaisen konstruktiin kohdalla on mahdollisuus todeta, että joko henkilöllä tai asialla on sitä koskeva ominaisuus tai sitten ei, esimerkiksi käsite *hyvä* voi olla olemassa vain vastakohtana käsitteelle *huono* tai *selvän käsialan* vastakohtana täytyy olla olemassa *epäselvä käsiala*. Myös sellaisille käsitteille tai konstrukteilte, joille ei ole kielellisesti suoraan osoitettavissa vastakohtaa, voidaan sellainen osoittaa epäsuorasti. Esimerkiksi ilmaus *suomalainen* on bipolaarinen siinä mielessä, että se erottaa suomalaisen muista kansallisuuksista mahdollisen sekaannuksen välttämiseksi (vrt. Bannister & Fransella 1977). Vastakohdat eivät siis ole välttämättä sanakirjamaisia vastakohtia (Stewart 2000b, 2).

Konstruktit siis ohjaavat sitä, miten me tulkitsemme ja hahmotamme tapahtumia ja miten me suhtaudumme niihin ja miten me ennakoimme niitä. Esimerkiksi jos pidämme jotakin ihmistä rehellisenä, lainaamme hänelle rahaa, koska uskomme hänen maksavan lainansa takaisin (vrt. Bannister & Fransella 1977). Samoin arvioijat arvioidessaan esimerkiksi kirjoittamista saattavat varsin yleisestikin olettaa, että hyvä teksti on merkki oppijan korkeasta koulutustasosta tai huono teksti koulutuksen puutteesta.

Niin nimeämiset, kuvaamiset, tulkinnat, arviot ja ennakoinnitkin ovat tavallaan merkitysverkkoja, joiden avulla henkilö näkee ja käsittelee tapahtumia. Tässä mielessä konstruktit viittaavat moniin psykologisiin tekijöihin, kuten persoonallisuuteen, asenteisiin, uskomuksiin ja informaation käsittelyyn (Fransella & Bannister 1977, 2). Valintaverkko-menetelmä on keino tutkia edellä mainittuja merkitysverkkoja, sitä, miten ne toimivat ja jäsentyivät. Metodien avulla voidaan nostaa tarkasteltavaksi sekä yksittäisiä konstrukteja että niiden välisiä suhteita (Stewart 2000). Itse asiassa gridejä voitaisiin kutsua ajatusverkoiksi, jotka koostuvat erilaisista konstrukteista.

Gredit, kuten tietysti konstruktitkin ovat sekä *yksilöllisiä* että *yhteisöllisiä*. Jos gridit ovat yhteisöllisesti sovittuja, yksilön täytyy kuitenkin antaa

niille ensin merkitys ja tehdä niistä osa omaa systeemiään (Fransella & Bannister 1977, 7). Ne eivät voi siis sellaisenaan ja tavallaan automaattisesti siirtyä yhteisöltä yksilölle, vaan yksilö konstruoi niiden merkityksen itse. Arviointia koskevissa yhteyksissä tämä voisi tarkoittaa sitä, että arvioija ei omaksu esimerkiksi arviointiyhteisön käytänteitä sellaisinaan vaan antaa niille merkityksen muun muassa omien näkemystensä, kokemustensa ja tietopohjansa kautta. Merkitykset ovat siis tässä mielessä yksilöllisiä, vaikka ne olisivatkin alkujaan yhteisöllisiä.

Yksilöllisyydestä huolimatta kahden ihmisen psykologiset prosessit voivat olla *samanlaisia*, jos he konstruivat kokemuksen tai asian samalla tavalla (Bannister & Fransella 1977, 7). Pintatasolla kahden konstruktin täsmälleen samanlainen yhteys voi osoittautua kuitenkin hyvin harhaanjohtavaksi sekä yksilöiden että ryhmien välillä. Esimerkiksi sekä oikeistolaiset että vasemmistolaiset puolueet voivat kannattaa oikeudenmukaista verotusta, joka kuitenkin merkitsee tarkemmin eriteltäessä eri asioita oikeistolaisille kuin vasemmistolaisille.

Samoin konstruktien sisäinen ja keskinäinen *hierarkia* voivat vaihdella; ne voivat olla toisiinsa nähden ylenevässä tai alenevässä suhteessa, esimerkiksi arvioijat voivat kiinnittää huomiota yleensä sanastoon, mutta yksi voi viitata sillä merkityksen tarkkuuteen ja toinen taas varioivuuteen sen mukaan, mitä itse kukin pitää ensisijaisena (Bannister & Fransella 1977, 8). Tässä mielessä myös yleiset ja yhteisölliset konstruktit, kuten yhteisesti käytetyn arviointiasteikon tulkinta tai laajemmin esimerkiksi tiede ja politiikka, ovat pohjimmiltaan henkilökohtaisia.

Valintaverkkoahaastattelun ja mittaamisen näkökulmasta konstruktii on mittaamisen yksikkö eli haastattelun avulla voidaan nostaa tarkasteltavaksi yksittäisiä konstruktteja. Konstruktien välisten suhteiden tarkastelu tuottaa taas matriisin, joka voidaan analysoida tilastollisesti (Stewart 2000b, 9). Metodina valintaverkko tarjoaa mahdollisuuden päästä lähemmäksi arvioijan ajattelua: millaisia merkityksiä he luovat arvioitavista suorituksista? Arvioijan työn näkökulmasta epävarmuutta aiheuttaa se, että käytettävien arviointikriteereiden ymmärtäminen ja hyväksyminen eivät vielä osoita, kuinka niitä sovelletaan. Kiinnostavaa on siis tietää, millaisilla merkityksillä tai missä määrin arvioija yleensä arvioi sitä, mitä hän sanoo arvioivansa. Käytännön pisteittäminen näkökulmasta tämä tarkoittaa myös sitä, että tehtävän konstruktii ei ole sama kuin arvion konstruktii (ks. Luoma 1997). Tehtävä heijastelee sitä, miten se on arvioitu, ja välillisemmin sitä, mitä se on suunniteltu arvioimaan.

Arviointitutkimuksessa ei ole kovin paljon selvitetty arviointi-prosessia tai sen aikana syntyviä ja käsiteltyjä merkityksiä, koska kiinnostus on kohdistunut etupäässä lopputulokseen eli itse arvioon. Tärkeää ei ole niinkään ollut tietää, miten arvioija päätyy ratkaisuunsa

vaan, miten yhdenmukainen arvio on muiden arvioijien antamien arvioiden kanssa. Arviointiprosessin syvällisempi tunteminen olisi kuitenkin tärkeää arviointikoulutuksen ja arviointikäytänteiden kehittämisen kannalta. Arvioijan käsitysmaailman selvittäminen ei ole kuitenkaan yksinkertaista. Miten selkeänä oikeastaan yksilön ajatusmaailmaa yleensäkin voidaan pitää ja miten tietoinen yksilö on itse antamistaan merkityksistä?

Tutkijan tekemä työ on sekin tulkintaa merkityksistä, joita hän poimii arvioijalta oman käsitysmaailmansa kautta; tutkijakin luo merkityksiä omien konstruktiensa pohjalta, kun hän haastattelee arvioijia ja tulkitsee keräämäänsä aineistoa. Konstrukteihin sitoutuessa on myös syytä muistaa, että niiden liian tarkka kuvaaminen voi rajoittaa ajattelua ja kaventaa näkökulmaa käsitysten tarkastelemiseen. Ja jos suhteellisuus ei ole lainkaan mukana tarkastelussa, voi siitä seurata se, että alamme uskoa kuvauksiimme. Valintaverkkoahaastattelu on työkalu, joten se ei kerro totuutta vaan kuvaa pelkkiä satunnaisia havaintoja tarkemmin sitä todellisuutta, jossa arvioija työskentelee, ja sitä, millaiset käsitykset saattavat ohjata hänen työtään tai ainakin vaikuttavat siihen.

Kyse on siis ennen kaikkea ymmärrystä lisäävästä metodista, mikä käy hyvin ilmi Fransellan ja Bannisterin tekstistä (1977, 4):

"the results of the grid have often been looked on as a map of the construct system of an individual, a sort of idiographic cartography as contrasted with, the nomothetic cartography grid gives us of an individual's construct system, it is probably about as accurate and informative as the maps which Columbus provided of the American cost line."

Tässä tutkimuksessa valintaverkkoahaastattelu nähdään yhtenä strukturoidun haastattelun erityismuotona (ks. esim. Fransella & Bannister 1977; Stewart 2000b). Se on siis yksi lisäkeino selvittää arvioijien käsityksiä ja ajattelua puolistrukturoitujen haastattelujen lisäksi. Toisekseen tutkimusahaastatteluna se ei ole pelkästään teoreettinen metodi, vaan se heijastelee yleisestikin tapaamme keskustella ja konstruoida merkityksiä.

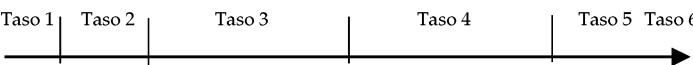
Konstruktioiden elisitointi edellyttää, että ärsykeaineisto on esitettävää haasteltavalle elementteinä, jotka ovat olennaista ainesta tutkimuksessa. Elementit voivat olla asioita, ilmiöitä, esineitä tai ihmisiä. Tässä tutkimuksessa elementit ovat oppijoiden kirjoittamia tekstejä. Tutkimukseen valittavien elementtien tulee olla järkeviä ja odotuksenmukaisia suhteessa tutkimuksen kohteena olevaan konstruktiin tai konstrukteihin. Elementtien tulee olla myös edustavia suhteessa siihen joukkoon tai kontekstiin, jota ne edustavat. Tässä yhteydessä kyse on siitä, että haastatteluun valittujen tekstien tulee edustaa kattavasti niitä tekstejä, joita arvioijat arvioivat tutkimuksen aikana. (Ks. Fransella & Bannister 1977; Solas 1992; Stewart 2000b.)

Edellisen lisäksi elementeiltä edellytetään, että niihin liittyvien konstruktioiden nimeämisen eli niiden verbaalistamisen täytyy heijastaa

myös tutkimuksen ulkopuolisia tilanteita. Tämän tutkimuksen elementit eli oppijoiden kirjoittamat tekstit valittiin niiden tekstien joukosta, joita arvioijat olivat arvioineet aikaisemmillä arviointikierroksilla. Kaikki valitut tekstit olivat vastauksia yhteen ja samaan tehtävään eli kuvasta kertomiseen. Valitsin tämän tehtävätyypin siksi, että se oli kaikille kirjoittajille yhteinen tehtävä ja sen arvioimisesta oli kaikilla arvioijilla kokemusta. Tämän perusteella elementit olivat ainakin jossakin määrin järkeviä ja odotuksenmukaisia sillä tavalla, että arvioijat pystyivät yhdistämään ne omiin käsityksiinsä ja merkitysmaailmaansa. (Bannister & Mair 1968, 41–42; Fransella & Bannister 1977, 13.) Tekstit ovat esillä liitteessä 10.

Haastattelussa käytettäviä tekstejä oli kaikkiaan yhdeksän, ja ne saivat kukin oman tunnistenumeron ykkösestä yhdeksään. Tekstit valittiin siten, että mukana olisi sekä sellaisia tekstejä, jotka arvioijat olivat arvioineet yhdenmukaisesti, että sellaisia tekstejä, joista he olivat olleet eri mieltä, mikä näkyy taulukon 19 mediaaneista ja keskiarvoista. Taulukon 19 yläriivi viittaa taitotasoon, jolle tai jonka tuntumaan tekstit oli arvioitu toisella arviointikierroksella. Tekstin id viittaa tekstin tunnistenumeroon. Taulukko kertoo siitä, mitä taitotasoa tai kohtaa 6-portaisesta asteikosta kukin yhdeksästä tekstistä edusti toisen arviointikierroksen arvioiden perusteella. Esimerkiksi teksti 7 oli saanut eniten taitotason 3 arvioita, ja muutamat arvioijat olivat arvioineet sen paremmaksi.

TAULUKKO 19 Haastattelun elementtien eli arvioitavien tekstien (N=9) jakaantuminen tasoittain.



Tekstin id	6	3	2	7	4	1	9	5	8
md	2	2,5	3	3	4	4	4	5	5
ka	1,9	2,5	2,9	3,4	3,8	4	4,3	4,4	5,1

Kuten taulukosta 19 näkyy, tekstit sijoittuvat tasoille 2–5. Heikoin teksti numero 6 sai yhden ykkösen ja muilta arvioijilta kakkosen, ja vahvin teksti numero 8 sai neljältä arvioijalta kuutosen mutta toisaalta taas kolmelta arvioijalta nelosen, minkä vuoksi sen keskiarvo jää viitosen tuntumaan. Suurin osa teksteistä sijoittuu tason 3 tuntumaan, mikä oli tyypillistä teksteille myös eri arviointikierroksilta kootun aineiston mukaan. Myös ensimmäisen ja toisen kierroksen haastatteluissa arvioijat raportoivat keskivälin tasojen olevan hankalia päätöksenteon kannalta,

koska heidän mukaansa monet suorituksista jäivät tasojen rajoille tai ne olisivat voineet olla esimerkiksi tasoista 2 ja 3 kummalla tasolla tahansa.

Kellyn mukaan valintaverkkohaastattelussa konstruktin minimi-konteksti on kolme asiaa, vaikka myöhempien käsitysten mukaan konstrukteja voi etsiä yhtä lailla pareilla tai joillakin muilla yhdistelmillä (Bannister & Mair 1968, 50; Fransella & Bannister 1977, 15–16). Käytän tässä tutkimuksessa triadeja eli kolmikkoja, mikä tarkoittaa kerrallaan kolmen tekstin vertailemista toisiinsa. Kolmikkojen esittämiseen haastateltavalle on kuusi erilaista tapaa, joista käytin tässä ehkä yleisintä eli minimikontekstin kordtimuotoa (the minimum context card form). Se tarkoittaa, että koehenkilölle esitetään kolme elementtiä eli kolme tekstiä ja häntä pyydetään sanomaan, mitkä kaksi tekstiä voi yhdistää samanlaisuutensa vuoksi yhteen niin, että ne eroavat kolmannelta (ks. Fransella & Bannister 1977, 14–15).

Kolmen sarjan elementit esitettiin taulukon 20 mukaisessa järjestyksessä. Järjestykseen ei ole olemassa yhtä mallia, vaan sitä voi vaihdella tutkimuksen tarkoituksen mukaan. Varsinaisesti merkitystä on läpikäytävien kolmikkojen kokonaismäärällä. Haastattelu jakaantui siis kahteen osaan, joista toisessa haastateltavat kävivät läpi kaikki yhdeksän taulukon 20 mukaista triadia ja miettivät jokaisen triadin kohdalla, mitkä kaksi tekstiä voi yhdistää samanlaisuutensa vuoksi yhteen niin, että ne eroavat kolmannelta. Saman triadin tekstejä voi yhdistää useammalla tavalla, esimerkiksi siten, että tekstejä 1 ja 2 yhdistää sanasto ja tekstejä 2 ja 3 samanlainen sisältö. Vertailussa esiin nousevien konstruktien kirjaamisessa käytin apuna liitteen 11 mukaista taulukkoa.

TAULUKKO 20 Triadien esittämisjärjestys.

Triadit	Triadissa esitettävät tekstit ja niiden id-numerot
1. triadi	1, 2 ja 3
2. triadi	4, 5 ja 6
3. triadi	7, 8 ja 9
4. triadi	1, 4 ja 7
5. triadi	2, 5 ja 8
6. triadi	3, 6 ja 9
7. triadi	1, 5 ja 9
8. triadi	2, 6 ja 7
9. triadi	3, 4 ja 8

Haastattelun aikana on tärkeää rohkaista haastateltavaa sanomaan kaikki, mikä tulee hänen mieleensä. Haastatteluissa voi olla tavallista, että haastateltava alkaa triadien toistuesssa epäröidä, että hän toistaa itseään kohdasta toiseen, koska sama teksti tulee hänen eteensä kolmesti. Haastateltavan on hyvä tietää, että se on tyypillistä tälle haastattelutavalle ja itse asiassa tärkeää kokonaisuuden kannalta. Osin ehkä tästä ja muistakin syistä haastateltava saattaa haastattelun aikana keinotekoisesti keksiä uusia konstruktteja. Se on tutkimusten mukaan kuitenkin aika epätodennäköistä, koska konstruktien on havaittu olevan siinä määrin pysyviä, että ne eivät ole keksittävisissä ja vaihdeltavissa kovin helposti. (Fransella & Bannister 1977, 17, 56–57.) Jos haastateltava keksiikin uusia, on hyvin todennäköistä, että se tulee esille viimeistään haastattelun toisen vaiheen aikana.

Haastattelun ensimmäisen vaiheen jälkeen siirryttiin haastattelun toiseen vaiheeseen, jossa haastateltavat järjestivät tekstit sekä kunkin esiin nostamansa konstruktin että yleistason mukaan heikoimmasta vahvimpaan. Tässä vaiheessa merkitsin kunkin haastateltavan esittämät tekstien järjestykset liitteen 11 mukaiseen taulukkoon.

Valintaverkkohaastatteluun osallistui kaikkiaan 15 arvioijaa, koska yksi arvioijista muutti ulkomaille juuri kolmannen arviointikierroksen ollessa käynnissä. Haastateltavien määrä on sikäli tarkoituksenmukainen, että Stewartin (2000b, 7) mukaan 15–20 haastateltavaa riittää silloin, kun haastateltavat ovat esimerkiksi ammatillisesti homogeeninen joukko ja kun haastattelun aihe on haastateltaville tuttu. Yleensä tällaisissa tapauksissa lisähaastattelut eivät juurikaan tuo enää uutta informaatiota.

Ennen kuin haastattelu aloitettiin, jokainen haastateltava tutustutettiin haastattelumetodiin esimerkkisanojen avulla. Esitin haastateltaville kolmelle erilliselle lapulle kirjoitettuna sanat: *juna*, *sakset* ja *tyyny*. Tämän jälkeen pyysin haastateltavia kertomaan, mitkä kaksi sanaa voi yhdistää samanlaisuutensa vuoksi yhteen niin, että ne eroavat kolmannelta. Osa haastateltavista pystyi heti yhdistelemään kaikkia sanoja eri tavoin, ja osa taas tarvitsi vähän enemmän miettimisaikaa joidenkin yhteyksien löytämiseen, ja osan mukaan puolestaan esimerkiksi sanoilla *sakset* ja *juna* ei ollut yhteisiä konstruktteja.

5.3.2 Haastattelun 1. vaihe: tekstejä erottelevat konstruktit

Haastateltavien aloitusstrategiat vaihtelivat jonkin verran: Esimerkiksi Pia ja Sanna lukivat aluksi tekstejä hitaasti ja tarkkaan. Pirjo puolestaan lähti liikkeelle siitä, että hän vertasi tekstejä omien oppilaittensa teksteihin ja mietti, millä tasolla kirjoittajat olisivat: ”tämä olisi niinkun meidän alkeistason teksti eli mun oppilaat kirjottasivat just tälläsen ykköskurssin

jälkeen”. Eroja oli myös siinä, mihin huomio alussa ankkuroitui. Joku luetteli luontevasti teksteissä olevia rakennevirheitä, kun taas joku toinen vertasi heti alussa tekstien sisältöjä ja niiden mielekkyyttä toisiinsa. Merja ja Pirkko kiinnittivät alussa huomiota tekstin sisällöllisiin yksityiskohtiin, kuten siihen, onko toisen tarinan tytöllä nimi ja tai onko jalka kipsissä vai ei.

Joillakin haastateltavista saattoi aivan haastattelun alussa olla vaikeaa löytää yhteisiä piirteitä tai kriteereitä teksteille; yhden tekstin tai sen jonkin tietyn piirteen kuvaaminen saattoi tuntua luontevammalta kuin miettiä, millä tavalla kaksi tekstiä ovat samanlaisia. Tosin muutaman triadin jälkeen haastattelun idea selvisi ja ohjaavia kysymyksiä ei tarvinnut enää ylipäänsä esittää. Myös joidenkin haastateltavien perustelut yksityiskohtaistuivat haastattelun edetessä; esimerkiksi ”kielivirheistä” tuli ”lauserakennevirheitä” tai ”epäselvästä tekstistä” tuli ”huonosti jäsennelty”. Niille taas, jotka jo haastattelun alussa ilmaisivat monia kriteereitä ja yhdistelymahdollisuuksia, saattoi haastattelun loppuosa tuntua toistavalta, koska uusia assosiaatiomahdollisuuksia ei enää syntynyt.

Yhdessä haastattelussa nousi esiin keskimäärin 8 arviointikonstruktiota eli erottelevaa/yhdistävää piirrettä, vähimmillään 6 ja enimmillään 10. Piirteeksi ei ole nostettu sellaisia kommentteja, joiden mukaan haastateltavat eivät olisi pystyneet järjestämään tekstejä heikoimmasta vahvimpaan. Arviointikonstrukteiksi on katsottu sellaiset piirteet, joiden perusteella tekstejä voi selkeästi erotella toisistaan akselilla parempi–huonompi. Sellaiset kommentit, jotka ovat havaintoja mutta eivät tässä mielessä arvottavia, eivät ole tässä yhteydessä konstruktiveja. Esimerkkeinä tällaisista havainnoista voi pitää haastateltavien kommentteja kirjoittajien taustoista, koska ne eivät varsinaisesti ole tekijä, jonka mukaan tekstit eroavat taitotasoltaan toisistaan.

Taulukkoon 21 on koottu kaikki haastateltavien esiin nostamat arviointikonstruktit ja niiden frekvenssit. Kukin piirre on laskettu vain kerran, joten esimerkiksi sanaston frekvenssi 14 tarkoittaa sitä, että viidestätoista arvostelijasta neljätöistä erotteli/yhdisteli tekstejä sanaston perusteella.

TAULUKKO 21 Haastattelussa esiintyneet piirteet ja niiden frekvenssit.

PIIRRE	f
1. SANASTO (<i>tarkka-epätarkka, laaja-suppea</i>)	14
2. TARINAMAISUUS (<i>tarinamainen-ei-tarinamainen</i>)	14
3. RAKENTEET (<i>hallitsee-ei hallitse, varioivaa-yksinkertaista, idiomaattista-epäidiomaattista</i>)	12
4. PITUUS (<i>pitkä-lyhyt</i>)	12
5. KÄSIALA (<i>selkeä-epäselvää</i>)	10
6. EMOTIONAALINEN VAIKUTELMA (<i>kiva lukea-ärsyttävä</i>)	8
7. TEHTÄVÄNANNON TÄYTTÄMINEN (<i>täyttää-ei täytä</i>)	7
8. TEKSTIN SISÄLTÖ (<i>laaja-suppea, yleinen-yksityiskohtainen</i>)	5
9. LAUSERAKENNE (<i>oikein-väärin, idiomaattinen-yksinkertainen</i>)	4
10. TEKSTIN LOOGINEN RAKENNE (<i>looginen-epälooginen</i>)	4
11. MIELIKUVITUKSEN KÄYTTÖ (<i>mielikuvituksellinen-asiallinen</i>)	4
12. TEHTÄVÄÄN PANEUTUMINEN (<i>huolellinen-huolimaton</i>)	4
13. KIRJOITTAJAN KYPSYYS (<i>aikuismainen-naivi</i>)	4
14. TEKSTIN YMMÄRRETTÄVYYS (<i>ymmärrettävä-ei-ymmärrettävä</i>)	3
15. IDIOMAATTISUUS (<i>idiomaattinen-ei-idiomaattinen</i>)	3
16. OIKEINKIRJOITUS (<i>oikein-väärin</i>)	3
17. TEKSTUAALINEN RAKENNE (<i>tekstuaalisesti toimii-ei toimi</i>)	3
18. VÄLIMERKKIEN KÄYTTÖ (<i>on käytetty-ei ole käytetty</i>)	2
19. TAIVUTUS (<i>taivutusta esiintyy systemaattisesti-ei taivutusta</i>)	2
20. KIRJOITTAMISEN KONVENTIOT (<i>konventionaalinen-ei-konv.</i>)	1
21. TUNTEIDEN KUVAUS (<i>tunteita kuvattu-ei ole kuvattu</i>)	1

Arvioijat (N=15) nostivat esiin kaikkiaan 21 erilaista arviointikonstruktiota. Kuten taulukosta 21 näkyy, kaikkein yleisimmät konstruktit olivat sanasto, tarinamaisuus, rakenteet ja pituus.

Sanastoa ei ole eritelty sen tarkemmin osiin, vaikka haastateltavat käyttivätkin sitä vähän toisistaan eriytyneissä merkityksissä. Arvioijat tarkoittivat sanastolla muun muassa sen laajuutta, tarkkuutta tai monipuolista käyttöä tilanteen mukaan. Sanaston yhteydessä viitattiin myös erikoisalan sanastoon lähinnä tekstin 9 kohdalla, joka oli kirjoitettu epikriisiin muotoon (esim. "ihorikko"). Sanasto on kuitenkin taulukon 21 mukaisesti vain yksi yksikkö, koska haastateltavat eivät viitanneet sanaston eri osa-alueisiin niin systemaattisesti, että niitä olisi ollut

tarkoituksenmukaista eritellä. Tavallisimmin haastateltavat viittasivat kommentteissaan sanaston kokoon.

”ja tää on sitten se kutonen on tää tämmönen kömpelö missä nyt sitten tässä on verbejä yks kas kolme neljä viis siis siis se käyttää niinku viittä eri verbiä plus toi broke se niinku käyttää viittä kuutta verbiä jota se ei tunne eikä se oo keksiny niinku siihen mitään...tällä on heikko sanavarasto” Tiina

”tää sanavarasto tietysti yhdistää kakkosta ja seiskaa koska nää pystyy sanomaan paljon enemmän kun toi kutonen” Sanna

”nää on sanastollisesti monipuolisempia kuin tämä just sen takia että tästä tästä heijastuu näitten sanojen osaava käyttö oikeessa kontekstissa niinku tässä toisessakin että täällä on ehkä enemmän semmosta tietyllä tavalla kömpelömpää ja yksinkertasempaa sanastoo siinä suhteessa” Kaisu

Sen sijaan konstruktia *rakenteet* käytettiin systemaattisesti eri merkityksissä: 12 haastateltavaa tarkasteli rakenteita yleisellä tasolla, esimerkiksi ”nää on kieliopillisesti hyvää kieltä” sekä/tai yksilöidymmin lauserakenteita (N=4), esimerkiksi ”sen erottaa niistä muista siit et siinä on just jotain tommosii partisiippeja ja lauseenvastikkeita haettu mitä ei oo näis muissa” ja/tai taivutusta (N=2), esim. ”et ne taivutukset ei niinku oo silleen selvii”. Kommenttien yleisyys tai yksityiskohtaisuus ei ollut mitenkään arvioijakohtaista vaan samakin arvioija tarkasteli jotakin triadia yleisesti rakenteellisesti tai sitten yksilöidymmin esimerkiksi vain imperfektin näkökulmasta.

Rakenteista yleisesti puhuneet haastateltavat saattoivat mainita idiomaattisuuden, oikeellisuuden tai varioivuuden jonkin tekstin kohdalla mutta eivät käyttäneet systemaattisesti yleistä ilmausta, kuten ”taivutusta” tai ”lauserakenteita”. Yksi arvioijista kiinnitti huomiota vain lauserakenteisiin eikä siis kommentoinut muita rakenteita lainkaan. Tässä mielessä arvioijien huomio rakenteiden kohdalla vaihteli varsin yleisestä varsin yksityiskohtaiseen tarkasteluun. Näkökulma rakenteisiin saattoi myös olla virhelähtöinen, jolloin virheitä tarkasteltiin määrällisestä näkökulmasta tai niiden tyyppin mukaan: ”seiskassa ja ykkösessä on samantyyppisiä kielivirheitä” tai ”tässä on enemmän kielivirheitä ku näissä kahessa muussa”.

Sanaston ohella frekventein konstruktiksi oli *tarinamaisuus*, mitä selittänee tehtävän luonne, koska sekä tehtävänanto (kuva + kronologisuuden vaatimus) että tekstityyppi (kirjoita tarina) vaativat ”et ne on kirjoitettu niinku tavallaan tarinan muotoon”. Tarinamaisuuden luonnetta saattoi korostaa myös se, että haastattelussa käytettävät tekstit vaihtelivat mielikuvituksellisesta ja dramaattisesta tarinasta epikriisiin toteavaan ja asialliseen tyyliin: ”toi yheksikkö ja vitonen kuuluu kuitenkin yhteen ni niinku niissä se on semmonen niinkun perustarina niinku sillä eikä

epikriisi kerrottuna.” Siinä missä sanasto ja rakenteet olivat kielellisesti peruskomponentteja, tarinamaisuuden voi nähdä tekstuaalisena peruskomponenttina. Tarinamaisuus viittasi myös tekstin sisältöön tehtävänannon kuvan välityksellä, koska kuva sääteli omalta osaltaan sitä, mitä tekstin sisällöltä odotettiin.

Rakenteiden ohella toiseksi yleisin konstrukti oli *tekstin pituus*. Haastattelun perusteella se näytti olevan hyvin konkreettinen ja helposti havaittava ero tekstien välillä (ks. myös Charney 1984; Huot 1990). Tekstien pituus vaihteli muutamasta rivistä lähes kahteen sivuun (kaksipuoleinen A4). Jos nämä kaksi ääripään tekstiä ovat rinnakkain, on oletettavaa, että pituus erottaa jo ensisilmäyksellä ne toisistaan. Arvioijat eivät kuitenkaan ottaneet pituuteen kantaa kovin arvottavasti: ”pituutta ei pitäis koskaan pelkäästään katsoo mut tuota kuitenkin että nää on pitempiä nytte”. Pituus liittyi yhtenä elementtinä myös tekstin yksinkertaisuuteen: ”nää kolmone ja kutonen tuntuu niinku silleen palik niinku yksinkertasemmilta ja (millä tavalla yksinkertasilta) silleen noh yksistään se jo heti et ne on lyhkäsempiä”.

Myös *käsialaa* voi pitää pituuden ohella hyvin konkreettisesti havaittavana konstruktina ja myös jokseenkin yleisenä kirjoittamiseen liittyvänä piirteenä (ks. esim. Vaughan 1990; Pollitt & Murray 1996). Näissä haastatteluissa arvioijat kiinnittivät käsialaan huomiota lähinnä vain silloin, kun se oli niin epäselvää, että sitä oli työlästä tai mahdotonta lukea tai jos se selkeästi erottui rinnalla olevasta selkeästä ja siististi kirjoitetusta tekstistä: ”kutosta ja seiskaa on huomattavasti helpompi lukea ku sitte tuota kakkosta.” Käsialaan vaikutti myös muu tekstin typologia, jos epäselvän käsialan lisäksi kirjoittaja ei ollut käyttänyt välimerkkejä eikä jaotellut tekstejä lainkaan kappaleisiin.

Käsialaa kommentoidessaan haastateltava saattoi esittää myös emotionaalisia kommentteja kirjoittajan kansallisuudesta tai käsiala saattoi muuten ärsyttää lukijaa, koska se oli ”inhottavaa”, tai miellyttää, koska se oli ”lennokasta”. Käsialan selvyyttä kommentoitiin myös sen mukaan, miten tottunut tai tottumaton kirjoittaja saattoi olla latinalaiseen kirjaimistoon. Arvioijien kommentteissa heijastui myös itsekritiikki, jos he kiinnittivät käsialaan huomiota, vaikka he olisivat periaatteessa olleetkin sitä mieltä, että sen ei pitäisi vaikuttaa tekstistä tehtäviin havaintoihin.

”...harjaantuneisuus että ettei oo läheskään niin harjaantunu käsiala esimerkiks kirjottamaa kun näillä...” Päivikki

”näissä on samanlainen käsiala (naurua) samantyylinen vähän täs tässä on erilainen kauhee enk mie nyt mitään muuta keksi tää on tämmönen pysty tää on tämmönen vähän lapsen lapsellisen näkönen käsiala tää nelosen...”
Tiina

”sitte tommonenki vaikuttaa sit siihen yleiskuvaan ihan kaikki tommone käsialat ja kaikki et tää on nyt sitte niinku tää vitonen...ihan ku ois viivottimen päällä kirjoitettu tai ollu apuviivat ja sitte taas toi kakkonen on tommosta vähän avuttoman näköstä söherrystä ja toi on tolleen et kyl semmonenkin vaikuttaa” Martti

Noin puolet haastateltavista esitti kommentteja, joihin sisältyi emotionaalinen lataus. *Emotionaalisuus* liittyi lähinnä sellaisiin tunnepohjaisiin kommentteihin, joiden mukaan joitakin tekstejä oli kiva lukea tai jokin teksti ärsytti esimerkiksi niin paljon, ettei sitä olisi enää halunnut edes katsoa. Emotionaalisuus täytti arviointikonstruktin kriteerit, jos haastateltava pystyi järjestämään tekstit esimerkiksi ärsyttävimmästä miellyttävimpään eli konstrukti täytti bipolarisuuden ehdot ja mahdollisti tekstien järjestämisen.

Emotionaalisille kommentteille näytti olevan tyypillistä sellainen kaksijakaisuus, että ne tulivat joiltakin haastateltavilta varsin spontaanisti jo haastattelun alussa ja joiltakin haastateltavilta vasta haastattelun loppupuolella. Yhdellä haastateltavista emotionaalisuus tuli esiin vasta haastattelun toisessa eli järjestelyvaiheessa ikään kuin puolivahingossa.

”tää on moraalinen...taas tää opettaa” Heidi

”tää vitosteksti jo ällöttää mie en haluais kattoo sitä...” Teija

”tää on musta oikeastaan hauska” Tarja

”mä inhoon tätä yhtä tekstiä musta tää on tää on jotenkin niin yök mä muistan tän niin hyvin tää oli just semmonen hankala ei mitään tää on musta jotenki sympaattinen tää teksti” Päivikki

”mua tökkii näissä molemmissa tässä ei töki...mut nelosta ei oo niin kiva lukea” Tiina

Yli puolet arvioijista nosti esiin vertailussaan edellä esitetyt kuusi yleisintä piirrettä. Taulukon muihin 15 arviointikonstruktiin viittasi seitsemästä yhteen arvioijaa. Nämä 15 arviointikonstruktiota voisi ryhmitellä sisältönsä mukaan kolmeen ryhmään: 1) tehtävään ja tekstiin viittaavat konstruktit (sisältö, tehtävänannon täyttäminen, tekstin looginen rakenne, tekstuaalinen rakenne, kirjoittamisen konventiot ja idioomaattisuus), 2) kirjoittajaan viittaavat konstruktit (mielikuvituksen käyttö, tehtävään paneutuminen, kirjoittajan kypsyyt, tunteiden kuvaus) ja 3) kieleen liittyvät piirteet (lauserakenne, oikeinkirjoitus, välimerkkien käyttö ja taivutus). Koska näistä 15 arviointikonstruktiosta monen frekvenssi on pieni, niitä on myös siksi tarkoituksenmukaista käsitellä ryhmittäin.

1) Tehtävään ja tekstiin viittaavat konstruktit

Tehtävään liittyvistä piirteistä tekstin sisältöä käsiteltiin useimmin akselilla suppea-laaja, jolla haastateltavat tarkoittivat sitä, miten yksityiskohtaisesti tai ylimalkaisesti kirjoittaja oli käsitellyt aiheita.

Yksityiskohtaisuuteen liittyi synonyymisesti myös tarkkuus eli se, miten tarkasti sisältöä kuvattiin ja käsiteltiin, esimerkiksi "yksityiskohtien tarkkuus on suurempi näissä seiskassa ja kasissa" tai "tää kasi on siinä epämääräisempi mitä siinä tapahtu".

Tekstin sisältöä kommentointiin myös sen suhteen, miten samanlainen tai erilainen se oli toisen tekstin kanssa. Nämä kommentit saattoivat liittyä aivan pieniin yksityiskohtiin tai sitten esimerkiksi johonkin tekstin osaan: "sit toisaalta kasissa ja ysissä tää loppu jos ajatellaan tää niinku näin sisällöllisesti on aika samantyyppinen ja vähän tämmönen niinku tietyllä tapaa mun mielestä onks se opettavainen mut kuitenkin." Tekstin sisältö liittyy myös läheisesti tekstityyppiin, esimerkiksi siihen, miten sisältö muodosti toimivan kokonaisuuden "tässä tää tarina on ehjä kokonaisuus".

Koska tehtävänä oli kirjoittaa tarina, lienee aika luonnollista, että myös tehtävänannon täyttäminen liittyi tarinamaisuuteen: teksti täytti tehtävänannon, jos kirjoittaja oli kirjoittanut tarinan ohjeen mukaisesti. Esimerkiksi tekstiä 1, joka oli kirjoitettu epikriisiin muotoon, moitittiin siitä, että se ei täyttänyt tehtävänantoa. Tarinamaisuuden lisäksi tehtävänanto täyttyi seuraavan kommentin mukaisesti: "mut ehkä nää ysi ja seiska että näissä on niinku enemmän just semmönen ennen nyt jälkeen tarina". Koska tehtävän annossa korostui ajallinen näkökulma, myös kronologisuuteen tai lähinnä aikamuotojen käyttöön kiinnitettiin huomiota: "kasissa ei oo otettu sitä mennyttä aikaa huomioon" tai "kolmonen ja ysi on luku ton tehtävänannon jollakin tavalla koska täs täs on toi mitä on tapahtunut ennen kuvaa mitä kuvassa tapahtuu mitä vois tapahtua sen jälkeen".

Tekstin looginen rakenne nousi esiin lähinnä sellaisen tekstin kohdalla, jossa teksti ei haastateltavan mielestä edennyt loogisesti: "tää ei etene kuitenkaan sillä tavalla loogisesti että ku tässä ei oo mitään kappaleita niin niinku tässä menee niinku putkeen tästä tapahtumasta kertominen ja mitä tapahtu ennen ja tällä hetkellä ja sitten on että minäki jouduin että tavallaan sen kokonaisuuden hahmottamisen jouin joutusin lukemaan joitakin kohtia kaks kertaa etku tää on niin looginen tai semmönen etenee järjestyksessä."

Edellisessä ja seuraavassa esimerkissä, kuten yleensäkin loogisuuteen liittyvissä kommentteissa, konstruktivi viittasi tekstiin 9, jossa kirjoittaja käänsi tekstin näkökulman yhtäkkiä itseensä, mikä tuntui haastateltavista epäloogiselta ja rikkoi tekstin rakenteen tai tarinamaisuuden rakenteen ja sen kautta myös tehtävänannon: "se vaan yhtä äkkiä vaihtaa tavallaan juonen siinä" tai "nää pysyttelee siinä kolmannessa persoonassa ku tää ysi sitte niinku kääntää kesken jutun niinku itteensä tän homman."

2) Kirjoittajaan viittaavat konstruktit

Kirjoittajaan liittyvistä piirteistä tehtävään paneutuminen erosi tehtävänannon täyttämisestä siinä, että sillä tarkoitettiin lähinnä kirjoittajan asennoitumista eli sitä, miten kirjoittaja oli suorittanut tehtävän loppuun ja miten huolellisesti hän oli tehnyt sen: ”tässä tehtävään on paneuduttu, tarinaa on kehitelty eteenpäin” tai ”se asenne siihen niinku miten se miten se tehään se homma ni on ihan erilainen ja tää niinku jotenki tunnollisemmin tekee tää nelonen on niinku tehny siitä kirjetelman ja nää kolmonen ja kasi niinku näyttäis siltä niinku ne ois vähemmän aikaa käyttäny”.

Kirjoittajan kypsyys, jolla arvioijat viittasivat tekstin naiiviin luonteeseen, nousi esiin neljän haastateltavan kommentteista. Kypsyys kriteerinä voi olla osin yhdistettävissä tarinamaisuuteen siinä mielessä, että tarinamaisista teksteistä voi tulla helposti naiiveja, varsinkin jos ne on kirjoitettu heikohkolla kielitaidolla. Tässä yhteydessä pohdittiin myös sitä, miten teksti heijastelee kirjoittajan ikää ja kansallisuutta, esimerkiksi ”et kolmonen ja kasi on samanikäsiä suunnilleen ja tää nelonen on jotenki nuorempi tää on vähän naiivi tää nelonen”.

Mielikuvitukseen viitattiin oikeastaan vain silloin, kun arvioijan mielestä sitä oli tekstissä: ”tää on niinku tässä on niinku käytetty hirveen paljon mielikuvitusta että tää on osannu niinku käyttäny paljon mielikuvitusta kirjottamiseen ja pystyny niinku päässään muodostamaan sellasen ison tilanteen missä mukava ulkomaalainen lääkäri hoitaa”. Mielikuvitukseen viitattiin myös käyttämällä ”luova”-ilmaisua: ”tätä vois sanoa että tää on niinku luova tää juttu.” Mielikuvituksen puutetta ei suoranaisesti kommentoitu, joskin tekstin toteavaa tai lyhyttä tyyliä kommentointiin. Arvioijien mielestä kirjoittajan käyttämät yksinkertaiset lauserakenteet tekivät tyylistä töksähtelevää ja toteavaa.

3) Muut kieleen viittaavat konstruktit

Muihin kuin edellä mainittuihin kielen piirteisiin kiinnittivät huomiota muutamat arvioijat. Kaksi arvioijaa nosti esiin taivutuksen, ja heistä kumpikin tarkasteli sitä melko systemaattisesti useampia triadeja vertaillessaan, esimerkiksi ”ja sitten tää verbintaivutus yleensäkin näitten sanojen taivutus ni hetkinen nää taivuttaa kuitenkin vähän samalla tavalla esimes voisko sanoa yksikön kolmannessa mut silti niinku aika paljon ei koko ajan mut aika paljon niinku sillä tavalla että mulle tulee mieleen täytyyverbien taivutus joka ei taivu ollenkaa et mun mielestä näitten verbintaivutus on paikoitellen semmosta niinku yksipersonaisen verbien taivutusta.”

Oikeinkirjoitukseen ja välimerkkien käyttöön puututtiin oikeastaan vain sellaisissa tapauksissa, joissa ne horjuivat tai vaikeuttivat lukemista

mutta silloinkaan niitä ei pidetty niin tärkeinä kuin jotakin muuta konstruktia: "jos nyt oikein ottaa punakynän täältäähä löytyy paljon sellasia kirjoitusvirheitä mut jotenki se ei se ei haittaa". Oikeinkirjoituksen ja välimerkkien kohdalla mietitytti myös se, missä määrin oikeinkirjoitus ja esimerkiksi taivutus menevät päällekkäin: "ehkä oikeinkirjoitusvirheitä tai et ne taivutukset ei niinku oo silleen selviä."

Monet esiin tulleista konstrukteista voisi ainakin osin yhdistää, sillä niiden merkitykset menevät joiltain kohdin limittäin. Esiin nostetut piirteet eivät siis ole merkitykseltään täysin selvärajaisia. Monista konstrukteista olisi vielä tarpeen neuvotella tarkemmin haasteltavien kanssa, koska he saattoivat käyttää samaa konstruktia useammassa eri merkityksessä tai käsittää konstruktit merkitykseltään sisäkkäisiksi. Esimerkiksi jollakin arvioijalla mielikuviutus saattoi sisältyä tarinamaisuuteen, eikä hän tullut sitä siksi maininneeksi. Mielikuviutus on näin ollen molemmilla olemassa oleva konstrukt, vaikka vain toinen nostaa sen esiin.

Yhdellä tai vain muutamalla haastateltavalla esiin nousset arviointikonstruktit voisi myös sisältönsä puolesta liittää johonkin muuhun konstruktiiin, esimerkiksi välimerkkien käytön osaksi oikeinkirjoitusta tai tekstuaalisen rakenteen tekstin loogiseen rakenteeseen: "seiska on sitte loogisempi ja niinkun kerronnallisesti paljon järkevämpi ja parempi."

Haastatteluissa esiin tulleiden arviointikonstruktioiden kirjo osoittaa, että ne kattavat ison alueen kirjoittamiseen liittyvästä metatekstistä, tyypillisistä piirteistä, joiden avulla toisen kielen kirjoittamista analysoidaan, opetetaan ja arvioidaan ja siitä puhutaan (esim. White 1986; Milanovich ym. 1996; Grabe & Kaplan 1996). Jos arviointikonstrukteja tarkastelee Grabe & Kaplanin (1996) esittämän kirjoittamisen taksonomian näkökulmasta, ne voisi luokitella taulukon 22 mukaisesti. Taulukossa 22 eivät ole mukana kaikki kyseisen taksonomian osa-alueet, mutta taksonomian mukaan erotellut konstruktit kuvaavat paremmin kuin taulukon 21 lista sitä, mihin ja miten monipuolisesti arvioijat kiinnittivät huomiota kirjoittamisen eri tekijöihin.

TAULUKKO 22 Valintaverkkohaastattelun tekstejä erottelevat/yhdistävät piirteet kirjoittamisen taksonomian mukaan (vrt. Grabe & Kaplan 1996).

Kirjoitustehtävä	Tarinamaisuus		T E K S T I N Y M M Ä R E T Ä V Y Y S
	Tehtävänannon täyttäminen		
Tekstityyppi	Mielikuvituksen käyttö		
Kirjoittajan intentio, asenne ja tavoite	Tehtävään paneutuminen		
	Kirjoittajan kypsyyden		
Kielellinen tieto	Sanasto		
	Rakenteet	Lauserakenne	
		Taivutus	
	Oikeinkirjoitus		
Välimerkit			
Diskurssitieto	Tekstin sisältö		
	Tekstin looginen rakenne		
	Kirjoittamisen konventiot		
	Tekstin rakenne		
Sosiolingvistinen tieto	Idiomaattisuus		
Lukija	Emotionaalisuus		
	Käsiala		
	Tekstin pituus		

Vaikka konstruktioiden merkitykset eivät ole täysin selvärajaisia suhteessa taksonomian eri luokkiin, konstruktit asettuvat kuitenkin melko hyvin niihin päällekkäisyyksistä huolimatta. Kategorisointia kannattaa kuitenkin tarkastella lähinnä informaatioarvon kannalta, koska kirjoittamisprosessin luonteen mukaisesti taksonomian eri luokat ovat päällekkäisiä ja keskinäisessä vuorovaikutussuhteessa. Esimerkiksi tekstin sisältö voi olla riippuvainen kirjoitusaiheesta, tehtävänannon selkeydestä, kirjoittajan kulttuuritaustasta tai hänen motivaatiostaan.

Kuten taulukosta 22 näkyy, olen sijoittanut ymmärrettävyyden muista konstruktista poiketen äärimmäisenä oikealla olevaan sarakkeeseen. Ymmärrettävyyden voisi nähdä eräänlaisena yleiskriteerinä, joka ei ole sijoitettavissa yksistään mihinkään taksonomian kategorioista. Ymmärrettävyyden luonne tuli esiin myös arvioijien kommentoissa,

esimerkiksi Tiinan, joka kolmen tekstin vertailua aloittaessaan totesi: "kyllähän nää ymmärtää kaikki". Ymmärrettävyys koostuu tavallaan kaikista muista konstruktista, tai ainakin ne kaikki vaikuttavat siihen joko sitä parantavasti tai heikentävästi.

Jos katsotaan kaikkia konstruktia suhteessa niiden frekvenssiin, arvioijat näyttäisivät painottaneen eniten kielellisiä piirteitä. Sen sijaan kuuden yleisimmän piirteen joukossa oli vain yksi tekstuaalinen konstrukt, mikä voi tuntua yllättävältä kirjoittamisen arvioinnissa. Toisaalta kielellisten piirteiden vahvuus on melko oletettavaa muiden kirjoittamisen arviointia koskevien tutkimusten mukaan (esim. Raimes 1985; Vaughan 1990; Pollitt & Murray 1996). Myös Komulaisen (1999) tutkimuksessa, jossa valintaverkkoa käytettiin englannin puhumisen arviointiin, lähes kaikki arvioijat nostivat esiin rakenteellisen tai sanastollisen konstruktin.

Kielelliset piirteet tuntuvat olevan konkreettisia siinä mielessä, että niitä on helpompi nimetä ja nostaa esille. Tässä mielessä ne ovat selkeämmin havaittavissa kuin muun muassa tekstuaaliset piirteet: esimerkiksi koheesio voi tuntua abstraktimmalta ja vaikeammalta kohdentaa johonkin tiettyyn kohtaan tekstiä tai sitä voi olla vaikeaa analysoida ilmiönä.

Kielellisten komponenttien vahva rooli voi toki myös johtua opettajien koulutuksesta, joka on ehkä perustunut formalistiseen kielikäsitteeseen, tai siitä, että kielelliset konstruktit ovat yksinkertaisesti eräänlaisia peruselementtejä, joiden avulla kirjoittamisen laatua hahmotetaan. Suomi toisena kielenä -kirjoittamisen tulevaisuuden haasteita onkin pohtia, miten pystymme valjastamaan kirjoittamiseen liittyvät erilaiset taitotekijät konkreettisiksi, opetettaviksi ja arvioitaviksi konstruktiksi. Puhukoon osaltaan tämän puolesta seuraava esimerkki, jossa haastateltava vastaa kysymykseen: "Mikä susta tekee siitä kerronnasta sujuvaa?"

"no se on että ne rakenteet ovat no ainaki kasissa on on hyvin semmone erityyppisiä lauseita käytetään ja pitkiä virkkeitä ja vaihtelevia virkerakenteita jotenki ja sitte varmaan ihan ihan tämmönenki että "pääasia että henki on säilyny" tää on ihan niinku suomalaista ja sitte välimerkkien käyttö monipuolista..."

Arvioijat nostivat esille monia kirjoittamisen eri puolia korostavia konstruktia, kuten taksonomian mukainen taulukko 22 osoittaa. Jos taas konstruktia katsotaan frekvenssien perusteella, yhteistä arvioijille on kielikeskeisyys ja tekstityypin huomioiminen. Konstruktia tarkasteltaessa on kuitenkin otettava huomioon niiden merkitysrajojen päällekkäisyys: tietynlainen hajanaisuus koskee kaikkia konstruktia yleisellä tasolla. Konstruktihin ei viitata yksioikoisen yksiselitteisesti ja syste-

maattisesti, mikä voi johtua monesta eri tekijästä, esimerkiksi itse konstruktin merkityksestä puhujalle, tehtävätyypin luonteesta, tekstin tasosta tai kirjoittajan tyylistä. Konstruktit kyllä viittaavat kohteisiinsa, mutta eivät sanakirjamaisen yksiselitteisesti.

5.3.3 Haastattelun 2. vaihe: suoritusten järjestäminen ”ykkösestä yhdeksään”

5.3.3.1 Tekstien järjestäminen yksittäisten konstruktien mukaan

Triadivaiheen jälkeen haastateltavia pyydettiin järjestämään tekstit sekä kunkin vertailuvaiheessa esiin nousseen konstruktin että yleistason mukaan. Ennen kuin tekstejä alettiin järjestellä, haastateltavan kanssa neuvoteltiin kunkin konstruktin merkityksestä ja siitä, että haastattelija oli kirjannut sen oikein ja ymmärtänyt sen haastateltavan edellyttämällä tavalla. Joitakin konstruktia hylättiin tässä vaiheessa, koska tekstien järjestäminen ei onnistunut niiden mukaan. Tällaisia konstruktia olivat esimerkiksi kirjoittajan äidinkieli tai kulttuuritausta, jotka liittyivät vertailuvaiheessa lähinnä tekstin kuvailemiseen ei-arvottavasta näkökulmasta: ”tämä kirjottaja on näitten virheitten perusteella arabi” tai ”sit tää kolmone vois olla hyvin jostain tuolta venäjältä päin”.

Tekstien järjestäminen heikoimmasta vahvimpaan konstruktien mukaan ei ollut kaikissa tapauksissa helppoa, koska tekstit näyttivät muodostavan enemmänkin ryhmiä. Tekstit pystyi siis jakamaan esimerkiksi kahteen tai kolmeen ryhmään, mutta ryhmän sisällä tekstejä oli mahdotonta laittaa järjestykseen. Esimerkiksi ymmärrettävyyden mukaan tekstit olisi voinut jakaa kolmen ryhmään: 1) tekstiä on mahdoton ymmärtää, 2) teksteissä on joitakin kohtia, joita ei ymmärrä ja 3) tekstit ovat täysin ymmärrettäviä. Tai esimerkiksi tehtävänannon mukaan kahteen ryhmään: 1) tekstit täyttävät tehtävänannon ja 2) tekstit eivät täytä tehtävänantoa. Myös käsiala oli sellainen konstrukt, jonka mukaan tekstit olisi voinut olla luontevaa jakaa vain kahteen ryhmään.

Sen sijaan konkreettisten kriteerien, kuten pituuden, mukaan tekstit oli helppo järjestää, koska pituus sisältää jo itsessään mittatikun. Tekstit pystyi järjestämään melko luontevasti myös rakenteiden mukaan, koska arvioijat kommentoivat rakenteita siitä näkökulmasta käsin, minkä rakenteiden hallintaa ja miten systemaattista hallintaa tietyn tasoilta oppijoilta voi odottaa. Yhdeksän tekstin rakenteiden analysointi ei kuitenkaan onnistu pelkästään silmäilemällä, joten tekstien järjestämisen vaatima analyttisyys olisi vaatinutkin monen arvioijan mielestä enemmän aikaa.

Siinä missä haastateltavat erosivat toisistaan haastattelun aloituksessa, he erosivat myös tekstien järjestämisvaiheessa. Esimerkiksi Merja ja

Päivikki järjestelivät tekstit nopeahkosti, kun taas esimerkiksi Sanna tutki tekstejä pitkään, ennen kuin järjestys oli valmis. Joku etsi ensin ääripäät ja sijoitti muut niiden väliin, joku taas otti yhden teksteistä maamerkiksi ja vertasi siihen muita tekstejä ja sijoitteli siten tekstit paikalleen maamerkin mukaan. Joku taas aloitti suoraan järjestelemällä tekstit heikoimmasta vahvimpaan ja joku toinen vahvimmasta heikoimpaan. Jonkun arvioijan mielestä järjestäminen tuntui lähes mahdottomalta, esimerkiksi Heidi kommentoi useampaan otteeseen järjestettyjen tekstien olevan enemmän ryhmiä kuin järjestystä ykkösestä yhdeksään.

Olipa arvioijan tekstien järjestämisstrategia mikä tahansa, lähes kaikki arvioijat poimivat järjestettävistä teksteistä ensin selvimmät tapaukset eli ääripäät tai toisen niistä, konstruktin osalta joko hyvin heikon tai vahvan suorituksen. Samantyyppinen ääripäiden helppous tai selvyys oli havaittavissa myös sekä itse määritellyn että valmiiksi määritellyn arviointiasteikon soveltamisessa; useimpien arvioijien mukaan ykköset ja kuutoset oli kaikkein helpoimpia suorituksia arvioida.

Toinen tekstien järjestämiseen vaikuttava seikka oli se, miten hyvin haastateltava oli selvittänyt itselleen esiin nostamansa konstruktin merkityksen. Esimerkiksi triadivaiheessa tekstin looginen rakenne voi tuntua selvältä kahden tekstin välillä, jos toinen teksteistä etenee epäjohdonmukaisesti ja toinen teksti sen rinnalla on rakenteeltaan looginen. Mutta kun yhdeksän tekstiä täytyy järjestää loogisen rakenteen mukaan, konstrukt ”looginen rakenne” ei ehkä tunnu yhtä selkeältä. Silloin kun teksteistä oli vaikea osoittaa, mikä siitä tekee rakenteeltaan loogisen, haastateltava saattoi alkaa epäillä konstruktia ja sitä, miten sen mukaan yleensäkin voi uskottavasti järjestellä tekstejä. Sama vaikeus koski myös tarinamaisuutta: se, millaisia elementtejä toimivalta tarinalta edellytetään, ei tuntunut olevan ollenkaan itsestään selvää.

Myös konstruktien eräänlainen suhteellisuus, sisäkkäisyys tai päällekkäisyys aiheuttivat ongelmia: miten oikeinkirjoituksen voi erottaa esimerkiksi pituudesta, koska on todennäköistä, että lyhyessä tekstissä kirjoitusvirheitä on vähemmän kuin pitkässä? Lyhyys voi olla myös strategia välttää kirjoitusvirheitä. Yksi haastateltava totesikin, että pituus voi olla merkki siitä, että rohkea kirjoittaja uskaltaa ottaa riskin kirjoittaa pitkän tekstin. Kriteerien päällekkäisyys tuli esiin varsinkin siinä vaiheessa, kun tekstejä järjesteltiin yleistason mukaan. Esimerkiksi Tarja koki hyvin vaikeana sen, mitä asioita painottaisi tekstiä järjestellessään: onko vaikkapa pituus tärkeämpi piirre kuin sanasto, jos teksti on kovin lyhyt?

5.3.3.2 Tekstien järjestäminen niiden yleistason mukaan

Haastateltavat järjestivät ensin tekstit kontruktikohtaisesti ja sen jälkeen yleistason mukaan, joka perustui arvioijan itsensä käsitykseen tekstien yleistasosta. Taulukko 23 havainnollistaa sitä, mille sijalle haastateltavat sijoittivat kunkin tekstin sen yleistason mukaan. Tekstit voitaisiin sijoitusten yksimielisyyden mukaan jakaa teksteihin, joista ollaan lähes samaa mieltä, ja teksteihin, joiden sijoituksesta ollaan eri mieltä. Yksimielisyydessä on suurimmillaan viiden yksikön poikkeama; esimerkiksi teksti 1 on sijoitettu parhaimmillaan toiseksi vahvimmaksi ja heikoimmillaan seitsemänneksi. Yksimielisyys vaihtelee myös sillä tavalla, että arvioijien sijoituksissa voi olla 4 yksikön poikkeama, mutta kaikki arvioijat ovat kuitenkin sijoittaneet tekstin samaan päähän asteikkoa. Tällaista edustaa esimerkiksi teksti 8.

TAULUKKO 23 Yleistason mukaan järjestettyjen tekstien sijat.

Järjestys	Tekstit								
	Teksti 1	Teksti 2	Teksti 3	Teksti 4	Teksti 5	Teksti 6	Teksti 7	Teksti 8	Teksti 9
1 (vahvin)									
2									
3									
4									
5									
6									
7									
8									
9									
Mediaani	4	7	8	4	2	9	4	2	5

Teksti 6 on hyvä esimerkki samanmielisyydestä, koska kaikki arvioijat yhtä lukuun ottamatta pitävät kyseistä tekstiä kaikkein heikoimpana (ks. liite 10). Itse asiassa teksti 6 oli ainoa teksti, josta arvioijat olivat kyseisessä määrin samanmielisiä. Teksti 6 oli sekä kielellisesti että tekstuaalisesti niukka, joten tekstissä ei ollut ilmeisestikään paljon hajontaa aiheuttavaa ainesta.

Arvioijat ovat melko samanmielisiä myös teksteistä 8, 2 ja 3. Näissä tapauksissa ainakin 11 arvioijaa 15:sta sijoitti tekstit vierekkäisille tasoille. Teksti 8 on näistä hyvin idiomaattinen, varsinkin kielellisesti, mutta aihetta oli käsitelty hyvin yleisellä tasolla ja melko lyhyesti, mikä on saattanut aiheuttaa variaatiota arvioijien välillä. Tekstit 2 ja 3 on sijoitettu pääasiassa sijoille 7 ja 8, eli molemmat tekstit olivat arvioijien mielestä heikohkoja. Yhteistä teksteille on se, että ne ovat lyhyehköjä. Niitä taas erottaa toisistaan se, että teksti 2 on kielellisesti vahvempi mutta se on

lyhyt, perusasiat kertova teksti, ja teksti 3 taas on dramaattisenakin pidetty mielikuvitusta sisältävä ehyt tarina, jossa on kuitenkin taajaan rakenteellisia ja oikeinkirjoituksellisia puutteita.

Tekstit 1, 7 ja 9 näyttivät aiheuttavan eniten erimielisyyttä. Triadivaiheen aikana teksti 9 kiinnitti joidenkin arvioijien huomiota negatiivisessa mielessä epäloogisen rakenteensa vuoksi. Teksti 9 oli pitkäkö ja kielellisesti melko vahva tarina, joten osa arvioijista kiinnitti pääasiassa huomiota tekstin vahvuuksiin. Tekstissä kuitenkin kirjoittajan näkökulma vaihtui kesken tekstin kirjoittajaan itseensä, mikä voi osin selittää erimielisyyttä. Samoin teksti 1 jakoi arvioijien mielipiteitä jo triadivaiheessa poikkeuksellisen tekstityyppinsä vuoksi. Teksti 1 oli epikriisin muotoon kirjoitettu teksti, jossa oli käytetty joitain erikoisalan sanoja, kuten *ihorikko*. Tekstien järjestyksessä saattaa näkyä nimenomaan se, miten hyvin teksti haastateltavien mielestä vastaa tehtävänantoa. Tekstinä numero 7 oli varsin neutraali, se täytti tehtävänannon ja oli tekstuaalisesti ja kielellisesti tässä sarjassa keskitason suoritus. Näiden tekstien sijoitukset osuvat koko lailla asteikon keskivälille, jota arvioijat pitivät hankalana. Toisin sanoen heidän mielestään järjestys asteikon keskivälillä olisi saattanut yhtä hyvin olla erilainen, mikä voi selittää näiden tekstien sijoitushajontaa.

Tarkastelen seuraavaksi sitä, miten eri arvioijien tekstien sijoitukset yleistason mukaan korreloivat keskenään. Taulukossa 24 on esitetty arvioijien (N=15) väliset sijoituskorrelaatiot yleistason mukaan. Korrelaatioiden laskemisessa on käytetty Spearmanin järjestyskorrelaatiokerrointa.

TAULUKKO 24 Arvioijien (N=15) väliset sijoituskorrelaatiot yleistason mukaan.

Arvioija	Kaisu	Teija	Tarja	Eeva	Tiina	Inkeri	Raija	Heidi	Pia	Päivikki	Pirkko	Martti	Merja	Sanna	Pirjo
Kaisu	1	.450	.800**	.933**	.767*	.817**	.850**	.550	.867**	.933**	.867**	.750*	.767*	.783*	.817**
Teija		1	.850**	.683*	.617	.683	.650	.550	.417	.517	.667*	.700*	.617	.667*	.683*
Tarja			1	.917**	.800**	.900**	.833**	.567	.733*	.833**	.900**	.883**	.800	.833**	.917**
Eeva				1	.800**	.850**	.883**	.683*	.783*	.917**	.883**	.817**	.800**	.817**	.833**
Tiina					1	.583	.967**	.567	.600	.700*	.667*	.617	1.00**	.517	.783*
Inkeri						1	.650	.450	.883**	.883**	.967**	.933**	.583	.967**	.917**
Raija							1	.600	.683*	.800**	.717*	.617	.967**	.583	.767*
Heidi								1	.250	.400	.617	.583	.567	.517	.433
Pia									1	.950**	.833**	.733*	.600	.833**	.833**
Päivikki										1	.850**	.750*	.700*	.817**	.833**
Pirkko											1	.950**	.667*	.950**	.917**
Martti												1	.617	.950**	.933**
Merja													1	.517	.783*
Sanna														1	.883**
Pirjo															1

* Korrelaatio on merkitsevä tasolla .05

** Korrelaatio on merkitsevä tasolla .01

Arvioijat, joiden sijoitukset korreloivat parhaiten, olivat Tarja, Eeva, Päivikki, Pirkko ja Pirjo. Nämä arvioijat näyttäisivät järjestäneen tekstit melko yhdenmukaisesti. Heidän järjestys näyttäisi korreloivan heikoimmin muiden arvioijien järjestyksen kanssa; korrelaatio on merkitsevä ainoastaan Eevan kanssa, joskin Eevan järjestys korreloi vähintään .05-tasolla kaikkien arvioijien kanssa. Myös Teijan järjestys korreloi melko heikosti muiden kanssa. Lisäksi Tiinan ja Merjan kohdalla on joitakin ei-merkitseviä korrelaatioita.

Korrelaatioita sinänsä ei kannata lähteä analysoimaan sen syvemmin, mutta ne kuitenkin osoittavat arvioijakohtaisia linjoja tekstien järjestämisessä. Esimerkiksi taulukossa 23 tekstin 2 toinen sija on Heidän antama ja tekstin 9 ensimmäinen sija on Teijan antama. Kaikkein heikoin korrelaatio on Pian ja Heidän välillä, joten he näyttävät asettaneen tekstit toisistaan eniten poikkeavaan järjestykseen. Sen sijaan Tiina ja Merja järjestivät tekstit täysin identtisesti, kuten taulukosta 24 näkyy. Kaiken kaikkiaan arvioijat näyttivät järjestäneen tekstit jossakin määrin samansuuntaisesti, lukuun ottamatta edellä mainittuja tapauksia.

5.3.3.3 Konstruktien vaikutus yleistason

Missä määrin arvioijien vaikutelmaan yleistasosta vaikuttivat yksittäiset konstruktit? Millainen yhteys yleistasosta syntyvän vaikutelman ja yksittäisten konstruktien välillä on? Taulukossa 25 on esitetty Spearmanin korrelaatiot sen välillä, miten arvioijat järjestivät tekstit yksittäisten konstruktien mukaan ja yleistason mukaan. Taulukkoon 25 on otettu mukaan vain tasolla .01 merkitsevät korrelaatiot.

TAULUKKO 25 Tasolla .01 merkitsevät korrelaatiot yleistason ja yksittäisten konstruktien välillä.

Konst.	Kaisu	Teija	Tarja	Eeva	Tiina	Ink.	Raija	Heidi	Pia	Päiv.	Pirkk.	Martti	Merja	Sanna	Pirjo	ka
Sanasto	.917**			.933**		.883**			.800**		.917**			.933**	.933	0.758
Tarinam.											.817**		.917**			0.495
Rakenteet		.867**	.950**	.983**		.950**	.967**			.950**	.950**		.817**	.833**	.983**	0.879
Pituus													.817**	.817**		0.632
Emot.															.950**	0.552
Sisältö													.850**			0.700
Lauser.				.933**								.933**				0.817
Loog. rak.										.817**	.800**					0.729
Teht. pan.								.100**						.900**		0.446
Ymmärr.	.983**			.817**						.817**						0.872
Idiom.						.950**						.917**				0.845
Tekstuaal.							.933**					.833**				0.805
Välimerkk		.817**														0.792
Taivutus									.817**							0.700
Konvent.	.950**															0.950

** Korrelaatio on merkitsevä tasolla .01

Kunakin arvioijan kohdalla yleistason merkittävästi (.01) vaikuttaneiden konstruktien määrä vaihtelee nolasta neljään. Yleisin yleistason kanssa korreloinut konstrukti on rakenteet, koska se korreloi merkitsevästi yleistason mukaisen järjestyksen kanssa 10 arvioijalla. Pian ja Martin kohdalla myös muiden rakenteellisten konstruktien (lauserakenteet ja taivutus) mukainen järjestys korreloi merkitsevästi yleistason järjestyksen kanssa. Itse asiassa taulukon 25 mukaan osan arvioijista, esimerkiksi Eevan, Teijan ja Tarjan, korrelaatiot antavat ymmärtää, että heidän yleistason mukainen järjestyksensä korreloi parhaiten kielellisten konstruktien eli rakenteiden ja sanaston kanssa. Esimerkiksi Tarjan tapauksessa rakenteiden lisäksi sanasto korreloi yleistason kanssa (.73**), kun taas muut konstruktit eivät korreloineet noin voimakkaasti yleistason kanssa.

Sanaston ja tarinamaisuuden nosti esille triadivaiheessa 14 haastateltavaa. Korrelaatioita tarkasteltaessa sanastolla on tilastollisesti merkitsevä yhteys yleistason seitsemän arvioijan osalta, mutta tarinamaisuus korreloi merkitsevästi yleistason kanssa vain yhden arvioijan kohdalla. Useimmissa tapauksissa tarinamaisuuden mukaisella järjestyksellä oli hyvin alhainen korrelaatio yleistason mukaisen järjestyksen kanssa, esim. -.39, -.03, .07, .19, .47. Tarinamaisuuden frekventtiyden vuoksi olisi ollut odotettavissa, että sen vaikutus olisi ollut vahvempi yleistason, kun nyt sen vaikutus oli jopa negatiivinen. Toisaalta tarinamaisuuden korrelaatioihin on voinut vaikuttaa myös sen kompleksinen luonne, mikä tuli jo edellä esille.

Myös käsiala ja pituus olivat frekventtejä konstruktia triadivaiheessa, mutta pituus korreloi merkitsevästi vain Merjan ja Sannan kohdalla. Näidenkään arvioijien kohdalla pituus ei ollut ainoa konstrukti vaan yksi kolmesta/neljästä. Sen sijaan käsialalla ei näyttänyt olevan kenenkään kohdalla merkitsevää korrelaatiota yleistason kanssa. Käsialaa koskevista korrelaatioista korkein oli .62, mutta muut olivat alhaisempia, esimerkiksi -.38, -.18, .13, .15, .22. Käsialan alhaiset korrelaatiot voi tulkita positiivisesti siinä hengessä, että käsialan ei pitäisikään vaikuttaa kirjoittamisesta tehtäviin arvioihin (ks. esim. Charney 1984; Hamp-Lyons 1991).

Jos arvioijia tarkastelee yksilötasolla, näyttäivät Tiinan ja Heidän korrelaatorivit muista poikkeavilta. Tiinan rivi taulukossa on tyhjä, eli hänen kohdallaan konstruktien mukaiset järjestykset eivät korreloineet merkitsevästi – ei tasolla .01 eikä myöskään tasolla .05 – yleistason järjestyksen kanssa. Tiina nosti esiin triadivaiheessa kaikkiaan seitsemän konstruktia, joista parhaiten korreloi yleistason kanssa rakenteet (.63). Tekstejä vertaaillessaan hän käytti taajimmin sanastoa, rakenteita ja tarinamaisuutta tekstejä erottelevina ja yhdistävinä konstruktina.

Heidin kohdalla on taas yksi merkitsevä (tasolla .01) korrelaatio eli 1.00** tehtävään paneutumisen kohdalla. Toisin sanoen Heidi on järjestänyt tekstit samaan järjestykseen tehtävään paneutuminen -konstruktin ja yleistason mukaan. Heidin muut merkitsevät korrelaatiot (tasolla .05) liittyivät idiomaattisuuteen ja emotionaalisuuteen. Tämä saattaa selittää sitä, miksi Heidin korrelaatiot suhteessa toisiin arvioijiin ovat niin alhaisia. Heidi näyttäisi painottaneen järjestämisessä eri piirteitä kuin muut arvioijat.

Kokonaisuudessaan taulukon 25 aineisto viestii siitä, että arvioijille jossakin määrin yhteisiä konstruktia olivat rakenteet ja sanasto. Toisaalta on hyvä muistuttaa siitä, että kyse on korrelaatioista eikä siitä, mitä arvioijien päässä liikkuu heidän arvottaessaan tekstejä. Korrelaatiot antavat pääasiassa vihjeitä siitä, millaiset tekijät myötävaikuttavat siihen, miten tekstejä arvotetaan. Kyseiset tulokset kaiuttavat myös jossakin määrin aikaisempia tutkimustuloksia ainakin kielellisten konstruktien osalta. Voidaanko edellisen pohjalta sitten väittää, että arvioijat ovat kielellisesti orientoituneita ja tarkastelevat kirjoittamista kielellisten linssien läpi?

Vaikka tekstuaaliset ja tehtävään liittyvät konstruktit olivat mukana arviointikeskustelussa, niiden mukaiset järjestykset eivät kuitenkaan korreloineet yleistason järjestyksen kanssa. Toisaalta nämä piirteet eivät myöskään olleet kaikkien konstruktien osalta kovin frekventtejä. Myöskään emotionaalisten tekijöiden mukainen järjestys ei osoittanut merkitsevää yhteyttä yleistason mukaiseen järjestykseen kuin yhdellä arvioijalla. Muut emotionaalisuuden ja yleistason väliset korrelaatiot vaihtelivat välillä .10 ja .70, mikä vaikuttaa arvioinnin luotettavuuden kannalta hyvältä. Toisaalta emotionaalisuus konstruktina on hyvin suhteellinen, koska se on kietoutunut myös kaikkiin muihin konstruktihin, esimerkiksi lukemistilanteeseen, kirjoittajasta syntyvään vaikutelmaan ja arvioijan vireystilaan. Se on myös aika hankalasti observeitavissa ja erotettavissa. Lisäksi se kuuluu osana inhimilliseen toimintaan, minkä vuoksi sen läsnäolo – ainakin jossakin määrin – on odotuksenmukaista.

5.3.4 Valintaverkkohaastattelun anti

Kokonaisuudessaan valintaverkkohaastattelu on joustava ja sopiva työkalu sen tarkastelemiseen, mitä arvioijan mielessä liikkuu arviointiprosessin aikana. Tosin valintaverkko on varsin vaativa haastattelu menetelmänä ja edellyttää toimiakseen mahdollisimman paljon kokemusta haastattelutyypin soveltamisesta. Haastattelijan näkökulmasta jokainen haastattelu toi jotakin uutta esiin sekä metodista että tutkittavasta sisällöstä. Haastattelut myös selittivät omalla tavallaan, miten monin eri

tavoin arvioija voi tarkastella tekstejä ja kielentää havaitsemiaan eroja. Vaikka valintaverkkohaastattelu oli mukana kokeilumielessä, sen anti ei jäänyt pelkäksi kokeiluksi. Se toimi odotusten mukaisesti eli lisäsi ymmärrystä arvioijien tavasta tarkastella tekstejä.

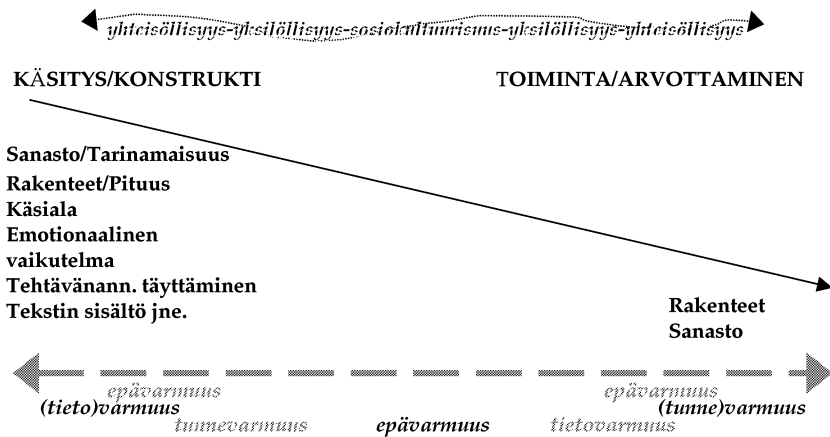
Eräissä mielessä isoltakin näyttävät erot esimerkiksi tavassa puhua jostakin tekstistä ja sen piirteistä voivat kuitenkin johtaa hyvin samantyyppiseen arviointiin, tässä yhteydessä tekstien järjestämiseen. Haastattelu myös osoitti, että värikkäätkin ja arvottavat kommentit esimerkiksi kirjoittajan taustasta eivät välttämättä vaikuta tekstistä tehtäviin havaintoihin tai tekstin arvottamiseen. Erot arvioijien välillä ja subjektiiviset kommentit eivät siis välttämättä tarkoita sitä, että tekstejä arvioidaan epäyhdenmukaisesti, epäloogisesti tai sattumanvaraisiin piirteisiin tukeutuen, vaikka tältäkin mahdollisuudelta ei tule sulkea silmiä.

Valintaverkkohaastattelu saa haastattelutyypinä pohtimaan sitä, mitä arvioija oikeastaan näkee tekstissä: katsooko hän etupäässä kielen sisälle, tekstin sisältöön vai jonkinlaiseen kokonaisuuteen? Jos arvioija sitten tarkastelee tekstiä pääasiassa jonkin tietyn piirteen osalta, missä määrin hän kiinnittää huomiota muihin tai pakottaa itsensä katsomaan myös muita piirteitä teksteistä? Ensimmäisen kierroksen haastatteluissaan esimerkiksi Pirjo ja Tiina molemmat kritisoivat omaa kieliooppipainotteista näkökulmaansa ja totesivat, että olisi pitänyt kiinnittää huomiota enemmän myös muihin seikkoihin, mitä he eivät omasta mielestään kuitenkaan tehneet tarpeeksi.

Missä määrin arvioija tietoisesti kiinnittää huomiota johonkin tiettyyn piirteeseen ja miten hyvin hän pystyy säätelemään toimintaansa? Ja jos hän tietoisesti ohjaa arviointiprosessia, mikä vaikuttaa prosessin luonteeseen ja kuinka hyvin hän pystyy motivoimaan itsensä esimerkiksi muutoksiin? Kyseiset seikat ovat aika vaikeasti havaittavissa, vaikka toisaalta tämänkin tutkimuksen eri vaiheiden tulokset osoittavat, että arvioijat kyllä suhtautuvat varsin analyyttisesti omaan arviointitapaansa. Matka siitä, mitä piirteitä teksteistä tarkastellaan, siihen, mitä piirteitä viime kädessä arvostetaan, on kuitenkin melko mutkainen.

Triadivaiheessa oli mielenkiintoista se, miten eri tavalla rinnakkain olevat tekstit saavat näyttämään toisensa erilaisilta. Kun yksi teksti on tiettyssä kolmikossa vähän kömpelö eikä niin idiomaattinen, se saattaa taas olla jossakin toisessa kolmikossa ehdottomasti vahvin ja sisällökkäin. Myös todellisissa arviointitilanteissa tekstit ovat satunnaisessa järjestyksessä, jossa eri tavalla heikot ja vahvat tekstit ovat sekaisin. Vaikutus on samantyyppinen, eli tekstejä tulee varmastikin verranneeksi toisiinsa. Tekstien keskinäinen suhde saattaa vaikuttaa sitä enemmän, mitä väljemmät tai monitulkintaisemmat arviointikriteerit ovat.

Tekstejä vertaillessaan haastateltavat olivat vapaita nostamaan esiin mitä tahansa konstruktaje, joskin vertailuvaiheen konstruktit olivat kaikki odotuksenmukaisia (ks. taulukot 8 ja 21). Ehkä tämä kertoo siitä, että teksteissä kiinnitetään huomiota niihin piirteisiin, joihin on opittu kiinnittämään huomiota. Kuten arvioijien omista asteikoista (ks. taulukot 14–18), niin myös tässäkin arvioijat nostivat yhdenmukaisesti esiin kielellisiä piirteitä. Vaikka tulokset eivät annakaan oikeutta väittää, että ne ohjaavat arviointia, näyttäisi niillä kuitenkin olevan näkyvin rooli tekstien laadun arvioinnissa. Kielellisten piirteiden vahvuus näkyi tässä haastattelussa varsinkin tekstien järjestämisessä yleistason mukaan, kuten kuviosta 25 näkyy.



KUVIO 25 Käsitusten ja toiminnan rajapinnalla.

Kuvio 25 havainnollistaa myös sitä, miten arvioijien varmuusaste vaihteli haastattelun kahdessa eri vaiheessa. Kaikkein varmimpia arvioijat olivat silloin, kun he vertailivat tekstejä eri konstruktien mukaan. Arvioijien varmuus saattoi liittyä sekä asiantuntijamaiseen fakta- ja tietopohjaiseen että irrationaalisen lukijan tunnepohjaiseen varmuuteen (ks. myös kuvio 24). Sen sijaan tekstien järjesteleminen sekä konstruktikohtaisesti että yleistason mukaan herätti epävarmuutta lähes kaikissa arvioijissa. Haastattelujen perusteella näytti siltä, että tekstin analysointi suhteessa toisiin teksteihin ei vielä synnyttänyt epävarmuutta, mutta tekstin arvottaminen, varsinkin jonkin yksittäisen konstruktin mukaan, nostatti esiin kommentteja arvottamisen vaikeudesta ja jopa mahdottomuudesta.

Tekstien järjestämiseen eli niiden arvottamiseen liittyvä yleinen epävarmuus saattoi osaltaan aiheuttaa myös sen, että arvioijien yleistason

mukainen järjestys oli samansuuntainen rakenteellisen konstruktin mukaisen järjestyksen kanssa. Ehkä epävarmuudessa oli yksinkertaisinta ja turvallisinta tukeutua rakenteisiin ja sanastoon. Ehkä myös muiden konstruktien eräänlainen epämääräisyys tai progressiomainen luokattomuus sai arvioijat järjestämään tekstit rakenteiden mukaan. Olipa syy mikä tahansa, tuntuu siltä, että arvioijat itse eivät ajatelleet järjestäessään tekstejä yleistason mukaan järjestävänsä niitä rakenteiden hallinnan mukaiseen järjestykseen.

Tuloksia tarkastellessa on syytä suhtautua kriittisesti myös siihen, miten pitkälle meneviä johtopäätöksiä aineisto sallii tehdä. Haastateltavilla oli myös tietyt odotukset muun muassa siitä, miten heidän oletetaan käyttäytyvän, mitä on järkevää sanoa kyseisessä tilanteessa, kuinka kuulostaa ammattimaiselta ja mikä on sallittua sanoa ääneen kuulostamatta epäeettiseltä tai epäkohteliaalta. Lisäksi ohjasin varmasti itsekin haastattelijana haastattelun ilmapiiriä omalla käyttäytymiselläni sekä tietoisesti että tiedostamatta.

Kaiken kaikkiaan kognitiivinen näkökulma arviointiin saa kysymään, miten psykologinen maailma oikeastaan on. Tai tarkemmin sanottuna miten kognitiivinen se on ja missä määrin toimintamme ja käyttäytymisemme riippuu kognitiosta? Kellyn mukaan kukin yksilö konstruoi oman maailman, mutta se, millainen maailmasta tulee, on jossakin määrin riippuvainen yksilöä ympäröivästä sosiaalisesta maailmasta. Jotta voisimme ymmärtää arvioijan toimintaa ja neuvotella arviointiin liittyvistä merkityksistä, meidän pitäisi pohtia taidon arvottamista siihen liittyvine kysymyksineen ensisijaisesti arvioijien mutta myös koko arviointiyhteisön kanssa. Vaughanin (1991) tutkimuksen mukaan arvioijat arvioivat suorituksia viime kädessä oman skeemansa mukaan. Juuri näiden omien skeemojen, konstruktien, käsitysten ja yleensäkin yksilöllisten erojen selvittäminen auttaa ymmärtämään paremmin arviointiprosessia ja siihen vaikuttavia tekijöitä.

6. KATSAUS TUTKIMUSPROSESSIIN

6.1 Tulosten kertomaa

Tässä tutkimuksessa on haettu vastausta toisaalta arvioijien yhdenmukaisuutta ja johdonmukaisuutta koskeviin kysymyksiin ja toisaalta siihen, millaisia arvioijien käsitykset ovat kirjoittamisesta taitona ja miten he raportoivat suhtautuneensa yleensä arviointiprosessiin. Tutkimuksen selittävä aineisto koostuu arvioijien antamista arvioista eli numeerisesta aineistosta ja ymmärtävä aineisto kolmesta eri haastattelusta. Mitä tutkimus sitten itse asiassa selittää ja mitä se ymmärtää; selittääkö se sitä, mitä se ymmärtää, ja ymmärtääkö se, mitä se selittää?

6.1.1 Arvioinnin johdonmukaisuus ja yhdenmukaisuus

Yleisesti ottaen tutkimuksen kvantitatiivisten tulosten mukaan arvioijat arvioivat suorituksia varsin yhdenmukaisesti, toisin sanoen arvioijien välillä ei ollut numeerisen aineiston perusteella suuria eroja millään arviointikierröksellä. Yhdenmukaisuuteen voi olla olemassa monenlaisia syitä. Yhtäältä tulos antaa olettaa, että määritelty arviointiasteikko ei vaikuttanut odotuksenmukaisella tavalla: arviointi ei ollut merkittävästi yhdenmukaisempaa toisella ja kolmannella arviointikierröksellä kuin ensimmäisellä arviointikierröksellä. Toisaalta hypoteesi saattoi olla turha. Ehkä ensimmäisen arviointikierröksen yhdenmukaisuus viestii yksinkertaisesti siitä, että arvioijilla on saamansa koulutuksen, alan yhtenäisyyden tai jonkin muun vastaavanlaisen syyn perusteella keskenään yhdenmukainen kirjoittamiskäsitys, joka ohjaa heitä samansuuntaisiin tasoarvioihin. Ehkä myös arvioivat tekstit edustivat eräänlaisia oppijoiden prototyyppejä, jotka olivat arvioijille tuttuja opetuskokemuksen perusteella.

Tulosten perusteella voisi myös päätellä, että arvioijien itsensä määrittelemät asteikot olivat sisällöltään samantyyppisiä määritellyn asteikon kanssa. Mahdollista on myös se, että arvioijat seurasivat omaa kirjoittamiskäsitystään määritellystä asteikosta välittämättä. Jos tilanne oli tämä, se saattoi johtua ainakin kahdesta syystä: 1) määritelty asteikko oli niin avoin, että arvioijan täytyi täydentää sitä omilla kriteereillään tai se oli niin huono sovellettavaksi, että arvioija tukeutui sen sijasta omaan

asteikkoonsa, tai 2) arvioijat arvioivat omaan asteikkoonsa tukeutuen riippumatta siitä, millainen määritely asteikko oli. Mikäli arvioijat nojasivat omiin kirjoittamiskäsityksiinsä, niiden täytyi yhdenmukaisuuden perusteella olla ainakin jossakin määrin samantyyppisiä. Yhdenmukaisuus edellyttää myös, että arvioijat toimivat sisäisesti johdonmukaisesti tai sitten linjassa epäjohdonmukaisesti.

Yleisesti ottaen kvantitatiivisia tuloksia tarkasteltaessa silmiin pistää se, että arvioijien sisäinen johdonmukaisuus tuntuu osoittaneen suurempia poikkeamia kuin heidän keskinäinen yhdenmukaisuutensa. Karkeasti sanoen arvioijat jakaantuivat johdonmukaisuuden perusteella sellaisiin arvioijiin, joilla määritely asteikko johdonmukaisti arviointia, ja sellaisiin, joilla määritellyllä asteikolla ei ollut vaikutusta sisäiseen johdonmukaisuuteen. Tässä mielessä arvioijien sisäisten poikkeamien on täytynyt olla niin pieniä, että ne eivät näy yhdenmukaisuutta kuvaavissa korrelaatioissa, tai sitten heidän poikkeamansa olivat toisiinsa nähden samansuuntaisia mittauskerrasta toiseen. Faktorianalyysin tulos ja eri mittauskertojen väliset pienet erot antaisivat kuitenkin aiheen olettaa, että arvioijat arvioivat jotakin samaa piirrettä jokaisella mittauskerralla riippumatta arvioinnissa käytettävästä asteikosta.

Toisaalta tällaisen tuloksen voi tulkita kertovan arvioijien ammattilaisuudesta ja siitä, että arvioijilla on yhdensuuntainen käsitys kirjoittamisesta taitona ja he soveltavat sitä yhdenmukaisesti. Jos tämä on johtopäätös tuloksista, se ei kuitenkaan kerro, millainen arvioijien kirjoittamiskäsitys on. Millaisista piirteistä, konstrakteista tai kriteereistä kirjoittamiskäsitys koostuu ja miten näiden eri piirteiden osalta kirjoittamistaitoa arvioidaan? Vai oliko yhdenmukaisuuden takana sittenkin jokin yksi muita painavampi kriteeri?

Ehkä kirjoittamiskäsitystä ei ole edes mielekästä ja tarkoituksenmukaista lähestyä analyttisestä näkökulmasta, koska arvioijat arvioivat suorituksia holistisesti. Ehkä heidän kirjoittamiskäsityksensä oli arviointitavan ohjaamana holistinen, jonka mukaan kirjoittaminen on jotakin enemmän kuin osiensa summa. Holistisesti ajatellen tuntuu odotuksenmukaiselta, että arvioijien yleisluonteiset käsitykset ja mielikuvat tekstien keskimääräisestä tasosta olisivat yhdenmukaisia ja samansuuntaisia pelkästään jo heidän koulutuksensa ja työkokemuksensa eli heidän ammatillisen osaamisensa perusteella.

Ehkä niin, mutta eikö tämä kuulosta liian yksinkertaiselta tai liian järkeenkäyvältä ollakseen totta. Miksi tuudittautua liian helppoon selitykseen: kaiken takana on holistinen kirjoittamiskäsitys? On kuitenkin pidettävä mielessä, että numeerinen aineisto ei riitä perusteluksi, kun on tarpeen vastata kysymykseen käsitysten laadusta ja luonteesta. Käsityksien osalta vastuu kuului haastatteluille, joiden tarkoituksena oli

vastata kysymykseen, millaisena taitona arvioijat näkevät kirjoittamisen. Millainen käsitys kirjoittamisesta ohjaa arviointia ja miten erilaiset arviointiprosessissa vaikuttavat tekijät ohjaavat käsityksen soveltamista?

6.1.2 Kirjoittamiskäsitysten ja arviointiprosessin tarkastelua

Charneyn (1984, 78) mukaan kirjoittamisen arviot heijastelevat arvioijien samanmielisyyttä vain silmiinpistävästä mutta pinnallisista kirjoittamisen piirteistä, kuten käsialan laadusta tai oikeinkirjoitusvirheistä. Charneyn toteamus tuntuu aika lohduttomalta. Joko se viestii siitä, että arviot ovat tavallaan kykenemättömiä välittämään syvempää informaatiota tai sitten ne peilaavat arvioijien varsin mekaanisia kirjoittamiskäsityksiä. Tämän tutkimuksen perusteella arvioijien kirjoittamiskäsitykset eivät kuitenkaan liittyneet pelkästään kirjoittamisen kannalta jokseenkin toisarvoisiin kriteereihin, kuten käsialaan ja oikeinkirjoitusvirheisiin. Nekin olivat kyllä mukana, mutta arvioijien kommentoiman mukaan annetut arviot heijastelivat myös muita osaitoja – tosin sillä varauksella, että kommenttien tarkkuus ja osuvuus riippuivat siitä, missä määrin käsitysten ja toiminnan taso vastasivat toisiaan.

Jokaisen kolmen haastattelun perusteella nousi yksi kriteeri varsin yhdenmukaisesti esiin: rakenteiden hallinta. Ensimmäisellä arviointikierroksella se oli frekventein ja systemaattisimmin perusteltu kriteeri, joka oli arvioijien asteikoissa mukana. Toisella arviointikierroksella arvioijat kaipasivat määriteltyyn asteikkoon ennen kaikkea kielen hallintaan liittyvää kuvausta. Kolmannessa haastattelussa arvioijat nostivat vertailtavista teksteistä esiin monenlaisia konstruktteja, mutta kuitenkin arvottivat samoja tekstejä kaikkein tuntuvimmin rakenteiden ja sanaston perusteella. Voi olla, että rakenteet olivat esiin tuoduista kriteereistä kaikkein systemaattisimmin etualalla jokaisella kolmella arviointikierroksella, minkä vuoksi arviointi oli niin yhdenmukaista asteikosta riippumatta.

Tällainen yleistys voi olla liian karkea, mutta toisaalta rakenteiden painokasta roolia ei sitäkään sovi kieltää. Osa arvioijistahan kertoi suoraan painottaneensa ennen kaikkea rakenteiden hallintaa; ehkä muutkin arvioijat tekivät niin, vaikka he sanoivatkin tukeutuneensa rakenteisiin lähinnä rajatapauksissa tai joissakin muissa epäselkeissä tilanteissa. Rakenteisiin tukeutuminen ei liene mitenkään epätavallista, vaan ne ovat tekstin piirre, johon on helppo ottaa kantaa niiden konkreettisuuden, hierarkkisen luonteen ja opetuksessa olevan aseman vuoksi.

Rakenteiden asema tuntuu korostuvan varsinkin arviointitilanteessa. Eräissä koulutustilaisuudessa opettajat sanoivat arvostavansa rakenteita

sen vuoksi, että kiinnittäessään huomiota rakenteiden hallintaan he tuntevat tekevänsä arviointia oikein. Kyse ei ole niinkään rakenteiden merkittävästä asemasta kielitaidossa, vaan niiden konkreettisuudesta ja tuttuudesta.

Rakenteiden korostuneeseen asemaan voi olla toki muitakin syitä. Ehkä tässä tutkimuksessa kyse ei ollutkaan rakenteiden yleensä vahvasta asemasta, vaan siitä, että käyttäjakeskeisyydessään ja holistisuudessaan määriteltä aseikko ajoi arvioijat tukeutumaan arviointia paremmin tukevaan ratkaisuun. Toisaalta arvioijien itsensä määrittelemät asteikot ja valintaverkkoahaastattelu puhuivat rakenteiden vahvan aseman puolesta. Uskaltaisiko tämän päättelyketjun tuloksena pitää tämän tutkimuksen arvioijien kirjoittamiskäsityksiä rakennelähtöisinä?

Ensimmäisen kierroksen haastattelujen sekä valintaverkkoahaastattelun perusteellahan näytti siltä, että arvioijien kirjoittamiskäsitys oli jotakin enemmän kuin se, millaisena se heijastui arviointiasteikoissa tai tekstien arvottamisessa yleistason mukaan. Tämän perusteella väittäisin, että arvioijien kirjoittamiskäsitys on suppeampi silloin, kun he arvottavat tekstejä. Toisin sanoen arvioijilla on ikään kuin kaksi käsitystasoa: 1) yleinen kirjoittamiskäsitys ja 2) arvottava kirjoittamiskäsitys. Kirjoittamiskäsityksistä nimenomaan arvottava on rakennekeskeinen. Arvioijien taitokäsitysten kahtiajakoisuus on sikäli odotuksenmukaista, että yksi arvioinnin ikuisuusongelmista liittyy nimenomaan taitokäsitysten mallinnettavuuteen ja konkretisoinnin ongelmallisuuteen. Miten tehdä oikeutta taidolle, kun se pilkotaan osiksi, kriteerikohtaisiksi tai taitotasokohtaisiksi kuvauksiksi?

Määritellyn asteikon osalta oli oletettavaa, että sen mukaan annetut arviot heijastelivat yleiseurooppalaisen viitekehyksen mukaista kirjoittamiskäsitystä sillä varauksella, että sekä kirjoittamisen tehtävät että arviointiasteikko olivat linjassa viitekehyksen kanssa ja että arvioijat arvioivat suorituksia viitekehyksen kirjoittamiskäsityksen mukaisesti. Arvioijathan eivät saaneet käyttöönsä muuta kuin määritellyn asteikon; he eivät tutustuneet viitekehyksen kielitaitokäsitykseen eivätkä saaneet koulutusta asteikon tulkintaan, joten kovin tarkasta vastaavuudesta arvioiden ja käsitysten välillä ei voi puhua.

Arvioijien esittämän kritiikin mukaanhan tehtävät ja asteikko eivät vastanneet toisiaan kaikilta osin. Myöskään asteikon tekstityyppi-painotteinen kirjoittamiskäsitys ei miellyttänyt arvioijia, koska se ei tukenut riittävästi arviointia. Asteikkoon kohdistunut kritiikki heijasteli osaltaan myös asteikon suhdetta sen taustalla olevaan kirjoittamiskäsitykseen. Yleiseurooppalaisen viitekehyksen (1996; 2001) mukaanhan kirjoittaminen määritellään pitkälti erilaisten tekstityyppien kirjoittamisena, mikä jää kirjoittamisen opettamisen ja oppimisen näkökulmasta

aika suppeaksi tai laveaksi näkemyksesi. Samaa suppeutta tai laveutta edustaa myös käsityksen pohjalta laadittu kirjoittamisen asteikko.

Vaikka määritely asteikko oli monella tavalla puutteellinen ja vaikka sitä käytettiin kylmiltään, en kuitenkaan ole varma, millaiseen kirjoittamiskäsitykseen perustuva asteikko olisi tulkittu yksiselitteisemmin. Palaan epäilyksieni kera takaisin käsitysten kahtiajakoisuuteen. Mielestäni se, että arvioijat arvottivat valintaverkkohaastattelussa tekstejä konstruktikohtaisesti, puhuu myös käsitysten kahtiajakoisuuden puolesta. Arvioijathan nostivat tekstejä vertaillen esiin monenlaisia konstruktia, mutta kokivat kuitenkin tekstien järjestelemisen vaikeaksi niiden mukaan, varsinkin keskivälin tekstien osalta. Konstruktien määrittelemisenäkään ei ollut selvää.

Vaikka kahtiajakoisuus saattoi johtua siitä, että arvioijat eivät pystyneet itse määrittelemään konstruktiaan kyllin tarkasti ja käyttämään niitä soveltaen, näkisin kuitenkin, että valmiiksi määriteltyihin kriteereihin voi liittyä sama ongelma. Vaikka kriteerit olisi kuvattu mahdollisimman yksiselitteisesti ja vaikka ne olisivat lähellä arvioijan omaa kielikäsitystä, niiden tulkinta ja soveltaminen teksteihin on monimutkaista. Monimutkaiseksi sen tekee ensinnäkin se, että suoritus on jotakin paljon enemmän kuin kriteeri pystyy kuvaamaan. Toiseksi kriteerin soveltaminen on sekä suorituskohtainen että arvioijakohtainen vuorovaikutusprosessi. Toisin sanoen suoritukset ovat erilaisia ja niin ovat myös kunkin suorituksen arvottamistilanteet kaikkine muuttujineen. Kyse on siis dynaamisesta prosessista, johon yritetään soveltaa staattisia määritelmiä. En tarkoita tällä tietenkään sitä, että kriteerit eivät voisi toimia, mutta niiden soveltaminen ei ole ongelmatonta eikä edes helppoa.

Myös kriteerit ovat erilaisia luonteeltaan. Toisista olemme varmempia ja toisista vähemmän varmoja. Toiset ovat abstrakteja ja toiset konkreettisia. Pituus eroaa koheesiosta ja rakenteet ymmärrettävyydestä, vaikka ne eivät olekaan toisistaan irrallisia. Paperille määriteltyinä kriteerit ovat selkeitä ja rajattuja, mutta eri suorituksissa ne voivat näyttää täysin erilaisilta ja ne voivat muodostaa erilaisia yhdistelmiä. Samantyyppinen ongelma liittyy tasokuvauksiin, joissa yritetään yhdistää eri kriteerien hallinta yleiseksi kuvaukseksi. Eräässä mielessä holistinen arviointi on oiva tapa välttää atomistisuus ja hallinnan harha, joka liittyy kriteerikohtaisuuteen; toisaalta se taas voi tarkoituksesta nostaa jonkin kriteerin valta-asemaan holistisuuden nimissä. Lisäksi holistisuuskin pakottaa lähestymään taitoa kriteerikohtaisesti: kirjoittaminen koostuu jostakin.

Arvioijien kirjoittamiskäsityksiä koskevaan tutkimuskysymykseen en siis pysty antamaan vastausta, joka olisi selvärajainen ja yksityiskohtainen. Arvioijan oma kirjoittamiskäsitys oli kuitenkin läsnä

koko ajan, ja se heijastuu epäsuorasti muun muassa määritellystä asteikosta esitetyissä kommentteissa. Mielestäni tulosten avulla pystyy kuitenkin hahmottamaan yleisiä linjoja, joita taulukot 14–18 ja kuvio 25 havainnollistavat arvottavien kirjoittamiskäsitysten ja kuvio 23 arviointiprosessin osalta. Siihen, miten eri osatekijät ovat etualalla, vaikuttavat myös kunkin arvioijan kulloinenkin vireys-, tahto- ja tunnetila.

Arviointiprosessin monitekijäisyydestä huolimatta arvioija ei toimi mielivaltaisesti poukkoillen. Esimerkiksi epävarmuuden hetkinä arvioija saattoi olla tietoinen ja ohjata omaa arviointiprosessinsa kulkua. Ratkaisut eivät kuitenkaan olleet aina tarkoituksenmukaisia tai järkeviä. Sitä, miten tietoisia arvioijat olivat tai missä määrin he ohjasivat oman arviointiprosessinsa kulkua, on vaikea sanoa. Kaiken kaikkiaan arviointiprosessille näyttäisi olleen luonteenomaista varmuuden ja epävarmuuden vuorottelu sen mukaan, millaiset tekijät olivat kulloinkin etualalla.

Arvioijien käsitysten tutkiminen osoittaa myös, että voi olla kovin keinotekoista erottaa toisistaan arvioijiin liittyvä reliabelius ja arvioinnin validius. Mitä syvemmin arvioijan toimintaa halutaan ymmärtää, sitä lähemmäksi tullaan validiuden eli sen tutkimista, arvioivatko arvioijat 1) sitä taitoa, mitä he sanovat arvioivansa itse määrittelemissään asteikossa ja 2) sitä taitoa, mitä heidän oletetaan arvioivan määritellyllä asteikolla. Vaikka tässä tutkittavat käsitykset eivät ole vastanneet näihin kysymyksiin suoraan, ne kuitenkin osoittavat, että kyseessä ei ole yksiselitteisen rajattu taitokonstrukti vaan itse asiassa erilaisten ja toisilleen jopa vastakkaistenkin käsitysten verkko. Kuten Marton (1994) toteaa, käsitykset ovat dynaamisia ja yksilö voi muuttaa niitä useasti varsin lyhyessäkin ajassa.

Toisaalta tästä herää hieman ristiriitainen ajatus siitä, missä määrin kyse on arvostuksista ja arvoista ja missä määrin käsityksistä. Haluaisin kuitenkin uskoa, että arvioijilla on yhtenäinen kielikäsite, josta yksittäiset, tilannesidonnaisemmat, käsitykset nousevat. Käsitykset voivat muuttua ja olla keskenään ristiriitaisiakin, mutta arvostuksia tuntuisi olevan vaikeampi muuttaa – ainakaan niiltä osin, mitä arvioijien valinnat osoittivat. Haastattelut osoittavat mielestäni hyvin sen, että käsitystasolla puhuttaessa voimme vaihtaa näkökulmaa ja perustella asioita toisaalta ja toisaalta, mutta valintoja tehdessä on nojattava johonkin arvostukseen: valintaa ei voi tehdä toisaalta ja toisaalta -periaatteella. Paradoksaalista kyllä, arvostamme sitä, mitä pidämme arvokkaana, vaikka se ei olisiakaan aina sopuosinnussa käsitystemme kanssa.

6.1.3 Selittävän ja ymmärtävän aineiston vuoropuhelua

Tämä tutkimus selittää yleisesti ottaen, että arvioijat arvioivat yhdenmukaisesti riippumatta käytettävästä asteikosta, joskin heidän sisäisessä johdonmukaisuudessaan on kuitenkin jonkin verran vaihtelua. Ymmärtävä puoli puolestaan viestii, että arvioijien kirjoittamiskäsitykset ovat yleisellä tasolla monipuolisempia kuin arvioivalla tasolla. Toisin sanoen arvioinnissa tuntui korostuneen rakennelähtöisyys, mikä tuli esiin arvioijien itsensä määrittelemissä asteikoissa, arvioijien kommentoissa määrittelystä asteikoista ja valintaverkkohaastattelun toisen vaiheen tekstien järjestelmissä.

Numeerisen aineiston osoittama yhdenmukaisuus antaa olettaa, että arvioijien arviot perustuvat jossakin määrin yhteiseen käsitykseen kirjoittamisesta. Haastatteluaineiston perusteella sosiaalisesti jaetussa käsityksessä – mitä muuta se sitten sisältääkään – korostui kaikkein voimakkaimmin kielen rakenteiden hallinta. Rakennelähtöisyys ei ole sikäli yllätyksellistä, että myös toisen kielen kirjoittamista koskevat tutkimukset tukevat rakenteiden vahvaa asemaa. Lisäksi formalistisen kielikäsitteiden vahva asema kielenopetuksessa puhuu puolestaan sekini; opettajilla on yleensä vahvaho käsitys siitä, miten rakenteet etenevät, missä järjestyksessä ne kannattaa opettaa ja missä järjestyksessä ne opitaan. Samaa ei voitane sanoa esimerkiksi tekstin rakenteesta tai sidosteisuudesta.

Rakenteiden vahva läsnäolo ei tarkoita sitä, että arvioijien kirjoittamiskäsitys on yhtä kuin rakenteet, kuten jo esitin tehdessäni eroa yleisen ja arvottavan kirjoittamiskäsityksen välillä. Rakenteiden dominoivuus näkyy nimenomaan arviointipäätöstä tehtäessä: ne näyttäisivät ohjaavan päätöstä arviosta rakenteiden hallinnan tasosta riippuen joko ylös- tai alaspäin. Varsinkin silloin, kun päätös arviosta tuntuu vaikealta tehdä tai kun suoritus on kahden tason rajalla, rakenteet saattavat olla ratkaiseva tekijä konkreettisuutensa vuoksi. Koska arvioijilla on oletettavasti jokseenkin yhdenmukainen käsitys rakenteiden hallinnasta eritasoisilla oppijoilla, huomion kiinnittäminen rakenteisiin voi osaltaan selittää numeerisen aineiston osoittamaa yhdenmukaisuutta.

Yhdenmukaisuuden taustalla voi toki olla yhtä lailla holistinen näkemys kirjoittamisesta, mitä selitin yksityiskohtaisemmin jo edellä. Jos näin olisi, voi olla, että analyttisellä arviointiasteikolla arvioituna arviot olisivat voineet näyttää kriteerikohtaisesti epäyhdenmukaisemmilta. Uskon, että ainakin jossakin määrin, asteikon luonne, kunkin taitotason kuvauksen tarkkuus, yksiselitteisyys ja sovellettavuus vaikuttavat siihen, millaisia arviointipäätöksiä suorituksista tehdään. Mitä tarkempia, yksiselitteisempiä ja lähempänä arvioijan omaa käsitystä nämä tekijät

ovat, sitä johdonmukaisempaa ja yhdenmukaisempaa arvioinnin voidaan odottaa olevan.

Tämän perusteluketjun mukaan tämän tutkimuksen tulokset ovat ristiriitaisia. Numeerisen aineiston perusteellahan voisi olettaa, että yhdenmukaisuuden näkökulmasta asteikkojen soveltaminen oli jokseenkin ristiriidatonta ja ehkä helppoakin, koska korrelaatiot ovat niin korkeita. Näin ei kuitenkaan ollut haastatteluaineiston perusteella, koska arvioijat valittivat itse määriteltävän asteikon epäjohdonmukaisuutta ja subjektiivisuutta sekä määritellyn asteikon monitulkintaisuutta ja epämääräisyyttä.

Molempien asteikkojen käyttämisessä pidettiin vaikeimpana asteikon keskitasojen, lähinnä 3–4, soveltamista. Arvioitava teksti saattoi siis kuulua kummalle tahansa vierekkäisistä tasoista riippumatta siitä, käyttivätkö arvioijat itse määrittelemäänsä vai heille annettua asteikkoa. Vastaavasti ääripäiden eli taitotasojen 1 ja 6 soveltaminen oli selkeintä molempien asteikkojen kohdalla. Välitasojen hankaluudesta kertoo myös valintaverkkohaastattelussa esiin tullut vaikeus asettaa tekstejä järjestykseen keskivälin osalta: arvioijien mielestä, esimerkiksi sanaston mukaan järjestettävät, tekstit olisivat voineet sijojen 4–7 osalta olla missä järjestyksessä tahansa.

Keskitasojen soveltamisen vaikeus saattaa liittyä osin siihen, että suurin osa suorituksista sijoittui näille tasoille. Arvioijien antamat tasoarviot edustivat enimmäkseen sekä keskiarvojen että raakadatan perusteella kolmessa ensimmäisessä tehtäväsarjassa tasoja 2–3 ja neljännessä tehtäväsarjassa tasoja 3–4 arvioijasta riippuen (ks. luku 5.1 ja liite 9). Tästä on vedettävissä kahdenlaisia johtopäätöksiä: 1) suoritukset olivat todella tämäntasoisia tai sitten 2) kyse oli mittausvirheestä eli asteikon keskivälin korostuneesta käyttämisestä (central tendency). Kirjoittajien itsearviot, arvioijien yhdenmukaisuus ja ennen kaikkea haastatteluaineisto puhuvat kuitenkin sen puolesta, että suoritukset olivat kutakuinkin tämäntasoisia. Suoritusten samantasoisuus sai arvioijat itsensäkin epäilemään ja pohtimaan omaa arviointilinjaansa.

Jos arvioijia tarkastellaan yksilöinä, vaikutelma yhdenmukaisuudesta muuttuu vähän erinäköiseksi: siinä on enemmän rosoja ja yksilöllisiä tapoja tehdä valintoja. Arvioijien sisäistä johdonmukaisuutta kuvaavat arvioijatyyppit (ks. kuvat 19–22) osoittavat, että jotkin arvioijat toimivat jossakin määrin odotuksenvastaisesti, esimerkiksi arvioijat 020030/Teija, 031234/Tarja ja 090230/Heidi.

Haastattelujen osalta yksilölliset erot näkyvät lähinnä siinä, millaisena taitona arvioija näkee kirjoittamisen ja millaisia tekijöitä hän raportoi omaan arviointiprosessiinsa liittyen. Näissä ei ole ehkä niinkään suuresti kyse suurista eroista vaan painotuksista, koska puheesta

välittyneet käsitykset olivat luonteeltaan niin monisäikeisiä ja moniulotteisia. Tässä mielessä numeerinen aineisto selittää yksioikoisemmin toiminnan laatua arvioiden välityksellä, kun taas haastattelut välittävät jollakin tavalla vaikeammin yksinkertaistettavaa tai tiivistettävää kuvaa.

Esimerkiksi edellä mainittujen kolmen arvioijan haastattelut eivät nosta heitä esiin muista haastateltavista ainakaan näiden analyysien perusteella. Haastattelut eivät tunnu antavan oikeutta perustella sitä, miksi he eivät toimineet odotuksenmukaisella tavalla johdonmukaisesti. Teija ja Tarja painottivat molemmat rakenteiden hallintaa järjestäessään valintaverkko-haastattelun tekstejä ja Heidi tehtävään paneutumista, mutta he eivät olleet ainoita arvioijia, jotka tekivät näin. Tarjan itsensä määrittelemä asteikko kuuluu rakennelähtöisiin ja Teijan sekä Heidin asteikko ymmärrettävyyden-/sisältölähtöisiin. Tosin Teijan asteikossa painottuu myös rakenteiden hallinta. Valmiiksi määritellyn asteikon kokivat sekä Heidi että Teija helpottavana esittämästään kritiikistä huolimatta. Kaikki kolme arvioijaa raportoivat myös siitä, miten heidän arviointiprosessiinsa vaikuttivat muut tekijät, kuten kirjoittajan taustatekijät, tekstin sisältö ja oma vireystila.

Haastatteluaineiston perusteella voi siis selittää yleisiä linjoja ja osoittaa sen, että arviointiprosessissa vaikuttavat monenlaiset tekijät ja raportoidun perusteella arvioijat itsekin tunnistavat niiden olemassaolon. Tutkimuksen haastatteluaineisto ei kuitenkaan paljasta sitä, miten nämä erilaiset tekijät vaikuttivat kullakin arvioijalla jonakin tiettyä hetkenä. Kyse on jälkikäteisestä reflektiosta. Esimerkiksi melkein kaikki arvioijat käydessään läpi arviointiprosessiaan viittasivat subjektiivisten vaikutelmien läsnäoloon, mikä osoittaa heidän olleen niistä tietoisia ainakin jälkikäteen. Eri asia on kuitenkin se, missä määrin ne vaikuttivat ja mitkä niistä vaikuttivat silloin, kun arvioija teki päätöstä arviosta.

Voin vain olettaa, että varsinkin sisäisen johdonmukaisuuden erot johtuvat lukuisista arviointiprosessissa vaikuttavista tekijöistä ja siitä, että nämä eri tekijät vaikuttavat eri tavalla eri aikoina. Toki numerot viestivät tietyistä systemaattisesta linjasta: esimerkiksi arvioija saattaa samalla asteikolla antaa toisella kerralla huomattavasti alhaisempia arvioita. Tämän tutkimuksen perusteella on kuitenkin mahdotonta selittää eksaktisti yksilöllisiä eroja, sillä ankaruuteen on voinut vaikuttaa esimerkiksi nopeastikin vaihtuvat erilaiset käsitykset, tunnetila, mieltymys tietynlaiseen aiheen käsittelyyn tai senhetkinen vireystila. Tämä ei kuitenkaan mielestäni vähennä käsitysten, arviointiprosessissa vaikuttavien tekijöiden ja arvostusten selvittämisen arvoa. Ne osoittavat sen, että arvioija ei toimi mekaanisesti tai yksilöotteisesti vaan inhimillisesti. Kirjoittamistaidon arviointi on paljon muutakin kuin sanaston, rakenteiden,

koheesion ja jäsentelyn tarkastelua. Numeerinen aineisto ja haastatteluaineisto yhdessä puolestaan osoittavat, että vaikka arviointiprosessissa vaikuttavat niin monenlaiset erilaiset muuttujat tai arviointiin luonnollisena osana liittyvät tekijät, arvioijat arvioivat jokseenkin johdonmukaisesti ja yhdenmukaisesti.

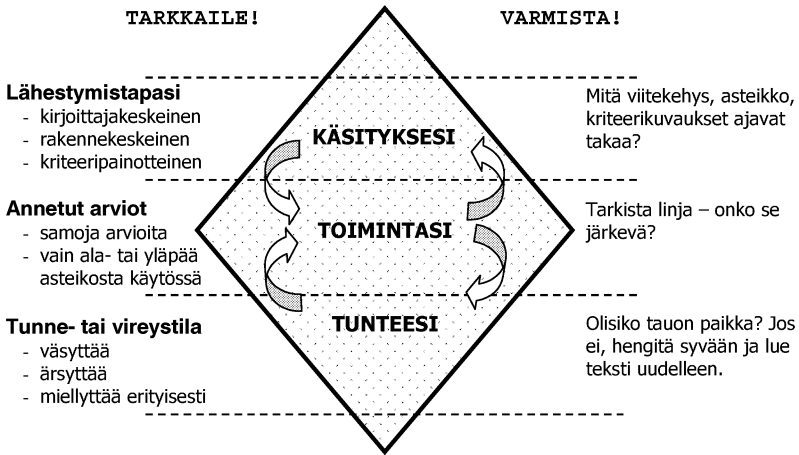
Vaikka arviointiprosessi näyttää käsitysten perusteella olleen monitekijäinen ja moniliikkeinen, se ei kuitenkaan ole johtanut kaoottiseen arvioiden mielivaltaisuuteen. Itse asiassa pinnalta katsoen tämän tutkimuksen arviot heijastelevat yhdenmukaisuudessaan sosiaalisesti jaettua käsitystä kirjoittamisesta, jota myös haastatteluissa esiin tulleet käsitykset tavallaan kuvaavat ja tukevat (vrt. Hornberger 1994; Savage 2000). Johdonmukaisuutta sen sijaan ei voine selittää minkään yhden tai kahden tekijän nimissä – ehkä ei yhdenmukaisuuttakaan. Yhteiskuva, joka tästä syntyy, antaa aihetta numeerisista tuloksista huolimatta pohtia sitä, miten arvioinnin johdonmukaistamiseen ja yhdenmukaistamiseen voisi käytännössä vaikuttaa.

Vaikka tässä tutkimuksessa arvioijat arvioivat yhdenmukaisesti ilman koulutusta ja jokseenkin huonoksi katsomillaan asteikoilla, on kuitenkin oletettavaa, että arviointiasteikot, -ohjeet ja -koulutus eivät itsessään riitä johdonmukaistamaan ja yhdenmukaistamaan arviointia, vaan yhdenmukaisuuden tavoittelu edellyttää sekä neuvottelua yhteisestä kielikäsituksesta että sen soveltamisesta suorituksiin. Arviointikoulutuksessa olisi käytävä keskustelua ja neuvoteltava arvioijan omasta kielikäsituksesta: miten hän verbalisoi itse sitä ja miten se ehkä eroaa asteikon käsityksestä? Liikkeelle on siis lähdettävä arvioijan omasta kielikäsituksesta, ei suoraan asteikosta.

Tutkimukseni haastatteluaineiston perusteella pidän tärkeänä, että arvioija on tietoinen siitä, millaiset tekijät hänen arviointiinsa vaikuttavat, ja että hän pystyy tunnistamaan ja kontrolloimaan niiden vaikutusta. Suurin osa arvioijista varmasti onkin tietoisia erilaisten tekijöiden vaikutuksesta, mutta huomattavasti vaikeampaa on luultavasti ohjata niitä työ- ja aikapaineen alla. On hyvin todennäköistä, että väsyessään tai kyllästyessään ihminen palaa hänelle ominaisimpaan tapaan tulkita asioita ja toimia, minkä vuoksi arvioijalla tulisikin olla välineitä tarkistaa oma linjansa säännöllisin väliajoin.

Johdonmukaisuuden varmistuskeinoja voivat olla esimerkiksi tauon pitäminen, linjan tarkistaminen tietyn tekstimäärän jälkeen, muistilistan tarkistaminen tai jonkin tarkistuskuvion, jollainen voisi olla esimerkiksi kuvio 26, katsominen. Kuvio 26 on esimerkkiahahmotelma, minkä vuoksi se on yksinkertainen ja osin vielä puutteellinenkin. Joka tapauksessa kuvion tehtävänä olisi pysähdyttää arvioija aika ajoin reflektoimaan toimintaansa ja muuttamaan sen suuntaa, jos toiminta ei tunnu

tarkoituksenmukaiselta. Pelkkä oman toiminnan laadun toteaminen tai epäileminen ei siis riitä.



KUVIO 26 Arviointilinjan tarkistus.

Varmuus omasta linjasta ja sen varmistaminen voisivat myös vähentää arviointiin liittyvää epävarmuutta. Epävarmuudellakin on hyvät ja huonot puolensa. Liiallisena se vie arvioinnista ilon, aiheuttaa ahdistusta ja voi saada arvioijan tekemään sellaisia ratkaisuja tai ylivarmistuksia, joita hän ei muuten tekisi. Positiivinen epävarmuus puolestaan ajaa kyseenalaistamaan omaa toimintaa ja tarkkailemaan sitä tietoisemmin. Epävarmuutta tulisi siis lähestyä analyttisemmin; kyse on sen tarkoituksenmukaisesta tulkinnasta ja sitä aiheuttavien syiden erittelemisestä.

Arviointiprosessin tuloksia eli arvioita tulisi niitäkin tarkastella analyttisesti. Vaikka yhdenmukaisuutta ja johdonmukaisuutta osoittavat numerot eivät olisikaan huolestuttavia, on kuitenkin syytä pitää mielessä kysymys siitä, mitä arvioijien väliseltä yhdenmukaisuudelta ja johdonmukaisuudelta odotetaan. Onko kyse tyydyttävästä korrelaatiosta, informaatiosta vai palautearvosta, jota arviot välittävät? Missä määrin arvioiden katsotaan heijastelevan arvioijien tulkintoja? Onko arvio totta ja missä määrin? Jos se on totta, kenen totuutta tai millaista totuutta se edustaa?

Arviointiprosessi ei päätty arvion antamiseen, vaan se jatkuu vielä jonkin aikaa: arvioija voi käydä keskustelua itsensä tai toisten arvioijien kanssa muun muassa tekijöistä, jotka vaikuttivat arvioiden antamiseen.

Tässäkään tutkimuksessa pelkkä arvioiden tarkasteleminen ei olisi kertonut, millainen arviointiprosessi on luonteeltaan ja millaisia valintoja arviosta päättämiseen liittyy. Itselleni arviointitutkijana oli tärkeätä päästä kurkistamaan nimenomaan arvioijien käsityksiin ja kokemuksiin arviointiprosessista. Toisin sanoen toisaalta olisi ollut yksipuolista ja puolisokeaksi heittäytymistä tuudittautua vain numeeriseen dataan ja toisaalta taas hämmentävää ja ristiriitaistakin nojata yksin haastatteluaineistoon.

6.2 Tutkimuksen rajoituksia

6.2.1 Tutkimusasetelma

Tutkimuksen aineiston keruu- ja arviointisuunnitelma ohjasivat pitkälti numeerisen aineiston tarkastelua. Näin jälkikäteen suunnitelmaan olisi kannattanut sisällyttää ei vain kaikille arvioijille yhteinen tehtävä vaan vähintään yksi, mielellään kaksi, sellaista tehtävää, jotka jokainen arvioija olisi arvioinut samoilta kirjoittajilta. Tässä käytetyn asetelman mukaan yhdenmukaisuuden tarkasteleminen oli mahdollista tehtäväsarjoittain ja kaikkien tehtävien osalta niiden neljän arvioijan kohdalla, jotka arvioivat kaikkien kirjoittajien suoritukset.

Toisin sanoen näin jälkikäteen tuntuu haaskaukselta, että on heitetty hukkaan mahdollisuus saada tietoa kaikkien mukana olleiden arvioijien yhdenmukaisuudesta edes yhden tehtävän osalta, koska tässä yhteydessä arvioitavien tekstien määrä oli niin suuri. Jossitellen jokainen arvioija olisi siis arvioinut vähintään sata samojen kirjoittajien kirjoittamaa tekstiä. Toisaalta tutkimusasetelman ansioksi voi lukea sen, että se antoi systemaattisesti yhdenmukaisuudesta tietoa 9–10 arvioijalta vähintään noin 300 tekstin osalta. Näin ollen päätelmät yhdenmukaisuudesta perustuvat tilastojen valossa uskottavaan määrään aineistoa.

6.2.2 Asteikkojen puutteita

Tutkimuksen ehdottomaksi puutteeksi voi katsoa sen, että arvioijien käyttämä asteikko oli käyttäjakeskeinen asteikko (vrt. Alderson 1991). Toisaalta se palveli tutkimuksen alkuperäistä ideaa, jonka tarkoituksena oli kerätä maamerkkisuorituksia DIALANG-hankkeeseen, mikä tietysti toteutuikin. Maamerkkikirjoitelmia on tarkoitus käyttää DIALANG-testin suoran kirjoittamisen osakokeessa, jossa kirjoittajat itse arvioivat kirjoittamisen taitoaan vertaamalla suoritustaan kommentoituun maamerkkikirjoitelmaan ja apukysymyksiin. Maamerkkikirjoitelmat valitaan

toisen ja kolmannen arviointikierroksen tulosten perusteella. Sitä, miten hyvin kirjoittamisen itsearviointi ja maamerkit toimivat, on jo hieman tutkittakin (Tarnanen & Luoma 2001; Luoma & Tarnanen tulossa).

Kaiken kaikkiaan asteikon soveltamisen tutkimiseen näin laaja ja raskaskin koeasetelma tuntui paikoin vähän liioitellulta, joskin tuloksilla on implikaatioita muihinkin yhteyksiin, esimerkiksi siihen, miten holistinen asteikko voi yleisesti ottaen toimia, miten arvioijat suhtautuvat tekstityyppilähtöiseen kirjoittamiskäsitykseen, miten määritelty asteikko ehkä peilaa arvioijien omia kirjoittamiskäsityksiä ja miten yksityiskohtaista asteikkoa arvioijat odottavat.

Määritellyn asteikon osalta tutkimuksen rajoituksena voidaan pitää sitä, että arvioijia ei mitenkään koulutettu tulkitsemaan asteikkoa eikä soveltamaan sitä. Tulokset ja tulkinnat olisivat voineet olla erilaisia, jos arvioijat olisivat saaneet koulutusta. Koulutus olisi voinut olla myös yksi tärkeä ja mielenkiintoinen muuttuja tutkimuksessa, joskin se olisi vaatinut erilaista koeasetelmaa ja suunnitelmaa koulutuksesta sekä resursseja sen toteuttamiseen. Koulutuksen merkitys olisi tärkeä varsinkin asteikosta tehtävistä tulkinnoista neuvottelemisessa; haastatteluaineisto tässä yhteydessä osoitti kyllä, miten tärkeää neuvottelu merkityksistä olisi.

Arvioijan itsensä määrittelemistä asteikoista olisi taas saattanut tulla yksityiskohtaisempaa tietoa, jos arvioijat olisivat saaneet kirjeen mukana tyhjän asteikkopohjan, jossa olisi ollut tyhjät laatikot tasokuvauksille, esimerkiksi liitteen 1 määritelty asteikko tekstit poistettuna. Toisaalta se olisi voinut ohjata arvioijia tekemään sinänsä tarkempia kuvauksia, mikä ei ehkä olisi kuitenkaan kuvannut samalla tavalla tai yhtä spontaanisti heidän käsityksiään. Asteikon kirjoittaminen olisi voinut saada heidät kaivamaan esiin jonkin käytössä olevan asteikon ja hakemaan tukea siitä enemmän kuin he nyt tekivät.

Sekä itse määriteltävän että määritellyn asteikon käyttöä täytyy ehdottomasti kritisoida myös siksi, että arvioijat joutuivat arvioimaan suorituksia ilman konkreettisesti esitettyä tavoitetasoa. Omassa opetuksessaan arvioijat arvioisivat suorituksia suhteessa opetuksen tavoitteisiin ja virallisessa testisysteemissä suhteessa testin viitekehykseen ja tuttuun asteikkoon. Nyt he joutuivat itse määrittelemään tavallaan sokkona taitotasoasteikkoa, joka ei suoranaisesti liittynyt mihinkään. Siksi arvioijien ratkaisu määritellä asteikot pitkälti opetuksessa käytettävien asteikkojen mukaan, tuntuu tästä näkökulmasta lähes ainoalta järkevältä ja luonteelta ratkaisulta.

Määritellyn asteikon osalta arvioijat tukeutuivat pelkkään asteikkoon ja toimivat pelkästään sen varassa. Heillä ei siis ollut tutkimuksen puolesta mahdollisuutta tutustua Kielten opetuksen ja oppimisen yleis-eurooppalaiseen viitekehykseen, joka olisi voinut auttaa asettamaan

asteikon tasokuvaukset laajempaan yhteyteen. Se olisi voinut selkiyttää asteikkoon rakennettua käsitystä kirjoittamisesta ja toisen kielen kirjoittamistaidon edistymisestä. Arvioijien täytyi muodostaa käsitys taidon progressiosta pelkästään tasokuvausten perusteella, jolloin suorituksen ja tasokuvauksen vastaavuus saattoi jäädä hämäräksi. Arvioijille ei siis saattanut olla pelkän asteikon perusteella selvää, millaiseen taitoon, tasokohtaiseen tavoitteeseen he vertasivat kutakin suoritusta.

6.2.3 Laadulliset analyysimenetelmät ja niiden taustaoletukset

Kuten jo aikaisemmin totesin, yleensäkin osa toiminnastamme on sellaista, jonka kielentäminen ei ole mahdollista tai kielentäminen ei anna oikeaa kuvaa toiminnasta. Myöskään yhteys oman toiminnan reflektointiin ja itse toiminnan välillä ei ole aina kovin suora. Voi olla, että tässä mielessä ääneenajattelu metodina olisi ollut tarkoituksenmukaisempi kuin etnografis-fenomenografinen ja valottanut paremmin annettujen arvioiden ja arvioijien mielen välistä yhteyttä. Toisaalta ääneenajattelu metodina ei olisi mahdollistanut yhtä monen haastattelun tekemistä eikä myöskään yhtä suuren tekstimäärän arviointia. Ääneenajattelemisen käyttäminen olisi myös muokannut tutkimuskysymyksiä erilaisiksi, joten spekulointi sen eduista ja haitoista on tässä mielessä turhaa.

Nyt käyttämäni analyysimenetelmän ongelmana näen edellisen lisäksi sen, että etnografis-fenomenografisella luokittelulla en ehkä päässyt niin syvälle aineistoon kuin olisin päässyt esimerkiksi diskurssi-analyyttistä metodia käyttämällä. Toisaalta käyttämäni analyysimenetelmä mahdollisti ison aineiston läpikäymisen, litteroinnin ja analysoinnin ja antoi melko hyvän läpileikkauksen arvioijien yksilöllistä ja yhteisöllistä käsityksistä. Aineisto ei kuitenkaan sallinut tehdä kovin syvälle meneviä johtopäätöksiä. Myös haastattelukysymykset muokkasivat aika pitkälti haastattelun sisältöä ja edelleen myös aineiston luokittelua ja analyysia. Näin ollen eduistaan huolimatta puolistrukturoitu teemahaastattelu valjasti aika pitkälle haastatteluaineiston sisältöä eteenpäin ja sitä, millaisia tulkintoja arvioijien käsityksistä tutkimuksessa nousi esille.

6.2.4 Tutkimuksen luotettavuus ja eettisyys

Tutkimuksen luotettavuutta arvioitaessa tarkastellaan yleensä tutkimuksen totuudellisuutta, pysyvyyttä, sovellettavuutta ja hyödynnettävyyttä (ks. esim. Lincoln & Guba 1985, 290; Eskola & Suoranta 1998, 209–235).

Luotettavuuden arvioinnissa tehdään kuitenkin ero kvantitatiivisen ja kvalitatiivisen tutkimuksen välillä paradigmaerojen mukaisesti. Tämän tutkimuksen osalta kvantitatiivisten tulosten luotettavuus viittaa mittauksen luotettavuuteen eli siihen, miten valideina, reliaabeleina, yleistettävinä ja objektiivisina tämän tutkimuksen kvantitatiivisia menetelmiä ja tuloksia voidaan pitää. Kvalitatiivisten tulosten luotettavuus puolestaan koskee koko tutkimusprosessia eli sitä, miten uskottavaksi, varmaksi, siirrettäväksi ja vahvistettavaksi tutkimus voidaan kvalitatiivisen otteen osalta arvioida. (Lincoln & Guba 1985, 290; Miles & Huberman 1994, 278–279.) Tutkimuksen luotettavuuden saavuttaminen edellyttää myös, että tutkimusprosessin eri vaiheissa tehdyt valinnat, tulkinnat ja päätelmät ovat eettisiä.

Käsillä olevaa tutkimusta voitaneen pitää pääosin validina eli tutkimuksessa on selvitetty arvioijien sisäistä johdonmukaisuutta ja keskinäistä yhdenmukaisuutta sekä tarkasteltu arvioijien käsityksiä kirjoittamisen arvioinnista ja tulkintoja arviointiprosessista. Menetelmällisten ratkaisujen osalta kvantitatiivisen ja kvalitatiivisen otteen yhdistäminen osoittautui hedelmälliseksi valinnaksi. Sen sijaan arvioijien sisäisen johdonmukaisuuden ja keskinäisen yhdenmukaisuuden tarkastelun luotettavuutta heikentää ennen kaikkea käyttäjäkohtaisen asteikon käyttö. Myös arvioijakoulutuksen ja tarkempien arviointiohjeiden puute on otettava luotettavuutta tarkasteltaessa huomioon. Nämä puutteet vaikuttavat siihen, millaisia päätelmiä johdonmukaisuudesta ja yhdenmukaisuudesta voidaan tehdä, että arvioijien toiminnasta tehtävät päätelmät kuvaisivat tarkasti tutkittavaa ilmiötä ja olisivat oikeudenmukaisia arvioijia kohtaan.

Laadullisen aineiston uskottavuus tulee parhaiten arvioiduksi tutkimusta lukiessa. Toisin sanoen olen yrittänyt kirjoittaa tutkimusprosessin auki siten, että lukijan ei tarvitsisi epäillä aineiston aitoutta eikä sen yhteyttä tutkimuksen teoreettisiin lähtökohtiin. Tutkimukseen ei ole kuitenkaan mahdollista ottaa mukaan koko haastatteluaineistoa eikä analysoinnin kaikkia luokitteluvaiheita, joten lukijan on luotettava tutkijan valintoihin ja esimerkkeihin, joissa kuuluu arvioijien oma ääni. Käsitusten laatua arvioitaessa on syytä kiinnittää huomiota myös tutkimuksen arviointikontekstiin eli esimerkiksi arvioitaviin teksteihin, arvioinnin tavoitteisiin ja ajankohtiin, jolloin tekstit arvioitiin, koska konteksti vaikuttaa osaltaan siihen, millaisia arvioijan käsitykset ovat. Tutkijana olen kiinnittänyt huomiota myös siihen, että käsityksiä ei ole ollut tarkoitus verrata toisiinsa hyvinä ja huonoina käsityksinä. Kyse ei ole siis ymmärryksen arrottamisesta vaan sen lisäämisestä.

Tutkimuksen pysyvyyttä varmistettiin tekemällä useampia erilaisia analyyseja sekä yhdenmukaisuudesta että johdonmukaisuudesta.

Analyysien samansuuntaiset tendenssit osoittavat, että johdonmukaisuudesta ja yhdenmukaisuudesta voidaan vetää edellä esitetyn kaltaisia päätelmiä. Tietysti analyysien riittävästä määrästä ja analyysimenetelmien valinnasta voi käydä keskustelua. Kun arvioijien johdonmukaisuutta ja yhdenmukaisuutta tarkastellaan keskiarvoina, prosentteina tai vaihtelun lineaarisina muutoksina, arvioita ei tarkastella oppijan näkökulmasta. Toisin sanoen vaikka numeerinen tulos osoittaa korkeaa arvoa, voi jonkin yksilön kohdalla poikkeama olla kolmekin tasoarviota. Tällainen yksikin tapaus on liikaa ja on ehdottomasti epäoikeudenmukaista oppijalle. Toisaalta epäoikeudenmukaisuudessaan poikkeamat osoittavat, että arvioijien toimintaan on tällaisen suorituksen kohdalla vaikuttaneet henkilökohtaiset tekijät, esimerkiksi vireystilan muutos tai mieltymykset.

Laadullisen aineiston puolestaan luokitteli yksi ulkopuolinen tutkija, jonka luokittelusta noin 80 prosenttia vastasi omaa luokitteluani. Hänellä oli käytössään valmiit merkityskategoriat. Tämän lisäksi luokittelin itse aineistoni pariin otteeseen. Luokittelun varmuutta arvioitaessa oma painoarvonsa on yleensäkin luokittelun tulkinnanvaraisuudella, johon vaikuttavat esimerkiksi luokittelijan perehtyneisyys aineistoon, intentiot, kullonienkin vireystila ja uskomukset. Laadullisen aineiston varmuutta vahvistanee kuitenkin se, että haastatteluja oli kolme eli kukin arvioija tuli kuulluksi kolme kertaa.

Tutkimustulosten sovellettavuusarvoa säätelevät tulosten yleistettävyys ja siirrettävyys. Arvioijien arvioimien tekstien suurta määrää voi kritisoida, mutta toisaalta suuren määrän vuoksi arviot eivät heijastele yksittäisiä tapauksia vaan keskimääräisiä linjoja. Myös maamerkkien keräämisen kannalta suoritusten suuri määrä on tarkoituksenmukaista, koska valituksi tulevien maamerkkien tulee olla edustavia esimerkkejä kuudesta taitotasosta. Lisäksi suuren tekstimäärän arvioinnin voi olettaa tuovan edustavammin esille arvioijien käsityksiä ja tuntemuksia siinä mielessä, että arviointiprosessi on erilaisten tehtävien ja kirjoittajien sekä suoritusten määrän vuoksi monimuotoisempi ja kestoaltaankin pitempi.

Toisaalta päätelmät arviointiprosessista viittaavat nimenomaan suuren tekstimäärän arviointiin eli massa-arviointiin eikä ehkä niinkään arvioijille tutumpaan omien oppilaiden arviointiin, jossa arvioitavat ovat tuttuja, arviointikriteerit kuvaavat opetuksen sisältöjä ja arvioinnin tavoitteet ovat tiiviissä yhteydessä opetuksen tavoitteisiin. Tutkimuksen tulosten voi kuitenkin jossakin määrin olettaa liittyvän mihin tahansa arviointitilanteeseen. Ne osoittavat myös, mihin asioihin arvioija-koulutuksessa tulisi kiinnittää huomiota varsinkin silloin, kun arvioijat arvioivat muilla kuin itse laatimillaan kriteereillä. Tulokset lisäävät myös ymmärrystä arviointiprosessista, mitä voidaan hyödyntää arvioijan oman toiminnan ohjaamisessa. Tulokset vahvistavat joiltakin osin aikaisempia

tutkimuksia ja ovat osittain hyödynnettävissä myös muiden osataitojen, varsinkin puhumisen, arvioinnissa ja yleensä vieraan kielen taidon arvioinnissa (vrt. esim. Vaughan 1990; DeRemer 1998; Smith 2000).

6.2.5 Tutkijan oma ääni – liian lähelläkö?

Tutkimukseni etuna ja haittana voi pitää sitä, että olin itse yksi arvioijista - tosin vain numeerisen aineiston osalta. Se saattoi olla haitta siksi, että omat motiivini ja käsitykseni ovat saattaneet vaikuttaa siihen, miten toimin haastatteluissa, esimerkiksi jatkokysymyksissä ja palautteenannossa, ja miten analysoin aineistoa ja millaisia tulkintoja siitä tein. Tutkimuksessani on vääjäämättä paljon subjektiivisia valintoja ja tulkintoja, joten niiden laatu ja määrä on voinut korostua roolini vuoksi. Itse arvelen, että kokemukseni yleensäkin arvioijana, tehtävänlaatijana ja arvioijakouluttajana on vaikuttanut kaikkeen tässä työssä, joten rehellisesti sanoen en pystyisi erittelemään, mikä on vaikuttanut mihinkin ja miten. Toisaalta se ei kai olisi tarkoituksenmukaistakaan, kun työn lähtökohta on hermeneuttinen.

Osallistumiseni tekstien arviointiin näen positiivisena nimenomaan hermeneuttisesta näkökulmasta. Keskustellessani arvioijien kanssa ja analysoidessani haastatteluaineistoa tunsin ymmärtäväni, mistä haastattelijat puhuivat: sekä kielteiset että myönteiset kokemukset tuntuivat tutuilta. Itsestäni tuntui siltä, että juuri kokemukseni kautta sain tuotua esiin myös arviointiprosessiin liittyviä negatiivisia tuntemuksia ja kokemuksia, vaikka tietysti yritin olla ylitulkitsematta arvioijien kommentteja omien kokemusteni mukaisesti. Omaan osuuteeni subjektiivisena tutkijana liittyivät mielestäni samat toimintaohjeet kuin arviointiinkin: pyrin olemaan toiminnastani tietoinen, ohjaamaan sitä tarkoituksenmukaiseen suuntaan ja siten välttämään parhaalla mahdollisella tavalla oman näkökulman korostamista muiden näkökulmien kustannuksella.

Sitä, miten siinä onnistuin, tuntuu vaikealta arvioida prosenttien tai skaalan mukaan. Tarkoitin tällä sitä, että omaan kokemusmaailmaan liittyvien asioiden ymmärtäminen voi vain olla helpompaa ja eräällä tavalla syvempää. Vaikka numeerinen ja laadullinen aineisto puhuvat puolestaan, ne eivät kuitenkaan tee sitä itsestään, vaan ne tarvitsevat tulkitsejan. Toisin sanoen tietynlainen kahtiajakoisuus on väistämätöntä. Kuten luvussa 1 esitin, tutkimukseni kautta yritän kuvata itsestäni riippumatonta ilmiötä, mutta siihen liittyvistä lainalaisuuksista ja ulkoisesta todellisuudesta huolimatta, eksaktia vastausta on mahdoton antaa, koska tutkimuksen kohteena on ihmisen toiminta rationaalisena, emotionaalisenä ja fyysisenä yksilönä.

6.3 Tutkimukseni tarjoamia jatkolinjauksia

Tutkimukseni herättää rajoitustensa ja tuottamiensa tulosten yhdistelmänä ideoita monista jatkotutkimusaiheista, jotka liittyvät yleensäkin suomi toisena kielenä -opetuksen ja -arvioinnin ajankohtaisiin haasteisiin, arvioijan käsityksiin ja toimintaan sekä kielitaidon arviointiin liittyvään problematiikkaan. Jatkotutkimuksen aiheet riippuvat myös siitä, minkä paradigman mukaisesti tutkimukseni tuloksia tarkastellaan; perinteisen arviointikulttuurin valossa reliaabeliuden ja validiuden käsite voidaan selkeästi erottaa toisistaan, mutta kvalitatiivisten menetelmin tarkasteltuna niiden välinen ero hämärtyy, varsinkin reliaabeliuden ja konstruktivaltiuden. Validiuden ja reliaabeliuden raja onkin tässä mielessä jokseenkin keinotekoinen. Tämä ei kuitenkaan tarkoita sitä, ettei numeerista tarkastelua tarvittaisi, vaan sitä, että reliaabeliuden kokonaisvaltainen tutkiminen edellyttää hermeneuttista näkökulmaa (ks. Moss 1994).

Hermeneuttisen lähtökohdan mukaan arvioijaa ei voitane pitää perinteisen testiarvioinnin mukaisesti virhelähteenä. Ei ainakaan virhelähteenä siinä mielessä, että virheettömät lähteet olisivat mahdollisia. Virhelähtöisen ajattelun mukainen suhtautuminen arvioijaan vähättelee myös itsearvioinnin käyttöarvoa; se ei tue millään tavalla itsearvioinnin uskottavuutta ja luotettavuutta. Hermeneuttisen näkökulman mukaan arvioinnissa ovat mukana myös inhimilliset tekijät, joiden varassa ja ehdoilla pyritään tekemään mahdollisimman eettistä ja tarkoitustaan palvelevaa arviointia.

Arvioinnin tutkimukseen liittyvän paradigman – yhä jatkuvassa – muutosvaiheessa olisi mielenkiintoista tarkastella sitä, mitä paradigman muutos tarkoittaa esimerkiksi testiarvioinnin osalta (ks. myös kuviota 1). Miten lähtöoletusten muuttuminen vaikuttaa reliaabeliuden ja validiuden käsitteiden muuttumiseen, millaisia konkreettisia muutoksia se edellyttäisi testijärjestelmiltä. Kun testijärjestelmiltä edellytetään eettisyyttä, kenen näkökulmasta eettisyydestä puhutaan ja miten taloudellisia, poliittisia ja koulutuksellisia näkökohtia arvioidaan?

Itse asiassa nämä kysymykset ovat erittäin ajankohtaisia suomi toisena kielenä -opetuksen ja -arvioinnin näkökulmasta. Maahanmuuttajien kielikoulutus ja siihen liittyvä arviointi peilaavat omalla tavallaan yhteisöimme sosiaalisia, poliittisia ja koulutuksellisia arvoja, asenteita, käsityksiä ja toimintatapoja. Huttunen (2002) kuvaa mielestäni osuvasti Suomen sosiaalista maailmaa kategorisoivaksi, millä hän tarkoittaa sitä, että maahanmuuttajia kohdellaan ennen kaikkea kategorioiden edustajina eikä heidän oman tahtonsa mukaisesti yksilöinä. Yksilöttömyys näkyy muun muassa siinä, että kaikki maahanmuuttajat halutaan kotouttaa

samalla tavoin, vaikka heillä on hyvin heterogeeniset taustat ja heidän sopeutumisensa tapoineen yksilöllistä.

Kategorisointi ja yksilöttömyys vaivannee myös osin suomen kielikoulutusta, koska Helsingin Sanomien uutisen mukaan Väestöliiton tutkija Annika Forsander toteaa suomen kielen opetuksen olevan aivan retuperällä (HS 27.1.2002). Hänen mukaansa kotouttamislaki on hyvä, mutta toteutuminen huono. Forsander on erityisen huolestunut huonosta kielitaidosta ja sen vaikutuksesta työllistymiseen. Huoli kielitaidon omaksumisesta liittyy myös inkeriläisiin, joilta on vuodesta 2001 edellytetty auttavaa suomen kielen taitoa ennen muuttoa Suomeen (HS 13.2.2002). Käytännössä auttavan kielitaidon omanneet eivät ole kuitenkaan osanneet suomea. Ainoana syynä huonoihin oppimistuloksiin mainitaan motivaation puute.

Maahanmuuttajiin liittyvässä keskustelussa on mielenkiintoista se, millaisia päätelmiä kielitaidon tasosta tehdään, millaisiin arviointeihin päätelmät ehkä viittaavat, mitä kielitaidolta odotetaan ja miten sen nähdään kehittyvän. Sanomalehtikeskustelua seurattaessa päällimmäiseksi mieleen jää henkilöttömyys siten, että ongelmien nähdään ratkeavan esimerkiksi resurssien lisäämisellä, vastuuministeriötä vaihtamalla, maahanmuuttajakiintiötä supistamalla tai yhtenäisen kielitaitotestistön luomisella. Keskusteluissa jää kaipaamaan mainintaa siitä, miten opetusta voitaisiin muuttaa sisällöllisesti, missä määrin kieltä opitaan itse asiassa luokkahuoneessa tai miten kielitaidon opetus ja arviointi ottavat yksilöt huomioon.

Miksi ongelmia tarkastellaan hierarkkisesti ja organisatorisesti ylhäältä päin eli esimerkiksi valtiosta käsin. Yhtä hyvin voisi kysyä, miten maahanmuuttaja itse arvioi oman kielitaitonsa tasoa ja miten hän itse kokee sen riittävyyden ja riittämättömyyden. Nousisiko motivaatio omista tarpeista, tarpeesta selviytyä jokapäiväisessä elämässä ja halusta kuulua suomalaiseen yhteisöön? Mikä on se suomen kielen taso, joka avaa portin työelämään? Miltä osin ja miten kielenoppiminen ja taidon arviointi voisivat perustua luonnollisemmin suomalaisen ja maahanmuuttajan vuorovaikutukseen? Miten yksilöt, heidän eri vaiheissa olevat kielenoppimis- ja sopeutumisprosessinsa voitaisiin ottaa paremmin huomioon esimerkiksi siirryttäessä lukioon, ammatilliseen koulutukseen, työharjoitteluun tai työelämään. Mikä on kielitaidon painoarvo sopeutumisessa ja selviytymisessä, miten kielitaitoa arvioidaan ja millaisen vastuun olemme valmiita antamaan oppijalle itselleen?

Isoja ja idealistisiakin kysymyksiä, joihin tuntuu olevan vaikea vastata. Vastaaminen on kuitenkin mahdotonta, elleimme kysy myös itseltämme, missä määrin tunnemme todellista tarvetta muutokseen ja missä määrin olemme valmiita siihen? Muutos edellyttäisi ennen kaikkea

taitokäsityksen ja arvioinnin valtasuhteiden miettimistä uudelleen. Uskommeko aidosti siihen, että oppija voi itse laatia kriteereitä, arvioida oppimistaan niillä ja asettaa itselleen uusia tavoitteita. Käytännössä tämä voisi tarkoittaa esimerkiksi sitä, että oppijat laatisivat itselleen opetus-suunnitelman tavoitteineen ja arviointikriteereineen joko työelämää tai opiskelua varten. Kyse ei olisi vain kielenopetuksellista tavoitteista vaan ennen kaikkea kielen asemasta esimerkiksi työelämän eri tehtäviä hoidettaessa tai reaaliaineita opiskeltaessa. Tällaiset muutokset ajaisivat myös pohtimaan sitä, millaisena kokonaisuutena kielitaito ymmärretään.

Myös kielitaidon testaaja joutuu kysymään itseltään, onko kielitaito itse asiassa autonominen yksikkö? Onko se pilkottavissa osiin? Miten tehtävyyppien luotettavuus yleensäkin määritellään ja kuka sen määrittelee? Millaiset tehtävät ja miten suoritettuina ovat tarpeeksi hyviä mittaamaan tai arvioimaan taitoa? Miten paljon uskotaan oppijan itsensä mahdollisuuksiin arvioida omaa taitoaan? Nämäkin ovat isoja ja monisäikeisiä kysymyksiä. Joka tapauksessa näitä kysymyksiä tulisi pohtia ja niihin tulisi yrittää löytää vaikka pienimuotoisiakin vastauksia. Se on selvää, että pelkkä kritiikki – joka alkaa olla jo niin tuttua – arviointijärjestelmiä kohtaan, ei ole rakentavaa ja eteenpäin vievää.

Jos siirryn lähemmäksi toiminnan tasoa, arvioijien eli opettajien käsityksiä olisi mielenkiintoista selvittää myös luokkahuonearvioinnissa, koska käsitykset vaikuttavat myös siihen, mitä opetetaan, ja osin myös siihen, miten opetetaan. Tutkimuksen aineisto jätti myös houkuttelevan avoimeksi sen, miten käsitykset voivat jakaantua yleisiin ja arvottaviin. Tuntuisi tärkeältä selvittää, miten nämä kaksi käsitystasoa saataisiin kohtaamaan toisensa. Arvioinnilla on ehkä edelleenkin, hyvistä muutoksista huolimatta, liian ulkopuolinen rooli opetuksen sisältöihin nähden. Arvioinnin tiiviimpi nivoutuminen osaksi opetusta edellyttäisi näiden oletettujen käsitystasojen sulautumista yhdeksi ja samaksi.

Tutkimusaihiot eivät lopu luokkahuoneeseen; myös valtakunnallisten testisysteemien, kuten Yleisten kielitutkintojen ja ylioppilastutkinnon, arvioinnin tutkiminen kaikista näkökulmista olisi erityisen tärkeää. Molemmilla arviointisysteemeillä on myös näkyviä yksilöllisiä ja yhteiskunnallisia seurauksia. Vaikka molempia systeemejä voisi tarkastella eri arviointivaiheiden (esim. viitekehys/opetussuunnitelma, tehtävänlaadinta, arviointitilanne, yksilön tulkinta osaamisestaan, palautearvo, systeemin ylläpito ja kehittäminen) kannalta, tämän tutkimuksen innoittamana olisi erityisen halukas selvittämään ylioppilastutkinnon äidinkielen kokeen arviointia sekä käytänteiden (1. arvioijana oma opettaja, 2. arvioijana sensori) että arviointiprosessin näkökulmasta. Ei ole yhdentekevää, mitä äidinkielen osaamisessa arvostetaan ja millaista

käsitystä siitä viestitään arvostusten välityksellä, puhumattakaan lähtöoletuksista, joihin arvostukset nojaavat.

Keskustelu, jota äidinkielenkokeesta on käyty viime vuosina, osoittaa myös ylioppilastutkinnon painoarvon, kun tehdään päätelmiä suomalaisten nuorten äidinkielen ja kirjoittamisen taidosta. Yllättävän vähän keskustelua käydään kuitenkin mittarista ja sen kyvystä osoittaa taitoa. Vaikka äidinkielen koetta on uusittu ja uudistetaan parhaillaan, pitäisi keskustelussa ottaa huomioon muun muassa se, miten onnistuneita tehtävät ovat, miten hyvin ne heijastelevat nykyisen yhteiskunnan lukemis- ja kirjoittamistarpeita, miten motivoiviksi ne koetaan ja millaisin periaattein tekstejä arvioidaan.

Kaiken kaikkiaan arviointitutkimuksessa, oli yksittäinen tutkimuskohde mikä tahansa, arvioinnin totuusarvoa ei saisi ottaa itsestäänselvyytenä. Arviointi ei ole myöskään pelkästään erilaisia tekniikoita selvittää osaamista, vaan erilaiset arviointitavat ja -käytänteet kantavat aina mukanaan omanlaistaan käsitystä osaamisesta ja sen mitattavuudesta. Arviointitutkimus voi valinnoillaan tukea arviointia, jossa lähtöoletukset, arviointitavat ja -käytänteet sekä arvioinnin palaute ovat keskenään mahdollisimman hyvin sopusoinnussa. Ja arviointihan ei tee itsestään mitään, vaan me annamme sille sen painoarvon, mikä sillä on.

Ehdottamani arviointitutkimuksen jatkolinjaukset ovat laveja ja suuria pensselin vetoja. Lisätäkseen vielä yhden vedon toteaisin, että kielten opettajien koulutuksessa on liian vähän arviointia koskevia sisältöjä. En kuitenkaan tarkoita, että haluaisin painottaa arviointia muiden tärkeiden sisältöjen kustannuksella. Arviointi on kuitenkin olennainen osa opettajan työtä; se on parhaimmillaan oppimista tukeva ja pahimmillaan sitä estävä tekijä. Arviointi voi olla kapeimmillaan pisteitä ja koekysymyksiä ja laajimmillaan koko opetuksen perustan pohtimista. Toisin sanoen tästäkin voisi avautua monta tutkimuksen aihetta, oli kyse sitten survey-tyyppisistä selvityksistä tai syvälle menevistä kvalitatiivisista analyysistä.

6.4 Loppusanat

Arvioijan sisäinen johdonmukaisuus ja arvioijien keskinäinen yhdenmukaisuus liittyvät olennaisesti käsityksiin arvioinnin luotettavuudesta ja eettisyydestä. Arvioijan käsitysten tutkiminen avaa luotettavuuteen ja eettisyyteen subjektiivisuuden ikkunan, joka paljastaa näiden käsitteiden inhimillisyyden. Inhimillisuus ei invalidisoi arviointia eikä tee ristiriitaisuudessaan siitä epätarkoituksenmukaisempaa tai vähäarvoisempaa. Se ei myöskään tarkoita sitä, etteivät ensisijaisesti eettisyys ja sen myötä

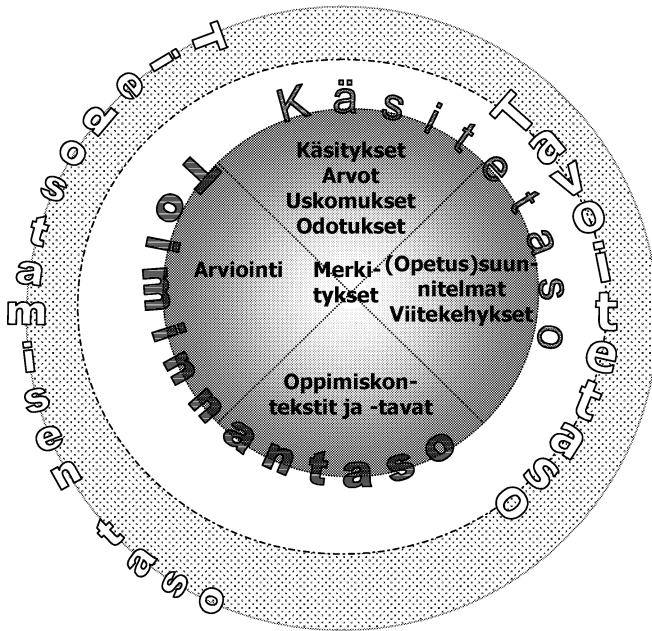
arvioijan sisäinen johdonmukaisuus ja arvioijien keskinäinen yhdenmukaisuus olisi jokaisen arvioijan tavoiteltava ominaisuus.

Johdonmukaisuuden ja yhdenmukaisuuden tavoitteluun ei kuitenkaan riitä vain numeerinen samankaltaisuus arvioijien kesken vaan johdonmukaisuus ja yhdenmukaisuus edellyttävät yhteisen taitokäsityksen jakamista. Johdonmukaisuus ja yhdenmukaisuus liittyvät läheisesti validiuteen: kirjoittamisen arvioinnissa arvioijan käsitys kirjoittamisesta ohjaa sitä, miten hän omaan taitokäsitykseensä nojaten arvottaa tekstejä ja tulkitsee ja soveltaa määriteltyä asteikkoa. Tässä tutkimuksessa arvioijien taitokäsityksistä nousi yhtenäisemmin esiin kielen rakenteiden vahva asema, mikä saa kysymään, missä määrin käsitykset ovat sosiaalisia ja kulttuurisia. Toisin sanoen arvioita ja niiden viestimiä arvostuksia on tarkasteltava myös sosiokulttuurisesta näkökulmasta. Arvioijien käsitykset ovat osa opetuskulttuuria, sen käsityksiä, arvotuksia, traditioita ja normeja. Johdonmukaisuus ja yhdenmukaisuus eivät liity vain arvioijiin yksilöinä, ne eivät myöskään ilmene vain numeroina vaan ne ovat osa laajempaa arviointikulttuuria ja edustavat sen arvostuksia. Olen myös itse osa tätä arviointikulttuuria ja opetus-traditiota. Pohdiskellessani arvioijien toimintaa motiiveineen ja seurauksineen pohdiskelen samalla omaa toimintaani. En halua ottaa kantaa siihen, miten oikeita tai vääriä tässä tutkimuksessa esiin tulleet käsitykset ja arvostukset ovat. Sen sijaan kirjoittamiskäsityksen mahdollisen yksipuolisuuden kriittinen pohdinta on paikallaan, nimenomaan arvottamisen kannalta. Arvioinnin validius ja eettisyys edellyttävät, että käsitystaso ja konkreettinen arvottaminen vastaavat toisiaan.

Johdonmukaisuus ja yhdenmukaisuus palvelevat ensisijaisesti oppijaa ja hänen oikeusturvaansa. Mielestäni on tärkeää pohtia myös sitä, missä määrin voidaan olettaa, että arvioijan, opettajan tai opetuksen odotukset arvioinnista vastaavat oppijan odotuksia. Varsinkin testauksessa näkökulma on usein yksisuuntainen, oppija on läsnä vain suorituksena, jota käytetään testin onnistuneisuuden indikaattorina. Itsearvioinnin yleistymisen myötä johdonmukaisuus ja yhdenmukaisuus liittyvät myös opettaja- tai testiarvioinnin ja oppijan oman arvioinnin vastaavuuteen. Tässä vuorovaikutussuhteessa on myös kyse käsitysten vastaavuudesta mutta myös erilaisuuden sallimisesta ja merkitysten neuvottelemisesta erilaisista lähtöoletuksista käsin.

Kaiken kaikkiaan arvioijan toimintaa pohtiessa ei tule kuitenkaan unohtaa arvioinnin perustarkoitusta. Se ei ole itsessään validius, reliaabelius tai eettisyys. Arvioinnin tehtävänä on ennen kaikkea oppimisen tukeminen ja palautteen antaminen eettisesti ja johdonmukaisesti. Arvioinnin tehtävänä on tarjota oppijalle tie itsereflektioon ja parhaimmillaan myös työkaluja oppimaan oppimiseen, uusien tavoitteiden

asettamiseen ja itsensä kehittämiseen. Olen tiivistänyt ajatukseni yksilöarvioinnin asemasta oppimisessa kuvioon 27, jolla haluan havainnollistaa arvioinnin tehtävää riippumatta siitä, millainen arviointi-konteksti tai arviointitapa on (vrt. Kolb 1984; Järvinen 2001).



KUVIO 27 Yksilöarviointi osana oppimista.

Tutkimukseni kaltainen arviointi, esimerkiksi taitotasoarviointi, ei yleensä liity mihinkään tiettyyn kurssiin tai oppimäärään, joten sen tavoitelähtöisyys ei ole yhtä selvää kuin opetuksessa. Opetuksessa on mahdollista asettaa opetussisältöihin ja niiden saavuttamiseen liittyvät ryhmän yhteiset tavoitteet ja muun muassa oppijan lähtötason huomioonottavat henkilökohtaiset tavoitteet. Testiarvioinnissa puolestaan tavoitteet kätkeytyvät viitekehykseen ja asteikon tasokuvauksiin; ne ovat ikään kuin epäsuoremmin esillä ja tulkittavissa. Niiden yhteys oppijan omaan taitotasoon ja omiin tavoitteisiin on myös epäsuorempi. Arvioinnin perustehtävä on kuitenkin samantyyppinen molemmissa yhteyksissä.

Kuviossa 27 on kyse vuorovaikutteisesta systeemistä, jossa vaikutukset koskevat prosessin tai ympyrän kaikkia osia. Opetussuunnitelmat, viitekehykset ja tavoitteet esitetään ja niitä tavallisesti

tarkastellaan käsitteellisesti, ja niillä onkin hallinnollinen ja ohjaava asemansa abstraktisina kuvauksina. Niiden vaikuttavuutta voi kuitenkin arvioida vasta sitten, kun ne muuttuvat toiminnan kautta merkitykselliseksi. Systeemin eri osasten symmetrinen ja systemaattinen toiminen edellyttää yhteisöllistä neuvottelua merkityksistä ja merkitysten soveltamista käytäntöön sovitulla tavalla. Käytännössä tämä tarkoittaa muun muassa sitä, että arviointi ei ole autonominen yksikkö, vaan se on osa systeemiä. Siksi arvioinnin toteuttaminen, kehittäminen ja arvioiminen ei koske vain arviointia, vaan myös systeemin muita osia.

Tämän tutkimuksen johtopäätökset arvioinnista osoittavat arvioinnin luonteen monikerroksisuuden ja inhimillisyyden, minkä vuoksi edellä tehdyt johtopäätökset ovat vasta päänavaus. Kuten Raippamaa (1965) toteaa: "Johtopäätös on ymmärryksen tulos. Se saattaa olla virheellinen, jos riittävä määrä asiantietoja puuttuu. Mutta senhetkiseen tilanteeseen se on ainoa mahdollinen virheellisyydestään huolimatta. Jos se tuottaa tuonnempana hankaluuksia, se on paras syy uusien asiantietojen hankkimiseksi." Arvioinnin vaikuttavuus, hyödyllisyys ja tarpeellisuus haastavat jatkossakin lisäämään ymmärrystä arvioinnin luonteesta ja toimivuudesta.

LÄHTEET:

- Abedi, J., E. L. Baker & H. Herl 1995. *Comparing reliability indices obtained by different approaches for performance assessments*. CSE Technical Report 401. University of California.
- AERA (American Educational Research Association), APA (the American Psychological Association) & NCME (the National Council on Measurement in Education) 1999. *Standards for educational and psychological testing*. Washington D.C: American Educational Research Association.
- Alasuutari, P. 1999. *Laadullinen tutkimus*. Tampere: Vastapaino.
- Alderson, J. C. 1991. Bands and scores. Teoksessa J. C. Alderson & B. North (toim.) *Language testing in the 1990s*. London: Modern English Publications, McMillan, 71—86.
- Alderson, J. C. & D. Wall 1993. Does washback exist? *Applied Linguistics* 14, 115—129.
- Anastasi A. 1986. Evolving concepts of test validation. *Annual Review of Psychology* 37, 1—15.
- Andelin, B. 1991. *Kemiön ruotsinkielisten lukiolaisten puhutusta ja kirjoitetusta kielestä. Peruslausetyyppianalyysi ja virheanalyysi*. Pro gradu -tutkielma. Turku: Åbo Akademi.
- Angoff, W. H. 1988. Validity: an evolving concept. Teoksessa H. Wainer & H. I. Braun (toim.) *Test validity*. Hillsdale: Lawrence Erlbaum, 19—33.
- Apajalahti, M. 1996. Peruskoulun oppilasarvioinnin arviointi. Teoksessa A. Räisänen & T. Frisk (toim.) *Silta uuteen opiskelija-arviointiin. Arviointia opiskelija-arvioinnista*. Arviointi 6. Helsinki: Opetushallitus, 83—93.
- Ariely, D. & D. Zakay 2001. A timely account of the role of duration in decision making. *Acta Psychologica* 108, 2, 187—207.
- Atkinson, P. & M. Hammersley 1994. Ethnography and participant observation. Teoksessa N. K. Denzin & Y. Lincoln (toim.) *Handbook of qualitative research*. Thousand Oaks: Sage Publications, 248—261.
- Atkinson, R., R. C. Atkinson, E. E. Smith, D. J. Bem & S. Nolen-Hoeksema 2000. *Hilgard's introduction to psychology*. Orlando: Harcourt College Publishers.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. 2000. Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17, 1, 1—42.
- Bachman, L. F., B. K. Lynch & M. Mason 1995. Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing* 12, 2, 238—257.
- Bachman, L. F. & A. S. Palmer 1996. *Language testing in Practice*. Oxford: Oxford University Press.
- Bailey, K. M. 1996. Working for washback: a review of the washback concept in language testing. *Language Testing* 13, 3, 257—279.
- Baker, C. 1997. Membership categorization and interview accounts. Teoksessa D. Silverman (toim.) *Qualitative research. Theory, method and practice*. London: Sage Publication, 130—143.

- Ballard B. & J. Glanchy 1991. Assessment by misconception: cultural influences and intellectual traditions. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 19—36.
- Bannister, D. & J. M. M. Mair 1968. *The evaluation of personal constructs*. London: Academic Press.
- Bannister, D. & F. Fransella 1971. *Inquiring man. The theory of personal constructs*. Harmondsworth: Penguin Books.
- Barnard, A., H. McCosker & R. Gerber 1999. Phenomenography: A qualitative research approach for exploring understanding in health care. *Qualitative Health Research*, 9, 2, 212—227.
- Barton, D. 1991. The social nature of writing. Teoksessa D. Barton & R. Ivanič (toim.) *Writing in the community. Written communication annual. An international survey of research and theory*. Newbury Park: SAGE Publications, 1—13.
- Barton D. 1994. *Literacy: an introduction to the ecology of written language*. Oxford: Blackwell.
- Beaufort, A. 2000. Learning the trade. A social apprenticeship model for gaining writing expertise. *Written Communication* 17, 2, 185—224.
- Bell, J. S. 1995. The relationship between L1 and L2 literacy: some complicating factors. *TESOL QUARTERLY* 29, 4, 687—704.
- Bereiter, C. & M. Scardamalia 1987. *The psychology of written composition*. Hillsdale: Erlbaum.
- Birenbaum, M. 1996. Assessment 2000: towards a pluralistic approach to assessment. Teoksessa M. Birenbaum & F. J. R. C. Dochy (toim.) *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston: MA, Kluwer, 3—29.
- Brannen, J. 1992. Combining qualitative and quantitative approaches: an overview. Teoksessa J. Brannen (toim.) *Mixing methods: qualitative and quantitative research*. Aldershot: Avebury, 3—37.
- Brown, A. tulossa. Legibility and the rating of second language writing: an investigation of the rating of handwritten and word-processed IELTS task. Teoksessa R. Tulloh (toim.) *Research Reports*.
- Bruner, J. 1986. *The culture of education*. Cambridge, Mass.: Harvard University Press.
- Bryman, A. 1988. *Quantity and quality in social research. Contemporary social research series*. London: Unwin Hyman Ltd.
- Burry-Stock, J. A. & D. G. Shaw 1996. Rater agreement indexes for performance assessment. *Educational and Psychological Measurement* 26, 2, 251—263.
- Cazden, C. B. 2000. Taking cultural differences into account. Teoksessa B. Cope & M. Kalantzis (toim.) *Multiliteracies. Literacy learning and the design of social futures*. London: Routledge, 249—266.
- Carlson, B. W. 1984. *Condensation theory: a process model of human decision making*. University of Michigan: University Microfilms International.
- Carlisle, R. & E. McKenna 1991. Placement of ESL/EFL undergraduate writers in college-level writing programs. Teoksessa L. Hamp-Lyons (toim.)

- Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 197—211.
- Carson, J. G. 1990. Reading-writing connections: toward a description for second language learners. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. Cambridge: Cambridge University Press, 88—101.
- Carson, J. G. 1993. Reading for writing: cognitive perspectives. Teoksessa J. G. Carson & I. Leki (toim.) *Reading in the composition classroom. Second language perspectives*. Boston: Heinle and Heinle Publishers, 85—104.
- Chafe, W. 1987a. What good is punctuation? *Occasional Paper No. 2. Center for the Study of Writing*. Berkeley: University of California.
- Chafe, W. 1987b. Punctuation and the prosody of written language. *Technical report no. 11. Center for the Study of Writing*. Berkeley: University of California.
- Charney, D. 1984. The validity of using holistic scoring to evaluate writing: A critical overview. *Research on the Teaching of English* 18, 1, 65—80.
- Chelimsky, E. 1997. The coming transformations in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.) *Evaluation for the 21st century. A handbook*. Thousand Oaks: Sage Publications, 1—26.
- Cohen, A. & M. C. Cavalcanti 1990. Feedback on compositions: teacher and student verbal reports. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. Cambridge: Cambridge University Press, 155—177.
- Cole, N. S. & P. A. Moss 1993. Bias in test use. Teoksessa R. L. Linn (toim.) *Educational measurement*. National Council on Measurement in Education, American Council on Education. Phoenix: The Oryx Press, 201—219.
- Clapham, C. 2000. Assessment and testing. *Annual Review of Linguistics*, 20, 147—161.
- Conger, A. J. 1985. Kappa reliabilities for continuous behaviors and events. *Educational and Psychological Measurement* 45, 861—868.
- Connor, U. 1991. Linguistic/rhetorical measures for evaluating ESL writing. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 215—226.
- Connor, U. 1996. *Contrastive rhetoric. Cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press.
- Connor, U. & P. Carrell 1993. The interpretation of tasks by writers and readers in holistically rated direct assessment of writing. Teoksessa J. G. Carson & I. Leki (toim.) *Reading in the composition classroom: second language perspectives*. Boston: Heinle and Heinle Publishers, 141—160.
- Cook-Gumperz, J. & J. Gumperz 1981. From oral to written culture: the transition to literacy. Teoksessa M. F. Whiteman (toim.) *Writing: the nature, development and teaching of written communication*. Volume 1. Variation in writing: functional and linguistic-cultural differences. Hillsdale: Lawrence Erlbaum Associates, 89—109.
- Cooper, C. R. 1977. Holistic evaluation of writing. Teoksessa C. R. Cooper & L. Odell (toim.) *Evaluating writing: Describing, measuring, judging*. Buffalo:

- State University of New York, 3—32.
- Cooper, C. R. & L. Odell 1977. Introduction. Teoksessa C. R. Cooper & L. Odell (toim.) *Evaluating writing: Describing, measuring, judging*. Buffalo: State University of New York, vii—xii.
- Cortina, J. M 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 78, 1, 98—104.
- Council of Europe 1996. *Modern languages: learning, teaching, assessment. A common European framework of reference*. Draft 2 of a framework proposal. Strasbourg: Council of Europe.
- Council of Europe, 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cozby, P. C. 1985. *Methods in behavioral research*. Palo Alto: Mayfield Publishing Company.
- Cramer, D. 1998. *Fundamental statistics for social research. Step-by-step calculations using SPSS for Windows*. New York: Routledge.
- Cronbach, L. J. 1990. *Essentials of psychological testing*. New York: Harper and Row.
- Cronbach, L. J. 1996. Coefficient alpha and the internal structure of tests. Teoksessa W. Ward, H. W. Stoker & M. Murray-Ward (toim.) *Educational measurement, origins, theories and explications*. Volume 1. Basic concepts and theories. Lanham: University Press of America, Inc., 257—269.
- Crocker, L. & J. Algina 1986. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston, Inc.
- Cumming, A. 1989. Writing expertise and second language proficiency. *Language Learning* 39, 1, 81—141.
- Cumming, A. 1990. Expertise in evaluating second language compositions. *Language Testing*, 7, 1, 31—51.
- Cumming, A. 1996. Introduction: The concept of validation in language testing. Teoksessa A. Cumming & R. Berwick (toim.) *Validation in language testing*. Clevedon: Multilingual Matters LTD, 1—14.
- Cumming, A. 1997. The testing of writing in a second language. Teoksessa C. Clapham & D. Corson (toim.) *Encyclopedia of language and education, 7: Language testing and assessment*. Dordrecht: Kluwer Academic Publishers, 51—63.
- Cumming, A. 1998. Theoretical perspectives on writing. *Annual Review of Applied Linguistics* 18, 61—78.
- Cushing Weigle, S. 1994. Effects of training on raters of ESL compositions. *Language Testing* 11, 2, 197—223.
- Cushing Weigle, S. 1999. Investigating rater/prompt interactions in writing assessment: quantitative and qualitative approaches. *Assessing Writing* 6, 2, 145—178.
- Davidson, F., Turner, C, E, and A. Huhta 1997. Language testing standards. Teoksessa C. Clapham & D. Corson (toim.) *Language testing and assessment 7. The encyclopedia of language and education*. Dordrecht: Kluwer Academic, 303—311.

- Davis, B. G., M. Scriven & S. Thomas 1987. *The evaluation of composition instruction*. New York: Teacher College Press.
- DeRemer, M. L. 1998. Writing assessment: raters' elaboration of the rating task. *Assessing Writing* 5, 1, 7—29.
- Doyle, K. O 1998. Research note: reliability and dimensionality of stimulus ratings. *Journal of Advertising Research* 38, 2, 45—52.
- Dunn, G. 1989. *Design and analysis of reliability studies: the statistical evaluation of measurement errors*. New York: Oxford University Press.
- Ebel, R. L. 1979. *Essentials of educational measurement*. New Jersey: Prentice-Hall Inc.
- Eggleston, J. 1990. School examinations – some sociological issues. Teoksessa T. Horton (toim.) *Assessment debates*. London: Hodder and Stoughton, 57–67.
- Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Engelhard, G. Jr. 1994. Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement* 31, 2, 93—112.
- Engelhard, G. Jr. 1996. Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement* 33, 1, 56—70.
- Engelhard G. Jr. & G. E. Stone 1998. Evaluating the quality of ratings obtained from standards-setting judges. *Educational and Psychological Measurement* 58, 2, 179–196.
- Enqvist, N. E. 1990. Seven problems in the study of coherence and interpretability. Teoksessa U. Connor & A, M. Johns (toim.) *Coherence in writing: research and pedagogical perspectives*. Arlington: TESOL, 9–28.
- Eskey, D. E. 1993. Reading and writing as both cognitive process and social behavior. Teoksessa J.G. Carson & I. Leki (toim.) *Reading in the composition classroom: second language perspectives*. Boston: Heinle and Heinle Publishers, 221–233.
- Eskola, J. & J. Suoranta 1998. *Johdatus laadulliseen tutkimukseen*. Tampere: Vastapaino.
- Everitt, B. S. & D. F. Hay 1992. *Talking about statistics. A Psychologist's guide to data analysis*. London: Edward Arnold.
- Faigley, L. 1985. Nonacademic writing: the social perspective. Teoksessa L. Odell & D. Goswami (toim.) *Writing in nonacademic settings*. New York: Guilford Press, 231 – 248.
- Faigley, L., R. D. Cherry, D. A. Jolliffe & A. M. Skinner 1989. *Assessing writers' knowledge and processes of composing*. Norwood: Ablex Publishing Corporation.
- Fairclough, N. 1995. *Media discourse*. London: Edward Arnold.
- Fatham, A. K. & E. Whalley 1990. Teacher response to student writing: focus on form versus content. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. Cambridge: Cambridge University Press, 178–190.
- Feldt, L. S. & R. L. Brennan 1993. Reliability. Teoksessa R. L. Linn (toim.) *Educational measurement*. National Council on Measurement in Education,

- American Council on Education. Phoenix: The Oryx Press, 105—146.
- Filer, A. 1995. Teacher assessment: social process and social product. *Assessment in Education: Principles, Policy & Practice* 2, 1, 23–39.
- Fitzgerald, J. 1992. Reading and writing stories. Teoksessa J. W. Irwin & M. A. Doyle (toim.) *Reading/writing connections. Learning from research*. Newark: International Reading Association, 81–95.
- Flower, L. S. & J. R. Hayes 1980. The dynamics of composing: making plans and juggling constraints. Teoksessa L. W. Gregg & E. R. Steinberg (toim.) *Cognitive processes in writing*. Hillsdale: Lawrence Erlbaum Associates, 31–50.
- Flower, L. 1987. Interpretive acts: cognition and the construction of discourse. *Occasional Paper No. 1. Center for the Study of Writing*. Berkeley: University of California.
- Flower, L. 1989. Cognition, context and theory building. *Occasional Paper No. 11. Center for the Study of Writing*. Berkeley: University of California.
- Flower, L. 1994. *The construction of negotiated meaning. A social cognitive theory of writing*. Southern Illinois University: Southern Illinois University Press.
- Flower, L., A. K. Schriver, L. Carey, C. Haas & J. R. Hayes 1989. Planning in writing: The cognition of a constructive process. *Technical Report No. 34. Center for the Study of Writing*. Berkeley: University of California.
- Forest, J. & C. Mehier 2001. John R. Commons and Herbert A. Simon on the concept of rationality. *Journal of Economic Issues* 35, 3, 591–606.
- Fox, S. & T. Caspy 1996. Variables affecting leniency, halo and validity of self-appraisal. *Journal of Occupational & Organizational Psychology* 67, 1, 45–57.
- Fransella, F. & D. Bannister 1977. *A manual for repertory grid technique*. London: Academic Press.
- Fredriksen, J. R. 1992. Learning to “see”: scoring video portfolios or “beyond the hunter-gatherer” in performance assessment. *Esitelmä American Educational Research Association -kongressissa* San Franciscossa.
- Freedman, D. 1989. Teacher training, development, and decision making: a model of teaching and related strategies for language teacher education. *TESOL QUARTERLY*, 23, 1, 27—45.
- Freedman, S. W. 1982. Some reasons for the grades we give compositions. *English Journal* 71, 7, 86–89.
- Freedman, S. W. & R. C. Calfee 1983. Holistic assessment of writing: experimental design and cognitive theory. Teoksessa P. Mosenthal, L. Tamor & S. A. Walmsley (toim.) *Research on writing: principles and methods*. New York: Longman, 75—98.
- Frisk, T. & A. Räisänen. Oppilas- ja opiskelija-arvioinnin valtakunnallisen ohjauksen arviointia. Teoksessa A. Räisänen & T. Frisk (toim.) *Silta uuteen opiskelija-arviointiin. Arviointia opiskelija-arvioinnista*. Arviointi 6. Helsinki: Opetushallitus, 27–41.
- Gajdusek, L. & D. VanDommelen 1993. Teoksessa J. G. Carson & I. Leki (toim.) *Reading in the composition classroom. Second language perspectives*. Boston: Heinle and Heinle Publishers, 197–217.
- Grabe, W. 2001. Notes toward a theory of second language writing. Teoksessa

- T. Silva & P. K. Matsuda (toim.) *On second language writing*. Mahwah: Lawrence Erlbaum Associates, 39—57.
- Grönholm, M. 1991. Ruotsinkielisten oppilaiden sanaston kehityksestä suomen kielessä. Aikakausjulkaisussa M. Grönholm & P. Muikku-Werner (toim.) *Finlance. The Finnish Journal of Learning and Language Teaching*, IX, 17—32.
- Grönholm, M. 1992. Suomen verbisysteemin omaksumisesta ruotsinkielisillä peruskoululaisilla. Aikakausjulkaisussa R. Markkanen (toim.) *Finlance. The Finnish Journal of Learning and Language Teaching*, X, 101—116.
- Grönholm, M. 1993. Tempusten ja modusten käyttö toisen kielen kirjoittamisessa. *Virittäjä* 97, 3, 410—420.
- Guba, E. G. 1990. The Alternative Paradigm Dialog. Teoksessa E. C. Guba (toim.) *The paradigm dialog*. Newbury Park: Sage Publications, 17—27.
- Guba, E. G. & Y. S. Lincoln 1989. *Fourth generation evaluation*. Newbury Park: Sage Publications.
- Guba, E. G. & Y. S. Lincoln 1994. Competing paradigms in qualitative research. Teoksessa N. K. Denzin & Y. Lincoln (toim.) *Handbook of qualitative research*. Thousand Oaks: Sage Publications, 105—117.
- Halliday, M. A. K. 1994. *An introduction to functional grammar*. London: Edward Arnold.
- Hamp-Lyons, L. 1990. Second language writing: assessment issues. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. Cambridge: Cambridge University Press, 69—87.
- Hamp-Lyons, L. 1991a. The writer's knowledge and our knowledge of the writer. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 51—70.
- Hamp-Lyons, L. 1991b. Reconstructing "academic writing proficiency". Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 127—154.
- Hamp-Lyons, L. 1991c. Pre-text: task-related influences on the writer. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 87—110.
- Hamp-Lyons, L. 1991d. Scoring procedures for ESL contexts. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 241—278.
- Hamp-Lyons, L. 1997a. Washback, impact and validity: ethical concerns. *Language Testing* 14, 3, 295—303.
- Hamp-Lyons, L. 1997b. Ethics in language testing. Teoksessa C. Clapham & D. Corson (toim.) *Language testing and assessment 7. The encyclopedia of language and education*. Dordrecht: Kluwer Academic, 223—333.
- Hamp-Lyons, L. 2001. Fourth generation writing assessment. Teoksessa T. Silva & P. K. Matsuda (toim.) *On second language writing*. Mahwah: Lawrence Erlbaum Associates, 117—127.
- Hamp-Lyons, L. & B. K. Lynch 1998. Perspectives on validity: a historical analysis of language testing conference abstracts. Teoksessa A. J. Kunnan (toim.) *Validation in language assessment: selected papers from the 17th*

- Language Testing Research Colloquim, Long Beach*. Mahwah, NJ: Lawrence Erlbaum, 253–276.
- Hansman, C. & A. L. Wilson 1998. Teaching writing in community colleges: a situated view of how adults learn to write in computer-based writing classroom. *Community College Review* 26, 1, 21–42.
- Harwell, M. 1999. Evaluating the validity of educational rating data. *Educational & Psychological Measurement* 59, 1, 25–38.
- Harvey, N. 1997. Confidence in judgement. *Trends in Cognitive Sciences* 1, 2, 78–82.
- Haswell, R. & S. Wyche-Smith 1994. Adventuring into writing assessment. *College Composition and Communication* 45, 220–236.
- Hatch, J. A., Hill, C. A. & J. R. Hayes 1993. When the messenger is the message: Readers' impressions of writers' personalities. *Written Communication* 10, 4, 569–598.
- Hayes, J. R. & L. S. Flower 1980. Identifying the organization of writing processes. Teoksessa L. W. Gregg & E. R. Steinberg (toim.) *Cognitive processes in writing*. Hillsdale: Lawrence Erlbaum Associates, 3–30.
- Hayes, J. R. & J. A. Hatch 1999. Issues in measuring reliability. *Written Communication* 16, 3, 354–368.
- Heikkilä, T. 1998. *Tilastollinen tutkimus*. Helsinki: Edita.
- Henning, G. 1987. *A guide to language testing. Development, evaluation, research*. Boston: Heinle & Heinle Publishers.
- Henning, G. 1991. Issues in evaluating and maintaining an ESL writing assessment program. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 279–292.
- Hogarth, R. M. 1980. *Judgement and choice. The psychology of decision*. Chichester: John Wiley & Sons.
- Holstein, J. A. & J. F. Gubrium 1994. Phenomenology, ethnomethodology and interpretive practice. Teoksessa N. K. Denzin & Y. Lincoln (toim.) *Handbook of qualitative research*. Thousand Oaks: Sage Publications, 262–272.
- Homburg, T. J. 1984. Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL QUARTERLY* 18, 1, 87–106.
- Hornberger, N. H. 1994. Ethnography. *TESOL QUARTERLY*, 28, 4, 688–690.
- House, E. R. 1995. Evaluoinnin futuurin perfekti. Teoksessa S. Takala (toim.) *Arviointi ja koulutuksen laadun kehittäminen*. Kasvatustieteiden tutkimuslaitos, Jyväskylä: Jyväskylä: Kasvatustieteiden tutkimuslaitos, 23–32.
- Huber, O. 1989. Information-processing operators in decision making. Teoksessa H. Montgomery & O. Svenson (toim.) *Process and structure in human decision making*. Chichester: John Wiley & Sons, 3–21.
- Huhta, A. 1993. Suullisen kielitaidon arviointi. Teoksessa S. Takala (toim.) *Suullinen kielitaito ja sen arviointi*. Teoriaa ja käytäntöä 77. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 143–225.
- Huhta, A. & S. Takala 1999. Kielitaidon arviointi. Teoksessa A. Piirainen-Marsh

- & K. Sajavaara (toim.) *Kielenoppimisen kysymyksiä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 179–228.
- Huot, B., 1990. The literature of direct writing assessment: major concerns and prevailing trends. *Review of Educational Research*, 60, 2, 237–263.
- Huot, B. 1997. Beyond accountability: reading with faculty as partners across the disciplines. Teoksessa K. B. Yancey & B. Huot (toim.) *Assessing writing across the curriculum. Diverse approaches and practices*. Greenwich: Ablex Publishing Corporation, 69–78.
- Huttunen, I. 1997. *Kimmoke. Kielten opetuksen ja oppimisen yleiseurooppalainen viittekehys*. Opettajan opas. Opetushallitus. Moniste 27/1997. Helsinki: Edita Oy.
- Huttunen, L. 2002. *Kotona, maanpaossa, matkalla: kodin merkitykset maahanmuuttajien omaelämäkertoissa*. Tampere: Tampereen yliopisto.
- Inkinen, I. 1992. *Lukiolaisten ainekirjoitusten kontrastiivista tarkastelua*. Pro gradu -tutkielma. Vaasan yliopisto.
- Ivanič, R. & W. Moss 1991. Bringing community writing practices into education. Teoksessa D. Barton & R. Ivanič (toim.) *Writing in the community. Written communication annual. An international survey of research and theory*. Newbury Park: SAGE Publications, 193—223.
- Jaakkola, H. 1997. *Kielitieto kielitaitoon pyrittäessä. Vieraiden kielten opettajien käsityksiä kielioinnin oppimisesta ja opettamisesta*. Jyväskylä Studies in Education, Psychology and Social Research 128. Jyväskylä: Jyväskylän yliopisto.
- Jahoda, M. 1986. Opening Address. The range of convenience of personal construct psychology - an outsider's view. Teoksessa F. Fransella & L. Thomas (toim.) *Experimenting with personal construct psychology*. London: Routledge & Kegan Paul, 1–14.
- James, C. & P. Garrett 1992. Language awareness. Teoksessa C. James & P. Garrett (toim.) *Language awareness in the classroom*. New York: Longman, 3–20.
- Johns, A. 1990. L1 composition theories: implications for developing theories of L2 composition. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. New York: Cambridge University Press, 24–36.
- Jokinen A., Juhila K. & E. Suoninen 1993. Diskursiivinen maailma. Teoreettiset lähtökohdat ja analyttiset käsitteet. Teoksessa A. Jokinen, K. Juhila & E. Suoninen (toim.) *Diskurssianalyysin aakkoset*. Tampere: Vastapaino, 17–47.
- Järvinen, P. T. 2001. Johtamisjärjestelmät. *Esitelmä Valmentavaa johtaminen -seminaarissa* 20.9.2001.
- Kalliokoski, J. 1996. Puhe, kirjoitus, tekstilajin normit ja toisella kielellä kirjoittaminen. Teoksessa H. Ruuska & S-M. Tuomi (toim.) *Moneja baareja: tiellä toimivaan kaksikielisyysyteen*. Helsinki: Äidinkielen opettajain liitto, 107–119.
- Kalliokoski, J. 2002. Features of oral and written discourse in second language writing. *Esitelmä Uralic Today -kongressissa Pariisissa* 14.3.2002.
- Kaplan, R. B. 1988. Contrastive rhetoric and second language learning: notes towards a theory of contrastive rhetoric. Teoksessa A. C. Purves (toim.) *Writing across languages and cultures. Issues in contrastive rhetoric*. Newbury

- Park: SAGE Publications, 275—304.
- Karjalainen, A. 2001. *Tentin teoria*. Oulu: Oulun yliopisto.
- Karlsson, G. 1989. Rules and strategies in decision making: A critical analysis from a phenomenological perspective. Teoksessa H. Montgomery & O. Svenson (toim.) *Process and structure in human decision making*. Chichester: John Wiley & Sons, 51–62.
- Keltikangas-Järvinen, L. 1994. *Hyvä itsetunto*. Helsinki: WSOY.
- Kern, R. G. & J. M. Schultz 1992. The effects of composition instruction on intermediate level French students' writing performance: some preliminary findings. *The Modern Language Journal* 76, 1, 1–13.
- Klaczynski, P. A., J. P. Byrnes & J. E. Jacobs 2001. The development of decision making. *Journal of Applied Developmental Psychology* 22, 3, 225–236.
- Kobayashi, H. & C. Rinnert 1996. Factors affecting composition evaluation in an EFL context: cultural rhetorical pattern and readers' background. *Language Learning* 46, 3, 397–437.
- Kolb, D. A. 1984. *Experimental learning*. London: Prentice-Hall International, Inc.
- Komulainen, S. 1999. *Characteristics of proficient speaking skills in foreign language: empirically-derived criteria for assessing speaking competence*. Pro gradu -työ. Jyväskylän yliopiston englannin kielen laitos.
- Konttinen, R. 1981. *Testiteoria. Johdatus kasvatus- ja käyttäytymistieteellisen mittaamisen teoriaan*. Helsinki: Oy Gaudeamus Ab.
- Korkiasaari, J. & I. Söderling 2001. Finnish emigration and immigration after world war II. Siirtolaisuusinstituutti. http://www.utu.fi/erill/instmigr/art/finn_migr.htm.
- Korkman, C. 1995. *Tvåspråkighet och skriftlig framställning: en undersökning av tvåspråkiga elevers uppsatser i den finlandssvenska grundskolan*. Helsingfors: Svenska litteratursällskapet i Finland.
- Korpinen, E. 1993. Uskooko oppilas oppivansa? Luki-oppilaan minäkokemukset ja minäkäsitys peruskoulun päätösvaiheessa. Teoksessa P. Linnakylä & H. Saari (toim.) *Oppiiko oppilas peruskoulussa? Peruskoulun arviointi 90 - tutkimuksen tuloksia*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 5–26.
- Korpinen, E. 1996. Kuka, mikä ja minkä arvoinen olen? Minäkuva ja arviointi. Teoksessa A. Räisänen & T. Frisk (toim.) *Silta uuteen opiskelija-arviointiin. Arviointia opiskelija-arvioinnista*. Arviointi 6. Helsinki: Opetushallitus, 69–80.
- Kroll, B. 1998. Assessing writing abilities. *Annual Review of Applied Linguistics* 18, 219–240.
- Kunnan, A. J. 1997. Connecting fairness with validation in language assessment. Teoksessa A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (toim.) *Current development and alternatives in language assessment. Proceedings of LTRC 96*. Jyväskylä: University of Jyväskylä.
- Kvale, S. 1995. Evaluation as construction of knowledge. Teoksessa J. Jussila & R. Rajala (toim.) *Rajanylityksiä: monipuolistuva kasvatustutkimus tieteiden kentässä*. XIV Kasvatustieteen päivät Rovaniemellä 24.–26.11.1994. Rovaniemi: Lapin yliopisto, 15–39.

- Kvale, S. 1996. *Interviews: an introduction to qualitative research interviewing*. Thousand Oaks: Sage Publications.
- Laine, E. & M-K. Pihko 1991. *Kieliminä ja sen mittaaminen*. Jyväskylä: Kasvatustieteiden tutkimuslaitos.
- Latomaa, S. 1996. Matkalla uuteen kieleen. Teoksessa H. Ruuska & S-M. Tuomi (toim.) *Moneja baareja: tiellä toimivaan kaksikielisyYTEEN*. Helsinki: Äidinkielen opettajain liitto, 97–105.
- Laufer, B. & P. Nation 1999. A vocabulary size test of controlled productive ability. *Language Testing* 16,1, 33–51.
- Lehman, R. 1993. Rating the quality of student writing – findings from the IEA study of achievement in written composition. Teoksessa A. Huhta, K. Sajavaara & S. Takala (toim.) *Language testing: new openings*. Jyväskylä: Institute of Educational Research, 186–204.
- Leimu, K. 1993. Bloomilaisuuden harhat – tiedollinen viattomuus ja sen menettäminen. Teoksessa E. Kangasniemi & R. Konttinen (toim.) *Lue, etsi, tutki. Tutkittua tietoa koulun kehittämiseksi*. Opetus 2000. Porvoo: WSOY, 10–31.
- Leiwo, M. & M-R. Luukka 2002. Lukutaidosta ja lukemisen vaikeudesta. Teoksessa I. Herlin, J. Kalliokoski, L. Kotilainen & T. Onikki-Rantajääskö (toim.) *Äidinkielen merkitykset*. Helsinki: SKS, 178–190.
- Leki, I. 1990. Coaching from the margins: issues in written response. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. New York: Cambridge University Press, 57–68.
- Leki, I. 1993. Reciprocal themes in ESL reading and writing. Teoksessa J. G. Carson & I. Leki (toim.) *Reading in the composition classroom: second language perspectives*. Boston: Heinle and Heinle Publishers, 9–32.
- Leki, I. 1997. Cross-talk: ESL issues and contrastive rhetoric. Teoksessa C. Severino, J. Guerra, and J. Butler (toim.) *Writing in multicultural settings*. New York: Modern Language Association, 234–244.
- Leki, I. 2000. Writing, literacy and applied linguistics. *Annual Review of Applied Linguistics* 20, 90–115.
- Levy, C. M. & S. Ransdell 1995. Is writing as difficult as it seems? *Memory & Cognition* 23, 6, 767–779.
- Lilis, T. & J. Tuner 2001. Student writing in higher education: contemporary confusion, traditional concerns. *Teaching in Higher Education* 6, 1, 57–68.
- Lincoln, Y. S. & N. K. Denzin 1994. The fifth moment. Teoksessa N. K. Denzin & Y. S. Lincoln (toim.) *Handbook of qualitative research*. Thousand Oaks: Sage Publications, 575–586.
- Linn, R. L. & N. E. Gronlund 1995. *Measurement and assessment in teaching*. New Jersey: Prentice Hall.
- Linn, R. L. 2000. Assessment and accountability. *Educational Researcher* 29, 2, 4–16.
- Linnakylä, P. 1993. Portfolio oppilaan itsearvioinnin ja itseohjautuvan oppimisen tukena: mahdollisuuksia ja rajoituksia. Teoksessa E. Kangasniemi & R. Konttinen (toim.) *Lue, etsi, tutki. Tutkittua tietoa koulun kehittämiseksi*. Opetus 2000. Porvoo: WSOY, 70–93.

- Linnakylä, P. 1994. Arviointi oppilaan oppimisen ja kasvun tukena. Teoksessa R. Jaku-Sihvonen, A. Räisänen & P. Väyrynen (toim.) *Virikkeitä koulutuksen arvioinnin kehittäjille. Arviointi ja seuranta 2*. Opetushallitus. Helsinki: Yliopistopaino, 29–41.
- Linnakylä, P. 1998a. Individual in the school picture exploring students' writing portfolios. Teoksessa H. Jokinen & J. Rushton (toim.) *Changing contexts of school development – the challenges to evaluation and assessment*. <http://www.jyu.fi/ktl/publications/electronic/001/linnakyla.html>.
- Linnakylä, P. 1998b. Koulutuksen kansallisen ja kansainvälisen arvioinnin haasteita. *Kasvatus* 2, 175–189.
- Lloyd-Jones, R. 1977. Primary trait scoring. Teoksessa C. R. Cooper & L. Odell (toim.) *Evaluating writing: Describing, measuring, judging*. The United States of America: the National Council of Teachers of English, 33–68.
- Lumley, T. J. N. 2000. *The process of the assessment of writing performance: the rater's perspective*. Melbourne: University of Melbourne.
- Lumley, T. & T. McNamara 1995. Rater characteristics and rater bias: implications for training. *Language Testing*, 12, 1, 54–71.
- Lunz, M. & J. A. Stahl 1994. Interjudge reliability and decision reproducibility. *Educational & Psychological Measurement* 53, 4, 913–926.
- Luoma, S. 1997. *Comparability of a tape-mediated and a face-to-face test of speaking. A triangulation study*. Lisensiaatin työ. Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus.
- Luoma, S. 2001. *What does your test measure? Construct definition in language test development and validation*. Jyväskylä: University of Jyväskylä.
- Luoma, S. & M. Tarnanen tulossa. Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*.
- Luukka, M-R. 1995. *Puhuttua ja kirjoitettua tiedettä. Funktionaalinen ja yhteisöllinen näkökulma tieteen kielen interpersonaalisiin piirteisiin*. Jyväskylä: Jyväskylän yliopisto.
- Luukka, M-R. 2000. Näkökulma luo kohteen: diskurssitutkimuksen tausta-oletukset. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kieli, diskurssi & yhteisö*. Soveltavan kielentutkimuksen teoriaa ja käytäntöä. Jyväskylä: Soveltavan kielentutkimuksen keskus, 133–160.
- Luukka, M-R. 2001. Elämää tekstien kanssa. *Kielikuvia* 2, 4–14.
- Lynch, B. K. 1997. In search of the ethical test. *Language Testing* 14, 3, 315–327.
- Lynch, B. K. 2001. Rethinking assessment from a critical perspective. *Language Testing* 18, 4, 351–372.
- Lähdemäki, E. 1995. *Suomenruotsalaisten suomen kielen ainek kirjoituksessa tekemistä virheistä ja niiden mahdollisista syistä*. Lisensiaattityö. Turku: Åbo Akademi.
- Lähteenmäki, M. & H. Dufva 1998. On meaning and understanding: a dialogical approach. *Dialogism. An International Journal of Bakhtin Studies* 1, 74–91.
- Marton F. 1994. Phenomenography. Teoksessa T. Husén & T. Neville (toim.) *The international encyclopedia of education*. Volume 8. Postlethwaite: Pergamon, 4 424–4 429.
- Marton, F. & P. M. Fai 1999. Two faces of variation. *Esitelmä 8th European Conference for Learning and Instruction -kongressissa*.

- Matsuda, P. K. 1998. Situating ESL writing in a cross-disciplinary context. *Written Communication* 15, 1, 99–122.
- Mauranen, A. 1993. *Cultural differences in academic discourse. A textlinguistic study*. Frankfurt am Main: Peter Lang.
- Mauranen, A. 2000. Kontrastiivinen retoriikka. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kieli, diskurssi & yhteisö*. Soveltavan kielentutkimuksen teoriaa ja käytäntöä. Jyväskylä: Soveltavan kielentutkimuksen keskus, 299–330.
- McDonald, R. P. 1999. *Test theory: a unified treatment*. New Jersey: Lawrence Erlbaum Associates.
- McNamara, T. 1996. *Measuring second language performance*. New York: Longman.
- McNamara, T. 2001. Language assessment as social practice: challenges for research. *Language Testing* 18, 4, 333–349.
- Meiron, B. E. 1998. *Understanding raters' thought process in holistic oral proficiency test scoring: an explanatory triangulated study*. A Thesis. Los Angeles: California State University.
- Messick, S. 1989. Validity. Teoksessa R. L. Linn (toim.) *Educational measurement*. National Council on Measurement in Education, American Council on Education. Phoenix: The Oryx Press, 13–103.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Research* 23, 2, 13–23.
- Messick, S. 1996. Validity and washback in language testing. *Language Testing* 13, 3, 241–256.
- Milanovic, M., N. Saville & S. Shuhong 1996. A study of the decision-making behavior of composition markers. Teoksessa M. Milanovic & N. Saville (toim.) *Performance testing, cognition and assessment: selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem*. Cambridge: Cambridge University Press, 92–114.
- Miles, M. B. & A. M. Huberman 1984. *Qualitative data analysis. A sourcebook of new methods*. Beverly Hills: Sage Publications.
- Moss, P. 1994. Can there be validity without reliability? *Educational Research*, 23, 2, 5–12.
- Mäkelä, J. 1995. Tilastollisen päättelyn, kvantitatiivisen menetelmän ja sosiaalitytökimuksen suhteesta. Teoksessa J. Eskola, J. Mäkelä & J. Suoranta (toim.) *Ihmistieteiden 1990-luvun metodologiaa etsimässä. Keskustelua kasvatus- ja sosiaalitytöiden 1990-luvun metodologiasta*. Lapin yliopiston kasvatustieteellisen julkaisuja. C 8 Katsauksia ja puheenvuoroja. Aikuiskasvatuksen ja jatkuvan koulutuksen julkaisuja 2. Rovaniemi: Lapin yliopisto, 59–73.
- Mäntysaari, M. 1999. Arviointitutkimuksen taustaoletukset. Teoksessa R. Eräsaari, T. Lindqvist, M. Mäntysaari & M. Rajavaara (toim.) *Arviointi ja asiantuntijuus*. Helsinki: Gaudeamus.
- Nation, P. 1990. *Teaching and learning vocabulary*. New York: Newbury House Publishers.
- Nation, P. & J. Newton 1997. Teaching vocabulary. Teoksessa J. Coady & T. Huckin (toim.) *Second language vocabulary acquisition*. Cambridge:

- Cambridge University Press, 238—254.
- Newkirk, T. 1982. How competent are the writing competency tests? Teoksessa J. C. Raymond (toim.) *Literacy as a human problem*. Alabama: The University of Alabama Press, 99–117.
- Niiniluoto, I. 1994. *Järki, arvot ja välineet*. Helsinki: Otava.
- Niskanen, K. 1995. Kvalitatiivisen ja kvantitatiivisen metodin yhteiskäyttö – onko se mahdollista? Teoksessa J. Eskola, J. Mäkelä & J. Suoranta (toim.) *Ihmistieteiden 1990-luvun metodologiaa etsimässä. Keskustelua kasvatustieteiden ja sosiaalitieteiden 1990-luvun metodologiasta*. Lapin yliopiston kasvatustieteellisen julkaisuja. C 8 Katsauksia ja puheenvuoroja. Aikuiskasvatuksen ja jatkuvan koulutuksen julkaisuja 2. Rovaniemi: Lapin yliopisto, 259–270.
- Norris, N. 1995. Koulutusohjelmien arviointi. Teoksessa S. Takala (toim.) *Arviointi ja koulutuksen laadun kehittäminen*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, 33—38.
- Norton, B. 1997. Accountability in language assessment. Teoksessa C. Clapham & D. Corson (toim.) *Language testing and assessment 7. The encyclopedia of language and education*. Dordrecht: Kluwer Academic, 313–322.
- Nystrand, M. 1986. *The structure of written communication. Studies in reciprocity between writers and readers*. Orlando: Academic Press, INC.
- Nystrand, M. 1987. The role of context in written communication. Teoksessa R. Horowitz & S. J. Samuels (toim.) *Comprehending oral and written language*. London: Academic Press Inc., 197–214.
- Nystrand, M. 1990. Sharing word: the effects of readers on developing writers. *Written Communication* 7, 13–24.
- Paribakht, T. S. & M. Wesche 1997. Vocabulary enchainment activities and reading for meaning in second language vocabulary acquisition. Teoksessa J. Coady & T. Huckin (toim.) *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 174–200.
- Passman, R. 2001. Experiences with student-centered teaching and learning in high-stakes assessment environments. *Education* 122, 1, 189–200.
- Patton, M. Q. 1990. *Qualitative evaluation and research methods*. Newbury Park: Sage Publications.
- Penny, J., R. L. Johnson & B. Gordon 2001. The effect of rating argumentation on inter-rater reliability. *Assessing Writing* 7, 2, 143–164.
- Pietikäinen, S. 2000a. *Discourses of differentiation. Ethnic representations in newspaper texts*. Jyväskylä: Jyväskylän yliopisto.
- Pietikäinen, S. 2000b. Kriittinen diskurssintutkimus. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kieli, diskurssi & yhteisö*. Soveltavan kielentutkimuksen teoriaa ja käytäntöä. Jyväskylä: Soveltavan kielentutkimuksen keskus, 191–217.
- Piirainen-Marsh, A. & A. Huhta 2000. Soveltava kielentutkimus ja tutkimuksen peruskysymykset. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Näkökulmia soveltavaan kielentutkimukseen*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 67–120.
- Pitkänen-Huhta, A. 1999. Vieraalla kielellä lukemisen tutkimus. Teoksessa K.

- Sajavaara & A. Piirainen-Marsh (toim.) *Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja käytäntöä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 259–288.
- Polio, C. 2001. Research methodology in second language writing research: the case of text-based studies. Teoksessa T. Silva & P. K. Matsuda (toim.) *On second language writing*. Mahwah: Lawrence Erlbaum Associates, 91–115.
- Pollitt, A. 1991. Response to Charles's Alderson's paper: "Bands and scores". Teoksessa J. C. Alderson & B. North (toim.) *Language testing in the 1990s*. London: Modern English Publications, McMillan, 87–94.
- Pollitt, A. & N. L. Murray 1996. What raters really pay attention to? Teoksessa M. Milanovic & N. Saville (toim.) *Performance testing, cognition and assessment: selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem*. Cambridge: Cambridge University Press, 74–91.
- Potter, J. & M. Weherell 1995. Discourse Analysis. Teoksessa . A. Smith, R. Harré & L. Van Langenhove (toim.) *Rethinking methods in psychology*. London: SAGE publications, 143–161.
- Powers, D. E., M. E. Fowles, M. Farnum & P. Ramsey 1994. Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement* 31, 253–233.
- Price, E. & S. Takala 1988. The narrative writing task. Teoksessa T. P. Gorman, A. C. Purves & R. E. Degenhart (toim.) *The IEA study of written composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press, 136–149.
- Purves, A. C. 1988. Introduction. Teoksessa A. C. Purves (toim.) *Writing across languages and cultures. Issues in contrastive rhetoric*. Newbury Park: SAGE Publications, 9–21.
- Purves, A.C. 1990. *The scribal society. An essay on literacy and schooling in the information age*. New York: Longman.
- Purves, A. C., T. P. Gorman & S. Takala 1988. The development of the scoring scheme and scale. Teoksessa T. P. Gorman, A. C. Purves & R. E. Degenhart (toim.) *The IEA study of written composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press, 41–58.
- Pälli, P. & S. Latomaa 1997. *Aikuisten maahanmuuttajien suomen kielen taito. Maahanmuuttajakoulutuksen arviointia*. Helsinki: Opetushallitus.
- Quellmaltz, E. S., F. Capell & C. P. Chou 1982. Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement* 19, 241–258.
- Quellmaltz, E. S. 1984. Toward successful large-scale writing assessment: where are we now? Where did we go from here? *Educational Measurement: Issues and Practice, Spring*, 29–32.
- Rafoth, B. A. 1988. Discourse community: Where writers, readers, and texts come together. Teoksessa B. A. Bennett & D. L. Rubin (toim.) *The social construction of written communication*. Norwood: Ablex Publishing Corporation, 131–146.

- Rafoth, B. A. & D. L. Rubin 1984. The impact of content and mechanics of judgements of writing quality. *Written Communication* 1, 4, 446–458.
- Raimes, A. 1985. What unskilled ESL students do as they write: a classroom study of composing. *TESOL QUARTERLY*, 19, 2, 229–256.
- Raimes, A. 1991. Out of the woods: emerging traditions in the teaching of writing. *TESOL QUARTERLY* 25, 3, 407–430.
- Raimes, A. 1998. Teaching writing. *Annual Review of Applied Linguistics* 18, 142–167.
- Raippamaa, O. 1965. Ymmärrys. http://www.pcuf.fi/~msiivola/para2000/artikkelit/orvon_index.htm
- Ramanathan, V. & R. B. Kaplan 2000. Genres, discourse communities: theory and application for (L1 and) L2 writing instructors. *Journal of Second Language Writing* 9, 2, 171–191.
- Raunio, K. 1999. *Positiivismi ja ihmistiede. Sosiaalitutkimuksen perustat ja käytännöt*. Helsinki: Gaudeamus.
- Rauste-von Wright, M-L & J. von Wright 1994. *Oppiminen ja koulutus*. Porvoo: WSOY.
- Reid, J. 1990. Responding to different topic types: a quantitative analysis from a contrastive rhetoric perspective. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. New York: Cambridge University Press, 191–210.
- Reid, J. 1993. Historical perspectives on writing and reading in the ESL classroom. Teoksessa J. G. Carson & I. Leki (toim.) *Reading in the composition classroom. Second language perspectives*. Boston: Heinle and Heinle Publishers, 33–60.
- Ross, K. N. Scientific research procedures. Teoksessa J. P. Keeves (toim.) *Educational research, methodology and measurement. An international handbook*. Oxford: Pergamon Press, 527–536.
- Rubin, D. L. 1988. Introduction: four dimensions of social construction in written communication. Teoksessa B. A. Bennett & D. L. Rubin (toim.) *The social construction of written communication*. Norwood: Ablex Publishing Corporation, 1–36.
- Rumelhart, D. E. 1977. *Introduction to human information processing*. New York.
- Ruth, L. & S. Murphy 1988. *Designing writing task for the assessment of writing*. New Jersey: Ablex Publishing Corporation.
- Saal, F. E., R. G. Downey & M. A. Lahey 1980. Rating the ratings: assessing the psychometric quality of rating data. *Psychological Bulletin* 88, 2, 413–428.
- Sajavaara, K. 1999. Kontrastiivinen kielentutkimus ja virheanalyysi. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja käytäntöä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, 103–128.
- Santos, T., D. Atkinson, M. Erickson, P. K. Matsuda & T. Silva 2000. On the future of second language writing: a colloquium. *Journal of Second Language Writing*, 9, 1, 1–20.
- Sarjala, J. 2000. Päätötkoe ei sovi peruskouluun. *Helsingin Sanomat* 12.8.2000.
- Sasaki, M. 2000. Toward an empirical model of EFL writing process: an

- exploratory study. *Journal of Second Language Writing* 9, 3, 259—291.
- Savage, J. 2000. Ethnography and health care. *British Medical Journal* 321, 2, 400—402.
- Sayer, A. 1992. *Method in social science. A realist approach*. London: Routledge.
- Schank, R. C. & R. P. Abelson 1977. Scripts, plans, and knowledge. Teoksessa P. N. Johnson-Laird & P. C. Wason (toim.) *Thinking. Readings in cognitive science*. Cambridge: Cambridge University Press, 421—432.
- Scribner, S. & M. Cole 1981. *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Scriven, M. 1991. *Evaluation thesaurus*. Newbury Park: Sage.
- Scriven, M. 1997. Truth and objectivity in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.) *Evaluation for the 21st century. A handbook*. Thousand Oaks: Sage Publications, 477—500.
- Segers, M. & F. Dochy 2001. New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in Higher Education*, 26, 3, 327—343.
- Selinker, L. 1992. *Rediscovering interlanguage*. London: Longman.
- Shaw, M. L. & B. R. Gaines 1992. Kelly's geometry of psychological space and its significance for cognitive modelling. *The New Psychologist*, October, 1—13.
- Shepard, L. A. 1993. The place of testing reform in educational reform: a reply to Cizek. *Educational Researcher* 22, 10—13.
- Shepard, L. A. 2000. The role of assessment in a learning culture. *Educational Research* 29, 7, 4—14.
- Shohamy, E., C. M. Gordon & R. Kraemer 1992. The effect of rater's background and training on the reliability of direct writing test. *The Modern Language Journal*, 76, 1, 27—33.
- Shohamy, E. 1993. The power of tests. The impact of language tests on teaching and learning. *Washington DC: NFLC Occasional papers*.
- Shohamy, E. 1997. Testing methods, testing consequences: are they ethical? Are they fair. *Language Testing* 14, 3, 340—349.
- Shohamy, E. 2001a. *The power of tests. A critical perspective on the uses of language tests*. Essex: Pearson Educated Limited.
- Shohamy, E. 2001b. Democratic assessment as an alternative. *Language Testing* 18, 4, 373—391.
- Silva, T. 1990. Second language composition instruction: developments, issues, and directions in ESL. Teoksessa B. Kroll (toim.) *Second language writing. Research insights for the classroom*. New York: Cambridge University Press, 11—23.
- Silva, T. 1993. Toward an understanding of the distinct nature of L2 writing: the ESL research and its implications. *TESOL QUARTERLY* 27, 4, 657—677.
- Silvonen, J. 1999. ATLAS/ti laadullisen analyysin välineenä. *Psykologia* 34, 398—404.
- Simoila, R. 1993. *Mallit, kontaktit, tiimit ja verkot: välineitä terveystieteiden analysointiin*. Helsinki: Sosiaali- ja terveystieteiden tutkimus- ja kehittämiskeskus VAPK.

- Smith, D. 2000. Rater judgement in the direct assessment of competency-based second language writing ability. Teoksessa G. Brindley (toim.) *Studies in immigrant English language assessment*. Volume 1. Research series 11. Sidney: Macquarie University, 159–189.
- Smith, J. A. 1995. Semi-structured interviewing and qualitative analysis. Teoksessa J. A. Smith, R. Harré & L. Van Langenhove (toim.) *Rethinking methods in psychology*. London: SAGE publications, 9–26.
- Solas, J. 1992. Investigating teacher and student thinking about the process of teaching and learning using autobiography and repertory grid. *Review of Educational Research* 62, 2, 205–225.
- Spack, R. 1997. The acquisition of academic literacy in a second language. *Written Communication*, 14, 1, 3–62.
- Sperber, D. & D. Wilson 1986. *Relevance: communication and cognition*. Oxford: Blackwell.
- Stansfield, C. W. & J. Ross 1988. A long-term research agenda for the Test of Written English. *Language Testing*, 5, 2, 160–186.
- Stewart, V. 2000a. Understanding George Kelly and personal construct theory. <http://www.enquirewithin.co.nz/skills2.htm>
- Stewart, V. 2000b. Teaching the skills of repertory grid interviewing and analysis. <http://www.EnquireWithin.co.nz/business.htm>
- Stotsky, S. 1987. A comparison of the two theories about development in written language: implications for pedagogy and research. Teoksessa R. Horowitz & S. J. Samuels (toim.) *Comprehending oral and written language*. London: Academic press Inc., 371–395.
- Street, B. V. 1994. Struggles over the meaning(s) of literacy. Teoksessa M. Hamilton, D. Barton & R. Ivancic (toim.) *Words of literacy*. Clevedon: Multilingual Matters Ltd, 15–20.
- Suni, M. 1993. *Maahanmuuttajapoikien suomen kielen taito peruskoulun päättövaiheessa*. Moniste 11/1996. Helsinki: Opetushallitus.
- Suomen kielen perussanakirja*, 1990. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Ensimmäinen osa A–K. Helsinki: Valtion painatuskeskus.
- Suoninen, E. 1993. Mistä on perheenäidit tehty? Haastattelupuheen analyysi. Teoksessa A. Jokinen, K. Juhila & E. Suoninen (toim.) *Diskurssianalyysin aakkoset*. Tampere: Vastapaino, 111–150.
- Suoranta, J. 1993. Metodologinen fundamentalismi ja sen viholliset. *Kasvatus* 24, 1, 38–46.
- Swartz, C. W., S. T. Hooper, J. W. Montgomery, M. B. Wakely, R. E. L. De Kruif, M. Reed, T. T. Brown, M. D. Levine & K. P. Kinnard 1999. Using generalizability theory to estimate reliability of writing scores derived from holistic and analytical scoring methods. *Educational & Psychological Measurement*, 59, 3, 492–507.
- Syrjälä, L. & M. Numminen 1988. *Tapaustutkimus kasvatustieteessä*. Oulun yliopiston kasvatustieteiden tiedekunnan tutkimuksia. Oulu: Oulun yliopiston kasvatustieteiden tiedekunta.
- Syrjälä, L., S. Ahonen, E. Syrjäläinen & S. Saari 1994. *Laadullisen tutkimuksen työtapoja*. Helsinki: Kirjayhtymä Oy.

- Takala, S. 1985. Kirjoittamisen käsitteellistä hahmotusta. *Kasvatus. Suomen kasvatustieteellinen aikakauskirja* 16, 2, 93–100.
- Takala, S. 1989. Sanaston opettamisen uudet haasteet. Teoksessa S. Takala (toim.) *Sanaston opettaminen ja oppiminen*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Jyväskylä: Jyväskylän yliopisto, 1–11.
- Takala, S. 1994. Arviointi – ongelma ja mahdollisuus. Teoksessa P. Linnakylä, P. Pollari & S. Takala (toim.) *Portfolio arvioinnin ja oppimisen tukena*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 1–8.
- Takala, S. 1996. Suoritusarviointi puntarissa. Mahdollisuuksia ja ongelmia valtakunnallisissa kokeissa. Teoksessa P. Pollari, M. Kankaanranta & P. Linnakylä (toim.) *Portfolion monet mahdollisuudet*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, 207–222.
- Takala, S. 1997. Vieraan kielen taidon kehittymisen arviointiperusteita peruskoulussa ja lukiossa. Teoksessa R. Jakku-Sihvonen (toim.) *Onnistuuko oppiminen*. Helsinki: Opetushallitus, 85–112.
- Tarnanen, M. & S. Luoma 2001. Kokeilu kirjoittamisen itsearviointista tietokoneella. Teoksessa M-L. Charles & P. Hiidenmaa (toim.) *Tietotyön yhteiskunta – kielen valtakunta*. AFinLAn vuosikirja n:o 59. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 132–149.
- Tashakkori, A. & C. Teddlie 1998. *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks: Sage Publications.
- Tesch, R. 1990. *Qualitative research. Analysis types & software tools*. New York: The Falmer Press.
- Thaiss, C. & T. M. Zawicki 1997. How portfolios for proficiency help shape a WAC program. Teoksessa K. B. Yancey & B. Huot (toim.) *Assessing writing across the curriculum. Diverse approaches and practices*. Greenwich: Ablex Publishing Corporation, 79–96.
- Thorndike, R. L. & E. Hagen 1977. *Measurement and evaluation in psychology and education*. New York: John Wiley & Sons.
- Thorndike, R. L. 1988. Reliability. Teoksessa J. P. Keeves (toim.) *Educational research, methodology and measurement: an international handbook*. Oxford: Pergamon Press, 330–343.
- Traub, R. E. & G. L. Rowley 1990. *An NCME instructional module on understanding reliability. ITEMS: Instructional topics in educational measurement*. National Council on Measurement in Education.
- Turunen, K. E. 1997. *Halut, arvot ja valta*. Jyväskylä: Gummerus.
- Tyndal, B. & D. M. Kenyon 1996. Validation on a new holistic rating scale using Rasch multi-faceted analysis. Teoksessa A. Cumming & R. Berwick (toim.) *Validation in language testing*. Clevedon: Multilingual Matters LTD, 39–57.
- Töttö, P. 1999. *Pirullinen positioivismi: kysymyksiä laadulliselle tutkimukselle*. JYJ julkaisusarja n:o 41. Jyväskylä: Kampus Kustannus.
- Upshur, J.A. & C. E. Turner 1999. Systematic effects in the rating of second-language speaking ability: test method and learner discourse. *Language Testing*, 16 (1), 82–111.
- Varto, J. 1992. *Laadullisen tutkimuksen metodologia*. Helsinki: Kirjayhtymä.

- Vaughan, C. 1991. Holistic assessment: What goes on in the rater's mind? Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation, 111–126.
- Veal, L. R. & S. A. Hudson 1983. Direct and indirect measures for large-scale evaluation of writing. *Research in the Teaching of English* 17, 285–296.
- Venezky, R. L. 1990. Definition of literacy. Teoksessa R.L. Venezky, D.A. Wagner & B. S. Ciliberti (toim.) *Toward defining literacy*. Newark: International Reading Association, 2–16.
- Ventola, E. & A. Mauranen 1996. *Academic writing: intercultural and textual issues*. Amsterdam: Benjamin.
- Vidich, A. J. & A. M. Lyman 1994. Qualitative methods. Their history in sociology and anthropology. Teoksessa N. K. Denzin & Y. S. Lincoln (toim.) *Handbook of qualitative research*. Thousand Oaks: Sage Publications, 23–59.
- Virta, A. 1999. *Uudistuva oppimisen arviointi. Mahdollisuuksia ja varauksia*. Turun yliopiston kasvatustieteiden tiedekunta. Julkaisusarja B:65. Turku: Turun opettajankoulutuslaitos.
- Väljjarvi, J. 1996. Arvioimalla uuteen kouluun. *Kasvatus* 4, 317–319.
- Waltman, K., A. Kahn & G. Koency 1998. *Alternative approaches to scoring: The effects of using different scoring methods on the validity of scores from a performance assessments*. CSE Technical Report 448. Los Angeles: University of California.
- Weigle, S. C. 1994. Effects of training on raters of ESL compositions. *Language Testing* 11, 2, 197–223.
- Weir, C. J. 1990. *Communicative language testing*. London: Prentice Hall.
- Weir, C. J. 1993. *Understanding & developing language tests*. Hertfordshire: Prentice Hall International.
- White, E. M. 1986. *Teaching and assessing writing. Recent advances in understanding, evaluating, and improving student performance*. San Francisco: Jossey-Bass Publishers.
- Wiggins, G. 1998. *Educative assessment. Designing assessment to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.
- Williams, J. D. 1993. Rule-governed approaches to language and composition. *Written Communication* 10, 4, 542–568.
- Wolfe, E. W. & C-W. Kao 1998. Cognitive differences in proficient and non-proficient essay scorers. *Written Communication* 15, 4, 465–293.
- Yleisten kielitutkintojen perusteet* 1995. Helsinki: Opetushallitus.
- Zamel, V. 1983. The composition process of advanced ESL students: six case studies. *TESOL QUARTERLY* 17, 2, 165–187.

KIRJOITTAMISEN ARVIOINTIASTEIKKO

TASO	
C2/6	Osaa kirjoittaa selkeitä ja sujuvia tyyliältään tilanteeseen sopivia tekstejä. Osaa kirjoittaa monimutkaisia kirjoituksia, raportteja tai artikkeleita, joissa on lukijaa helpottava looginen, tekstin pääasiaa korostava rakenne. Osaa kirjoittaa tiivistelmiä ja katsauksia ammatillisista aiheista. Tekstissä ajatukset on muotoiltu tarkasti ja yksiselitteisesti. Tekstissä ei ilmene lainkaan merkkejä siitä, etteikö kirjoittaja pystyisi ilmaisemaan sitä, mitä hän haluaa ilmaista.
C1/5	Osaa ilmaista itseään selkeästi ja rakenteellisesti hyvin. Osaa kirjoittaa erilaisia tyyllillisesti ja muodollisesti hallittuja tekstejä, jotka on osoitettu tietyille lukijalle. Osaa kirjoittaa mm. yksityiskohtaisesti kuvailevan kirjeen monimutkaisesta aiheesta sekä omia näkökulmia korostavan esseen tai raportin. Kirjoittaja pystyy ilmaisemaan sanottavansa ilman rajoituksia.
B2/4	Osaa kirjoittaa selkeitä ja yksityiskohtaisia tekstejä monenlaisista henkilökohtaisesti kiinnostavista aiheista. Osaa kirjoittaa mm. informoivan tai argumentoivan esseen tai raportin. Osaa kirjoittaa tekstejä, esim. kirjeen itselleen merkittävistä tapahtumista ja kokemuksista. Pystyy ilmaisemaan itseään ilman huomattavia, esim. sanastollisia ja rakenteellisia puutteita.
B1/3	Osaa kirjoittaa yksinkertaisia mutta sidosteisia tekstejä tutuista tai henkilökohtaisesti kiinnostavista aiheista. Osaa kirjoittaa mm. henkilökohtaisen kirjeen omista kokemuksistaan ja tuntemuksistaan. Kielitaito riittää kuvailemaan yllättäviä tilanteita, selittämään pääidean tai ongelman abstraktisesta tai kulttuuriaiheesta, esim. musiikista tai elokuvista sekä kertomaan itseään läheisistä aiheista, esim. perheestä, harrastuksista tai työstä, joskin esim. sanastolliset puutteet aiheuttavat toistoa.
A2/2	Osaa kirjoittaa lyhyitä ja yksinkertaisia tekstejä, esim. kirjelappuja ja viestejä jokapäiväisistä asioista sekä hyvin yksinkertaisen henkilökohtaisen kirjeen, esim. kiitoskirjeen. Teksti koostuu ainakin osin yksinkertaisista lauseiden ja fraasien jaksoista, joita yhdistävät tavallisimmat konjunktiot: <i>ja, mutta, koska</i> , joskin väärinymmärrykset ja lukurytmin katkeaminen ovat tavallisia.
A1/1	Osaa kirjoittaa hyvin lyhyitä ja yksinkertaisia tekstejä, esimerkiksi postikortin ja/tai täyttää lomakkeeseen itseään koskevat perustiedot. Teksti koostuu toisistaan irrallisista yksinkertaisista, ehkä ulkoa opituista sanaryhmistä, lauseista ja tavallisimmista fraaseista.

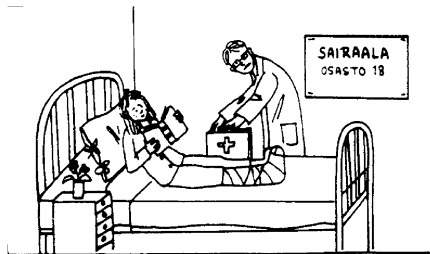
KIRJOITUSTEHTÄVÄT

Tehtävä X/ Tarina

Katso alla olevaa kuvaa. Kirjoita lyhyt tarina, jossa kerrot,

- mitä on tapahtunut ennen kuvaa
- mitä kuvassa tapahtuu
- mitä voisi tapahtua sen jälkeen.

Huom. Voit kertoa ihan mitä haluat, kunhan käytät annettua kuvaa tarinassasi.



TA1 / Ostoslistaviesti

Suomalainen ystäväsi on luonasi kylässä. Hän haluaa käydä kaupassa, kun sinä olet työssä. Kirjoitat hänelle lyhyen viestin, jossa on ostoslista. Kerro viestissä,

- mitä hän ostaa (6 asiaa)
- kuinka paljon.

TA2 / Kirje ystävälle

Kirjoitat kirjeen suomalaiselle ystävällesi. Kerro kirjeessä seuraavat asiat:

- Mitä sinulle kuuluu?
- Miten sinun perheesi voi?
- Millainen sää on?
- Mitä teet lomalla, joka alkaa ensi viikolla?

TB1 / Ajo-ohjeet

Olet muuttanut uuteen asuntoon. Suomalainen ystäväsi tulee ensimmäistä kertaa käymään sinun luonasi. Hän ei tiedä, missä sinä asut. Sinulla ei ole karttaa, joten kirjoitat hänelle ohjeet, miten hän löytää asutosi, kun hän tulee

bussipysäkillä. Kerro ohjeessa,

- miten hän kävelee bussipysäkillä
- mikä auttaa häntä löytämään kotisi
- millainen ympäristö on
- millaisessa talossa asut.

TB2 / Kirje

Luet lehdestä ilmoituksen, jossa etsitään kirjeystävää. Päättät vastata ilmoitukseen ja kirjoitat kirjeen hänelle. Kerro kirjeessäsi,

- kuka olet
- miksi kirjoitat
- millainen olet ja mistä pidät

Muista kirjeen aloitus ja lopetus.

Hei sinä nuori tai vanha, nainen tai mies, joka pidät kirjeiden kirjoittamisesta. Kirjoita samanlaiselle naiselle Pohjois-Suomeen. Pidän eläimistä, luonnossa liikkumisesta ja ennen kaikkea keskustelemisesta. Olen myös matkustellut paljon, mutta nyt haluaisin olla vähän aikaa aloillani täällä Suomessa. Minusta on mielenkiintoista tutustua uusiin ihmisiin kirjeitse.

TB3 / Lyhyt esittely

Sinua on pyydetty puhumaan omasta kulttuuristasi suomalaisille kuulijoille. Haluat lähettää heille etukäteen lyhyen esittelyn, jossa kerrot muutamia tärkeitä asioita kulttuuristasi. Kirjoita lyhyt esittely, jossa

- kerrot, mikä kulttuurissasi voisi olla kiinnostavaa suomalaisesta, anna
2 - 3 esimerkkiä
- kuvaillet ja esittelet valitsemiasi esimerkkejä
- selität, miksi valitsemasi esimerkit ovat kiinnostavia.

TC1 / Mieli-pidekirjoitus

Haluat tuoda esiin jonkin sinua häiritsevän epäkohdan. Päättät kirjoittaa jonkin lehden yleisönosastoon aiheesta. Esitä kirjoituksessasi selkeästi seuraavia asioita:

- epäkohta, josta kirjoitat
- miksi asia ei ole kunnossa
- mitä asia voi aiheuttaa
- mihin asia vaikuttaa
- miten asiaa voisi muuttaa (parannusehdotus)
- miksi juuri sinun ehdotuksesi on hyvä.

TC2 / Suruvalittelu

Ystäväsi äiti on kuollut vähän aikaa sitten. Olet poissa kotoa ulkomailla ja kuulet asian puhelimesta. Ystäväsi ei jaksanut jutella kanssasi juuri nyt, mutta haluat auttaa häntä surun keskellä. Kirjoitat hänelle kirjeen, jossa lohdutat häntä. Kirjoita hänelle,

- mitä ajattelet tapahtuneesta
- mitä ajattelet kuolemasta tällaisena hetkenä
- jokin positiivinen asia, joka voi lohduttaa häntä
- kuinka hän voi jaksaa eteenpäin
- tarjoa apuasi.

TC3 / Työpaikkahakemus

Lue oheinen työpaikkahakemus. Kirjoita sitten noin yhden sivun mittainen hakemus valitsemaasi työpaikkaan. Käsittele hakemuksessa ilmoituksessa pyydettyjä asioita.

Makumaa-ravintola on suomalainen ravintolaketju, jonka ravintoloita on Helsingissä, Turussa, Tampereella, Kuopiossa, Oulussa ja Rovaniemellä.

Jos olet energinen, osaava ja sopeutuva, saatat sopia kasvavaan joukkoomme, joka etsii seuraavia uusia työntekijöitä:

- **ravintolapäällikkö**
- **tarjoilijoita**
- **kokkeja**
- **siistijöitä**

Tarjoamme viihtyisän ja osaavan työympäristön sekä koulutuksen työhön. Alan koulutus ja aikaisempi työkokemus on eduksi.

Kirjalliset hakemukset 3.3. mennessä osoitteella Maakatu 4, 842841 Helsinki.

KIRJOITTAJIEN TAUSTATIELOMAKE

Ole ystävällinen ja vastaa seuraaviin kysymyksiin, ennen kuin teet tehtävät.

Nimi _____

nainen mies

Minkä ikäinen olet? alle 20 21 - 30 31 - 40 41 - 55 yli 55

Mikä/mitkä on äidinkielenesi? _____

Mitä vieraita kieliä osaat? _____

Mikä on koulutuksesi?

Ei koulutusta Peruskoulu tai sitä vastaava koulutus Ammattikoulutus

Korkeakoulututkinto Muu, mikä _____

Kuinka kauan olet asunut Suomessa? _____

Kuinka monta tuntia olet opiskellut suomea? Laita rasti (X) sen tuntimäärän kohdalle, joka on lähellä sinun opiskeluaikaasi.

- noin 100 tuntia tai vähemmän
- noin 200 tuntia
- noin 400 tuntia
- noin 800 tuntia
- noin 1000 - 1200 tuntia
- enemmän kuin 2000 tuntia

Miten usein käytät suomen kieltä seuraavissa tilanteissa?

- Ystävien kanssa
 - lähes aina
 - joskus
 - ei juuri koskaan
- Työssä
 - lähes aina
 - joskus
 - ei juuri koskaan
- Asioita hoitaessa, esim. pankki, posti
 - lähes aina
 - joskus
 - ei juuri koskaan

Arvioi omaa suomen kirjoittamisen taitoasi seuraavan taulukon mukaan. Laita rasti (X) siihen kohtaan, joka kuvaa hyvin sinun kirjoittamisen taitoasi.

Laita rasti	OSAAN KIRJOITTA...
	Osaan kirjoittaa itsestäni tietoja, esim. oma nimi, osoite, kansallisuus sekä hyvin lyhyitä ja yksinkertaisia viestejä, esimerkiksi postikortin.
	Osaan kirjoittaa lyhyitä viestejä, esim. lyhyen kirjeen ystävälle sekä kirjoittaa itsestäni, esimerkiksi perheestäni, harrastuksistani tai siitä mistä pidän ja en pidä.
	Osaan kirjoittaa erilaisia lyhyitä tekstejä tutuista aiheista, esimerkiksi kirjoittaa lyhyen ilmoituksen tai kertoa kirjeessä kokemuksistani tai mielipiteistäni.
	Osaan kirjoittaa erilaisia ja eripituisia tekstejä erilaisista aiheista, esimerkiksi kirjeen, mielipidetekstin tai muun lyhyen tekstin, jossa osaan esimerkiksi sekä esittää että perustella mielipiteeni tai kuvata selkeästi erilaisia tapahtumia.
	Osaan kirjoittaa erilaisia ja eripituisia tekstejä vaikeista ja monimutkaisistakin aiheista, esimerkiksi kirjoittaa virallisia kirjeitä, laajoja ja perusteltuja raportteja.
	Osaan kirjoittaa erityyppisiä, erilaisiin tilanteisiin täysin sopivia tekstejä hyvin vaikeistakin ja minulle uusista aiheista, esimerkiksi erityyppisiä virallisia tekstejä.

ARVIOIJIEN TAUSTATIETOLOMAKE

Nimi _____

Sukupuoli nainen mies

Minkä ikäinen olet? alle 20 21 - 30 31 - 40 41 - 55 yli 55

Mikä on äidinkieleni? suomi ruotsi muu, mikä? _____

Mikä on koulutuksesi?

peruskoulu tai sitä vastaava koulutus

ammattillinen kouluaste

ammattillinen opistoaste

ammattikorkeakoulu

yliopisto suomen kielen opintoja kyllä appr. cl l
ei

muu mikä? _____

Millaista kokemusta sinulla on suomesta vieraana/toisena kielenä?

ei kokemusta

tutkimuskokemusta alle vuoden 1 - 2 v. 3 - 4 v. 5 v. tai enemmän

opetuskokemusta millä tasolla?

alkeistaso alle vuoden 1 - 2 v. 3 - 4 v. 5 v. tai enemmän

keskitaso alle vuoden 1 - 2 v. 3 - 4 v. 5 v. tai enemmän

edistyneet alle vuoden 1 - 2 v. 3 - 4 v. 5 v. tai enemmän

muuta kokemusta mitä ja missä? _____

Kuinka paljon olet arvioinut kirjoittamista?

en lainkaan jonkin verran, satunnaisesti paljon, säännöllisesti

Millä tavalla olet arvioinut kirjoittamista?

kirjaimin ilman sanallisia kuvauksia missä? _____

kirjaimin, jotka kuvattu sanallisesti missä? _____

pistein ilman sanallisia kuvauksia missä? _____

pistein, jotka kuvattu sanallisesti missä? _____

muulla tavalla miten ja missä? _____

ESIMERKKI HAASTATTELUAINEISTON ANALYYSISTA

Enemmän haastattelu (samat vaiheet koskevat myös toista haastattelua)

VAIHE 1.

Ennen haastattelujen litterointia kuuntelin ensimmäisen ja toisen kierroksen haastattelut ja kirjoitin kullekin haastateltavalle profiilin haastattelun perusteella.

Esimerkki ensimmäisen haastattelun profiilien tiivistelmästä:

Tiinan haastattelu tehtiin Kotkassa hänen työpaikallaan. Tiina oli vähän hermostunut tullessaan haastatteluun, mutta rentoutuu kuitenkin haastattelun alkaessa. Tiina puhuu hyvin vuolaasti ja helposti pitkiäkin aikoja ”oma-aloitteisesti”. Tiina kuvaa ensin asteikkoaan varmemmin, mutta sitten epäilee vähän enemmän. Sitä ymmärtää mitä vaan jos haluaa, vaikka käsialojakin. Hän tuntuu olevan aika innostunut, joskin myös reflektoi itseään, esim. hän ajattelee olevansa vähän puolustuskannalla, koska hänestä kielioppi on niin tärkeää.

Tiivistelmä:

”Itku kun ei oo mitään kriteereitä, tää on vaan musta tuntuu” Tiina aloitti lukemaan ensin ja luki alussa paperit kahteen kertaan mutta sitten vain yhteen. Kuutonen oli Tiinan mielestä 100-prosenttisesti oikein kuten suomalaisen. Alussa Tiinan mielestä kolmonen ja nelonen oli vaikea erottaa toisistaan mutta sitten hän päätti verbit on oltava nelosessa oikein. Kriteerit oli Tiinalla vain päässä. Kriteerit muuttu lopussa, koska siellä oli parempia.

Kriteerien puuttuessa tekstejä tuli enemmän vertailtua toisiinsa. Tiina arvioi suoritukset kolmessa erässä eikun neljässä. Tiinaa kauhistutti, että onko kaikki niin huonoja, koska hän odotti, että koko skaala on käytössä.

Kopiot oli Tiinan mielestä hyviä, mutta ongelma oli ohjeiden puuttuminen eli kriteerien puuttuminen. Ja se annetaanko kaikille jotakin, vaikka olisikin kuinka lyhyt ja huono. Niukastakin tekstistä on voinut saada ihan hyvän, jos tekstistä näkee, että kirjoittaja ninku kirjoittaa ihan hyvä. Alussa on saattanut olla vaikea hahmottaa kakkosta. Tiina painottaa omasta mielestään kieliooppia paljon, esim. verbit ja rekiot ovat tärkeitä. Vitosta ei voi saada, jos perusjutuissa on virheitä. Tiinan on tottunut paljon lukemaan erilaisia käsialoja, vain pari vaikeasti luettavaa ja niihin jaksoi paneutua, koska niitä oli niin vähän.

Tiinan mielestä yksi on hyvä, koska siinä on kriteerit ettei tartte tekstejä toisiinsa, kurssilla Tiina ei anna arvioita paperiin, koska ei oo vielä kriteerejä. Tiina laittaa korjatessaan merkin, missä on virhe. Takaisin antaessa harjoitellaan niitä asioita, joissa on eniten virheitä. Tiina painottaa paljon kieliooppia, koska opiskelijat ei törmää siihen oikeastaan missään muualla. Esim. kukaan ei pysty selittämään opiskelijoille, mikä on mm. passiivi. Tiinalla on tunne, että hän puolustautuu ikään kuin hän olisi vanhanaikainen ja kommunikatiivisuutta vastaan.

Tiinan mielestä kielioppi on hyvä kriteeri, koska harva kirjoittaa kieliopillisesti täyttä soopaa. Yleensä soopaa kirjoitetaan huonoilla rakenteilla. Toisaalta voihan niitä olla sellaisiakin, joissa kielioppi on oikein ja siiti soopaa. Massa oli vaikeaa mutta ei mahdotonta eikä rankkaa.

VAIHE 2.

Litteroin haastattelut. Pidin litteroinnin aikana nousseista ajatuksista eräänlaista muistikirjaa, josta esimerkkinä seuraava tekstikatkelma:

- ensimmäinen haastattelu antaa tilaisuuden puretua opettajien omaan kielitaitokäsityksen, miten he hahmottavat kirjoittamisen taidon, lähinnä sen mukaan millaisia kriteerejä he käyttävät
- mielenkiintoista on se, miten annettu asteikko eroaa heidän käsityksistään ja miten ensimmäisen asteikon aikana esitetyt käsitykset eroavat/ovat samanlaisia toisesta asteikosta esitetyn kritiikin kanssa
- Tuulan kommentti massa-arvioinnista tuo kivasti esille sen, miten emootioista voi tulla tietoisemmaksi verrattuna pienen määrän arviointiin, jota tekee yleensä saman mielentilan vallitessa
- moni opettaja kuvaa sitä, miten selkeät kriteerit (rakennepohjaiset) kuvaavat eri tasoja, mutta ne eivät kuitenkaan tunnu olevan se ainoa mitta, koska jos kirjoitus on muuten ansiokas, ko. tasolta vaaditun kriteerin rikkominen ei pudotakaan tasoa alemmaksi -> ensivaikutelmaltaan voimakas rakennelähtöisyys ei olekaan niin rakennelähtöistä -> mikä on sitten itse asiassa opettajalle varmaa? Siis miten paljon kriteereistä on rakennetta siksi, koska se on niin selvä ja konkreettinen kriteeri verrattuna tekstuaalisuuteen ja muuhun vähän ei niin selvärajaiseen kriteeriin mutta joka kuitenkin vaikuttaa kirjoituksen kokonaisuuteen (oman havaintoni mukaanhan oli ihan käsittämättömän yllättävää, että niin huonoillakin rakenteilla saattoi kirjoittaa

Code: -kriittiki/kriteerien puute
Created: 14.09.99 08:34:43 by Super
Modified: 24.09.99 11:30:06

Code: -kriittikki/aika
Created: 15.09.99 08:15:03 by Super
Modified: 24.09.99 13:52:01

Code: -kriittikki/arviointi yleensa
Code: -kriittikki/kirjoittaja
Code: -kriittikki/oma työ
Code: -kriittikki/raha
Code: -kriittikki/tehtävien järjestys
Code: -kriittikki/tunne
Code: +-t/mielipide
Code: +/-kriittikki/selkeys
Code: +kriittikki/arviointinopeus
Code: +kriittikki/holistisuus
Code: +kriittikki/mielenkiintoisuus
Code: +kriittikki/oma työ
Code: +kriittikki/tasot
Code: +kriittikki/tunne
Code: aloitus/selailu
Code: teksti/kiva+hauska
Code: teksti/kopio huono
Code: teksti/kulttuuriseikat
Code: teksti/niukkuus
Code: teksti/ohi tehtävänannon
Code: teksti/puutteellinen sisältö
Code: teksti/tylsä
Code: teksti/tyyli
Code: tt-helppo/hakemus
Code: ratkaisu/virheiden rooli
Code: ratkaisu/ymmärtää
Code: skaala/dialang
Code: skaala/epävarmuus
Code: skaala/helppoa
Code: ratkaisu/niukkuus
Code: ratkaisu/ohi tehtävänannon

Code: ratkaisu/tunne
Code: ratkaisu/tyyli
Code: skaala/oma paa
Code: skaala/opetus
Code: skaala/opetustasot
Code: skaala/pienempi
Code: skaala/yki
Code: teksti/kasiala
Code: tt-helppo/kirje
Code: tt-helppo/survalittelu
Code: tt-helppo/tarina
Code: tt-vaikea/ajo-ohje
Code: tt-vaikea/hakemus/muoto
Code: tt-vaikea/kirje
Code: tunne/kiinnostava
Code: tuntemus/nakemys
Code: työ/vasymysCode: aloitus/suoraluku
Code: aloitus/tasojako
Code: arviointi/epavarmuus
Code: arviointi/epavarmuus
Code: arviointi/nollan puuttuminen
Code: arviointi/saman tasoisia
Code: arviointi/samat virheet
Code: arviointi/virheiden rooli/laatu
Code: arviointi/ kieli
Code: arvioija/ymmärtää
Code: asteikko/paa
Code: asteikko/paa+paperi
Code: asteikon kuvailu
Code: kirjoittajan tausta
Code: kokemus/kuvaukset
Code: kriittikki/tehtavat
Code: neutraali/kirjeenvaihtoilmoitus
Code: oma työ/ kriittikki
Code: ratkaisu/kasiala
Code: ratkaisu/kirjoittajan tausta
Code: ratkaisu/kopio huono

VAIHE 5.

Tarkastelin aineistoa koodattuina teemoina ja aloin hahmottaa niistä käsitysten verkostoa ensin lukemalla aineistoa teema-/koodikohtaisesti. Sitten pyrin selvittämään kokonaiskuvaa kategorioista kirjoittamalla niistä ensimmäisen tekstiversion, joka oli täysin kategoriapohjainen. Sen jälkeen pohdiskelin teemakategorioiden yhteyttä lähtöoletuksiin ja teoriaan, minkä pohjalta alkoi syntyä yleisempiä lähestymistapoja. Seuraavassa tekstiversiossa tarkastelinkin analysoidun aineiston sisältöjä yleisemmällä tasolla.

Esimerkki teema-/koodikohtaisesta luettelosta:

Code: kirjoittajan tausta {33-0}

Codes: [-kriittikki/kirjoittaja] [kirjoittajan tausta]

No ehk just ongelma oli niinku se nuo arabit toisaalta koska ne nyt tunnistaa tekstist ja muuta jotenki niinku sill tavalla ett et m mietin ett kolmone no ei nyt sitte no kakkone

Codes: [kirjoittajan tausta]

ni sitte tota sit siellä oli sellasia että jotka ei ollut vastannu että et siellä oli hirveen hieno lause että se oli varmasti venäinen just jotenki että m en nyt muista mit mut m annoin kakkosen sille vaikka se oli min en halua vastata tñ tai joku m en muista mik se oli mut kuitenkin että siellä oli joku tyytyläuse tysin oikein ja tleen nin ni m pistin kakkosen sille vaikka se ei ollut siihen koko tehtv

Codes: [kirjoittajan tausta] [teksti/niukuus]

sit oli jotain semmosia että mit ne nyt kirjotti et oli et oli vaan luetteltu niinku jotain et mit se oli se tehtv et mit omasta kotimaastasi haluat no venjll on pushkin ja tolstoj ja dostojevski ja ni ei se niinku kerro oikeestaan mitn

Codes: [-kriittikki/kirjoittaja] [kirjoittajan tausta]

noh noh se oli nyt vaan tehtv en m nyt osaa sanoa jollakin lailla siin vahvistu ennako odotukset että semmosii tietyn tasosii ihmisiin on hirveen paljon jotka osaa niinku kutakuinkin jotakin mut sit semmosii huippuja on hirveen vñ

Codes: [kirjoittajan tausta]

et tota se tietenki yks asia että et ku vñ on ollut tekemiss noitten suomea vieraana kielen opiskelevien kanssa nin se kuva mik ni suomesta niist vltty niin tietysti jollain lailla on painottunu varmaan mut tota aika synkk semmone vñ niinku vñ semmone ikv olo tuli vñll

Codes: [kirjoittajan tausta] [tuntemus/nakemys]

mietin sit monta kertaa että et kuinka niin sanotusti Insimainen tekstilaji esimerkiksi joku mielipidekirjoitus on et paljonko se nkyy ts että joku ei oo saanu tarpeeks tai saanu ja saanu mut hankkinu tarpeeks kokemusta siit siit lajista tm kulttuurin tavalla ja joissaki nkyy selvsti semmone kohteliaisuus esimerkiksi että niinku kaikki on hyvin mulla ei oo mitn valittamista et on oikeasta vñ tmmne ja tmmne asia mut loppujen lopuksi olen kiitollinen tai jotain tmmst että et se idea että on joku yks epkohta mihin ha haluaa hirveesti vaikuttaa ni tota ei se totetutunu monenkaan kohalla ollenkaan ja monesti tuli kyll mieleen että se ei tajaa olla nyt ihan pelkstm siit kirjottamisen taikka kielentajosta kyse

Codes: [kirjoittajan tausta]

no kyll siellä oli joitaki semmosia esimerkiksi just semmosia kohteliaita tapauksia jotka kasialan perusteella tulkitsein venäläiseksi että tuota et ne niinku ihan hirveen hyvin oli kirjottanu sen mit oli kirjottanu et ne oli semmosta viitosen kuutosen tasoa ehk mut sitten niinkun koska se tehtv oli jny tekemtt ni m vejin sinne kakkoseen esimerkiksi tai ykkseen koska jos siin oli sill lailla et oli siin oli vaan niinku luettu muutama muutama asia mut sitte kuitenkin vejetty matto niitten alta että no em m valittakaa nyt mitaan tasta asiasta ni tai mistaan ni tota sehan meni sitte ajatus pieleen

Codes: [kirjoittajan tausta] [teksti/tyls]

mutte enimmäkseen nehan on aika odotuksennukaisia aika tylsi tai semmosia noh rajallisella kielitaidolla niitä harvoja luo siellä oli hyvin vahan semmosia luovia ratkasuja mitka on sitte niinku semmosia tavallaan niinku sen lukijan palkintoja ehka ehka ne sisällöllisestiki aika tylsia

Codes: [kirjoittajan tausta] [tuntemus/nakemys]

eik se että m nn et t on venäinen t on arabipoika t on somalipoika t on t on noh ruottalainen ihan selvsti ni et m yritin niinku sit ensin niinku etten ota sit huomioon tmmssi omia noh sympatioita antipatioita

Codes: [kirjoittajan tausta] [arviointi/virheiden rooli/laatu] [ratkaisu/virheiden rooli]

nelosta ei voi venäläinen kirjottaja joka ei ymmarra aikamuotoja ni ei voi saada kovin herkästi

Codes: [-kriittikki/tunne] [kirjoittajan tausta] [tuntemus/nakemys] [arviointi/virheiden rooli/laatu]

et must joitaki tmmsti noh se on nyt vñ rajusti sanottu et joku tietty ryhm ei saa niille ei anneta mutta että se jos se nkyy niin systemaattisesti siit että sill ei oo mitn hajuja imperfektist eik mistn muustakaan menneen ajan muodosta ni tota ei se musta kyll oo kielest oivatunu sillon sit yht oleellista pointtia semmosta kontrastiivista havaintoo jostaki kie samalla lailla m voin sanoa et jos ei arabipojat osaa kirjottaa isoja kirjaimia pieni erikseen se rikkoo viel niin rajusti sit

Codes: [arviointi/epavarmuus] [kirjoittajan tausta]

No just se että ku sittiä tienny sit niinku niitte lahtotasoo siis siin mieless että et tuntu epaoikeudenmukaselta se arviointi ihan vaan semmoseen johonki abstraktiin m tason mrittelyy et ois halunnu ottaa huomioon vñ sit että olis tienny mit niille o opetettu ja ois tienny vñ että mist se on lahteny että jos ihmine on ollut vaikka aluks ihan luku ja kirjoitustaidoton tai niinku ni et sitte et sehan on aika hyvin et se saa jo jotain sahlattyy

Codes: [-kriittikki/tunne] [kirjoittajan tausta] [tuntemus/nakemys]

et sit tuntu pahalta niinku silleen et ajatteli että jos m tuntisin ton tyyppin ni m vois in antaa sille niinku yrityksest ettei pystyny arvioimaan sit yrityst koska ei tienny sit oppilasta et m koin sen arvostelun siis semmosena et voishan sit tietysti ajatella niin että arvioi vaan abstraktii suoritusta ja sillonha sil ei oo merkityst mut m se tuntu musta hijlyt et se tuntu ongelmalliselta

Codes: [arviointi/virheiden rooli/laatu] [kirjoittajan tausta] [ratkaisu/virheiden rooli]

noh em m oikeesta rat m en vaan voinu ottaa sit huomioon niih niih et kyl m nyt sit vhn kattelin niinku ett ku siell oli jotain jotka oli selvsti joko vietnamilaisii tai kiinalaisii ku niill oli morfologia ei ollu ollenkaan ja rrt ja llt oli sekasin ni m tein tn johtoptksen ni sit m niit nyt m kyl vlill sitte ehk annoin niille vhn enemmn anteeksi niinku just niit kielioppiongelmia mut em tietenk kokonaisuutena voinu et se ji vhn kaivelee ja sit sit m en tienny mit niille oli opetettu nin ollen m en tienny mit olis voinu kohtuudella vaatia

Codes: [-kriitikki/arviointi yleensa] [kirjoittajan tausta]

et sit koska sill arvioijalla ei oo henkilökohtast suhdetta nihi tekstintuottajii nii se vois silleen olla sit objektiivine ja et siit tulis sit tuikka rutiini et sill taval vois tietyl taval saada niinku jos ajattelee semmosta et olis joku valtakunnallinen taso ni parempii mut sit taas oppijoiden kannalta m pidn sit aivan merkitysetmn siis niinku et em m usko et se tietysti jotkut ihmiset arvostaa niit numeroita mut niinku oppijoiden kannalta m en pida hyvana semmosta

Codes: [-kriitikki/kirjoittaja] [-kriitikki/tunne] [kirjoittajan tausta]

mutta toisaalta niinku siin mielessa aika tylsii piirteit aika tylsaa alko olla niinku ties tasan mit kaikki venalaiset kirjottaa ja mita kaikki somalit kirjottaa et niinku niill on niin samat argumentitki et se oli ihan uskomatonta

Codes: [kirjoittajan tausta] [skaala/opetus]

ja sitten tavallaan noh tassa t oli kuitenkin ilmassa sillon et jos m niinkun opetukseen liittyvi tekstej arvioin ni ni m jotenkin arvioin niit aina suhteessa johonkin johonki semmoseen niinku mit opiskeltu tai tai tota mitk on ollu sen tehtvn tavoitteet et miks m oon ottanu semmosen

Codes: [kirjoittajan tausta]

ja muuten yks arviointiongelmia oli kyll ihan selvsti n niinku tmmset mm karjalaiset jotka kirjotti sellasta sellasta tota tai venajan puolelta jotka kir jotka niinku kirjotti semmosta lhell suomee olevaa joka oli idiomaattisen olosta mut sit kuitenkin tuntu et ei ne saa sanottua paljon mitn ni niist tuntu aina et niille pits kuitenkin antaa aika hyv mut kuitenkin yht aikaa et noh eihn niille voi antaa hyv kun ei ne saa sanottua mitn mut kuitenkin se oli idiomaattista se mit siin on et se on jotenki se oli kans hirveen vaikee arviointiongelmia

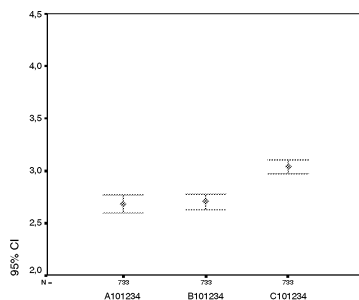
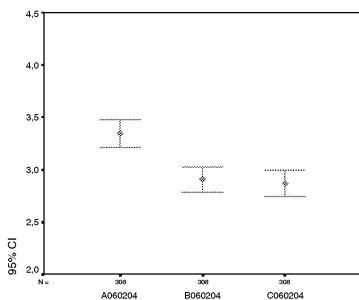
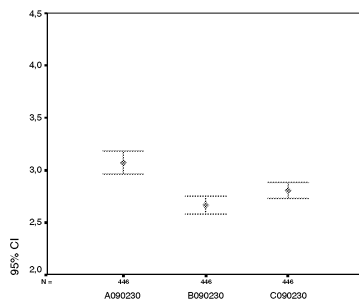
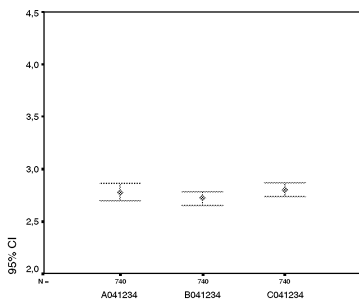
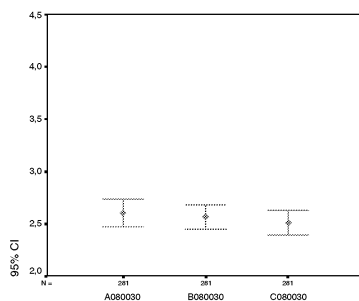
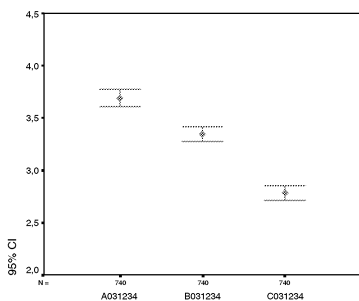
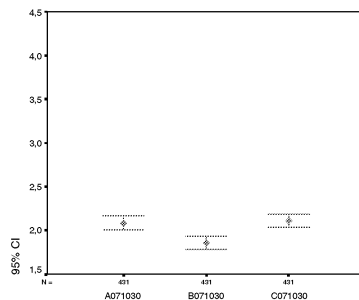
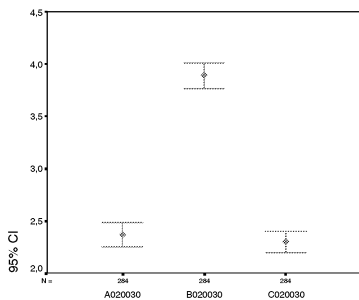
Codes: [-kriitikki/tunne] [kirjoittajan tausta]

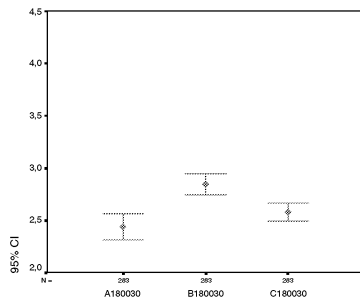
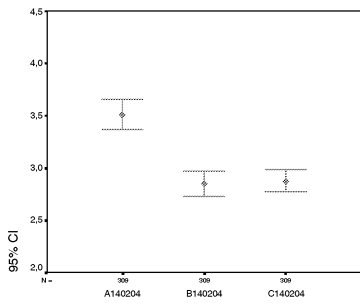
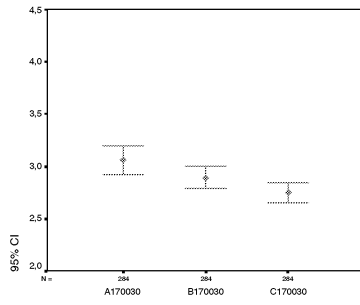
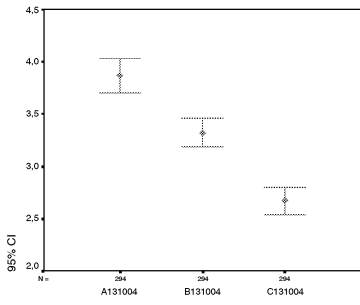
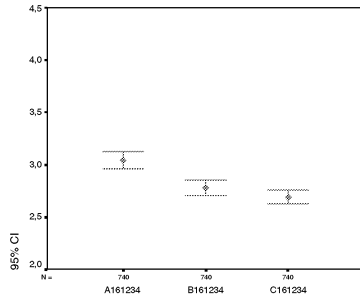
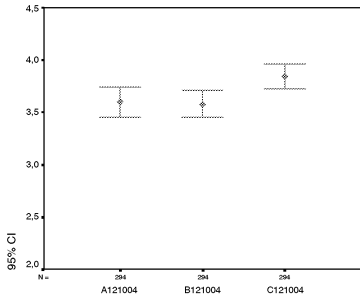
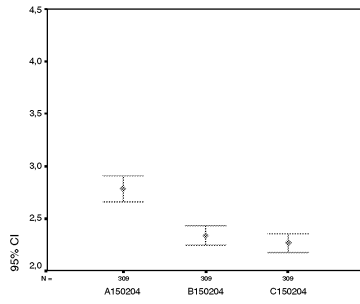
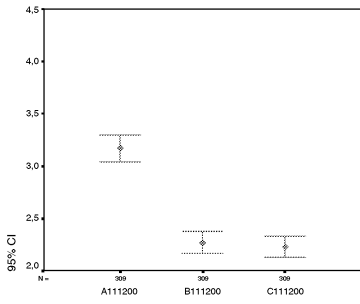
etku niit tuli siin perakanaa ni sit varmaan niinku nyppiinty vlill ja vlill sakotti ja vlill ei

Codes: [kirjoittajan tausta]

on kuitenkin se on kuitenkin meille niin lpinkyv mut sitte kuitenkin niinku joskus tuntuu ett tai se on niin vaikee panna joskus samalle venlis ja somalikakkoset ne on tai sill tavalla et vaikka niinku tuntuu et ne saa yht paljon silti sanottua noin niinku sislllisesti et n kulttuuriset kulttuuriset seikat on kans hirnu niinku ongelmallisia tn arvioinnin kannalta ja sit se on hirveet ku semmosista venlisist nkyv kaikki niitten koulujrjestelmt l niin hirveesti lpi ja se et miten tmmseen on tapana vastata ja mit asioita omasta kulttuurista on tapana esitell siis se oli kyl niinku se on hirven ristiriitoja hertty ja ja siin kyll

ARVIOIJIEN ANTAMIEN TASOARVIOIDEN KESKIARVOT MITTAUSKERTAKOHTAISESTI





ARVIOIJAKOHTAISTEN PEARSONIN KORRELAATIOIDEN PARITTAIS- VERTAILUT

r1-r2 viittaa mittauskertojen AB-AC, jossa A=1. mittauskerta, B=2. mittauskerta ja C=3. mittauskerta, vertailuun.

r1-r3 viittaa mittauskertojen AB-BC, jossa A=1. mittauskerta, B=2. mittauskerta ja C=3. mittauskerta, vertailuun.

r2-r3 viittaa mittauskertojen AC-BC, jossa A=1. mittauskerta, B=2. mittauskerta ja C=3. mittauskerta, vertailuun.

ARVIOIJA 011200

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= 0.006; t= .34018 (p=.63302; 1-p=.367)

r1-r3= 0.012; t= .68149 (p=.75196; 1-p=.248)

r2-r3= 0.006; t= .3413 (p=.63345; 1-p=.3666)

all with 306 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 020030

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= -0.005; t= -.16199 (p=.43572; 1-p=.5643)

r1-r3= -0.013; t= -.42018 (p=.33734; 1-p=.6627)

r2-r3= -0.008; t= -.25819 (p=.39823; 1-p=.6018)

all with 281 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 031234

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= -0.054; t= -3.48869 (p=0.00026; 1-p=.9997)

r1-r3= -.161; t= -10.09038 (p=0; 1-p=1)

r2-r3= -.107; t= -6.59923 (p=0; 1-p=1)

all with 737 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 041234

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= -0.004; t= -.36195 (p=.35875; 1-p=.6413)

r1-r3= 0.001; t= 0.09061 (p=.53609; 1-p=.4639)

r2-r3= 0.005; t= .45256 (p=.6745; 1-p=.3255)

all with 737 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 060204

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= 0.016; t= .76784 (p=.77841; 1-p=.2216)

r1-r3= -0.026; t= -1.23314 (p=.10923; 1-p=.8908)

r2-r3= -0.042; t= -2.00102 (p=0.02314; 1-

p=.9769)

all with 306 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 071030

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= -0.033; t= -1.63369 (p=0.05153; 1-p=.9485)

r1-r3= -0.01; t= -.4983 (p=.30926; 1-p=.6907)

r2-r3= 0.023; t= 1.13537 (p=.87157; 1-p=.1284)

all with 428 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 080030

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= 0.074; t= 3.04121 (p=.99871; 1-p=0.0013)

r1-r3= 0.063; t= 2.58093 (p=.99482; 1-p=0.0052)

r2-r3= -0.011; t= -.46019 (p=.32287; 1-p=.6771)

all with 279 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 090230

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= -0.035; t= -2.57426 (p=0.00518; 1-p=.9948)

r1-r3= -0.024; t= -1.77049 (p=0.03867; 1-p=.9613)

r2-r3= 0.011; t= .80374 (p=.78901; 1-p=.211)

all with 443 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJA 101234

dependent (single sample) differences the correlations must be: rxy, rxz & ryz
t-values for all differences:

r1-r2= 0.028; t= 1.82077 (p=.96548; 1-p=0.0345)

r1-r3= -0.015; t= -.96349 (p=.16781; 1-p=.8322)

r2-r3= -0.043; t= -2.7843 (p=0.00275; 1-p=.9972)

all with 735 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 111200

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.02; t= .84041 (p=.79933; 1-p=.2007)

r1-r3= -0.07; t= -2.86682 (p=0.00222; 1-p=.9978)

r2-r3= -0.09; t= -3.70744 (p=0.00012; 1-p=.9999)

all with 306 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 121004

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.011; t= .49362 (p=.68903; 1-p=.311)

r1-r3= -0.015; t= -.66821 (p=.25226; 1-p=.7477)

r2-r3= -0.026; t= -1.16184 (p=.12313; 1-p=.8769)

all with 291 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 131004

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.054; t= 3.38018 (p=.99959; 1-p=0.0004)

r1-r3= 0.011; t= .68049 (p=.75163; 1-p=.2484)

r2-r3= -0.043; t= -2.69963 (p=0.00367; 1-p=.9963)

all with 291 degrees of freedom (multiply p-values with 2 for double sided testing)

Confidence Intervals for the difference between r1 and r2

ARVIOJA 140204

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.021; t= 1.12311 (p=.86886; 1-p=.1311)

r1-r3= 0.011; t= .58667 (p=.72107; 1-p=.2789)

r2-r3= -0.01; t= -.53644 (p=.29602; 1-p=.704)

all with 306 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 150204

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= -0.051; t= -2.3772 (p=0.00903; 1-p=.991)

r1-r3= -0.093; t= -4.28441 (p=0.00001; 1-p=1)

r2-r3= -0.042; t= -1.90685 (p=0.02874; 1-p=.9713)

all with 306 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 161234

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.028; t= 2.67271 (p=.99615; 1-p=0.0038)

r1-r3= -0.021; t= -1.97785 (p=0.02416; 1-p=.9758)

r2-r3= -0.049; t= -4.65066 (p=0; 1-p=1)

all with 737 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 170030

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= -0.045; t= -1.50326 (p=0.06695; 1-p=.9331)

r1-r3= -.148; t= -4.79754 (p=0; 1-p=1)

r2-r3= -.103; t= -3.29329 (p=0.00056; 1-p=.9994)

all with 281 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOJA 180030

dependent (single sample) differences the correlations must be: rxy, rxz & ryz

t-values for all differences:

r1-r2= 0.021; t= .7958 (p=.78659; 1-p=.2134)

r1-r3= -0.063; t= -2.33031 (p=0.01025; 1-p=.9898)

r2-r3= -0.084; t= -3.12628 (p=0.00098; 1-p=.999)

all with 281 degrees of freedom (multiply p-values with 2 for double sided testing)

ARVIOIJIEN SAMANMIELISYYSPROSENTTIEN PARITTAISVERTAILU

Arvioitsijoiden samanmielisyyssprosenttien parittaisvertailu McNemar-testillä (kahden riippuvan otoksen testi). Arvioitsijoiden tunnistenumerot suluissa.

A viittaa 1. ja 2. mittauksen samanmielisyyteen

B viittaa 1. ja 3. mittauksen samanmielisyyteen

C viittaa 2. ja 3. mittauksen samanmielisyyteen

Muuttujien arvot: 0 = arviot eriävät kahdella mittauksella, 1 = arviot samat

A(01) & B(01)

	B(01)	
A(01)	0	1
0	69	39
1	45	155

A(01) & C(01)

	C(01)	
A(01)	0	1
0	44	64
1	45	155

B(01) & C(01)

	C(01)	
B(01)	0	1
0	50	64
1	39	155

Test

	A(01) &	B(01) &	A(01) &
N	308	308	308
Chi- ^a	,298	5,592	2,972
Asymp.	,585	,018	,085

a. Continuity

A(02) & B(02)

	B(02)	
A(02)	0	1
0	115	149
1	12	8

A(02) & C(02)

	C(02)	
A(02)	0	1
0	251	13
1	12	8

B(02) & C(02)

	C(02)	
B(02)	0	1
0	114	13
1	149	8

Test Statistics

	A(02) & B(02)	B(02) & C(02)	A(02) & C(02)
N	284	284	284
Chi-Square ^a	114,882	112,500	
Asymp. Sig.	,000	,000	
Exact Sig. (2-tailed)			1,000 ^b

a. Continuity Corrected

b. Binomial distribution used.

A(03) & B(03)

	B(03)	
A(03)	0	1
0	292	87
1	218	142

A(03) & C(03)

	C(03)	
A(03)	0	1
0	204	175
1	218	142

B(03) & C(03)

	C(03)	
B(03)	0	1
0	335	175
1	87	142

Test Statistics

	A(03) & B(03)	B(03) & C(03)	A(03) & C(03)
N	739	739	739
Chi-Square ^a	55,410	28,889	4,489
Asymp. Sig.	,000	,000	,034

a. Continuity Corrected

A(04) & B(04)

A(04)	B(04)	
	0	1
0	169	113
1	108	349

A(04) & C(04)

A(04)	C(04)	
	0	1
0	127	155
1	108	349

B(04) & C(04)

B(04)	C(04)	
	0	1
0	122	155
1	113	349

Test Statistics

	A(04) & B(04)	B(04) & C(04)	A(04) & C(04)
N	739	739	739
Chi-Square ^a	,072	6,272	8,046
Asymp. Sig.	,788	,012	,005

a. Continuity Corrected

A(06) & B(06)

A(06)	B(06)	
	0	1
0	110	41
1	49	109

A(06) & C(06)

A(06)	C(06)	
	0	1
0	69	82
1	49	109

B(06) & C(06)

B(06)	C(06)	
	0	1
0	77	82
1	41	109

Test Statistics

	A(06) & B(06)	B(06) & C(06)	A(06) & C(06)
N	309	309	309
Chi-Square ^a	,544	13,008	7,817
Asymp. Sig.	,461	,000	,005

a. Continuity Corrected

A(07) & B(07)

A(07)	B(07)	
	0	1
0	62	80
1	60	228

A(07) & C(07)

A(07)	C(07)	
	0	1
0	90	52
1	60	228

B(07) & C(07)

B(07)	C(07)	
	0	1
0	70	52
1	80	228

Test Statistics

	A(07) & B(07)	B(07) & C(07)	A(07) & C(07)
N	430	430	430
Chi-Square ^a	2,579	5,523	,438
Asymp. Sig.	,108	,019	,508

a. Continuity Corrected

A(08) & B(08)

A(08)	B(08)	
	0	1
0	75	40
1	49	120

A(08) & C(08)

A(08)	C(08)	
	0	1
0	58	57
1	49	120

B(08) & C(08)

B(08)	C(08)	
	0	1
0	67	57
1	40	120

Test Statistics

	A(08) & B(08)	B(08) & C(08)	A(08) & C(08)
N	284	284	284
Chi-Square ^a	,719	2,639	,462
Asymp. Sig.	,396	,104	,497

a. Continuity Corrected

A(09) & B(09)

A(09)	B(09)	
	0	1
0	144	66
1	37	199

B(09) & C(09)

B(09)	C(09)	
	0	1
0	57	124
1	66	199

A(09) & C(09)

A(09)	C(09)	
	0	1
0	86	124
1	37	199

Test Statistics

	A(09) & B(09)	B(09) & C(09)	A(09) & C(09)
N	446	446	446
Chi-Square ^a	7,612	17,100	45,938
Asymp. Sig.	,006	,000	,000

a. Continuity Corrected

A(10) & B(10)

A(10)	B(10)	
	0	1
0	249	100
1	186	204

B(10) & C(10)

B(10)	C(10)	
	0	1
0	246	189
1	100	204

A(10)&C(10)

A(10)	C(10)	
	0	1
0	160	189
1	186	204

Test Statistics

	A(10) & B(10)	B(10) & C(10)	A(10) & C(10)
N	739	739	739
Chi-Square ^a	25,262	26,796	,011
Asymp. Sig.	,000	,000	,918

a. Continuity Corrected

A(11) & B(11)

A(11)	B(11)	
	0	1
0	206	22
1	30	50

A(11) & C(11)

A(11)	C(11)	
	0	1
0	67	161
1	30	50

B(11) & C(11)

B(11)	C(11)	
	0	1
0	75	161
1	22	50

Test Statistics

	A(11) & B(11)	B(11) & C(11)	A(11) & C(11)
N	308	308	308
Chi-Square ^a	,942	104,066	88,482
Asymp. Sig.	,332	,000	,000

a. Continuity Corrected

A(12) & B(12)

A(12)	B(12)	
	0	1
0	91	54
1	60	88

A(12) & C(12)

A(12)	C(12)	
	0	1
0	75	70
1	60	88

B(12) & C(12)

B(12)	C(12)	
	0	1
0	81	70
1	54	88

Test Statistics

	A(12) & B(12)	B(12) & C(12)	A(12) & C(12)
N	293	293	293
Chi-Square ^a	,219	1,815	,623
Asymp. Sig.	,640	,178	,430

a. Continuity Corrected

A(13) & B(13)

A(13)	B(13)	
	0	1
0	153	17
1	79	44

A(13) & C(13)

A(13)	C(13)	
	0	1
0	99	71
1	79	44

B(13) & C(13)

B(13)	C(13)	
	0	1
0	161	71
1	17	44

Test Statistics

	A(13) & B(13)	B(13) & C(13)	A(13) & C(13)
N	293	293	293
Chi-Square ^a	38,760	31,920	,327
Asymp. Sig.	,000	,000	,568

a. Continuity Corrected

A(14) & B(14)

A(14)	B(14)	
	0	1
0	148	31
1	40	90

A(14) & C(14)

A(14)	C(14)	
	0	1
0	66	113
1	40	90

B(14) & C(14)

B(14)	C(14)	
	0	1
0	75	113
1	31	90

Test Statistics

	A(14) & B(14)	B(14) & C(14)	A(14) & C(14)
N	309	309	309
Chi-Square ^a	,901	45,563	33,882
Asymp. Sig.	,342	,000	,000

a. Continuity Corrected

A(15) & B(15)

A(15)	B(15)	
	0	1
0	113	31
1	32	133

A(15) & C(15)

A(15)	C(15)	
	0	1
0	47	97
1	32	133

B(15) & C(15)

B(15)	C(15)	
	0	1
0	48	97
1	31	133

Test Statistics

	A(15) & B(15)	B(15) & C(15)	A(15) & C(15)
N	309	309	309
Chi-Square ^a	,000	33,008	31,752
Asymp. Sig.	1,000	,000	,000

a. Continuity Corrected

A(16) & B(16)

	B(16)	
A(16)	0	1
0	229	60
1	118	332

A(16) & C(16)

	C(16)	
A(16)	0	1
0	83	206
1	118	332

B(16) & C(16)

	C(16)	
B(16)	0	1
0	141	206
1	60	332

Test Statistics

	A(16) & B(16)	B(16) & C(16)	A(16) & C(16)
N	739	739	739
Chi-Square ^a	18,253	79,041	23,361
Asymp. Sig.	,000	,000	,000

a. Continuity Corrected

A(17) & B(17)

	B(17)	
A(17)	0	1
0	84	40
1	38	122

A(17) & C(17)

	C(17)	
A(17)	0	1
0	66	58
1	38	122

B(17) & C(17)

	C(17)	
B(17)	0	1
0	64	58
1	40	122

Test Statistics

	A(17) & B(17)	B(17) & C(17)	A(17) & C(17)
N	284	284	284
Chi-Square ^a	,013	2,949	3,760
Asymp. Sig.	,910	,086	,052

a. Continuity Corrected

A(18) & B(18)

	B(18)	
A(18)	0	1
0	108	54
1	34	88

A(18) & C(18)

	C(18)	
A(18)	0	1
0	66	96
1	34	88

B(18) & C(18)

	C(18)	
B(18)	0	1
0	46	96
1	54	88

Test Statistics

	A(18) & B(18)	B(18) & C(18)	A(18) & C(18)
N	284	284	284
Chi-Square ^a	4,102	11,207	28,623
Asymp. Sig.	,043	,001	,000

a. Continuity Corrected

ARVIOIJA- JA MITTAUSKERTAKOHTAISET TAITOTASOFREKVENSsit

A01120

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	77	10,4	24,9	24,9
2	111	15,0	35,9	60,8
3	77	10,4	24,9	85,8
4	32	4,3	10,4	96,1
5	11	1,5	3,6	99,7
6	1	,1	,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

A080030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	54	7,3	19,0	19,0
2	89	12,0	31,3	50,4
3	77	10,4	27,1	77,5
4	46	6,2	16,2	93,7
5	18	2,4	6,3	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

A020030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	50	6,8	17,6	17,6
2	124	16,8	43,7	61,3
3	76	10,3	26,8	88,0
4	25	3,4	8,8	96,8
5	8	1,1	2,8	99,6
6	1	,1	,4	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

A090230

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	37	5,0	8,3	8,3
2	107	14,5	24,0	32,3
3	158	21,4	35,4	67,7
4	90	12,2	20,2	87,9
5	40	5,4	9,0	96,9
6	14	1,9	3,1	100,0
Total	446	60,3	100,0	
Missing System	294	39,7		
Total	740	100,0		

A031234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	26	3,5	3,5	3,5
2	77	10,4	10,4	13,9
3	222	30,0	30,0	43,9
4	225	30,4	30,4	74,3
5	156	21,1	21,1	95,4
6	34	4,6	4,6	100,0
Total	740	100,0	100,0	

A101234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	122	16,5	16,5	16,5
2	235	31,8	31,8	48,3
3	229	30,9	31,0	79,3
4	89	12,0	12,0	91,3
5	50	6,8	6,8	98,1
6	14	1,9	1,9	100,0
Total	739	99,9	100,0	
Missing System	1	,1		
Total	740	100,0		

A041234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	86	11,6	11,6	11,6
2	243	32,8	32,8	44,5
3	222	30,0	30,0	74,5
4	134	18,1	18,1	92,6
5	49	6,6	6,6	99,2
6	6	,8	,8	100,0
Total	740	100,0	100,0	

A111200

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	19	2,6	6,1	6,1
2	65	8,8	21,0	27,2
3	115	15,5	37,2	64,4
4	72	9,7	23,3	87,7
5	31	4,2	10,0	97,7
6	7	,9	2,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

A060204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	23	3,1	7,5	7,5
2	41	5,5	13,3	20,8
3	105	14,2	34,1	54,9
4	95	12,8	30,8	85,7
5	34	4,6	11,0	96,8
6	10	1,4	3,2	100,0
Total	308	41,6	100,0	
Missing System	432	58,4		
Total	740	100,0		

A121004

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	16	2,2	5,4	5,4
2	44	5,9	15,0	20,4
3	68	9,2	23,1	43,5
4	94	12,7	32,0	75,5
5	58	7,8	19,7	95,2
6	14	1,9	4,8	100,0
Total	294	39,7	100,0	
Missing System	446	60,3		
Total	740	100,0		

A071030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	112	15,1	26,0	26,0
2	198	26,8	45,9	71,9
3	99	13,4	23,0	94,9
4	17	2,3	3,9	98,8
5	5	,7	1,2	100,0
Total	431	58,2	100,0	
Missing System	309	41,8		
Total	740	100,0		

A131004

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	446	60,3	60,3	60,3
1	17	2,3	2,3	62,6
2	40	5,4	5,4	68,0
3	55	7,4	7,4	75,4
4	77	10,4	10,4	85,8
5	63	8,5	8,5	94,3
6	42	5,7	5,7	100,0
Total	740	100,0	100,0	

B01200

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	56	7,6	18,1	18,1
2	118	15,9	38,2	56,3
3	99	13,4	32,0	88,3
4	33	4,5	10,7	99,0
5	2	,3	,6	99,7
6	1	,1	,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

A140204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	25	3,4	8,1	8,1
2	44	5,9	14,2	22,3
3	75	10,1	24,3	46,6
4	98	13,2	31,7	78,3
5	47	6,4	15,2	93,5
6	20	2,7	6,5	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

B020030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	3	,4	1,1	1,1
2	23	3,1	8,1	9,2
3	71	9,6	25,0	34,2
4	107	14,5	37,7	71,8
5	66	8,9	23,2	95,1
6	14	1,9	4,9	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

A150204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	26	3,5	8,4	8,4
2	111	15,0	35,9	44,3
3	113	15,3	36,6	80,9
4	33	4,5	10,7	91,6
5	15	2,0	4,9	96,4
6	11	1,5	3,6	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

B031234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	12	1,6	1,6	1,6
2	134	18,1	18,1	19,7
3	270	36,5	36,5	56,2
4	246	33,2	33,2	89,5
5	67	9,1	9,1	98,5
6	11	1,5	1,5	100,0
Total	740	100,0	100,0	

A161234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	61	8,2	8,2	8,2
2	154	20,8	20,8	29,1
3	314	42,4	42,4	71,5
4	130	17,6	17,6	89,1
5	66	8,9	8,9	98,0
6	15	2,0	2,0	100,0
Total	740	100,0	100,0	

B041234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	76	10,3	10,3	10,3
2	231	31,2	31,2	41,5
3	280	37,8	37,8	79,3
4	135	18,2	18,2	97,6
5	17	2,3	2,3	99,9
6	1	,1	,1	100,0
Total	740	100,0	100,0	

A170030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	18	2,4	6,3	6,3
2	70	9,5	24,6	31,0
3	121	16,4	42,6	73,6
4	40	5,4	14,1	87,7
5	22	3,0	7,7	95,4
6	13	1,8	4,6	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

B060204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	31	4,2	10,0	10,0
2	67	9,1	21,7	31,7
3	146	19,7	47,2	79,0
4	43	5,8	13,9	92,9
5	12	1,6	3,9	96,8
6	10	1,4	3,2	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

A180030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	46	6,2	16,2	16,2
2	127	17,2	44,7	60,9
3	73	9,9	25,7	86,6
4	22	3,0	7,7	94,4
5	12	1,6	4,2	98,6
6	4	,5	1,4	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

B071030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	151	20,4	35,0	35,0
2	204	27,6	47,3	82,4
3	64	8,6	14,8	97,2
4	10	1,4	2,3	99,5
5	2	,3	,5	100,0
Total	431	58,2	100,0	
Missing System	309	41,8		
Total	740	100,0		

B080030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	39	5,3	13,8	13,8
2	98	13,2	34,6	48,4
3	104	14,1	36,7	85,2
4	32	4,3	11,3	96,5
5	10	1,4	3,5	100,0
Total	283	38,2	100,0	
Missing System	457	61,8		
Total	740	100,0		

B090230

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	46	6,2	10,3	10,3
2	138	18,6	30,9	41,3
3	195	26,4	43,7	85,0
4	56	7,6	12,6	97,5
5	10	1,4	2,2	99,8
6	1	,1	,2	100,0
Total	446	60,3	100,0	
Missing System	294	39,7		
Total	740	100,0		

B101234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	87	11,8	11,8	11,8
2	225	30,4	30,5	42,3
3	296	40,0	40,1	82,4
4	100	13,5	13,6	95,9
5	20	2,7	2,7	98,6
6	10	1,4	1,4	100,0
Total	738	99,7	100,0	
Missing System	2	,3		
Total	740	100,0		

B111200

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	63	8,5	20,4	20,4
2	130	17,6	42,1	62,5
3	93	12,6	30,1	92,6
4	17	2,3	5,5	98,1
5	5	,7	1,6	99,7
6	1	,1	,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

B121004

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	8	1,1	2,7	2,7
2	39	5,3	13,3	16,0
3	95	12,8	32,3	48,3
4	92	12,4	31,3	79,6
5	46	6,2	15,6	95,2
6	14	1,9	4,8	100,0
Total	294	39,7	100,0	
Missing System	446	60,3		
Total	740	100,0		

B131004

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	14	1,9	4,8	4,8
2	65	8,8	22,1	26,9
3	89	12,0	30,3	57,1
4	78	10,5	26,5	83,7
5	34	4,6	11,6	95,2
6	14	1,9	4,8	100,0
Total	294	39,7	100,0	
Missing System	446	60,3		
Total	740	100,0		

B140204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	35	4,7	11,3	11,3
2	73	9,9	23,6	35,0
3	124	16,8	40,1	75,1
4	59	8,0	19,1	94,2
5	17	2,3	5,5	99,7
6	1	,1	,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

B150204

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	49	6,6	15,9	15,9
2	123	16,6	39,8	55,7
3	124	16,8	40,1	95,8
4	10	1,4	3,2	99,0
5	3	,4	1,0	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

B161234

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	70	9,5	9,5	9,5
2	227	30,7	30,7	40,1
3	284	38,4	38,4	78,5
4	119	16,1	16,1	94,6
5	35	4,7	4,7	99,3
6	5	,7	,7	100,0
Total	740	100,0	100,0	

B170030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	18	2,4	6,3	6,3
2	60	8,1	21,1	27,5
3	158	21,4	55,6	83,1
4	31	4,2	10,9	94,0
5	16	2,2	5,6	99,6
6	1	,1	,4	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

B180030

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	12	1,6	4,2	4,2
2	75	10,1	26,4	30,6
3	153	20,7	53,9	84,5
4	37	5,0	13,0	97,5
5	3	,4	1,1	98,6
6	4	,5	1,4	100,0
Total	284	38,4	100,0	
Missing System	456	61,6		
Total	740	100,0		

C011200

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	56	7,6	18,1	18,1
2	107	14,5	34,6	52,8
3	109	14,7	35,3	88,0
4	33	4,5	10,7	98,7
5	4	,5	1,3	100,0
Total	309	41,8	100,0	
Missing System	431	58,2		
Total	740	100,0		

C020030

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	43	5,8	15,1	15,1
	2	144	19,5	50,7	65,8
	3	72	9,7	25,4	91,2
	4	19	2,6	6,7	97,9
	5	6	,8	2,1	100,0
	Total	284	38,4	100,0	
Missing	System	456	61,6		
Total		740	100,0		

C031234

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	61	8,2	8,2	8,2
	2	221	29,9	29,9	38,1
	3	300	40,5	40,5	78,6
	4	1	,1	,1	78,8
	4	137	18,5	18,5	97,3
	5	18	2,4	2,4	99,7
	6	2	,3	,3	100,0
	Total	740	100,0	100,0	

C041234

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	63	8,5	8,5	8,5
	2	204	27,6	27,6	36,1
	3	306	41,4	41,4	77,4
	4	155	20,9	20,9	98,4
	5	11	1,5	1,5	99,9
	6	1	,1	,1	100,0
	Total	740	100,0	100,0	

C060204

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	29	3,9	9,4	9,4
	2	92	12,4	29,8	39,2
	3	113	15,3	36,6	75,7
	4	52	7,0	16,8	92,6
	5	14	1,9	4,5	97,1
	6	9	1,2	2,9	100,0
	Total	309	41,8	100,0	
Missing	System	431	58,2		
Total		740	100,0		

C071030

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	100	13,5	23,2	23,2
	2	198	26,8	45,9	69,1
	3	118	15,9	27,4	96,5
	4	15	2,0	3,5	100,0
	Total	431	58,2	100,0	
Missing	System	309	41,8		
Total		740	100,0		

C080030

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	49	6,6	17,4	17,4
	2	88	11,9	31,2	48,6
	3	109	14,7	38,7	87,2
	4	25	3,4	8,9	96,1
	5	11	1,5	3,9	100,0
	Total	282	38,1	100,0	
Missing	System	458	61,9		
Total		740	100,0		

C090230

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	28	3,8	6,3	6,3
	2	119	16,1	26,7	33,0
	3	222	30,0	49,8	82,7
	4	66	8,9	14,8	97,5
	5	11	1,5	2,5	100,0
	Total	446	60,3	100,0	
Missing	System	294	39,7		
Total		740	100,0		

C101234

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	34	4,6	4,6	4,6
	2	144	19,5	19,6	24,2
	3	363	49,1	49,4	73,6
	4	157	21,2	21,4	95,0
	5	31	4,2	4,2	99,2
	6	6	,8	,8	100,0
	Total	735	99,3	100,0	
Missing	System	5	,7		
Total		740	100,0		

C111200

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	68	9,2	22,0	22,0
	2	131	17,7	42,4	64,4
	3	85	11,5	27,5	91,9
	4	21	2,8	6,8	98,7
	5	4	,5	1,3	100,0
	Total	309	41,8	100,0	
Missing	System	431	58,2		
Total		740	100,0		

C121004

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	5	,7	1,7	1,7
	2	16	2,2	5,4	7,1
	3	84	11,4	28,6	35,7
	4	122	16,5	41,5	77,2
	5	49	6,6	16,7	93,9
	6	18	2,4	6,1	100,0
	Total	294	39,7	100,0	
Missing	System	446	60,3		
Total		740	100,0		

C131004

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	43	5,8	14,6	14,6
	2	93	12,6	31,6	46,3
	3	98	13,2	33,3	79,6
	4	43	5,8	14,6	94,2
	5	12	1,6	4,1	98,3
	6	5	,7	1,7	100,0
	Total	294	39,7	100,0	
Missing	System	446	60,3		
Total		740	100,0		

C140204

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	24	3,2	7,8	7,8
	2	71	9,6	23,0	30,7
	3	142	19,2	46,0	76,7
	4	64	8,6	20,7	97,4
	5	7	,9	2,3	99,7
	6	1	,1	,3	100,0
	Total	309	41,8	100,0	
Missing	System	431	58,2		
Total		740	100,0		

C150204

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	56	7,6	18,1	18,1
	2	133	18,0	43,0	61,2
	3	105	14,2	34,0	95,1
	4	13	1,8	4,2	99,4
	5	2	,3	,6	100,0
	Total	309	41,8	100,0	
Missing	System	431	58,2		
Total		740	100,0		

C180030

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	12	1,6	4,2	4,2
	2	123	16,6	43,5	47,7
	3	126	17,0	44,5	92,2
	4	17	2,3	6,0	98,2
	5	5	,7	1,8	100,0
	Total	283	38,2	100,0	
Missing	System	457	61,8		
Total		740	100,0		

C161234

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	68	9,2	9,2	9,2
	2	250	33,8	33,8	43,0
	3	298	40,3	40,3	83,2
	4	94	12,7	12,7	95,9
	5	27	3,6	3,6	99,6
	6	3	,4	,4	100,0
	Total	740	100,0	100,0	

C170030

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	24	3,2	8,5	8,5
	2	71	9,6	25,0	33,5
	3	146	19,7	51,4	84,9
	4	39	5,3	13,7	98,6
	5	4	,5	1,4	100,0
	Total	284	38,4	100,0	
Missing	System	456	61,6		
Total		740	100,0		

VALINTAVERKKOHAASTATELUS KÄYTETYT 9 TEKSTIÄ

TEKSTI 1

25-vuotias nainen joka oli autokolarissa.
 Primaaristi hoidettu lähimmässä terveyskeskuksessa ja ~~sii~~
 sitä kautta sairaalaan.
 Kolarin
 tässä kuvassa lääkäri tarkistaa potilasta, jolla
 on jalan ja polven osassa ihorikko,
 lääkäri sanoo että polven liikkeet ovat kunnossa
 ja jalan röntgen-kuvassa ei näy mitään vakavaa.
 Sen jälkeen 25-vuotias nainen pääsi kotiin
 leväämään. Kontrolli sovittu kahden viikon päästä.
 Potilas käynyt jälleen sairaalassa.
 lääkäri sanoo että jalka ja polvi on koko ajan
 paranemassa, ja se tarkoittaa että ei tarvitse ~~tehdä~~
 käydä kontrollissa.

TEKSTI 2

Pieni tyttö oli koulussa, se oli kerskivikko. Hänen kolmas tunti loppui. Tyttö ja hänen ystävät menevät poltto. Kukaan vanhempi ei tiennyt että heidän tytär pötki. He poltti ja yhtiökä hestien opettaja käveli koulusta. Tytöt tiesivät että jos opettaja näkisi että he polttivat saavot se vanhemmille. He juoksi tien toiseen puoleen mutta auto tuli ja tappoi kaksi tyttöä mutta kolmas vain rikkoi hänen jalkansa. Nyt lääkäri auttoi tyttöille mottiini koska se köpeä paljon. Kumpi tyttö on parempi hän mene takaisin koulun. Ei kukaan tiedä, kuka tappoi tei kaksi tyttöä koska auto ei pysähtä.

Loppu.

TEKSTI 3

Kerran tyttö meni koulun.
Hänellä oli kova kiire.
Hän juokse tie poikki, sen takia
tapahtyi liikeonnettomuus. Hän
joutui sairaalloon. Hänellä jalka
mene poikki, koska hänelle laita
kipsi jalaan. Hän oli sairaalassa
paranemisen asti.

TEKSTI 4

Se oli kaksi vuotta sitten. Tyttö istui kotona ja katseli televisiota. Ovikello soi ja hän meni avata ovi. Oven takana seiso i mies hänen naapurinsa. Mies sano i että kymmenen minuuttia sitten näkii sinun isäs i hän ajo i auttolla ja törmäsi toiseen autoon. Sitten tuli auto isä istui autoon ja he ajoivat suoraan sairaalaan.

Tyttö nopeasti ajo i polkupyörällä sairaalaan. Hänellä oli kiire ja kun kääntyi vasemalle myös törmäsi, mutta puuhun. Hän halusi taas ajoa mutta ei voi jalka oli poikki. Kaksi miestä auttoivat. Miehet tytön kanssa myös ajoivat sairaalaan. Kun tyttö tuli sairaalaan hän heti kysyi: "Sairaalassa on minun isäni, mitä tapahtu i, onko hän tervena?" lääkäri sano i että kaikki on hyvin isäsi halusi mennä jo kotiin mutta nyt sinä olet sairaana ja isäsi seisoo ulkona ja haluaisi puhua sinun kanssa.

Tyttö oli sairaalassa noin seitsemän päivää mutta nyt tyttö ja sänsä istuvat kotona ja nyt heillä on kaikki hyvin.

TEKSTI 5

Yhden kevään päivän Maija asatteli aava polku-
pyörällä. Hän söi reseptensä kanssa lounaan, autoi äidille
tiskata. Noin kello neljatoista Maija laitoi vaatteita pääl-
le, otti avaimen polkupyörään lukosta ja talusi lähtee kotaa
pois. Mutta Maijan äiti näki, että Maija haluaa mennä sokon-
kin. Äiti kysyi: "Mihin sinä menet?" "Minä menen avamaan polku-
pyörää." - tyttö vastasi. "Kuuntele nyt minua! Ulkona on hel-
misen liukasta ja sen lisäksi tänään on lauantai, se tarkoittaa,
että ulkona on paljon autoja!" "Äiti, lopeta jo!" - Maija
suuttui. "Minä en ole joku pieni tyhmä tyttö, olen 13 jo! Minun
kanssa ei tarvittu mitään rahaa." "Tee mitä haluat! Mutta ole
varovainen!" - sanoi äiti ja meni huoneen. Maija oli oikein
iloinen, että hänet jätettiin rauhaan.

Maija oli hakenut polkupyörään varastosta. Hän rakkasi,
että aava vähän polkupyörällä ja sitten menee kaverin luok-
se. Maija oli hyvin vihainen: "Miksi äiti asattelee, että minä
en osaa ajattella itse. Minä olen jo aikuinen, minun kanssa
ei voi tarvittua mitään tahansa!" Maijan asattelessa tätä hän
ei huomannut tuleva auto... liukas tie, hidastava reaktio, ei
pareutunut auto - Maija makasi kadulla. Hän oli sokissa.
Autonkuljettaja oli hyvin pelästynyt. Hän kysyi tytöltä monta
kerta, onko hänellä jotain kipua. Mutta tyttö ei vastannut mitään.
Autonkuljettaja rakkasi, että parempi soittaa ambulanssin. Onneksi
hänellä oli kännykkä.

Vierasta minuuttia päästä tuli ambulanssi. Tyttö asetettiin
auton ja vietiin sairaalaan. Maijan selviytyä sokista hän kehti osoi-
→

leensa. Sairaalassa sairaanhoitaja soitti Maijan vanhemmille. Puolen tunnen päästä he tulivat sairaalaan.

Maijan lääkäri oli mukava ulkomaalainen mies. Hän tuli Venäjältä, mutta puhui hyvin suomea. Hän oli ottanut Maijan salasta röntgenkuuvia. Ilmaantui, että Maijalla on oikea jalka roikki. Jalkaan laitettiin kipsin.

Nyt Maijan täytyy olla sairaalassa kaksi viikkoa. Joka päivä äiti ja isä ovat käyneet tevehtimässä Maijan. Autonkuljettaja myös tuli, hän oli tuonut Maijalle lahjaksi kirjaa. Nyt Maijalle ei ole tylsä. Mutta hän oli sveulinen, että hänen polkupyöränsä ei ole. Äiti ja isä sanoivat Maijalle, että he ostavat hänelle uuden polkupyörän... mutta sillä ehdolla, että Maija kuntelee vanhempia.

TEKSTI 6

Tuula broke hän: jalka mannantaina. Hän on nyt sairaalassa.

Lääkäri sanoo hän voi mennä kotiin tänään mutta hänen täytyy mennä sankkyyntä päivälle. Tuula on itäinen että hän voi mennä kotiin.

TEKSTI 7

Minka on katsikymmentakuusi vuotias. Hän on työssä
Wärtösilässä, työskentelee sihteeriksi. Joka päivä, Minna
ajaa polkupyörää työhön. Mutta monta päivä sitten hän
~~huomasi että~~ hän riideli hänen poikaystävänsä kanssa niin
kavasti, että mies oli vihainen ja meni juomaan viiniä
pubissa. Koko yö, hän ei tullut takaisin kotumaan.
Aamulla Minna taas ajoi polkupyörää työhön, mutta
hän oli niin hajamielinen että lepusti hän ajoi polkupyörää
~~to~~ autoon, josta tuli vastaan päin. Onneksi, ei mitään hän
vakavaa ei luokattu niin vakavasti. Vain oikei jala oli
poikki. Nyt hän ei voi mennä työhön ja ei ~~virkkaan~~ ^{työhön} voi
riidellä poikaystävänsä kanssa. Hän ei tiedä kuka hoitaa
hänen työtä firmassa ja kuka hänen tulla voi. Onko
hän vielä humalassa. Lääkäri tulee katsomaan häntä
ja panna uuden lääket hänen jalaan. Lääkäri kertoi
että sama aika ^{poistava} hänelle tulee ~~tulla~~ katsi nuoria potilasta.
tänään aamuulla

Toinen on vain mies, ja tämä on nainen. Miehen vasemmi
jala oli luokattu ja hänen nimi on Matti Koskinen. Minna
sanoi: "Ei ole totta!" Mies on hänen poikaystävä.
Molemmat joutuivat sairaalaan 1 viikkoa.

TEKSTI 8

Miia on ollut polkuopiräoneitomuudessa.
Jalka on lytä nuistaut, mutta
onneksi ei ole tapahtunut mitään
vakavampaa. Päästä, että heusi on
säilynyt!

Lääkäri Pitsänea on juuri
tarkastanut Miian jalkaa ja
näyttää siltä, että Miia pääsee
nikonloppuna jo kotiin. Polkuopirällä
hän ei pääse ajamaan - ei ainakaan
tenä eelänä, mutta ^{ette} velyllä, tai
enri eelällä, lupaus kettä? Aika
menee nopeasti.....

TEKSTI 9

Juuso on uusi nainen, tyttö.
 Hän ajoi polkupyörällä koulun,
 valitettavasti hän jäi auton alle.
 Oli soitettu ambulanssiin ja kutsuttu
 auttamaan. Hän sai ensi apua ja viety häntä
 sairaalaan kirurgian osastolle N 18.
 Oikea jalka on murtunut ja panta kipsiin.
 Tällä hetkellä tilante ei paranna,
 kipu on vähentynyt ja tyttö voi vähän
 viihtyä, pysty lukemaan romaania.
 Ainakin joutuin ihan samanlaisiin
 tilanteeseen pääsiäisenä, sain ensi
 apua (esimerkiksi huulet on ommeltu...)
 ja lähtenyt pois sairaalasta kotiin, koska
 niin makaa kotonaakin ihan samalla
 tavalla kuin sielläkin. Ityt minulla
 on vähän pääkipu ja opettaja huomaa
 että minä olen vähän erikoinen
 (kuunnollinen) ammatin jälkeen.
 Toivotaan että kaikki menee parimmalle
 tavalla. Jos ajan pyörällä täytyy olla
 varovainen.

LIITE 11**VALINTAVERKKOHAASTATTELUN KOKOAMISLOMAKKEET**

ARVIOIJAN NIMI: _____

TRI-ADI	TEKSTIT			PERUSTELUJA
I	1	2	3	
II	4	5	6	
III	7	8	9	
IV	1	4	7	
V	2	5	8	
VI	3	6	9	
VII	1	5	9	
VIII	2	6	7	
IX	3	4	8	

J Ä R J E S T Y S	KRITEERIT										Y L E I N E N
I											
II											
III											
IV											
V											
VI											
VII											
VIII											
IX											

KOMMENTTEJA: