

Eeva Tuokko

Mille tasolle perusopetuksen englannin opiskelussa päästään?

Perusopetuksen päättövaiheen kansallisen
arvioinnin 1999 Eurooppalaisen viitekehysten
taitotasoihin linkitetyt tulokset







ABSTRACT

Tuokko, Eeva

What level do pupils reach in English at the end of the comprehensive school? National assessment results linked to the Common European Framework

Jyväskylä: University of Jyväskylä, 2007, 337 p.

(Jyväskylä Studies in Humanities,

ISSN 1459-4331; 69)

ISBN 978-951-39-2767-7 (PDF), 978-951-39-2701-1 (nid.)

Diss.

In several EU countries the National Core Curricula for foreign languages are being revised to describe the target language proficiency in terms of the Common European Reference Levels¹. The Finnish National Core Curriculum for Basic Education 2004 is already based on them. Standards like that make it possible to compare test results not only nationally but also internationally, and thus increase the transparency of marks and grades, for example. For the linkage between test results and skills levels to be considered reliable the European Council's Language Policy Division has produced instructions which, among other things, call for a detailed documentation of the whole assessment process. Although other procedures can be applied, this set of instructions is widely used today.

The focus of this study is the evaluation of the English language skills as expressed in the Framework Curriculum for the Comprehensive School 1994. The aim was to link the results of the 1999 National Evaluation of the English Language Skills of Finnish 9th form pupils to the CEF scale (A1, A2, B1, B2, C1 and C2), and the OPS 2004² with the levels of A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2 and C1 from the lowest to the highest. To assess the productive skills one more scale, A2001³ (0-5) was applied. To find the correspondence, a test-centered standard setting method was applied. Standard setting for the receptive skills tests was based on the application of IRT analysis of the empirical data, the results of which are expressed in θ values. They indicate both the skills of the pupil and the item difficulty, which makes it possible to compare the difficulty of items in different tests. This cannot be done by using other methods. In the OPLM analysis programme used the θ values range between -3 (easy items, low ability) and +3 (difficult items, high ability). Item discrimination indicates the quality of the item and is shown on a scale of 1-15, the values normally achieved being 2-3, like in this study.

The results of the study showed that the average θ values of both the whole receptive skills test as well as those of the listening, reading and grammar tests were 0,27, which indicates good skills. The levels of the productive skills tests were rated on level descriptor scales. When setting the standards, the rater judgements and the empirical results were combined to establish the link between the test results and skills levels. The average CEF level attained was B1 and the OPS 2004 level B1.1, which means that the targeted Finnish Framework level in receptive skills was achieved. The medians of the speaking and writing tests were B1 on the CEF scale, B1.1 on the OPS 2004 scale and 4 on the A2 001 scale. The corresponding targeted Finnish level for both is A2.2. Thus, the targeted levels of the productive English skills were exceeded by about one level.

Key words: evaluation, assessment, English, skills scales, standard setting

¹ CEF; Common European Framework of Reference for Languages

² The Finnish National Core Curriculum for Basic Education 2004

³ The National Board of Education levels of productive language proficiency 2001

Author's address

Eeva Tuokko
e-mail:tuokko@pp.inet.fi
tel 040 5511 538

Supervisors

Researcher Ari Huhta
Department of Applied Linguistics
University of Jyväskylä

Docent Sauli Takala
Department of Applied Linguistics
University of Jyväskylä

Reviewers

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

Docent Heini-Marja Järvinen
Department of Teacher Education
University of Turku

Opponents

Ph.D. Taina Juurakko-Paavola
Docent Heini-Marja Järvinen

ESIPUHE

Kiinnostukseni arviointia kohtaan juontaa juurensa 1960-luvun lopulle, jolloin vieraiden kielten järjestelmällistä arviointia oppimistulosten seurannan välineenä suoritettiin silloisen Kouluhallituksen ja Kasvatustieteiden tutkimuslaitoksen yhteistyönä. Opettajat 11:ssä peruskoulukokeilukunnassa panivat toimeen kokeet, joiden laadintaan osallistuivat muutamat kokeilukoulujen opettajat (mm. tämän tutkimuksen tekijä) ja tutkimuslaitoksen tutkija (ohjaajani Sauli Takala). Oppilaiden vastaukset toimitettiin Jyväskylään analysoitaviksi ja edelleen tiedoksi ja toimenpiteitä varten Helsinkiin. Kussakin kokeilukunnassa oli vieraiden kielten ohjaava opettaja, ja erityisesti toisiaan lähellä olevien kuntien ohjaajat kehittivät ja järjestivät yhteisiä kokeita ja laativat arviointiasteikkoja.

Kolmen vuosikymmenen kuluttua minulla oli ilokseni tilaisuus palata arvioinnin pariin Opetushallituksen englannin ja ruotsin kielten kansallisen arvioinnin erikoissuunnittelijana. Ohjausryhmän jäsen Sauli Takala ehdotti heti väitöskirjan laatimista mielestään sekä kansallisesti että kansainvälisesti tärkeästä aiheesta. Silloinen esimieheni, fil. tri Ritva Jakku-Sihvonen, tuki hankettani myöntämällä minulle mahdollisuuden osallistua SOLKIn jatko-opiskelijaseminaareihin, joita johtivat laitoksen karismaattinen johtaja, edesmennyt professori Kari Sajavaara, ja lukuisien vieraiden kielten tutkijoiden esikuva ja oppisä, tutkimusprofessori Sauli Takala. Hänen uhrautuvaisuutensa monien väitöskirjaversioideni lukemisessa ja kommentoinnissa on ollut uskomaton. Samoin on hänen tiedossaan ja omistuksessaan oleva lähteiden määrä, joista viime mainituista monet ovat päätyneet hyllyihini useiksi vuosiksi. Sauli Takalalle olen suuren kiitoksen velkaa.

Väitöskirjatyöni kokoamisen aktiivisin vaihe, jolloin työskentelin pääasiassa kotoa käsin, osui aikaan, jolloin SOLKIn johtajaksi tuli professori Minna-Riitta Luukka. Huolimatta siitä, että emme olleet tavanneet käytännön opiskelun parissa, hän hahmotti nopeasti aiheeni puutteet ja ansiot ehdottaen uutta näkökulmaa, jota tuli hänen ehdotuksestaan tukemaan SOLKIn tutkija, fil. lis. Ari Huhta. Kiitokset heille neuvoista, kommentoinnista ja kannustuksesta.

Fil. maist. Liisa Havola paneutui tunnetulla tarmollaan ja tarkkuudellaan produktiivisen kielitaidon tuotosten arviointityöhön. Sofian yliopiston tutkimusmetodologian professori Felianka Kaftandjieva, joka on yksi harvoja taitotasojen asettamiseen perehtyneitä asiantuntijoita, vastasi tilastollisista analyyseistä. Opetushallituksen Arviointivirkamiehet -ryhmä suhtautui työstä pois jäämisenikin jälkeen myönteisesti ja kannustavasti työhöni. Erityisen kiitoksen ansaitsevat työtoverini Aulikki Etelälahti, jonka ”Eeva” -tiedostosta ja muistista löytyy yhä kaikki, mitä tarvitaan, Tuija Koskela, jonka hyvin järjestetty arkisto oli korvaamaton etsittäessä aineistoa, jota en alkuaan ollut huomannut pyytää, ja Sirpa Ropponen, jonka tietokoneella syntyvät kuviot käden käänteessä. Entinen kollegani, kuvataidelehtori Eija Tukiainen tuki työtäni havainnollistamalla tutkimukseni aiheen kansikuvaksi. Kiitos, Eija. Parhaat kiitokset myös fil. maist. Gudrun Ericksonille Göteborgin yliopistosta tiivistelmän sisällöllisestä ja ruot-

sin kielen tarkistamisesta sekä Tampereen Kääntäjänkoulutuslaitoksen emerituslehtori Gerard McAlesterille englanninkielisen yhteenvedon kielentarkistuksesta. Tamperelaisen Ristinarkun koulun positiivinen ja työstäni kiinnostunut opettajakunta puolestaan on korvannut pääosin yksin puurtaneelle amatööri-tutkijalle tutkijayhteisön, johon aluksi kuuluneet jatko-opiskelijat ovat jo joko valmistuneet tai keskeyttäneet.

Työni loppuvaiheessa saamistani korjaus- ja muokkausehdotuksista osoitan kiitokset työni esitarkastajille, fil. tri Taina Juurakko-Paavolalle ja dosentti Heini-Marja Järviselle. Työn tausta, tavoite ja tulokset hahmottuvat varmasti lukijalle selkeämpinä heiltä saamani asiantuntevan palautteen johdosta. Ilman Jyväskylän yliopiston julkaisuyksikön julkaisusihteerin Marja-Leena Tynkkysen joustavuutta ja ammattitaitoa työni ei taas olisi nähnyt päivänvaloa toivomallani aikataululla.

Taloudellisesti työtäni on tukenut Suomalainen Konkordia -liitto, jonka apurahan turvin olen voinut teettää produktiivisen kielitaidon koetuotosten arviointia sekä vaativia tilastollisia analyysejä. Jyväskylän yliopiston rehtorin myöntämä Konnevesi -stipendi antoi minulle tilaisuuden keskittyä kahden viikon ajan tilastollisten analyysien tulkitsemiseen täydellisessä maaseudun rauhassa biologisen tutkimuslaitoksen sihteerin Tarja Hultin ja ehtoisan emännän Raija Tiuhosen huolenpidossa.

Tutkimukseni kokoaminen vei pitkän ajan, ja se on erityisesti koetellut sosiaalisesti suuntautunutta aviomiestäni Ossia, jolle olen liian usein joutunut vastaamaan kieltävästi hänen ehdotettuaan vieraiden kutsumista tai vierailulle lähtemistä. Hänen omia sanojaan lainatakseni hän on "kestänyt, mutta ei kuitenkaan estänyt" tämän tutkimuksen tekemistä. Jo aikuiset lapsemme ovat myös ymmärtäneet työtäni, mutta he ovatkin vuosien varrella joutuneet tottumaan äitinsä monenlaisiin projekteihin. Tutkimustyö on varmasti yksi tapa testata perheen yhtenäisyys.

Viime aikoina on vilkkaan yhteiskunnallisen keskustelun kohteena ollut naisten uralla etenemistä estävä lasikatto. Se ei ole tutkijoillekaan tuntematon. Vaikka vuonna 2003 lähes puolet tohtorin tutkinnon suorittajista oli naisia, professorikunnassa heitä on vain viidennes. Tutkimuksen tekemisen tulisi kuitenkin palkita paitsi korkeakoulumaailmassa myös yritys-elämässä. Tämän vuoksi omistan työni reippaille koululaisille, tyttäreni tyttäreille, Kaarinalle, ja poikani tyttäreille, Marialle, toivoen heidänkin aikanaan löytävän tutkimuksesta niin iloa kuin hyötyäkin.

Tampereella 13.12.2006

Eeva Tuokko

KUVIOT

KUVIO 1	Tutkimuksen rakenne	23
KUVIO 2	CIPP -arviointimalli	34
KUVIO 3	Englannin puheen ymmärtäminen eri koulumuodoissa ja oppilasryhmissä vuosina 1971, 1979 ja 1991, keskiarvot	52
KUVIO 4	Puheen ymmärtämisen kokeen tulokset seitsemässä Euroopan unionin maassa	53
KUVIO 5	Tekstin ymmärtämisen kokeen tulokset seitsemässä Euroopan unionin maassa	53
KUVIO 6	Kieliopillisen kompetenssin kokeen tulokset seitsemässä Euroopan unionin maassa	54
KUVIO 7	Kirjoittamisen kokeen tulokset seitsemässä Euroopan unionin maassa	54
KUVIO 8	Kielellisen tiedon alueet	65
KUVIO 9	Kielen käytön ja kielikokeessa osoitetun taidon komponentit	67
KUVIO 10	Kielen käytön ulottuvuudet	69
KUVIO 11	Kuulijan auditiivisen syötöksen prosessointi	73
KUVIO 12	Kirjoitetun tekstin prosessointi	75
KUVIO 13	Puheen prosessointi	84
KUVIO 14	Puhujan prosessointikaavio	85
KUVIO 15	Tiedon kertomismallin rakenne	90
KUVIO 16	Tiedon muuntamismallin rakenne.....	91
KUVIO 17	Kirjallisen esityksen yleinen malli.....	92
KUVIO 18	Kirjoittamisprosessi	93
KUVIO 19	Protokieli ja kielen suhde ekososiaaliseen ja fyysiseen ympäristöön	97
KUVIO 20	Kieliopin ja pragmatiikan teoreettinen määritelmä	99
KUVIO 21	Kielioppitiedon, kieliopillisen taidon ja kielellisen suorituksen välinen suhde	100
KUVIO 22	Kokeen laadinta	106
KUVIO 23	Kriteeri- ja normiviitteisen arvioinnin jatkumo	116
KUVIO 24	Suullisen kielitaidon tyypit	135
KUVIO 25	Sanaston arvioinnin dimensiot	146
KUVIO 26	Linkittämisprosessi	152
KUVIO 27	Kahden osaotoksen muuttujien välinen yhteys	163
KUVIO 28	Tehtäväsarjan 1 faktoreiden ominaisarvot	165
KUVIO 29	Tehtäväsarjan 2 faktoreiden ominaisarvot	165
KUVIO 30	Tehtäväsarjan 3 faktoreiden ominaisarvot	166
KUVIO 31	Linkittämisen kokonaiskuvaus	168
KUVIO 32	Kielitaidon arvioinnin vaiheet	173
KUVIO 33	Kielenkäyttöalueet	175
KUVIO 34	Reseptiivisen kielitaidon ratkaisuprosenttien (P) jakauma.....	197
KUVIO 35	Reseptiivisen kielitaidon koeosoiden θ -arvojen jakauma (b)	198
KUVIO 36	Reseptiivisen kielitaidon osiokohtainen θ -arvojen jakauma	199

KUVIO 37	Reseptiivisen kielitaidon koeosoiden erottelukykyindeksien (D) jakauma	199
KUVIO 38	Reseptiivisen kielitaidon koeosoiden erottelukykyindeksien (a) jakauma	200
KUVIO 39	Tehtäväsarjan 1 reseptiivisen kielitaidon koeosoiden IRT -vaikeustasojakauma	201
KUVIO 40	Tehtäväsarjan 2 reseptiivisen kielitaidon koeosoiden IRT -vaikeustasojakauma	201
KUVIO 41	Tehtäväsarjan 3 reseptiivisen kielitaidon koeosoiden IRT -vaikeustasojakauma	202
KUVIO 42	Reseptiivisen kielitaidon kokeiden tehtäväsarjojen vaikeustason vertailu	202
KUVIO 43	Reseptiivisen kielitaidon osa-alueiden kokeiden vaikeustason vertailu	205
KUVIO 44	Tekstin ja puheen ymmärtämisen sekä kieliopillisen kompetenssin koeosoiden θ -arvot	206
KUVIO 45	Kieliopillisen kompetenssin, puheen ja tekstin ymmärtämisen kokeiden keskiarvot ja niiden 95 % luottamusväli	207
KUVIO 46	Asiantuntijoiden CEF -taitotasoasteikoilla suorittama reseptiivisen kielitaidon koeosoiden vaikeustason arviointi	208
KUVIO 47	Reseptiivisen kielitaidon kokeen osioiden CEF -taitotasoarvioiden keskiarvot ja 95 % luottamusväli	209
KUVIO 48	Reseptiivisen kielitaidon kokeiden yksittäisten koeosoiden sijoittuminen CEF -taitotasoille asiantuntijoiden arvioinnin perusteella	210
KUVIO 49	CEF -taitotasoille asettuneiden osioiden tarkistetut θ -arvojen jakaumat	211
KUVIO 50	Reseptiivisen kielitaidon yksittäisten koeosoiden tarkistettu jakautuminen CEF -taitotasoille	212
KUVIO 51	Tarkistetun CEF -taitotasojen määrittämisen jälkeen eri taitotasoille asettuneiden reseptiivisen kielitaidon koeosoiden keskiarvot ja 95 % luottamusväli	212
KUVIO 52	Asiantuntijoiden OPS 2004 -taitotasoarvioiden mukainen reseptiivisen kokeen yksittäisten osioiden sijoittuminen taitotasoille	214
KUVIO 53	Asiantuntijoiden OPS 2004 -taitotasoarvioiden mukainen reseptiivisen kokeen osioiden sijoittuminen taitotasoille	214
KUVIO 54	Asiantuntijoiden arvioiden mukaiset OPS 2004 -taitotasojen θ -keskiarvot ja niiden 95 % luottamusväli	215
KUVIO 55	Reseptiivisen kielitaidon kokeen yksittäisten osioiden tarkistettu jakaantuminen OPS 2004 -taitotasoille	217
KUVIO 56	Eri OPS 2004 -taitotasoille asettuneiden osioiden tarkistetut θ -arvojen jakaumat	217
KUVIO 57	Tarkistetun OPS 2004 -taitotasojen määrittämisen jälkeen eri taitotasoille asettuneiden reseptiivisen kielitaidon koeosoiden θ -keskiarvot ja niiden 95 % luottamusväli	218
KUVIO 58	Koko reseptiivisen kielitaidon kokeen tulos (θ)	220

KUVIO 59	Tekstin ymmärtämisen kokeen tulos (θ)	220
KUVIO 60	Puheen ymmärtämisen kokeen tulos (θ)	221
KUVIO 61	Kieliopillisen kompetenssin kokeen tulos (θ)	222
KUVIO 62	Reseptiivisen kielitaidon osakokeiden tulosten (θ) vertailu	222
KUVIO 63	Reseptiivisen kielitaidon osakokeiden keskiarvot (θ) ja niiden 95 % luottamusväli	223
KUVIO 64	Koko reseptiivisen kokeen tulos CEF -taitotasoasteikolle sijoitettuna	225
KUVIO 65	Koko reseptiivisen kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna	225
KUVIO 66	Tekstin ymmärtämisen kokeen tulos CEF -taitotasoasteikolle sijoitettuna	226
KUVIO 67	Tekstin ymmärtämisen kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna	227
KUVIO 68	Puheen ymmärtämisen kokeen tulos CEF taitotasoasteikolle sijoitettuna	228
KUVIO 69	Puheen ymmärtämisen kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna	229
KUVIO 70	Kieliopillisen kompetenssin kokeen tulos CEF -taitotasoasteikolle sijoitettuna	230
KUVIO 71	Kieliopillisen kompetenssin kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna	231
KUVIO 72	Puhumisen kokeen A2001 ja OPS 2004 -taitotasojen korrespondenssianalyysi	235
KUVIO 73	Puhumisen kokeen OPS 2004 - ja CEF -taitotasojen korrespondenssianalyysi	235
KUVIO 74	Kirjoittamisen kokeen A2001- (sujuvuus) ja OPS 2004 -taitotasojen korrespondenssianalyysi	237
KUVIO 75	Kirjoittamisen kokeen OPS 2004- ja CEF -taitotasojen korrespondenssianalyysi	237
KUVIO 76	Puhumisen kokeen tulosten ekstrapoloitu jakauma CEF -taitotasoasteikolle sijoitettuna	239
KUVIO 77	Puhumisen kokeen tulosten ekstrapoloitu jakauma OPS 2004 -taitotasoasteikolle sijoitettuna	240
KUVIO 78	Puhumisen kokeen tulosten ekstrapoloitu jakauma A2001- (sujuvuus) taitotasoasteikolle sijoitettuna	241
KUVIO 79	Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma CEF -taitotasoasteikolle sijoitettuna	245
KUVIO 80	Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma OPS 2004 -taitotasoasteikolle sijoitettuna	246
KUVIO 81	Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma A2001 -taitotasolle sijoitettuna	247

TAULUKOT

TAULUKKO 1	Eurooppalaisen viitekehyksen (CEF) yleisen kielitaidon taitotasojen A1 ja C2 kuvaukset	20
TAULUKKO 2	Arvioinnin ja tutkimuksen välisiä eroja	27
TAULUKKO 3	Arvioinnin yleisimmät lähestymistavat, niihin liittyvä demokratiakäsitys, tunnetuimmat edustajat ja pääteos	29
TAULUKKO 4	Arviointiprosessia kuvaavia luokitteluja	30
TAULUKKO 5	Suomen opetustoimen arviointijärjestelmä	38
TAULUKKO 6	OPS 2004:n kuvaus hyvän oppilaan englannin osaamisesta 6. ja 9. luokan päättyessä	41
TAULUKKO 7	OPS 2004:n kuvaus viestintästrategioiden hallinnasta	41
TAULUKKO 8	Ohjelma-arviointistandardien yhteenveto	42
TAULUKKO 9	Opetushallituksen kahden viimeisen englannin arvioinnin (1993, 1999) tulosten vertailun keskeiset johtopäätökset	50
TAULUKKO 10	Tekstin ymmärtämisen kokeen tehtävien keskimääräiset ratkaisuprosentit teksteittäin vuosina 1971-83	52
TAULUKKO 11	Carrollin kielitaidon testaamisen luokitusjärjestelmä	61
TAULUKKO 12	Strategisen kompetenssin käyttö	66
TAULUKKO 13	Kommunikatiivisen kielitaidon komponentit	68
TAULUKKO 14	Puheen ymmärtämisen taitojen kognitiivinen luokittelu Valetten mukaan	76
TAULUKKO 15	Puheen ymmärtämisen taitojen kommunikatiivinen luokittelu	77
TAULUKKO 16	Kuuntelijan taidot	78
TAULUKKO 17	Tekstin ymmärtämisen taidot	79
TAULUKKO 18	Lukemisen toiminnot	81
TAULUKKO 19	Suullisen ja kirjallisen tuottamisen vertailua	83
TAULUKKO 20	Informaation järjestämismenetelmät	86
TAULUKKO 21	Makro- ja mikrofunktiot	87
TAULUKKO 22	Puhumisen konstruktin määrittelemineen	87
TAULUKKO 23	Kirjoittamisen taksonomia	94
TAULUKKO 24	Sanastolliset tiedot ja taidot	102
TAULUKKO 25	Sanan tuntemisen komponentit	103
TAULUKKO 26	Arviointimenetelmiä	114
TAULUKKO 27	Koetyypin valinnan ja arviointitarkoituksen välinen suhde	115
TAULUKKO 28	Produktiivisen kielitaidon suoritusten pisteittämisvaihtoehdot	118
TAULUKKO 29	Reliaabeliuden vähenemisen suhde arviointiasteikon porrastuksen vähenemiseen	121
TAULUKKO 30	Holististen ja analyyttisten asteikkojen vertailu suhteessa kokeen käyttökelpoisuuden ominaisuuksiin	122
TAULUKKO 31	Kysymysten kognitiivisen vaikeustason jatkumo	127
TAULUKKO 32	Ymmärtämisen kokeen tehtävätyyppien ansioita ja puutteita	129
TAULUKKO 33	Puheen ymmärtämisen helpottamiskeinoja	130

TAULUKKO 34	Puheen ymmärtämisen tehtävissä huomioon otettavia asioita	131
TAULUKKO 35	Suullisen tuottamisen laadulliset kriteerit A2 -tasolla	134
TAULUKKO 36	Taitotasoasteikkojen valitseminen kriteerien arviointiin	135
TAULUKKO 37	Suullisen tuottamisen tehtävätyyppien vertailua	138
TAULUKKO 38	Esimerkki yleisen kirjoittamisen taidon testispesifikaatioista ..	140
TAULUKKO 39	Suorien kirjoittamistehtävien dimensiot	141
TAULUKKO 40	Esimerkkejä kieliopillista kompetenssia mittaavista tehtävistä	143
TAULUKKO 41	Kielen käytön kokeen yleiskatsaus	143
TAULUKKO 42	Kieliopillisen taidon analyttinen arviointiasteikko	144
TAULUKKO 43	Puheen ymmärtämisen kokeen arvioinnissa käytetyt taitotasoasteikot	159
TAULUKKO 44	Tekstin ymmärtämisen kokeen arvioinnissa käytetyt taitotasoasteikot	159
TAULUKKO 45	Puhumisen kokeen arvioinnissa käytetyt taitotasoasteikot	161
TAULUKKO 46	Reseptiivisen kielitaidon kokeen eri tehtäväsarjojen pääkomponenttianalyysin tulos	164
TAULUKKO 47	Faktorimallin sopivuuden tarkastelu χ :n neliötestin avulla	164
TAULUKKO 48	Englannin kansallisen arvioinnin tavoitteet ja toimenpiteet	171
TAULUKKO 49	Ymmärtämisen operaatiot	176
TAULUKKO 50	Kirjallisiin kokeisiin osallistuneiden oppilaiden otostaminen.....	179
TAULUKKO 51	Esikokeilun kirjoittamisen kokeen aihe-ehdotusten lukumäärät ja vaihtoehdot	181
TAULUKKO 52	Tekstin ymmärtämisen koeversioiden sisältö	184
TAULUKKO 53	Tekstin ymmärtämisen kokeen aihepiirit ja viestintätehtävät	185
TAULUKKO 54	Tekstin ymmärtämisen kokeessa edellytetyt operaatiot	185
TAULUKKO 55	Puheen ymmärtämisen koeversioiden sisältö	186
TAULUKKO 56	Puheen ymmärtämisen kokeen aihepiirit ja viestintätehtävät	186
TAULUKKO 57	Puheen ymmärtämisen operaatiot	187
TAULUKKO 58	Kieliopillisen kompetenssin kokeen ankkuri- ja uniikkiosoiden lukumäärät	189
TAULUKKO 59	Puhumisen kokeen tehtävien jakaantuminen eri tehtävävihkoihin	190
TAULUKKO 60	Kirjoittamisen kokeen aiheiden kuvaus	191
TAULUKKO 61	Reseptiivisen kielitaidon kokeen vaikeustaso ja erottelukyky	197
TAULUKKO 62	Reseptiivisen kielitaidon kokeen tehtäväsarjojen osioiden vaikeustason ja erottelukyvyn vertailu	200
TAULUKKO 63	Eri tehtäväsarjojen ja niiden osakokeiden keskiarvo, -hajonta ja reliaabelius	203
TAULUKKO 64	Reseptiivisen kielitaidon koeosioiden lukumäärät erottelukykyindeksiryhmittäin	204

TAULUKKO 65	Reseptiivisen kielitaidon kokeen ja sen osa-alueiden koeosoiden vaikeustaso ja erottelukyky	204
TAULUKKO 66	Asiantuntijoiden arvioima koeosoiden jakaantuminen CEF -taitotasoasteikoilla.....	209
TAULUKKO 67	Reseptiivisen kielitaidon CEF -taitotasoasteikon tarkistettut välykset (θ)	210
TAULUKKO 68	Reseptiivisen kielitaidon koeosoiden tarkistettu CEF -taitotaso (θ).....	211
TAULUKKO 69	Asiantuntijoiden OPS 2004 -taitotasoasteikon mukaan arvioimat reseptiivisen kielitaidon koeosoiden jakaumat	213
TAULUKKO 70	Reseptiivisen kielitaidon kokeiden OPS 2004 -taitotasojen tarkistettut välykset (θ)	215
TAULUKKO 71	Reseptiivisen kielitaidon koeosoiden tarkistettu OPS 2004 -taitotasoasteikko (θ)	216
TAULUKKO 72	CEF- ja OPS 2004 -taitotasoasteikkojen vastaavuus reseptiivisen kielitaidon kokeita arvioitaessa	218
TAULUKKO 73	Koko reseptiivisen kielitaidon kokeen ja sen osakokeiden tulokset (θ)	219
TAULUKKO 74	Koko reseptiivisen kielitaidon kokeen ja sen osakokeiden korrelaatiot ja niiden merkitsevyyden testaus (2 -suuntainen, N = 4854)	223
TAULUKKO 75	Reseptiivisen kielitaidon kokeen CEF- ja OPS 2004 -taitotasoasteikkojen mediaanit, moodit ja kvartiilit	224
TAULUKKO 76	Kielitaidon yleiskuvaus CEF -asteikon B1 -tasolla ja OPS 2004:n kuvaus perusopetuksen päättövaiheen kielitaidosta	226
TAULUKKO 77	Oppilaiden tekstin ymmärtämisen taidon kuvaus CEF -asteikon B1 -tasolla ja OPS 2004 -asteikon tasolla B1.1	227
TAULUKKO 78	Puheen ymmärtämisen taidon kuvaus CEF -asteikon B1 -tasolla ja OPS 2004 -asteikon tasolla B1.1	
TAULUKKO 79	Tulosten jakaumat kokeissa (%) CEF- ja OPS 2004 -asteikoilla	229
TAULUKKO 80	Puhumisen kokeen A2001- ja OPS 2004 taitotasojen frekvenssien ristiintaulukointi	232
TAULUKKO 81	Puhumisen kokeen taitotasojen vastaavuus	234
TAULUKKO 82	Kirjoittamisen kokeen taitotasojen vastaavuus	236
TAULUKKO 83	Saavutetun puhumisen keskimääräisen taitotason B1.1 ja tavoitetason A2.2 kuvaus OPS 2004 -asteikolla	240
TAULUKKO 84	Puhumisen kokeen mediaanit ja moodit pyöristettyinä keskiarvoina CEF-, OPS 2004- ja A2001-taitotasoasteikkojen mukaan	243
TAULUKKO 85	Puhumisen kokeen taitotasoasteikkojen vastaavuus sekä kunkin tason suoritusten ekstrapoloidut frekvenssijakaumat ja niiden prosenttiosuudet	244
TAULUKKO 86	Kirjoittamisen taitotasojen A2.2 ja B1.1 kuvaukset OPS 2004 -asteikolla	246
TAULUKKO 87	Kirjoittamisen kokeen mediaanit ja moodit CEF-, OPS-2004- ja A2001 -taitotasoasteikoilla	248

TAULUKKO 88	Kirjoittamisen kokeen ekstrapoloitu frekvenssijakauma ja prosenttiosuudet CEF-, OPS 2004- ja A2001-taitotasoilla	248
TAULUKKO 89	Reseptiivisen kokeen frekvenssijakauma ja prosenttiosuudet (oppilaat) CEF -taitotasoasteikolla arvioituina	250
TAULUKKO 90	Reseptiivisen kokeen frekvenssijakauma ja prosenttiosuudet (oppilaat) OPS 2004 -taitotasoasteikolla arvioituina.	250
TAULUKKO 91	Puhumisen arvioinnin reliabeliuskerroin (α) ja kirjoittamisen kokeen arvioijien välinen korrelaatio	252
TAULUKKO 92	Taitotasojen validointi	256
TAULUKKO 93	Eri taitotasoasteikkojen väliset korrelaatiot asiantuntijoiden suorittamassa puhumisen ja kirjoittamisen arvioinnissa	258

ESIMERKIT

ESIMERKKI 1	CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava tekstin ymmärtämisen tehtävä	228
ESIMERKKI 2	CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava puheen ymmärtämisen tehtävä	230
ESIMERKKI 3	CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava kieliopillisen kompetenssin tehtävä	231
ESIMERKKI 4	Transkriboitu keskimääräistä puhumisen tasoa kaikilla taitotasoasteikoilla (CEF, OPS 2004 ja A2001) edustava tuotos	242
ESIMERKKI 5	Kaikilla taitotasoilla (CEF, OPS 2004 ja A2001) keskimääräiseksi arvioitu kirjoittamisen tuotos	247

SISÄLLYS

ABSTRACT

ESIPUHE

KUVIOT JA TAULUKOT

ESIMERKIT

SISÄLLYS

1	JOHDANTO	19
2	ARVIOINTI	25
2.1	Koulutuksen arviointi	31
2.1.1	Kansallinen koulutuksen arviointi	34
2.2	Koulutuksen arvioinnin etiikka ja standardit	42
2.3	Yhteenveto	45
3	ENGLANNIN KIELEN OPPIMISTULOSARVIOINTEJA.....	47
3.1	Suomen perusopetuksen englannin kielen arviointeja	47
3.2	Kansainvälinen arviointiyhteistyö	51
3.3	Kansalliset arvioinnit muissa Euroopan maissa	54
3.3.1	Ruotsin kansallisen englannin ainekokeen kehittäminen ...	55
3.3.2	Kansallinen englannin päättöarviointi Hollannissa	57
3.4	Yhteenveto	58
4	KIELITAIDON ARVIOINNIN LÄHTÖKOHTIA.....	60
4.1	Testauksen perustana käytettyjä kielitaitokäsityksiä	60
4.2	Prosessi ja tuotos kielitaidon arvioinnin kohteina	71
4.2.1	Ymmärtäminen	71
4.2.2	Tuottaminen	82
4.3	Kieliopillinen kompetenssi	95
4.3.1	Lauseoppikeskeinen kielioppikäsitys	96
4.3.2	Kommunikatiivinen kielioppikäsitys	96
4.4	Sanastollinen kompetenssi	100
4.4.1	Sanan ´osaaminen`	101
4.5	Yhteenveto	103
5	KIELITAIDON KOKEEN LAADINTA JA ARVIOINNIN TOIMEENPANO	105
5.1	Kokeen laadinnan malli	105
5.1.1	Suunnitteluvaihe	107
5.1.2	Operationaalistaminen	109
5.1.3	Toimeenpaneminen	111
5.1.4	Kokeen käyttökelpoisuuden piirteiden arviointi	111

5.2	Arviointimenetelmät	113
5.3	Arviointiasteikon valinta	117
5.4	Yhteenveto	122
6	KIELITAIDON OSA-ALUEIDEN ARVIOINTI	124
6.1	Ymmärtäminen	124
6.1.1	Kysymysten vaikeusasteen hierarkia	125
6.1.2	Puheen ymmärtämisen arviointi	129
6.1.3	Tekstin ymmärtämisen arviointi	131
6.2	Tuottaminen	134
6.2.1	Puhumisen arviointi	135
6.2.2	Kirjoittamisen arviointi	139
6.3	Kieliopillinen kompetenssi	142
6.4	Sanastollinen kompetenssi	145
6.5	Yhteenveto	147
7	TUTKIMUSAINEISTO JA -METODI	149
7.1	Tutkimuksen tausta	149
7.2	Tutkimuskysymykset ja tutkimusaineisto	150
7.3	Taitotasojen asettamisen teoreettinen tausta	152
7.4	Klassinen ja uusi testiteoria (Item Response Theory = IRT)	155
7.5	Taitotasojen asettamisen käytännön toimenpiteet	157
7.5.1	Puheen ja tekstin ymmärtämisen sekä kieliopillisen kompetenssin kokeiden taitotasojen arviointi	157
7.5.2	Puhumisen ja kirjoittamisen kokeiden taitotason arviointi	160
7.5.3	Aineistojen tilastollinen käsittely	162
7.6	Luotettavan linkittämisprosessin edellytykset	167
7.7	Yhteenveto	168
8	PERUSKOULUN 9. VUOSILUOKAN ENGLANNIN OPPIMISTULOSTEN KANSALLINEN ARVIOINTI	170
8.1	Alkuperäisen arvioinnin tausta ja tavoite	171
8.2	Arvioinnin suunnittelu	172
8.3	Varsinaisen arvioinnin valmistelu	178
8.4	Kokeet	183
8.5	Arvioinnin toimeenpano	191
8.6	Yhteenveto	194
9	TULOKSET	196
9.1	Reseptiivisen kielitaidon kokeet	196
9.1.1	Koko kokeen vaikeustaso ja erottelukyky	197
9.1.2	Eri tehtäväsarjojen osioiden vaikeustason ja erottelukyvyn vertailua	200

9.1.3	Kielitaidon osa-alueiden koeosioiden vaikeustason ja erottelukyvyn vertailua	203
9.2	Reseptiivisen kielitaidon kokeiden taitotasojen asettaminen	207
9.2.1	Asiantuntijoiden suorittamien taitotasoarviointien tulokset CEF -asteikoille sijoitettuina	208
9.2.2	Asiantuntijoiden suorittamien taitotasoarviointien tulokset OPS 2004 -asteikolle sijoitettuina	213
9.3	Reseptiivisen kielitaidon kokeiden tulokset taitotasoittain	219
9.3.1	Koko koe	224
9.3.2	Tekstin ymmärtämisen koe	226
9.3.3	Puheen ymmärtämisen koe	228
9.3.4	Kieliopillisen kompetenssin koe	230
9.4	Produktiivisen kielitaidon kokeet	232
9.4.1	Produktiivisen kielitaidon kokeiden taitotasojen asettaminen	233
9.5	Produktiivisen kielitaidon kokeiden tulokset taitotasoittain	238
9.5.1	Puhumisen kokeen tulokset	238
9.5.2	Kirjoittamisen kokeen tulokset	244
9.6	Yhteenveto	249
10	TUTKIMUKSEN LUOTETTAVUUDEN ARVIOINTI	251
10.1	Reliaabelius	251
10.1.1	Reseptiivisen kielitaidon kokeen reliaabelius	252
10.1.2	Produktiivisen kielitaidon kokeen reliaabelius	252
10.2	Validius	253
10.2.1	Kokeen sisältö- ja käsitevalidius	254
10.2.2	Taitotasojen asettamisen validius	255
10.3	Tutkimuksen ongelmia	258
10.4	Yhteenveto	259
11	JOHTOPÄÄTÖKSET	261
11.1	Tutkimuksessa todettua	261
11.2	IRT -menetelmien käytön edut	265
11.3	Arvioinnin kehittäminen	267
11.4	Kansainvälinen arviointiyhteistyö	268
	SUMMARY	270
	SAMMANFATTNING	277
	TIIVISTELMÄ	278
	LÄHTEET	279
	LIITTEET	292

1 JOHDANTO

Suomessa astui 1.1.1999 voimaan uusi perusopetuslaki, joka poikkesi edeltäjästään mm. siinä, että se edellytti opetuksen järjestäjiä osallistumaan ulkopuolisiin arviointeihin¹. Opetushallituksesta tuli tällöin kansallisten oppimistulosarviointien toimeenpanija.

Tässä tutkimuksessa, joka kohde on luonteeltaan ns. ohjelma-arviointia (*program evaluation*), käsitellään vuoden 1999 keväällä järjestettyä, vuoden 1994 perusopetuksen opetussuunnitelman perusteiden pohjalta laadittua perusopetuksen päättövaiheen englannin kielen kansallista arviointia. Se oli monessa suhteessa uraa uurtava. Se oli ensimmäinen ohjelma- ja oppimistulosten arvioinnin teorioiden periaatteita systemaattisesti noudattamaan pyrkivä Opetushallituksen Arviointivirkamiehet -ryhmän alaisuudessa toimeenpantu vieraiden kielten kansallinen arviointi. Sen oppilasotos, lähes 6 000, oli suurin Suomessa perusopetuksen arviointiin osallistunut. Otos oli suuri myös kansainvälisen mittapuun mukaan. Kokeiden laadintaa valvoi arvioinnin ja opetuksen asiantuntijoista koottu ohjausryhmä, analyyseistä vastasi arviointiin perehtynyt metodikko, arvioinnin laadintaprosessi dokumentoitiin, ja koulut saivat palautteen oppimistuloksistaan.

Kokeeseen sisältyi puheen ja tekstin ymmärtämisen, suullisen ja kirjallisen tuottamisen sekä kieliopillisen kompetenssin (kieliopilliset rakenteet) koe. Sanastoa testattiin soveltuvin osin ymmärtämiskokeiden yhteydessä. Suullisen kokeen järjestäminen tiedettiin vaativaksi, mutta se haluttiin sisällyttää arviointiin siksi, että sen toivottiin lisääntinä tuovan opetukseen positiivista takaistusvaikutusta. Peruskoulun opetussuunnitelman perusteissa on jo 1970-luvulta lähtien asetettu suullinen kielitaito kielenopetuksen ensisijaiseksi tavoitteeksi, mutta sen opettaminen ja erityisesti testaaminen ei ole koulun tasolla edennyt monista syistä johtuen yhtä voimakkaasti kuin muilla kielitaidon osa-alueilla.

Vuoden 2004 syksyllä Suomessa otettiin käyttöön uudet perusopetuksen opetussuunnitelman perusteet, joiden mukaan kielitaitoa arvioidaan taitotasosteikoilla (OPS 2004). Nämä taitotasot perustuvat Euroopan neuvoston

¹ Ulkopuolisella arvioinnilla tarkoitetaan muiden tahojen kuin kuntien itsensä toimeenpanemia arviointeja.

kielipolitiikan jaoksen julkaisemiin Eurooppalaisen viitekehyksen (Common European Framework of Reference for Languages; (CEF(R)) taitotasoasteikkoihin, joita on muokattu ja muokataan parhaillaan useissa Euroopan unionin maissa kansallisten oppimistulosarviointien pohjaksi. Jaos on laatinut myös ohjeet vastaavuuden määrittämiseksi CEFin ja erilaisten kansallisten taitotasoasteikkojen välillä sekä tutkinto- ja koetulosten linkittämisestä CEFiin. Niissä edellytetään mm. tutkinnon keskeisen sisällön ja kommunikatiivisten aktiiviteettien, kokeen laadintaprosessin, pisteityksen ja tulosten sekä kokeen analysoinnin kuvailua. Näin toivotaan voitavan löytää vastaavuus eri maissa annettujen vieraiden kielten arvosanojen välillä.

Linkittämisellä haetaan siis vastaavuutta erilaisten asteikkojen taitotasojen välillä. Ymmärtämiskokeiden yhteydessä tämä edellyttää tekstien ja tehtävien vaikeustason arvioimista sekä näiden arvioiden ja todellisten koetulosten (= taitojen) yhdistämistä. Tuottamistaidot (puhuminen ja kirjoittaminen) arvioidaan suoraan taitotasojen kuvauksia käyttäen. Näin mahdollistetaan taitotasojen asettaminen eli pisterajojen katkaisukohtien löytäminen niin, että näiden väliset tasot vastaavat toisiaan.

Taitotasoilla tarkoitetaan portaittain vaikeutuvia kielitaidon kuvauksia, joilla ilmaistaan, mitä kielenkäyttäjän tulisi osata kielellä tehdä osoittaakseen olevansa tietyllä tasolla. Kokonaiskielitaitoa kuvaava CEF -taitotasoasteikko koostuu kuudesta tasosta: A1, A2, B1, B2, C1 ja C2. Taulukossa 1 esitetään taitotasokäsitteen selventämiseksi CEF -asteikon yleisen kielitaidon taitotasokuvausten ääripäät.

TAULUKKO 1 Eurooppalaisen viitekehyksen (CEF) yleisen kielitaidon taitotasojen A1 ja C2 kuvaukset. (Lähde: Eurooppalainen viitekehys 2003:48.)

A1	C2
Ymmärtää ja käyttää tuttuja arkipäivän ilmauksia ja perustason sanontoja, joiden tavoitteena on yksinkertaisten, konkreettien tarpeiden tyydyttäminen. Pystyy esittäytymään ja esittelemään muita. Pystyy vastaamaan itseään koskeviin kysymyksiin ja kysymään vastaavia kysymyksiä muilta, esimerkiksi, missä he asuvat, keitä he tuntevat ja mitä heillä on. Pystyy käymään yksinkertaisia keskusteluja, jos puhekumppani puhuu hitaasti ja selvästi ja on valmis auttamaan.	Ymmärtää yleensä vaikeuksitta kaikenlaista puhuttua ja kirjoitettua kieltä. Osaa yhdistellä tietoja erilaisista puhutuista ja kirjoitetuista lähteistä ja rakentaa niissä esitetyistä perusteluista ja selostuksista sisällöllisesti yhtenäisen esityksen. Pystyy ilmaisemaan ajatuksiaan spontaanisti, erittäin sujuvasti ja täsmällisesti. Pystyy erottamaan merkitysvivahteet mutkikkaisinkin tilanteissa.

Tämän tutkimuksen päätavoitteena on linkittää Suomessa vuonna 1999 toimeenpannun perusopetuksen päättövaiheen kansalliseen englannin arviointiin osallistuneiden oppilaiden tulokset Eurooppalaisen viitekehyksen taitotasoihin Euroopan neuvoston kielipolitiikan jaoksen määrittämien standardien mukaisesti. Toiseksi tavoitteeksi asetettiin tulosten linkittäminen myös vuoden 1994 opetussuunnitelman perusteiden (OPS 2004) taitotasokuvauksiin. Näiden kahden tavoitteen lisäksi päätettiin täydentää vuoden 1999 arvioinnin raportointia (Tuokko 2000) produktiivisen kielitaidon arviointitulosten osalta sisällyttämällä tähän tutkimukseen myös

A2001 -taitotasoasteikkoja² käyttäen saadut tuottamistaitojen kokeiden tulokset ja pyrkimällä ilmaisemaan myös tällä asteikolla saadut tulokset CEF- ja OPS 2004 -taitotasoasteikoilla.

Taitotasojen vastaavuuden selvittäminen eli taitotasorajojen asettaminen on tärkeää useasta syystä. Euroopan unionin yksi pääperiaatteista on taata asukkaiden mahdollisuus liikkuvuuteen, mikä sisältää ajatuksen opiskelijoiden kotimaassaan saavuttamien tutkinto- ja koetulosten hyväksymisestä toisessa unionin maassa. Linkittämistoimenpiteitä, jotka perustuvat CEF -taitotasoasteikkoihin, on meneillään useissa maissa. Suomessa Jyväskylän yliopiston hallinnoimat Yleiset kielitutkinnot, ylioppilastutkinnon A-englannin ja B-ruotsin sekä ammattikorkeakoulujen kielitaidon arviointiasteikot (AMKKIA) on linkitetty CEFiin. Myös perusopetuksen kielitaidon tilasta pohjana jatko-opinnoille on tarpeen saada tietoa. Suomi Euroopan unionin jäsenmaana tulee Eurooppa-neuvoston vuonna 2002 tekemän päätöksen mukaan osallistumaan kansalaistensa vieraiden kielten taidon selvittämiseen (European Indicator of Language Competence), jolloin taitokuvauksina tullaan käyttämään CEFiä. Tämä tutkimus antaa pohjaa Suomen panokselle kyseisen indikaattorin kehittämisessä.

Taitotasojen asettamismenettely antaa kokeiden laadintaan lisäulottuvuuden. Tässä tutkimuksessa käytetty menetelmä edellyttää ns. uuden testiteorian mukaista IRT -analyysien käyttöä. Näillä analyyseillä saadaan selville sekä koeosoiden vaikeustaso että oppilaiden taidot samalla asteikolla. Kun käytettyjen osioiden vaikeustaso on tiedossa, niitä ankkureina käyttäen voidaan laatia myös uusia kokeita, joiden vaikeustaso myös tiedetään jo ennakolta. Tätä tutkimusta varten suoritettavat analyysit palvelevat siis myös tulevien kansallisten englannin arviointien laadintaa. Lisäksi IRT -analyysiä käyttämällä voidaan eri kokeissa saavutettuja tuloksia verrata keskenään, jos aineisto täyttää analyyysien pohjana olevat tilastolliset oletukset, ts. jos aineisto sopii malliin. Se siis mahdollistaa esimerkiksi puheen ja tekstin ymmärtämisen tulosten välisen vertailun.

Vuoden 1999 englannin kansallisen arvioinnin tulokset analysoitiin ja raportoitiin vuonna 2000 käyttämällä klassisen testiteorian menetelmiä, jotka soveltuvat luontevimmin normiviitteiseen arviointiin, mikä yksinkertaistettuna tarkoittaa oppilaiden sijoittamista paremmuusjärjestykseen koetulosten perusteella. Vastaavanlaisissa opetussuunnitelmien perusteiden tavoitteisiin perustuvissa kansallisissa arvioinneissa on kuitenkin siirrytty yhä enemmän analysoimaan oppimistuloksia uuden testiteorian menetelmillä. Ne soveltuvat paremmin ns. kriteeriviitteiseen arviointiin, jolla tarkoitetaan sitä, että testattavien taitoja arvioidaan suhteessa tiettyihin kriteereihin eli opittaviin asioihin. Ko-keessa testattavien asioiden tulee olla edustava otos opetussuunnitelman perusteiden sisällöstä, jotta tulokset voidaan yleistää koskemaan sitä. Vuoden 1999 kansallisen englannin arvioinnin mittarit oli jo laadittu kriteerimittauksen pe-

² A2001 -taitotasoasteikot ovat Opetushallituksessa vuonna 2001 laaditut puhumisen ja kirjoittamisen arviointiin tarkoitetut asteikot, joita käytettiin vuosina 2001-2003 ruotsin, finskan, ranskan, saksan ja venäjän kansallisissa arvioinneissa. Näin saadut englannin tulokset ovat vertailukelpoisia noiden kielten kansallisten arviointitulosten kanssa.

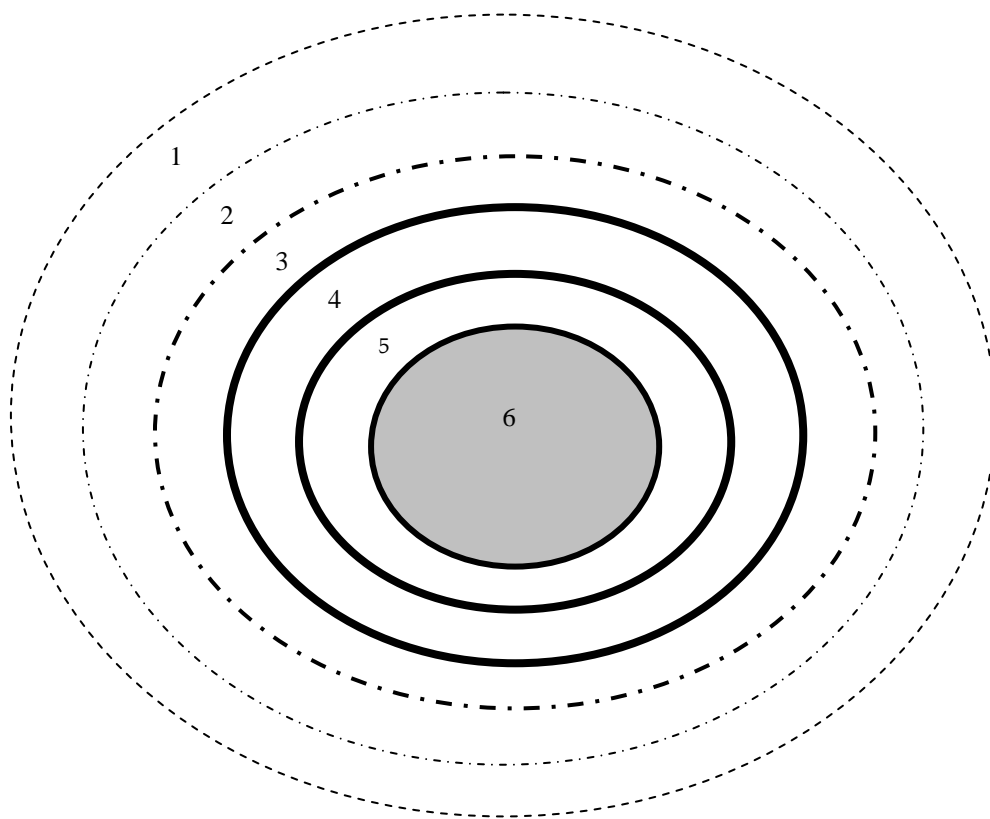
riatteita noudattaen, joten on perusteltua tutkia, millaisiksi tulokset muodostuvat analysoituina tehokkailla IRT -menetelmillä. Jotta voidaan osoittaa kokeen olleen edustava otos opetussuunnitelman perusteiden osioaltaasta, on myös esitettävä evidenssiä kokeen laadinnan perusteista ja kielitaidon eri osa-alueiden arvioinnista.

Edellä esitetyt tavoitteet ja pohdinnat johtivat seuraaviin tutkimuskysymyksiin:

1. Millaisia ovat reseptiivisen kielitaidon (puheen ja tekstin ymmärtämisen sekä kieliopillisen kompetenssin) kokeiden osiot klassisen ja uuden testiteorian menetelmillä analysoituina?
2. Miten CEF- ja OPS 2004 -taitotasoasteikot vastaavat toisiaan reseptiivistä kielitaitoa arvioitaessa?
3. Mikä perusopetuksen päättövaiheen oppilaiden reseptiivinen kielitaito oli CEF- ja OPS 2004- taitotasoasteikoilla ilmaistuna?
4. Miten CEF-, OPS 2004- ja A2001-taitotasoasteikot vastaavat toisiaan produktiivista kielitaitoa arvioitaessa?
5. Mikä perusopetuksen päättövaiheen oppilaiden produktiivinen kielitaito oli CEF-, OPS 2004- ja A2001 -taitotasoasteikoilla ilmaistuna?

Nämä tutkimuskysymykset ovat luonteeltaan deskriptiivisiä. Täten ne edellyttävät laadukasta kuvausta. Sen sijaan saavutettua taitotasoa ei pyritä selittämään esimerkiksi koulua ja opetusta tai yksilöllisiä piirteitä kuvaavilla muuttujilla.

Tässä raportoitava tutkimus on monitahoinen ja edellytti huomattavan laajaa selontekoa. Lukemisen helpottamiseksi seuraavassa kuvataan tutkimuksen rakenne ja esityksen eteneminen sisäkkäisinä ympyrän kehinä (kuvio 1). Sisimmän ympyrän, taitotasojen asettamisen ja tulosten ilmaisemisen, laatu, luotettavuus ja yleistettävyyys riippuvat siitä, miten hyvin sitä ympäröivien laajempien kehien sisällöt on sisäistetty ja otettu huomioon koko arviointiprosessissa.



KUVIO 1 Tutkimuksen rakenne.

Ympyrällä olevat numerot viittaavat seuraaviin tutkimuksessa käsiteltäviin teemoihin:

1. Arvioinnin monimuotoisuus (luvut 2 ja 3)
2. Kielitaidon arviointi (luku 4)
3. Kokeen laadinta ja sisältö (luvut 5 ja 6)
4. Tutkimusaineisto ja -metodi (luku 7)
5. Kansallinen englannin kielen arviointi (luku 8)
6. Tulokset (luvut 9, 10, 11)

Uloimmalle kehälle 1 sijoittuvat aiheet, jotka luovat taustaa sovelletulle arviointimenetelmälle. Arviointi -käsitettä lähestytään eri tutkijoiden näkökulmista tulkiten heidän näkemyksiään myös arviointiprosessin monimuotoisuudesta. Samoin sille asettuu ohjelma-arvioinnin piiriin kuuluva tämän tutkimuksen kontekstiin liittyvä koulutuksen arviointi, johon opetustoimen arviointijärjestelmässä sisältyvät mm. kansalliset arvioinnit. Nykykäsitykseen arvioinnin luotettavasta toteuttamisesta liittyvät eettiset normit ja standardit, joista luvussa 2 esitetään eräiden organisaatioiden näkemyksiä. Luvussa 3 kuvaillaan aikaisempia, kansallisten arviointien ryhmään luettavia englannin arviointeja sekä esitetään vertailun vuoksi kaksi yksityiskohtaisempaa esimerkkiä muiden Euroopan unionin maiden kansallisiksi katsottavista arvioinneista. Liitteessä 1 kuvaillaan lisäksi taulukon muodossa muutaman muun Euroopan maan vieraiden kielten arvioinnin nykytilannetta. Nämä teemat täsmentävät edelleen tämän tutkimuk-

sen kohteena olevan arvioinnin kuvausta ja niiden käsittelyllä tuotetaan tutkimukseen relevanttia kontekstualisointia.

Tutkimuksen etenemisen kannalta kehälle 2 asetuvat kielitaidon arviointia koskeva teoreettiset pohdinnat. Kielikokeen laadinnan lähtökohtana on kielitaito -käsitteen määrittäminen. Sen lisäksi luvussa 4 käsitellään arvioinnin kohteita, tuotosta ja prosessia, ja esitellään erilaisia kielioppikäsitteitä ja sanastollisen kompetenssin koostumusta. Tämä luo perustaa käsillä olevan tutkimuksen kohteena olevassa arvioinnissa sovelletuille näkemyksille. Näihin liittyvät läheisesti myös mittaamiseen kehitetyt typologiat ja taksonomiat.

Kehällä 3, eli luvuissa 5 ja 6, esitellään nykyisin yleisesti suositeltavana pidettävä vieraiden kielten arviointiprosessi, joka perustuu yleiseen arvioinnin toimeenpanokäsitykseen (kehä 1). Arviointimenetelmä on ratkaistava ennen kokeen laadinnan aloittamista, ja laadinnan yhteydessä on otettava kantaa erilaisten arviointiasteikkojen ansioihin ja puutteisiin. Kielitaidon arvioinnin keskeinen osa, kielitaidon osa-alueiden testauksessa käytettävät tehtävätyypit, asettuvat tähän kontekstiin.

Kehälle 4 sijoittuvat tutkimusaineiston ja käytettyjen tutkimusmetodien kuvaukset (luku 7). Edellinen ennakoi seuraavan luvun 8 sisältöä, missä kuvataan kansallisen englannin arvioinnin toimeenpano, sen kohderyhmä ja kokeet. Tärkeimpien tutkimusmetodien, IRT -analyysin ja taitotasojen asettamisen teoreettisen taustan kuvaus pohjustaa ytimen, kehän 6 arviointituloksia. Kehällä 6 arvioidaan myös tutkimuksen luotettavuutta, ja sille sijoittuvat myös tutkimuksen tuloksena tehtävät johtopäätökset.

Kukin luku alkaa johdannolla, jossa esitetään tiivistetysti luvun sisältö ja perustellaan sisällön valinta. Lukujen yhteenvetoon on koottu niiden keskeinen sisältö.

2 ARVIOINTI

Tämän tutkimuksen kohteena oleva englannin arviointi toteutettiin ajankohtana, jolloin Suomessa toimeenpantiin ensimmäiset Opetushallituksen uusien koululakien mukaisesti hallinnoimat kansalliset oppimistulosarvioinnit. Vakiintuneita arviointikäytänteitä ei vielä ollut. Siksi tässä tutkimuksessa on tarpeen luoda katsaus arvioinnin teoriataustaan, joka ohjaa ohjelma-arviointien, kuten koulutuksen arvioinnin, toimeenpanoa. Alkuvaiheensa jälkeen 1960-luvulla Yhdysvalloissa ohjelma-arviointi puolestaan on kehittynyt ja monipuolistunut ja sen soveltaminen on tuottanut monia uusia malleja.

Suomessa opetussuunnitelmien arviointiin on valittu yksi käytäntö. On kuitenkin paikallaan olla selvillä siitä, mitä vaihtoehtoja on olemassa. Arviointien toimeenpanemisen ammattimaisuuden / asianmukaisuuden takaamiseksi eri arviointiyhteisöt ovat kehittäneet eettisiä normeja ja standardeja. Oppimistulosten ja Eurooppalaisen viitekehyksen linkittämisen laadukkuuden varmistamiseksi edellytetään arviointiprosessin huolellista dokumentointia. Prosessin tulee luonnollisesti olla yhteensopiva yleisesti hyväksytyjen arviointiteorioiden ja käytänteiden kanssa. Tämän luvun tarkoitus on siis antaa teoriapohja luvussa 8 esitetyllä toteutuneen arvioinnin prosessikuvaukselle.

Yhteiskunnan tarkoituksenmukaisen toiminnan takaamiseksi on laadittu eri alojen ohjelmia ja projekteja, joiden tarkoitus on tukea ja parantaa yhteiskunnan jäsenten hyvinvointia. Smith (1989) määrittelee ohjelman joukoksi suunniteltuja toimintoja, joiden tarkoituksena on saada aikaan muutoksia tietyissä kohderyhmissä (ks. esim. Owen & Rogers 1999:24). Cronbach (1981:4; 1982:2-3) kuvaa ohjelmaa pysyväksi järjestelmäksi, toimintasuunnitelmaksi tai ohjeeksi jonkin palvelun parantamiseksi. Tässä tutkimuksessa tarkasteltava ohjelma on vuoden 1994 opetussuunnitelman perusteiden englannin kielen opetusta koskeva osa. Perinteisesti ohjelma-arviointi ymmärrettiin sen selvittämiseksi, missä määrin ohjelman avulla saavutetaan asetetut tavoitteet. Yhä enemmän on kuitenkin alettu kohdistaa huomiota itse ohjelmaan lähtien siitä, keiden ehdoilla ohjelma laaditaan ja päätyn koko ohjelman käyttökelpoisuuden arviointiin.

Arviointi liittyy laajimmassa merkityksessään arvottamiseen (*valuing*), arvon (*worth*) ja ansion (*merit*) määrittämiseen (Gagné 1970:105; Scriven 1980:7 ja 1991:139; Joint Committee on Standards 1981; Chelmsky 1997:VII; Owen & Ro-

gers 1999:14). Patton (1981:137) käyttää termiä *value* termin *worth* sijasta. *Worth* ja *value* viittaavat arviointikohteen ulkoiseen arvoon eli siihen, miten hyvin arviointikohde toteuttaa ulkopuolisten ryhmien, instituutioiden tai yhteiskunnan edustajien eli arvioinnin asianosaisten ajatuksia ja edistää heidän näkemyksiään. *Merit* taas viittaa kohteen sisäiseen arvoon eli esimerkiksi siihen, missä määrin tietyt tavoitteet todennäköisesti saavutetaan, miten hyvä arviointikohde on tietyin standardein mitattuna. Raivola (2000:14) toteaa, että ei riitä, että se, mitä arvioitavassa tilanteessa tehdään, on tehokasta ja ansiokasta (*merit*), vaan myös toiminnan tuloksilla on oltava käyttäjälle arvoa (*worth*).

Ansion määrittäminen on periaatteessa suhteellisen yksinkertaista: kuvataan ansion dimensiot eli arviointikohteen tavoitteet ja sen standardit eli taso, jonka perusteella suoritus voidaan arvioida (Scriven 1980:18; Raivola 2000:199). Ulkoisen arvon määrittämisen ongelmaksi muodostuu kuitenkin se, kenelle tuloksilla on oltava arvoa eli kenen arvoja käytetään johtopäätöksiä tehtäessä (Patton 1997:23-25; Owen & Rogers 1999:15; ks. myös Raivola 2000:199). Stufflebeam (1994:323) varoitti hämärtämästä arvioinnin olemusta, arvottamista (*value assessment*), käyttötarkoituksia korostamalla. Tämä taas on Pattonin mielestä arvioinnin tärkein tehtävä.

Käsitepari faktat ja arvot arvioijan roolissa (*fact/value dichotomy*) on pitkään aiheuttanut ristiriitoja arviointiteoreetikkojen välillä. Vastakkainasettelu liittyy kysymykseen, onko arvioija arvovapaa kuvatessaan eri ryhmien intressien kohteita vai tulisiko hänen antaa joidenkin demokraattisten perusarvojen vaikuttaa arvioinnin luonteeseen. Perinteisesti on katsottu, että arvioija kuvaa faktoja, jotka on koottu tieteellisesti perusteltavin menetelmin. Tällaisen lähestymistavan tavoite on erottaa faktat ja arvot ja lisäksi todistaa, että mikään ulkopuolinen, esimerkiksi politiikka, ei ole vaikuttanut tuloksiin (Karlsson 2003:435). Esimerkiksi Campbell, Stake ja Lincoln ovat olleet sitä mieltä, että arvot ovat subjektiivisia, kun taas Scrivenin mielestä ne ovat objektiivisia (House 2003:10-12). House ja Howe (1999) lähtevät siitä, että koska arvioijat ovat tekemisissä arvoja sisältävien (*value-laden*) arvioinnin kohteiden kanssa, faktoja ja arvoja ei voi erottaa toisistaan, vaan ne ovat jatkumon ääripäitä, jotka sekoittuvat toisiinsa tilanteesta riippuen jatkumon eri kohdissa. Tunnettu ja arvostettu filosofi Hilary Putnam (2002) on myös hiljattain esittänyt, että kyseinen dikotomia on romahtanut. Arviointia määriteltäessä käydään usein myös rajaa arviointitutkimuksen ja arvioinnin välillä. Tutkimusta tehdään teorioiden testaamiseksi ja yleistämiseksi yli ajan ja paikan. Arvioinnilla taas on selkeä kohde, esimerkiksi tietty ohjelma, jonka toiminnasta, toimivuudesta ja vaikutuksista tehdään päätöksiä tässä ja nyt poliittisten realiteettien puitteissa. Tutkimuksen ja arvioinnin eroa onkin kutsuttu Lee Cronbachin (1987) jaottelun mukaisesti johtopäätös-keskeiseksi ja päätös-keskeiseksi tutkimukseksi (ks. myös Leimu 2004a:29). Hyödyllinen arviointi tukee toimintaa, ja jokaisessa tietojenkeruutilanteessa se, missä määrin kukin tiedon käyttäjä on kiinnostunut hyödyllisyydestä, yleistetävyydestä ja tiukasta tieteellisyydestä, vaihtelee (Patton 1997:23-25). Tasapaino näiden välillä tulisi löytää ennen arviointityöhön ryhtymistä.

Taulukossa 2 kuvaillaan arvioinnin ja arviointitutkimuksen välisiä eroja. Se perustuu mm. Smith ja Glassin (1987), Norrisin (1990) Yrjölän (1998) ja Rairvolan (2000) ryhmittelyihin. Siitä, ovatko ne toisistaan selvästi erotettavia, valitsee kuitenkin erimielisyys. Esimerkiksi Norris lähtee siitä, että niitä ei ole välttämättä syytä erottaa.

TAULUKKO 2 Arvioinnin ja tutkimuksen välisiä eroja.

Arviointi	Tutkimus
Tutkii, mitä tapahtuu (kuvaileva).	Tutkii, miksi jotakin tapahtuu (selittävä).
On yhteydessä ohjelman arvoon (<i>worth</i>); punnitsee ohjelmia päätöksenteon perustaksi.	Välttää lausumia ohjelman arvosta, pyrkii teorian kehittämiseen.
Kohdistuu tiettyyn tilanteeseen.	On tilanteesta riippumaton.
Asiakas tekee aloitteen.	Tutkija tekee aloitteen.
Arvioija vastuussa pääasiassa toimeksiantajalle.	Arvioija vastuussa pääasiassa tutkijakollegoille.
Auttaa toimeksiantajan toimintaa.	Auttaa tiedeyhteisön toimintaa.
On naturalistista ja käytännöllisesti suuntautunutta.	Tieteenala ohjaa.
Tulokset liittyvät annettuun tehtävään.	Tulokset ovat yleistettävissä.
Päätavoite on tukea asiakasta.	Päätavoite on lisätä tietoa.
Ajallisesti rajattu.	Ajallisesti rajaamaton.
Arvioi suoraan koulutuksellista ja yhteiskunnallista hyötyä.	Arvioi koulutuksellista ja yhteiskunnallista hyötyä epäsuorasti.
Saattaa olla riippuvaista toimeksiantajasta.	Riippumatonta, autonomista.

Arvioinnin teoriapohjaan on alettu kiinnittää viime vuosikymmeninä yhä enemmän huomiota. Shadish ym. (1991:67) jakavat arviointiteorioiden kehittämisen kolmeen vaiheeseen, vaikka huomauttavatkin painokkaasti, että jako on suuressa määrin yksinkertaistettu:

- 1) Etsittäessä tehokkaita ratkaisuja yhteiskunnallisiin ongelmiin korostettiin totuuteen pyrkimistä.
- 2) Syntyi useita vaihtoehtoisia arviointiteorioita, jotka perustuivat tietoon siitä, kuinka julkisen sektorin organisaatiot toimivat pyrkiessään poliittisesti ja yhteiskunnallisesti hyödyllisten tulosten tuottamiseen.
- 3) Syntyi teorioita, jotka pyrkivät integroimaan edellisten vaiheiden painotukset.

Shadish ym. (1991:30-31) määrittelevät käyttökelpoisen arviointiteorian seuraavasti:

”The ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront. It would (a) clarify the activities, processes, and goals of evaluation; (b) explicate relationships among evaluative activities and the processes and the goals they facilitate; and (c) empirically test propositions to identify and address those that conflict with research or other critically appraised knowledge about evaluation.”

Cook (1997:47) pahoittelee, että vaikka edellä siteeratussa teoksessa kuvataankin arviointiteorian laatiminen, tekijät eivät ole sitä kuitenkaan tehneet, vaan ovat pikemminkin laatineet arviointiteorian teorian. Shadish ym. ovat kuitenkin

muokanneet teoriakäsityksensä keskeiset kohdat seuraaviksi kysymyksiksi, joihin arviointiteorian tulee heidän mielestään ottaa kantaa:

- 1) *Miten yhteiskunnalliset ohjelmat laaditaan?* Millä tavoin ne kehittyvät ja parantuvat erityisesti suhteessa muuttuviin yhteiskunnallisiin ongelmiin?
- 2) *Miten tietoa hankitaan?* Millä tavoin tutkijat saavat tietoa yhteiskunnan toiminnasta?
- 3) *Miten arvot voidaan liittää ohjelman kuvauksiin?*
- 4) *Miten saatua informaatiota käytetään muokkaamaan ohjelmia?*
- 5) *Millaista arviointikäytäntöä sovelletaan?* Minkälaisia strategioita arvioijat noudattavat toiminnassaan ja miten he käsittelevät työssään kohtaamiaan rajoituksia? (Shadish ym. 1991:32.)

Scriven (1980:5) erottaa sekä arvioinnin alueita että tyyppejä. Alueita ovat esimerkiksi koulutus ja terveydenhuolto. Arviointityypit hän jakaa kokonaisarviointi- ja metodologisiin tyyppeihin. Edellisiä ovat ohjelma-, tuote- ja henkilöstöarviointit, jälkimmäisiä taas esimerkiksi formatiivinen, joka kohdistuu prosessiin, ja summatiivinen, joka kohdistuu lopputulokseen. Näihin kuuluu myös tavoitevapaa arviointi, jossa ei kiinnitetä huomiota ennalta asetettuihin tavoitteisiin, jotka saattavat olla asiakkaan kannalta epäolennaisia, vaan pyritään vastaamaan hänen tarpeisiinsa. Ryhmät eivät kuitenkaan ole toisensa pois sulkevia.

Lienee varsin yleisesti hyväksytty käsitys, että arvioinnin – samoin kuin tutkimuksen – tarkoituksen tulisi rajata käytettävät menetelmät. Perinteisesti ohjelma-arviointi ymmärrettiin sen selvittämiseksi, missä määrin ohjelman avulla saavutetaan asetetut tavoitteet. Nykyään lähdetään tavallisesti siitä, että ohjelma-arviointiin sisältyvät arvioinnin toimeenpano, ohjelman toiminnan seuraaminen ja dokumentointi, odottamattomiin seurauksiin varautuminen ja sekä välittömien että pitkän aikavälin vaikutuksien huomioon ottaminen (Patton 1997:23-25). Owen ja Rogers (1999:27-28) jakavat ohjelmat niiden laajuuden mukaan mega-, makro- ja mikrotasoihin muistuttaakseen, että eri tasojen arvioinnit edellyttävät erilaisia arviointitekniikoita ja lähestymistapoja, vaikka he toteavatkin, että välimuotoja on olemassa. He esittävät taulukon muodossa viisi eri arviointimuotoa, niiden orientaation, kysymykset, joihin etsitään vastauksia sekä lähestymistavan, joka sopii käytettäväksi valitun arviointimuodon yhteydessä (Owen & Rogers 1999:53). King (2003:727) toteaa, että arvioinnin tarkoitukseen liittyviä lähestymistapoja on hyvinkin yli kaksikymmentä. Esimerkkeiksi tästä monimuotoisuudesta on valittu taulukkoon 3 eräitä yleisesti esiintyviä lähestymistapoja. Toisen sarakkeen otsikko ”Keskeinen näkemys” kuvaa paitsi arvioinnin tarkoitusta myös sen demokratiakäsitystä eli sen, missä määrin asianosaiset osallistuvat arviointiin. ”Edustajat” -sarakkeeseen on sulkuihin merkitty pääasiallinen lähde.

TAULUKKO 3 Arvioinnin yleisimmät lähestymistavat, niihin liittyvä demokratiakäsitys, tunnetuimmat edustajat ja pääteos.

Lähestymistapa	Keskeinen näkemys	Edustajat	Pääteokset
Eksperimentaalinen	Löytää kausaaliyhteyksiä ohjelmien ja oppimistulosten välillä satunnaisotksin koottujen esikokeilujen avulla.	Campbell, Donald & Stanley, Julian (Shadish ym. 1991:120, 125)	Experimental and quasi-experimental designs for research, 1966
Päätöskeskeinen, yleensä kehittämistä palveleva	Arvioinnin tarjonnan tulisi kohdistua suoraan päättäjiin.	Stufflebeam, Daniel	The CIPP Model, 1971
Tavoitevapaa (<i>goal-free</i>), kustannusnäkökohdista riippumaton (<i>cost-free</i>), tarveperusteinen (<i>needs-based</i>)	Ohjelman arviointi tavoitteista tietämättä, jotta löydettäisiin vastauksia niihinkin kysymyksiin, joita ei osattu kysyä.	Scriven, Michael (Shadish ym. 1991:73)	The Logic of Evaluation, 1980
Yleensä kehittämistä palveleva	Ohjelman arvon toteaminen. Arvioinnin päämääräksi ei riitä, että tavoitteena on sen parantaminen, vaan ohjelmaa tulee voida muokata jo arvioinnin aikana.	Cronbach, Lee J. (Cronbach, Lee J. 1981:16-17)	Toward Reform of Program Evaluation, 1981
Pragmatistinen	Edistää arviointitulosten käyttöä.	Weiss, Carol (Shadish ym. 1991:179)	Utilization of Evaluation: Toward Comparative Study, 1972
	Edistää arviointitulosten käyttöä identifioimalla todennäköiset käyttäjät, joille arviointi suunnataan.	Patton, Michael Quinn (Patton 2003:225)	Utilization-Focused Evaluation, 1997
Responsiivinen	Arvioijan velvollisuus on ottaa asianosaisten näkemykset huomioon.	Stake, Robert E. (Shadish ym. 1991:270; Stake 2003:63)	Evaluating the Arts in Education, 1975
Neuvotteleva	Arvioija on vastuussa arvioinnista ja löydöksistä, mutta asianosaisten näkemykset on arvioinnin kuluessa otettava huomioon.	House, Ernest R. & Howe, Ken (House, Ernest & Howe, Kenneth R. 2003:79)	Values in evaluation and social research, 1999
Konstruktivistinen	Arvioija rakentaa tietoa asianosaisten kanssa, jotka tulevat olemaan tekemisessä myös tulosten kanssa.	Guba, E. G. ja Lincoln, Y. S. (Lincoln, Yvonna S. 2003:69-78)	Fourth generation evaluation, 1989
Osallistava	Arviointiasiantuntijat ja asianosaiset osallistuvat yhdessä arvioinnin suorittamiseen.	Cousins, J. Bradley (Cousins, J. Bradley 2003:245)	Participatory evaluation in education: Studies in evaluation use and organizational learning, 1995
Valtuttava	Eri intressiryhmien ottaminen mukaan arviointiin arvioijan toimiessa ikään kuin konsulttina.	Fetterman, D. M. (Morris 2003:316)	Empowerment evaluation: Knowledge and tools for self-assessment and accountability, 1996

Yleinen trendi on ollut arvioinnin kohteiden, intressiryhmien ja käyttötarkoitusten monipuolistuminen.

Owen ja Rogers (1999:24) jakavat arvioinnin kahteen osaan, dokumentoituun suunnitelmaan ja toimintaan, joka on yhdensuuntainen suunnitelmassa annetun informaation kanssa. Tästä yhteydestä, toisin sanoen arviointiprosessista, he käyttävät nimitystä "theory of action". Vastaavaan yhteyteen suunnitelman ja toteutuksen välillä Scriven (1980) viittaa termillä "the logic of evaluation", joka on yksi hänen arviointiteoriaansa peruskäsitteistä. Taulukkoon 4 on koottu eräiden tutkijoiden näkemyksiä arviointiprosessiin kuuluvista toiminnoista. Ralph Tylerin malli on vaikuttanut voimakkaasti arviointinäkemysten kehittymiseen samoin kuin hänen opetussuunnitelmien laadinnan mallinsa (1942) on vaikuttanut näkemyksiin opetussuunnitelmien laadinnan perusteista. (Sauli Takala, suullinen tiedonanto, 25.1.2006).

TAULUKKO 4 Arviointiprosessia kuvaavia luokitteluja

Tutkija	Arviointiprosessi
Tyler (Smith & Tyler ym. 1942; 15-25)	<ol style="list-style-type: none"> 1) Tavoitteiden muotoilu. 2) Tavoitteiden luokittelu. 3) Tavoitteiden määrittäminen käyttäytymisen muotoina. 4) Niiden tilanteiden osoittaminen, joissa tavoitteiden saavuttaminen ilmenee. 5) Lupaavien arviointimenetelmien valitseminen ja kokeileminen. 6) Arviointimenetelmien kehittäminen ja parantaminen. 7) Tulosten tulkitseminen.
Rich (1979: 246)	<ol style="list-style-type: none"> 1) Onko käytettävissä asianmukaisia menetelmällisiä työkaluja ja käytetäänkö niitä? 2) Millaisia arviointitutkimuksen antamien tulosten mukaisia koulutusohjelmia tulisi kehittää? 3) Muuttuvatko tutkimuksen tulokset käytännöksi? 4) Mitä on tehtävissä arviointitietojen väärinkäytön vähentämiseksi? 5) Mitä eettisiä normeja arvioijille tulisi laatia ja minkä tahon tulisi valvoa niiden toteuttamista?
Knapp (1979:227-8)	<ol style="list-style-type: none"> 1) Henkilö tai ryhmä, joka tulee tavallisesti ohjelman ulkopuolelta, nimitetään arvioijaksi. 2) Ohjelma, joka on jo toiminnassa tai vaihtoehtoisesti vasta alkamassa, valitaan arvioinnin kohteeksi. 3) Arvioija koordinoi systemaattista tietojen keruuta selvittäen ohjelman panostuksen, toiminnan ja tuotoksen. 4) Tiedot kootaan sellaisessa muodossa, että toiminta voidaan tarvittaessa toistaa. 5) Tiedot tulkintoineen toimitetaan asianosaisille.
Rea-Dickins & Germaine (1992:74)	<ol style="list-style-type: none"> 1) Kenelle informaatio on tarkoitettu ja miksi arviointi toimeenpannaan? 2) Mitä arvioidaan, opetusta, oppimista vai molempia ja mistä näkökulmasta? 3) Mitä kriteereitä arvioinnissa käytetään ja kuka päättää kriteereistä? 4) Milloin ja missä opiskelun vaiheessa arviointi toimeenpannaan? 5) Mitä menetelmiä käytetään tietojen saamiseksi? 6) Keneltä / Mistä tietoa saadaan? 7) Ketkä osallistuvat arviointiin? 8) Miten arviointi toimeenpannaan?

	<p>9) Mihin saatua tietoa käytetään? Tämä riippuu</p> <ol style="list-style-type: none"> arvioinnin tarkoituksesta. arvioinnin luonteesta. arvioinnin hyödyntäjistä. arviointipalautteen luonteesta. arviointitietojen raportointitavasta. <p>10) Mitkä rajoitukset ja ongelmat on tarpeen ottaa huomioon suunniteltaessa ja toimeenpantaessa arviointia? Arviointiprosessin arviointi: ovatko arvioinnit vaikuttavia?</p>
Owen & Rogers (1999:20)	<ul style="list-style-type: none"> - neuvottelu arviointisuunnitelmasta - aineiston koonta ja analysointi arviointitulosten tuottamiseksi - tiedottaminen tietyille kohderyhmille arviointikohteen ymmärtämiseksi tai arvioiden ja päätösten tekemiseksi arvioinnin kohteesta
Scriven 1980:18	<ol style="list-style-type: none"> Valitaan arvon (<i>merit</i>) kriteerit. Asetetaan suoritusstandardit. Mitataan suoritus. Annetaan tulosten perusteella arviointi.

Taulukossa esitellyissä käsityksissä on paljon yhteistä, mutta se osoittaa myös, miten eri tutkijat painottavat arvioinnin eri vaiheita. Samalla taulukko havainnollistaa arviointiprosessin monitahoisuutta.

Kuten arviointiprosessin kuvauksesta käy ilmi, systemaattinen tiedon keruu on keskeinen osa arviointia. Patton (1981:18; 1997:23-25) lähtee siitä, että 1) tietoa kootaan systemaattisesti 2) monenlaisista asioista 3) tietyn kohdejoukon käyttöön (ks. myös Cronbach ym. 1981:14; Weir & Roberts 1994:4). Enää eivät riitä vastaukset kysymyksiin, miten hyvä jokin on ja toimiiko se, vaan tarvitaan yksityiskohtaisempaa analyysia. Owen ja Rogers (1999:3) tarjoavat arvioinnin määritelmäksi prosessia, jossa ensiksi tulee neuvotella arviointisuunnitelman perusteista (ks. taulukko 3, "Neuvotteleva arviointi"), minkä jälkeen siirrytään aineiston koonta- ja analyysivaiheeseen arviointitulosten tuottamiseksi. Tulokista tiedottaminen on tärkeä osa arviointia. Tavoitteista riippuen sitä on tarkoitus käyttää joko kuvaamaan tai ymmärtämään arvioinnin kohdetta, tekemään arvioita kohteen käyttökelpoisuudesta tai kohdetta koskevaan päätöksentekoon. Tutkijat eivät myöskään enää hyödynnä pelkästään kvantitatiivista dataa, vaan viimeisten kahdenkymmenen vuoden kuluessa kvalitatiivisesta arvioinnista tai näiden kahden yhdistelmästä on tullut tavallisia ja hyväksyttäviä (King 2003:729).

2.1 Koulutuksen arviointi

Opetussuunnitelman perusteiden arviointi on koulutuksen arviointia. Sitä käsiteltäessä on ensinnä määriteltävä, mitä koulutuksella tarkoitetaan ja mitkä sen tavoitteet ovat. Franke-Wikberg ja Lundgren (1982:139) katsovat 1900-luvulla noudatetun pedagogisen tutkimuksen jakautuvan kolmeen linjaan: perustan

kehittäminen koulutuksen muutoksille, oppimis- ja ajattelututkimukset sekä koulutuksen kehittäminen historiallisessa ja yhteiskunnallisessa kontekstissa.

Aikaisemmassa panos – tuotos -ajattelussa panopiste oli panoksissa. Oletettiin, että esimerkiksi mahdollisimman monen nuoren koulutukseen pääsy, hyvät ulkoiset olosuhteet, tarkoituksenmukainen oppimateriaali ja monipuolinen opettajainkoulutus takaavat oppilaiden oppimisen. Painotus siirtyi 1900-luvun loppuvuosikymmenillä tuotoksiin mm. siksi, että tutkimuksissa ei ollut havaittu vahvaa yhteyttä koulun resurssien ja oppimistulosten välillä (Kellaghan 2003:874). Paineita koulutusjärjestelmän arviointiin lisäsivät paitsi kansainvälisistä koulutusalan kongresseista 1990-luvulla kiirineet tiedot siitä, miten vähän tiedettiin koulutuksen tuloksista suhteessa panoksiin, myös liikkeenjohdollisen ajattelutavan siirtyminen hallintoon (emt. 874-5).

Monissa länsimaissa julkinen sektori joutui sopeutumaan myös väheneviin resursseihin 1990-luvulla. Muutoksen sisältöä kuvaa käsite NPM (New Public Management), jonka mukaan resursseja allokoimalla eli kohdentamalla pyritään kustannusten leikkaamiseen, budjettikatkoihin, suurempaan läpinäkyvyyteen, byrokraattisten organisaatioiden toimintojen ulkoistamiseen, virkojen määräaikaistamiseen ja palkkojen muuttamiseen suoritusperustaisiksi (Karlsson 2003:432-3). NPM-käsite soveltui hyvin tukemaan informaation keräämistä ja kriteerien luomista tätä informaatiota käsiteltäessä. Sieltä saatiin mallit esimerkiksi strategiseen ja operationaaliseen suunnitteluun, jatkuvan kasvun (tässä tapauksessa tulosten paranemisen) odotuksiin, suoritusindikaattoreihin, laadunvalvontaan ja tuloksiin perustuviin tulosvastuujärjestelmiin. Tarkoituksena oli tuottaa päättäjille nopeaa ja jatkuvaa tietoa mm. resurssien suuntaamiseksi.

Useissa maissa oli todettu, että koulutusjärjestelmien tuloksistakin tuli saada tietoa. Ohjelma- ja opetussuunnitelma-arviointi kansallisten arviointien muodossa näytti tarjoavan tähän ratkaisun. Vähitellen siirryttiin pelkkien oppimistulosten esittämisen sijasta tutkimaan muitakin tuloksia, suunniteltiin vaihtoehtoja perinteisille ohjelmille ja pohdittiin arvioinnissa käytettyjen toimien riittävyttä. Vastatakseen arvioinnin uusiin vaatimuksiin arvioijat käyttivät tietysti edelleen mittaus- ja psykometristä asiantuntemustaan, mutta laajensivat sekä kvantitatiivisten että kvalitatiivisten menetelmien varastoaan ja sisäistivät suunnittelun ja päätöksentekoa palvelevan arvioinnin välisen yhteyden (Kellaghan ym. 2003:2). Karlsson (2003) erottaa kansallisen ja ylikansallisen arvioinnin kehittymisessä kaksi vaihetta. 1960- ja 70-luvuilla arviointi suuntautui koulutusohjelmien kehittämiseen, ja tässä ensimmäisessä vaiheessa mukana olivat hänen mukaansa esim. Iso-Britannia, Saksa ja Ruotsi. Toisessa vaiheessa, 1980- ja 90-luvuilla arviointi suuntautui kontrollin vaikuttavuuden lisäämiseen ja julkisten menojen vähentämiseen. Suomi mainitaan tässä vaiheessa mukaan tulleen maana yhdessä Italian ja Espanjan kanssa. Karlssonin (2003) mukaan e.m. maat "increased their evaluation capacity in the 1990s under pressure from the European Union (EU). The focus in this wave was on results, the measurement of effectiveness, and the control of quality, activities often associated with programs of regional and technical development." (Karlsson 2003:430.) Karlssonin käsitys tuntuu sängen oudolta sitä maassamme jo 1960-luvulla harjoitettua

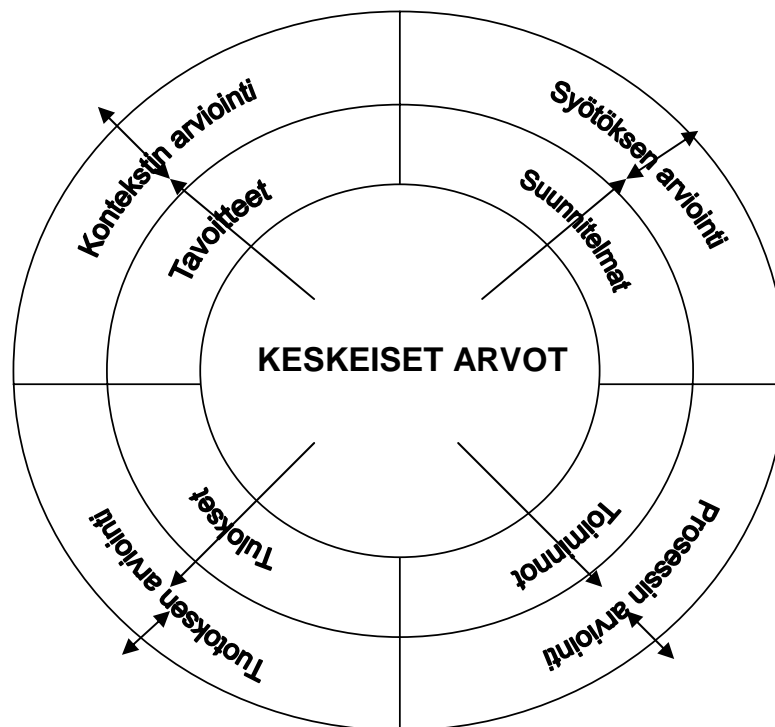
vilkkaan arviointitoiminnan taustaa vasten, jota on kuvattu tässä tutkimuksessa (luku 3). Tietämättömydestäkään ei voine olla kyse, sillä Pohjoismaiden kesken harjoitettiin esimerkiksi 1970-luvulla tiivistä yhteistyötä vieraiden kielten opetuksen ja arvioinnin kehittämiseksi mm. pyrkimällä harmonisoimaan oppimääriä (Pohjoismaiset mietinnöt 1973).

Koulutuksen arviointiin sisältyy laaja kirjo erilaisia arviointitoimia: oppilasarviointi, mittaus, testaus, ohjelma/opetussuunnitelma-arviointi, koulun henkilöstön arviointi, jopa koulujen toimintalupien vahvistamisedellytysten arviointi. Suomen kielen arviointi-sanalla on englannin kielessä kaksi vastinetta, *evaluation* (evaluaatio) ja *assessment*. Edellisellä tarkoitetaan yleensä systemaattista informaation keräämistä ohjelmista päätöksentekoa varten (esim. Nitko 2001:4; Lynch 2003:23) parantamaan ohjelman vaikuttavuutta (*effectiveness*) ja pohjaksi seuraaville ohjelmille (Patton 2003:224). Jälkimmäinen viittaa tiedon keräämiseen yksilöistä, heidän taidoistaan, tuotoksistaan ja suorituksistaan (Nitko 2001, Lynch 2003). Scriven toteaa, että *assessment*-sanaa käytetään usein *evaluation*-sanan sijasta, mutta katsoo, että edellinen liittyy suuremmassa määrin testaukseen (1981:12). Linn ja Gronlund (1995:5) myöntävät käyttävänsä teoksessaan sanaa *assessment* samassa merkityksessä kuin sanaa *evaluation* edellisessä teoksessaan.

Measurement, mittaus, määritellään prosessiksi jossa tietyistä piirteistä tai ominaisuudesta annetaan numeerinen arvo (*score*), joka osoittaa, missä määrin kyseistä piirrettä tai ominaisuutta omataan (Linn & Gronlund 1995, McNamara 2000, Nitko 2001, Lynch 2003). Sanan *test*, testi, koe, Linn ja Gronlund (1995:6) määrittävät välineeksi tai systemaattiseksi prosessiksi, jonka avulla mitataan käyttäytymistä antamalla eri osallistujille vastattavaksi samoja kysymyksiä. Koska testi on arvioinnin muoto, testit myös vastaavat kysymykseen, kuinka hyvin testattava suoriutuu – joko suhteessa toisiin testattaviin tai suhteessa kielenkäyttöalueelta otostettuihin suoritusta edellyttäviin tehtäviin. Arviointitermiä käytetään tässä tutkimuksessa suomen kielen mukaisesti viittaamaan yllä mainittuihin aspekteihin. Konteksti määrittelee sen, mikä merkitys kulloinkin on kyseessä.

Laadituista ohjelma-arviointimalleista Stufflebeamin CIPP-malli, jonka perusta on peräisin jo 1960-luvun loppupuolelta, soveltuu erittäin hyvin esimerkiksi koulutuksen arvioinnin kohteiden eri vaihtoehtoista, koska se on laadittu opetussuunnitelman arviointiin ja on laajasti käytetty, käytännöllinen arviointimalli. Stufflebeam toteaa CIPP-mallinsa pyrkivän sekä tilivelvollisuuden osoittamiseen että ohjelman kehittämiseen (2003:34) (ks. myös Weir & Roberts 1994:5). CIPP-malli koostuu neljästä komponentista: tavoitteista, suunnitelmista, toimenpiteistä ja tuloksista, joita kutakin arvioidaan omiin arviointityyppihinsä kuuluvilla menetelmillä, jotka kohdistuvat kontekstiin, panoksiin, prosessiin ja tuloksiin (Context, Input, Process, Product = CIPP). Tässä tutkimuksessa näistä komponenteista korostuvat konteksti ja tulosten arviointi. Kuvio 2 havainnollistaa arviointikohteen ja arviointityypin välistä suhdetta kaksipäisin nuolin ja osoittaa, että arvioinnin tulee lähteä keskeisistä arvoista (*core values*). Näitä saattavat olla esimerkiksi opetussuunnitelman tavoitteet, joita täsmentä-

mällä päästään selville siitä, mitä tietoa tarvitaan, mikä taas luo pohjaa arviointi-instrumenttien valinnalle.



KUVIO 2 CIPP -arviointimalli (Stufflebeam 2003:33).

Kontekstiarviointi tuottaa tietoa arvioitavan alueen tarpeista, ongelmista ja mahdollisuuksista. Löytämällä ne voidaan asettaa tavoitteet. Panosten arvioinnissa vertaillaan eri lähestymistavoin tehtyjä suunnitelmia, vaihtoehtoisia strategioita ja niiden edellyttämiä taloudellisia resursseja. Näiden pohjalta suunnitellaan perusteltavissa olevat toteuttamiskelpoiset parannusmenetelmät. Prosessia arviotaessa tarkkaillaan, arvioidaan ja dokumentoidaan ilmenneitä ohjelman parantamistarpeita ja -mahdollisuuksia. Tuotoksen arviointi kohdistuu lyhyen ja pitkän tähtäyksen sekä toivottuihin että yllättäviin tuloksiin, joiden pohjalta voidaan tehdä päätöksiä toimenpiteiden jatkamisesta, lopettamisesta tai muokkaamisesta (Stufflebeam 2003:31-55). Stufflebeam korostaa, että arvioinnissa tulisi soveltaa useita lähestymistapoja ja koota sekä kvantitatiivista että kvalitatiivista tietoa. Arvioinnin tulee myös olla tasokasta ja sitä pitäisi arvioida sekä sisäisen että ulkoisen metaevaluaation avulla (Stufflebeam 2003:58).

2.1.1 Kansallinen koulutuksen arviointi

Kansalliset arvioinnit suunnitellaan kuvaamaan koulutusjärjestelmän tuottamia saavutustasoja. Amerikkalainen NAEP (National Assessment of Educational Progress) on uranuurtajia tässä suhteessa, ja Ralph Tyler oli merkittävä vaikuttaja myös sen luomisessa (Jones 2003:883). IEA:n oppimistulosten kansainväliset vertailututkimukset ovat olleet toinen merkittävä vaikutuslähde. (Sauli Takala,

suullinen tiedonanto 25.1.2006). Ne kaikki etsivät vastauksia yhteen tai useampaan seuraavista kysymyksistä:

1. Kuinka hyvin oppilaat oppivat koulutusjärjestelmässä suhteessa yleisiin odotuksiin, opetussuunnitelman tavoitteisiin, elämään valmistautumiseen?
2. Onko todisteita erityisistä vahvuuksista tai puutteista oppilaiden tiedoissa ja taidoissa?
3. Suoriutuvatko jotkin väestön ryhmät heikosti? Onko esimerkiksi eroja poikien ja tyttöjen, kaupungeissa tai maaseudulla asuvien, äidinkieleltään erilaisten, etnisten ryhmien tai maan eri osissa asuvien oppilaiden saavutusten välillä?
4. Mitkä tekijät ovat yhteydessä oppimistuloksiin? Missä määrin tulokset vaihtelevat oppimisympäristöstä (koulun resurssit, opettajien suorittama oppituntien valmistelu ja heidän pätevyytensä, koulutyyppi) tai oppilaan koti- tai yhteisötaustasta johdettujen?
5. Muuttuvatko oppilaiden tulokset ajan myötä? (Kellaghan 2003:876-877).

Menettelytavat vaihtelevat etsittäessä vastauksia edellä esitettyihin kysymyksiin. Saatetaan käyttää koulujen tai oppilaiden otantaa, tai vaihtoehtoisesti arvioidaan kaikki tietyn ikäiset tai tietyn luokkatason oppilaat. Koe saattaa olla analyttinen, jolloin se suoritetaan ohjeiden mukaan tiettyyn aikaan tai holistinen ja luokkaopetukseen integroitu, jolloin sitä suoritetaan useana kertana opettajien valvonnassa. Opetussuunnitelman tavoitteiden kattamisen (kriteeriviitteisen arvioinnin) takaamiseksi oppilas voi suorittaa vain osan laajasta tehtäväjoukosta, jolloin säästetään aikaa, kuten käsillä olevassa tutkimuksessakin on tehty. Arviointiin saattaa liittyä myös esimerkkejä hyvistä arviointikäytännöistä sekä palautteena tietoa yksittäisten oppilaiden, opettajien, koulujen tai alueiden suoritustasosta (Kellaghan 2003:877).

Arviointitarkoituksen perusteella arvioinnit voidaan ryhmitellä seuraavasti:

- Tilivelvollisuutta palveleva funktio, jolloin niillä on päätösten seuraamuksiin liittyvät tehtävät,
- Päätöksentekoa palveleva funktio, jolloin niillä on suunnitteluun ja toiminnan ohjaukseen liittyviä tehtäviä koulutuksen toimeenpanossa,
- Johtopäätöksiin tähtäävä tehtävä, jolloin arviointien tuloksia käytetään opetus- ja oppimistapahtuman tutkivaan selvittelyyn (Leimu 2004a:9).

Käsillä oleva tutkimus palvelee jossakin määrin kaikkia edellä mainittuja tavoitteita. Tiedottaessaan opetussuunnitelman perusteisiin pohjautuvat oppimistulokset se palvelee tilivelvollisuutta koulua kohtaan. Selvimmin se palvelee päätöksentekoa tuottamalla tuloksia, jotka saattavat johtaa aikaisempien päätösten muuttamiseen. Toivottavasti se myös tuottaa johtopäätöksiä opetus- ja oppimistapahtuman selvittelytarpeesta.

Kansallinen koulutuksen arviointi Suomessa

Suomessa 1960-luvulla alkanut peruskoulu-uudistus nojaa juuri yhteiskunnalliseen muutokseen. Koulutuksen tavoitteeksi otettiin kaikille yhtäläisten koulutusmahdollisuuksien takaaminen yhä eriytyneemmässä yhteiskunnassa. Koulutuksen tuli ratkaista kaupungistumisen, kaupallisuuden ja teollistumisen mukanaan tuomat ongelmat (POPS I 1970:14). Koulutusohjelmien arvioinnin kaikille yhteisessä peruskoulutuksessa voidaan katsoa alkaneen peruskoulu-

uudistuksen valmistelun myötä 1960-luvun loppupuolella. Kouluhallitus, joka silloin oli normeja antava koulutushallinnon organisaatio, teetti toimeksiantotutkimuksina Kasvatustieteiden tutkimuslaitoksella oppimistulosarviointeja eräissä peruskoulun oppiaineissa. Koottu tieto oli pääosaltaan kvantitatiivista eikä kaikkea koottua runsasta aineistoa kyetty raportoimaan sisältöä perusteellisemmin tulkiten.

Peruskoulua toteutettiin siirtymäkauden 1972-77 aikana koulutuksellisen tasa-arvon hengessä, joka määriteltiin koulutukseen pääsyn, koulutusjärjestelyjen ja koulutuksen tulosten yhdenvertaisuuden kolmioppina (Rinne & Vuorio-Lehti 1996:49; Rinne 2001). Vuosikymmenen loppupuolella tasa-arvokeskusteluun nousi vielä uutena painopistealueena kysymys sukupuolten välisestä tasa-arvosta (Rinne - Vuorio-Lehti 1996:61). Se, että tietoa aikaisemmista oppimistuloksista ei vielääkään hyödynnetty tarpeellisessa määrin, johtuu vakiintuneen, pysyvän arviointiorganisaation puutteesta. Jo 1970-luvulla oli nimittäin laadittu muistio keskushallintojohtoisen eli kansallisen koulukoe- ja arviointitoiminnan kehittämiseksi. Siinä hahmoteltiin selkeä malli toiminnan toteuttamiseksi (Leimu & Saari 1976:103-109). Tällaista arviointitutkijoiden esittämää ajatusta systemaattisesta arviointitiedon keruusta ja sen hyödyntämistä ei liene kuitenkaan arvostettu tuolloin riittävästi. Koululainsäädännössä ei myöskään ollut mitään vaatimuksia kansallisten arviointien toimeenpanolle. Ainoa arvioinnin muoto, joka esiintyi lakitekstissä, oli oppilasarvostelu, so. todistusarvosanojen antaminen.

Kun kansantalouden tila oli huonontunut maailmanlaajuisen öljykriisin ja nopeasti nousseen työttömyyden myötä, ja valtionhallinnossa ilmeni pyrkimyksiä muun maailman tapaan hajauttaa päätösvaltaa toimijoille, syntyi tarve uudelle koululainsäädännölle. Kouluhallituksen tehtäväksi tuli peruskoululain (476/83) mukaan antaa kunnan opetussuunnitelman laadintaa ja oppiaineiden opetusta koskevat yleiset ohjeet ja päättää valtakunnallisista oppimääristä. Oppimäärien tarkistamiseen "ovat vaikuttaneet pääasiassa opetussuunnitelman laadintaa koskevien periaatteiden muuttuminen, koulutuspoliittiset ratkaisut, oppiaineiden taustalla olevien tieteenalojen kehitys ja nykyisen opetussuunnitelman toteutumisesta saatu palaute" (Kouluhallitus 1985:8). Opetussuunnitelman perusteita päätettiin myös tarkistaa määrävuosittain "kulloisenkin kehittämistarpeen ja koulutuspoliittisen tavoitteenasettelun perusteella" (Kouluhallitus 1985:8). Arvioinnin kehittämisen kannalta merkittävää oli siirtyminen valvovasta ohjauksesta tavoiteohjaukseen, minkä todettiin edellyttävän tavoitteiden selkeää esittämistä opetussuunnitelman perusteissa ja kunnan opetussuunnitelmassa (Kouluhallitus 1985:7). Desentralisointi siis viitoitti tietä arvioinnin kehittämiseksi, tarkkojen tavoitteiden asettaminen taas panos-tuotos-tyyppisten arviointimallien käytölle.

Suomessa hallinnollinen keskusvirasto Kouluhallitus muuttui vuonna 1991 Opetushallitukseksi. Sen myötä normipohjainen ohjaus muuttui informaatio-ohjaukseksi, ja viraston ainoaksi normatiiviseksi tehtäväksi jäi opetussuunnitelman perusteiden antaminen. Opetushallituksesta annetussa laissa (182/1991) Opetushallituksen tehtäväksi tuli koulutuksen tuloksellisuuden

edistäminen ja koulutuksen järjestämisen seuraaminen. Sitä varten tarvittiin arviointikäsitteistö. Vuonna 1993 valmistui ARVI -projekti ja vakiinnutettiin käsitteet tehokkuus, taloudellisuus ja vaikuttavuus. Keväällä 1994 alkoi ARMI -projekti (opetustoimen tuloksellisuuden arvioinnin metodinen kehittäminen). Lausuntokierrosten ja seminaarien jälkeen ns. ARMI -malli hyväksyttiin Opetushallituksen arviointityön pohjaksi keväällä 1995 (Yrjölä 1998:2).

Opetussuunnitelmien laatiminen pysyi kunta- ja koulutasolla, minkä seurauksena tuntimäärien ja oppiaineksen painotuksen kuntakohtainen vaihtelu ja profilointi saattoi johtaa hyvin erilaisiin opetussuunnitelmiin. Samaan aikaan tuli voimaan laki valtiosuusjärjestelmän muutoksesta (705/92), joka poisti tietyille hallinnonaloille ”korvamerkityt” varat (Lehtisalo & Raivola 1999:135), mikä johti siihen, että eri kunnissa saatettiin panostaa koulutukseen hyvin eri tavoilla. Osittain näiden haittojen välttämiseksi luotiin siis oppimistulosten arviointijärjestelmä, ja arviointien toimeenpanemisesta oli tuleva eräs Opetushallituksen tärkeistä tehtävistä.

Opetussuunnitelman perusteiden tarkistamisen aika koitti vuonna 1994. Tarve kiteytettiin koulutuksen laadun parantamispyrkimyksiksi, jotka määriteltiin

- koulutuksen tason nostamiseksi,
- koulutussisältöjen uudistamiseksi ja
- opetuksen yksilöllistämiseksi ja valinnaisuuden lisääntymiseksi. (OPH 1994:10).

Ensimmäistä kertaa perusteisiin sisältyi erillinen luku arvioinnista; siihen mennessä ohjeita oli annettu vain (oppilas)arvostelusta eli arvosanojen ja todistusten antamisesta. Ohjeet koskivat kuitenkin vain oppilaitoksen sisäistä arviointia, ei ulkoista, ulkopuolisen tahon suorittamaan arviointiin osallistumista. Uudistustarpeen perusteluissa todetaan lisäksi, että opetussuunnitelman on dynaamisen luonteensa vuoksi reagoitava mm. arviointituloksiin.

Suomen kansallisen arvioinnin perusteet sovittiin vuonna 1996 Opetusministeriön ja Suomen Kuntaliiton neuvottelujen ja opetusministeriön kutsuman työryhmätyön pohjalta laaditussa Koulutuksen arviointistrategiassa. Koulutuksen arvioinnin tavoitteeksi tuli mm. ”tuottaa tietoa kansallisen koulutuspolitiikan toteutumisesta sille asetettuihin tavoitteisiin nähden...” (Koulutuksen arviointistrategia 1996:1.2). Arvioinnin ensisijaisiksi käyttäjiksi määriteltiin opetusviranomaiset, poliittiset päättäjät, oppilaitokset ja niiden ylläpitäjät. Lisäksi: ”Luonnollisesti myös kansalaiset ovat kiinnostuneita siitä, miten koulutusjärjestelmä toimii ja millaisia tuloksia saadaan aikaan” (emt. 1996:1.2) Arvioinnin toteuttamiselle asetettiin myös eettisiä vaatimuksia toteamalla, että ”yksittäiset arviointitoimet eivät saa millään muotoa vahingoittaa arvioinnin kohteita tai hankaloittaa niiden toimintaa tavoitteiden saavuttamiseksi.” (emt. 1996:1.2).

Kansallisten arviointien tavoite on kahtalainen. Toisaalta niiden tehtävänä on antaa päättäjille ja kehittäjille tietoa opetussuunnitelman perusteissa asetettujen tavoitteiden toteutumisesta sekä siitä, onko niissä todettavissa eroja sukupuolten, alueiden tai kieliryhmien välillä. Toisaalta taas kunnan ja koulun tasolla kansalliset oppimistulokset tarjoavat vertailukohtaan omien koulujen tulosten tarkasteluun ja kehittämisalueiden määrittämiseen tavoitteena se, että erot kou-

lujen vaatimustason ja arvosanojen antamisperusteiden välillä vähenisivät. Vä-
lillisesti arviointi antaa tukea myös oppilasarvosteluun, koska kouluille toimi-
tettujen palautteiden tuloksia voidaan verrata oppilaiden saamiin arvosanoihin.
Kansallisen arvioinnin tuloksia ei ole kuitenkaan sellaisinaan tarkoitettu ar-
vosanan antamisen perustaksi.

Kiinnitettäessä huomiota siihen, miten lähellä itse opetustapahtumaa ope-
tustointia suunnitellaan tai toimeenpannaan, voidaan erottaa seuraavat tasot:

- Koulutuspoliittinen järjestelmätaso,
- Opetussuunnitelmallinen ohjelmataso ja
- Opetus/oppimisprosessien yksilötason toiminta (Leimu 2004a:10).

Taulukko 5 kuvaa Suomen opetustoimen arviointijärjestelmää ja osoittaa, mikä
taho arvioi mitäkin aluetta tai tasoa, mikä arvioinnin tavoite on ja miten se to-
teutetaan. Sen ensimmäisen rivin, Opetushallituksen ja opetusministeriön vas-
tuualueen Miksi? -sarakkeen tavoitteet viimeistä lukuun ottamatta ovat niitä,
joihin kansallisilla arvioinneilla haetaan vastauksia.

TAULUKKO 5 Suomen opetustoimen arviointijärjestelmä. (Lähde: OPH 1998b:54.)

KUKA	MITÄ	MIKSI	MITEN
Opetusministeriö/ Opetushallitus	- koulutuksen vaikutta- vuutta, tehokkuutta ja taloudellisuutta - koulutuspolitiikan kehitystrendejä ja muutoksia opetustoi- messa - ajankohtaisia temaatti- sia painoalueita - tilannetta kansainväli- sesti katsottuna - koulutuksen ja muun yhteiskunnan suhteita	- koulutuspolitiikan toteutuminen - opetussuunnitelman toteutuminen - tieto kansallisesta tieto- ja taitotasosta - uudistusten seuraaminen - kehittämistarpeiden tiedostaminen - koulutuksen tason parantaminen - kansalaisten koulutuk- seen liittyvien mielipitei- den tiedostaminen	- arviointitutkimuksia ja hankkeita - tilannekuvauksia, vertailevia ja temaat- tisia arviointeja - tilastotietoja ja indikaattoreita - otantaperusteisia kansallisia kokeita - kansainvälisiä vertailuja - asiantuntijalausun- toja - valintojen seuraamista - meta-arvioinnit - Gallupit (Koulutusbarometrit)
Aluetaso	- alueen koulutuksen tarvetta ja tarjontaa - alueen oppilaitosver- kon toimivuutta	- alueen koulutus- ja kehittä- mistarpeiden tyydyt- täminen - kansalliset tarpeet	- kartoituksia ja vertailuja - selvityksiä - seurantaa
Opetuksen järjestäjä: Kunta/ Kuntayhtymä	- paikallisten koulutus- palvelujen tehokkuutta, vaikuttavuutta ja taloudellisuutta	- kunnan/kuntayhtymän tarpeet - aluetason tarpeet	- osallistuminen ulko- puoliseen arviointiin - itsearviointi - kartoituksia ja vertailuja - selvityksiä - seurantaa

Koulu / Oppilaitos	<ul style="list-style-type: none"> - toiminnan tuloksellisuutta - koulukohtaisen OPSin toimivuutta - koulun toimintaideaa ja profiilia - oppimistuloksia (omien erityispiirteiden huomioon ottaminen) - johtamiskulttuuria ja työpaikkailmastoa - koulussa viihtymistä 	<ul style="list-style-type: none"> - tulosten ja toiminnan vertaileminen ajallisesti, omiin tavoitteisiin ja vastaavanlaisiin oppilaitoksiin nähden - kehittämistarpeiden tiedostaminen - toiminnan kehittäminen - uudistusten seuraaminen - paikalliset ja aluetason tarpeet 	<ul style="list-style-type: none"> - itsearviointi - osallistuminen ulkopuoliseen arviointiin - oppimistulosten vertaileminen - ulkoinen auditointi, bench-marking - kehittämiskeskusteluja - kyselyjä - tutkimuksia - kokeita - kansallisella tasolla tehdyn arvioinnin hyödyntäminen
Opettaja/Opettaja-tiimi	<ul style="list-style-type: none"> - omaa opetustaan (sisältö ja menetelmät) - oppimisympäristöjä ja materiaaleja (tilat ja laitteet) - oppimistavoitteiden saavuttamista 	<ul style="list-style-type: none"> - oman kehityksen tukeminen - oman opetuksen kehittäminen - oppilaiden oppimisen parantaminen 	<ul style="list-style-type: none"> - oma reflektointi, itsearviointi - kyselyjä ja kurssiarviointeja - kokeita - oppilasarviointeja
Oppilas/Opiskelija	<ul style="list-style-type: none"> - omaa kehitystä - omia oppimistuloksia - oppimisympäristön mielekkyyttä - opetusmenetelmien mielekkyyttä 	<ul style="list-style-type: none"> - oman oppimisen tukeminen - oppimisympäristön parantaminen - opetusmenetelmien kehittäminen 	<ul style="list-style-type: none"> - portfoliot ja itsearviointilomakkeita - oppilasarviointeja yksin ja ryhmissä - kehittämiskeskusteluja (opettaja + oppilas + vanhemmat)

Suomessa toimeenpantavat perusopetuksen kansalliset arvioinnit perustuvat 1.1.1999 voimaan tulleeseen Perusopetuslakiin, jonka Koulutuksen arviointi-luvussa todetaan: "Opetuksen järjestäjän tulee arvioida antamaansa koulutusta ja sen vaikuttavuutta sekä osallistua 3 momentin mukaiseen ulkopuoliseen toimintansa arviointiin" (628/1998). Oppilaitoksen toiminnan ulkopuolisen arvioinnin järjestämisestä todetaan, että "Opetushallitus huolehtii asianomaisen ministeriön päättämien perusteiden mukaisesti arvioinnin kehittämisestä ja ulkopuolisten arviointien toimeenpanosta" (emt. 5 luku, 21 §). Arviointien toimeenpano perusopetuksessa määritellään antamalla ohjeita tehtävien suunnittelusta, otannasta, kansallisten kokeiden järjestämisestä ja koesuoritusten tarkistamisesta (OPH 1998a, 11-13).

Ainekohtaisessa oppimistulosten arvioinnissa keskitytään opetussuunnitelman perusteissa tavoitteeksi asetettujen valmiuksien saavuttamisen arviointiin (OPH, 1998a:6). Koska perusteet sisältävät ainekohtaisten tavoitteiden lisäksi myös muita tavoitteita, kuten esimerkiksi kiinnostuksen herättämisen oppiainetta kohtaan (OPH 1994:68), arviointiin sisällytetään yleensä myös oppilas-, opettaja- ja rehtorikysely. Rehtorikyselyllä pyritään selvittämään esimerkiksi koulun toiminnan taloudellisia edellytyksiä, opettajakyselyillä opetussuunnitelman perusteiden, kunnan/koulun opetussuunnitelman ja oppimateriaalin yhteensopivuutta sekä opetustilojen ja -välineiden käyttömahdollisuuksia. Oppilaskyselyillä selvitetään esimerkiksi oppilaiden suhtautumista arvioitavaan oppiaineeseen sekä heidän käsitystään omasta kielitaidostaan.

Opetussuunnitelman perusteiden ainekohtaiset tavoitteet A1-englannin opetuksessa

Vuoden 1994 Perusopetuksen opetussuunnitelman perusteet, joiden tavoitteiden toteutumista tässä tutkimuksessa tarkasteltiin, antavat vieraan kielen opetukselle seuraavat yleistavoitteet:

Oppilas

- selviytyy opiskeltavalla kielellä jokapäiväisen elämän viestintätehtävissä ja -tilanteissa
- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja,
- saa tietoja kielialueen maista, kansoista ja kulttuureista ja
- asennoituu ennakkoluulottomasti eri kulttuureihin ja niiden edustajiin (OPH 1994:68).

Opetussuunnitelman perusteissa kielenopetuksen yleistavoitteet määritellään taitoina, tietoina ja asenteina. Niissä korostetaan mm. konstruktiiivista yhteistoimintaa sekä vastuun ottamista omasta opiskelusta ja opiskelutaitojen kehittämisestä. Oppimaan oppimisen tavoitteet ilmaistaan seuraavasti:

- kehittää opiskelutaitojaan yksin ja ryhmässä,
- kehittää itsearviointikykyjään ja oppia ottamaan vastuuta oppimisestaan,
- kokea opetus ja opiskelu mielekkäänä, elämyksellisenä ja haasteellisenä sekä
- kiinnostua vieraista kielistä ja kulttuureista (OPH 1994:68).

Peruskoulun yläasteen päättövaiheen tavoitteiden mukaan oppilas

- ymmärtää jokapäiväisiä asioita käsittelevää normaalitempoista puhetta,
- pystyy osallistumaan tavallisia asioita käsittelevään keskusteluun soveltaen luonnollista ja sujuvaa ääntämistä, painotusta, rytmisiä ja intonaatiota,
- ymmärtää helpohkoa yleisiä asioita käsittelevää kirjoitettua kieltä,
- on omaksunut kielen keskeistä sanastoa, keskeisiä sanontoja ja perusrakenteita,
- kykenee tuottamaan lyhyen, esimerkiksi kertovan tai kuvailevan, kirjoitetun tekstin tarvittaessa apuneuvoja käyttäen,
- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja sekä
- on omaksunut tietoja kielialueen maista, kansoista ja kulttuureista (OPH 1994:69).

Osaamisessa painottuvat funktionaalisuus (toiminnallisuus) ja kommunikatiivinen kompetenssi eli viestin sisällön ja välittämistavan hallinta niin lähettäjänä kuin vastaanottajanakin. Formaalisuuden (muotoon perustuva) vaatimusta ei erikseen korosteta. Opiskeltavien kielellisten rakenteiden (kieliopillisen kompetenssin) hallintaan viitataan antamalla opetuksen perustaksi niiden viestintäarvo ja esiintymistiheys. Aihepiirejä ei yksilöidä. Ne on ilmeisesti ajateltu kuuluviksi koulukohtaisissa opetussuunnitelmissa määriteltäviksi viestinnällisyyden ja opetussuunnitelman perusteissa esiteltyjen aihekokonaisuuksien pohjalta. Metodologisia ohjeita on vähän. Niissä esitetään kiinnitettäväksi huomiota esimerkiksi oman mielipiteen ja tunteiden ilmaisemisen sekä huomion kiinnittämiseen kohde- ja äidinkielen eroihin. Tienä tehokkaaseen kielenoppimiseen nähdään tunnistaminen, ymmärtäminen, omaksuminen ja omien oppimistavoitteiden määrittäminen, joiden kautta edetään kielen itsenäiseen ja aktiiviseen soveltamiseen (OPH 1994:71-74.)

Opetussuunnitelman perusteet eivät anna taitotasoehdotuksia, joten tasot on ollut määriteltävä koulukohtaisissa opetussuunnitelmissa ja kansallisia arvioiteja varten Opetushallituksessa. Taitotasojen määrittämisen tarve oli kuitenkin jo havaittu, ja tätä työtä tukemaan ilmestyivät vuonna 1999 Perusopetuksen päättöarvioinnin hyvän arvosanan (8) kriteerit. Tutkimuksen kohteena olevan oppilasikäluokan kouluarvioinneissa niitä ei siis vielä ollut kuitenkaan ehditty soveltaa.

Vuoden 2004 Opetussuunnitelman perusteet (OPS 2004) eroaa edeltäjästään selkeimmin siihen sisällytettyjen taitotasokuvausten ja -tavoitteiden osalta. Ne on pyritty laatimaan sekä soveltuviksi Suomen kielenopetukseen että samalla yhteensopiviksi Eurooppalaisen viitekehyksen taitotasojen kanssa. Tässä tutkimuksessa verrataan vuoden 1999 kansallisen kokeen englannin kielen oppimistuloksia molempiin em. taitotasoihin samalla kun myös selvitetään, missä määrin ne vastaavat toisiaan.

Hyvän oppilaan osaaminen (arvosana 8) on OPS 2004:ssä määritelty 6.:n ja 9.:n luokan päättyessä taulukon 6 mukaiseksi. Kirjain- ja numeroyhdistelmät viittaavat kuvaimiin, jotka ovat opetussuunnitelman perusteiden liitteenä.

TAULUKKO 6 OPS 2004:n kuvaus hyvän oppilaan englannin kielen osaamisesta 6. ja 9. luokan päättyessä.

Vuosiluokka	Kuullun ymmärtäminen	Puhuminen	Tekstin ymmärtäminen	Kirjoittaminen
6.	A2.1 Peruskielitaidon alkuvaihe	A1.3 Toimiva alkeiskielitaito	A2.1 Peruskielitaidon alkuvaihe	A1.3 Toimiva alkeiskielitaito
9.	B1.1 Toimiva peruskielitaito	A2.2 Kehittyvä peruskielitaito	B1.1 Toimiva peruskielitaito	A2.2 Kehittyvä peruskielitaito

Tietyn taitotason hallinnan lisäksi odotetut kulttuuritaidot ja opiskelustrategiat on kirjattu. Kommunikatiivisen kielenopetuksen lisääntynyt painotus näkyy mm. siinä, että sekä 6. että 9. luokan päättövaiheen hallittavat viestintästrategiat on koottu omien otsikoidensa alle (taulukko 7).

TAULUKKO 7 OPS 2004:n kuvaus viestintästrategioiden hallinnasta.

6. lk	9. lk
- puheen ja kirjoitetun tekstin pääasioiden tunnistaminen	- kielellinen tai tilanneviheisiin perustuva päättely viestin sisällön selvittämiseksi
- rajatun tiedon löytäminen tekstistä ja puheesta	- vuorovaikutustilanteessa saadun palautteen hyödyntäminen
- omien viestien suunnittelu	- puuttuvan kielitaidon kompensoiminen likimääräisellä ilmaisulla
- puhekumppanin apuun ja nonverbaaliseen viestintään tukeutuminen suullisessa vuorovaikutuksessa	- oman kielenkäytön tarkkailu
- kirjallisiin apuneuvoihin tukeutuminen tekstin tuottamisessa ja tulkitsemisessä	- joidenkin suulliselle vuorovaikutukselle ominaisten ilmausten, kuten puheenvuoron aloitukseen ja lopetukseen sekä puheenvuoron ottamiseen ja ylläpitämiseen ja palautteen antamiseen liittyvien ilmaisujen käyttö

Opiskelustrategiat ja kulttuuritaidot oli sisällytetty jo vuoden 1994 Opetussuunnitelman perusteisiin, mutta nyt ne on ryhmitelty selkeästi erikseen. Rakenteiden oppimistavoitteet samoin kuin keskeiset sisällöt eli tilanteet ja aihepiirit ovat myös uusia ja tervetulleita ryhmiä.

2.2 Koulutuksen arvioinnin etiikka ja standardit

Arviointia luonnehdittaessa tuodaan joissakin yhteyksissä esiin näkemys, että arviointi on vallankäytön väline. Valtaan liittyy aina myös vastuu. Arvioinnilla on aina seuraamuksia ja mitä merkityksellisempiä arvioinnin tulokset ovat arvioitaville (*high-stakes tests*), sitä suurempi on arvioijan vastuu. Arviointivastuuta on pyritty ohjeistamaan eettisillä säännöillä ja standardeilla. Artikkelissaan "Ethical Considerations in Evaluation" Morris määrittelee arvioinnin etiikan ja standardien välisen yhteyden toteamalla etiikan käsittelevän hyvää ja pahaa sekä moraalisia velvollisuuksia ja velvoitteita. Hän jatkaa, että yksi arvioijien ammattikunnan käyttämä strategia, jolla pyritään takaamaan se, että sen jäsenet menettelevät oikein, on kehittää standardeja, periaatteita ja/tai toimintaohjeita (*codes*). (Morris 2003:303) Etiikka ja eettiset pohdinnat luovat näin ollen perustaa menettelytapaohjeille ja standardeille.

Koulutuksen arvioinnin eettiset kysymykset ovat melko myöhäsyntyisiä, mikä johtuu siitä, että arviointi professiona alkoi kehittyä ja jäsenyää laajemmin vasta 1980-luvulla. Edelläkävijänä tässä asiassa oli vuonna 1975 perustettu the Joint Committee on Standards for Educational Evaluation, jonka tehtävänä on koulutuksen arvioinnin jatkuva kehittäminen ja sen standardien muokkaaminen ja kehittäminen (Stufflebeam 2003:280) Komitea julkaisi vuonna 1981 "Standards for Evaluations of Educational Programs, Projects and Materials", joka lyheni vuoden 1994 versioon mennessä muotoon "The Program Evaluation Standards". Standardit on jaettu neljään kategoriaan, joiden sisältö määritellään yhteensä 30:llä kriteerillä. Taulukko 8 on yksinkertaistettu käännös em. standardeista. Kategorioiden käännökset ovat Jakku-Sihvosen ja Heinosen (2001).

TAULUKKO 8 Ohjelma-arviointistandardien yhteenveto. (Lähde: Stufflebeam 2003:284-286.)

Kategoria	Kriteerit
<i>Hyödyllisyys</i> (Utility)	<p>- takaavat, että arviointi palvelee aiottujen käyttäjien tarpeita.</p> <ol style="list-style-type: none"> 1. Arvioinnin tulee määrittää asianosaiset. 2. Arvioijan tulee olla uskottava. 3. Annettavan informaation laajuus ja valintaperusteet tulee määrittää. 4. Arvot tulee määrittää. 5. Raportilta tulee vaatia selkeyttä. 6. Raportoinnin tulee olla nopeata ja oikea-aikaista, ja sen vastaanottajat tulee määrittää. 7. Raportin vaikutus tulee pyrkiä varmistamaan kaikin tavoin.

<p>Sovellettavuus (Feasibility)</p>	<p>- edellyttävät, että arvioinnin tulee olla realistista, viisasta, diplomaattista ja taloudellista.</p> <ol style="list-style-type: none"> 1. Arviointitoimenpiteiden tulee olla käytännöllisiä. 2. Arviointi tulee suunnitella eri eturyhmien mahdolliset vastakkainasettelut huomioon ottaen. 3. Arvioinnin tulee olla kustannustehokasta.
<p>Asianmukaisuus ja oikeudellisuus (Propriety)</p>	<p>- korostavat, että arviointi tulee suorittaa lakia noudattaen, eettisesti ja huomioon ottaen kaikkien niiden edut, joihin arviointi vaikuttaa.</p> <ol style="list-style-type: none"> 1. Arvioinnin tulee palvella asianosaisia. 2. Velvoitteet arvioinnin muodollisia osapuolia kohtaan tulee kirjata. 3. Arvioinnit tulee suunnitella ja toteuttaa kunnioittaen ja suojellen ihmisoikeuksia. 4. Arvioijien tulee kunnioittaa ihmisarvoa toimiessaan arvioinnin kanssa tekemisissä olevien henkilöiden kanssa. 5. Arvioinnin tulee olla kokonaisvaltaista ja rehellistä ja kirjata ohjelmien vahvuudet ja heikkoudet niin, että vahvuuksia voidaan tukea ja ongelmat ratkaista. 6. Tulokset tulee saattaa asianosaisten käyttöön. 7. Eturistiriidat on käsiteltävä avoimesti ja hienotunteisesti niin, että ne eivät vaaranna arviointiprosessin tuloksia. 8. Arvioinnin kustannusten tulee olla läpinäkyviä ja menojen tarpeellisuus tulee pystyä perustelemaan.
<p>Tarkkuus (Accuracy)</p>	<p>- takaavat, että arviointi välittää teknisesti riittävää informaatiota piirteistä, joilla ohjelman arvot ja ansiot on määritetty.</p> <ol style="list-style-type: none"> 1. Ohjelma tulee kuvata selvästi ja tarkasti. 2. Ohjelman konteksti tulee kuvata riittävän yksityiskohtaisesti, jotta sen vaikutukset ohjelmaan voidaan määrittää. 3. Ohjelman tavoitteiden toteutumista tulee seurata ja kuvailla niitä mahdollisimman yksityiskohtaisesti. 4. Ohjelman arvioinnin tietolähteitä tulee seurata ja kuvailla niitä mahdollisimman yksityiskohtaisesti. 5. Informaation keräämismenetelmät tulee valita niin, että ne takaavat validit tulokset. 6. Informaation keräämismenetelmät tulee valita niin, että ne takaavat reliabelit tulokset. 7. Koottu, käytetty ja raportoitu tieto tulee tarkistaa ja virheet korjata. 8. Kvantitatiivinen tieto tulee analysoida asianmukaisesti ja systemaattisesti niin, että vastaukset evaluaatiokysymyksiin saadaan. 9. Kvalitatiivinen tieto tulee analysoida asianmukaisesti ja systemaattisesti niin, että vastaukset evaluaatiokysymyksiin saadaan. 10. Johtopäätökset tulee perustella selkeästi, jotta asianosaiset voivat arvioida ne. 11. Raportoinnissa tulee varoa henkilökohtaisten tunteiden tai harhaisuuden aiheuttamaa vääristymistä niin, että arviointiraportti heijastaa oikeudenmukaisesti arviointituloksia. 12. Itse arviointi tulee arvioida formatiivisesti ja summatiivisesti suhteessa näihin ja muihin vastaaviin standardeihin niin, että se on asianmukaisesti suoritettu ja että sen päätyttyä asianosaiset voivat tutustua sen vahvuuksiin ja heikkouksiin.

Samoihin aikoihin em. standardien kehittämisen kanssa the Evaluation Research Society (ERS) julkaisi 55-kohtaiset standardit, joita edelleen päivitetään AEA:n (American Evaluation Association) toimesta nimellä AEA Guiding Principles for Program Evaluators ja jotka on suunnattu kaikkeen, ei vain koulutuksen, arviointiin (Stufflebeam 2003:280).

Kielitaidon arvioinnin etiikkaa on käsitellyt mm. McNamara (2000). Hän erottaa kaksi näkökulmaa, jotka molemmat ottavat huomioon arvioinnin yhteiskunnallisen ja poliittisen roolin. Toinen, eettinen näkemys, korostaa arvioi-

jan henkilökohtaista vastuuta eettisten menetelmien käytöstä ja johtopäätösten tekemisestä. Toinen, kriittinen arviointi, pitää arviointia yhteiskuntapoliittisena konstruktina, joka tulee voida alistaa samanlaiselle kritiikille kuin yhteiskunnan muutkin rakenteet (2000:71-72)

Eettistä arviointia voidaan katsoa laajemmasta tai rajatummasta näkökulmasta. McNamara (2000) kutsuu edellistä yhteiskunnallisesti vastuulliseksi näkemykseksi ja jälkimmäistä perinteiseksi. Yhteiskunnallisesti vastuullinen arviointi lähtee siitä, että testaus ei ole vain tieteellistä ja teknistä toimintaa, vaan että kokeen seurausvalidiuden tueksi osoitetaan, että on otettu huomioon niin mieluisat kuin epämieluisatkin tulokset. Eettisen arvioinnin toteuttajan tulee siis ottaa huomioon kokeen vaikutukset, joita ovat tilivelvollisuus sekä testattavaa että arvioinnin toimeenpanijaa kohtaan, takaistusvaikutus opetukseen ja vaikutus, jolla tässä tarkoitetaan arvioinnin vaikutusta koulutuksen kehittämiseen (McNamara 2000:72).

Kielten arvioinnin etiikkaa käsiteltiin laajasti vuonna 1997 kollokviumissa "Fairness in Language Testing", ja samana vuonna Language Testing-lehti julkaisi erillisen, Alan Daviesin (1997) toimittaman numeron, joka käsitteli kielten arvioinnin etiikkaa. National Council on Measurement in Education on koonnut dokumentin the "Code of Fair Testing Practices in Education" ja International Language Testing Association (ILTA) on julkaissut omat eettiset ohjeensa "Code of Ethics for ILTA", jotka hyväksyttiin vuonna 2000 ILTA:n vuosikokouksessa. "The ILTA Code of Practice" -luonnos hyväksyttiin kesällä 2005. Myös muutama vuosi sitten perustettu EALTA (European Association for Language Testing and Assessment) hyväksyi hyvän käytännön periaatteet kolmannessa vuosikongressissaan Krakovassa toukokuussa 2006.

Jakku-Sihvonen ja Heinonen (2001) toteavat, että Suomessa on koulutuksen arvioinnin etiikasta kirjoitettu melko vähän. He viittaavat Koulutuksen arviointistrategian periaatteellisiin ohjeisiin sekä Koulutuksen tuloksellisuuden arviointimallin avoimuuden korostamiseen (OPH 1998b:21-22;49). Raivola (2000) esittelee Program Evaluation Standardsin neljä kategorialaajaa tulkintoihin ja Laukkanen (1994, 1998) useiden ulkomaalaisten koulutuksen etiikkaa ja standardeja pohtineiden teoreetikkojen käsityksiä. Linnakylä (2004) otti virkaantujaisesityksessään "Koulutuksen arvioinnin eettiset ehdot" esille monia edellä käsiteltyjä teemoja, joiden sisällyttäminen arviointikeskusteluun olisi varsin suotavaa.

Eettiset ohjeet ja arviointistandardit on pääasiassa laadittu muiden maiden kuin Suomen lainsäädäntöä myötäileviksi ja niissä esille tulleiden arvioinnin ongelmien välttämiseksi. Siksi niistä tulisi valita kulloisenkin arviointiin niiden erityistarpeisiin soveltuvimmat kohdat, ja näin on tässäkin arviointitutkimuksessa pyritty tekemään. Ongelmat arvioijien ja asianosaisten välillä ovat eri tasoisia, ja arviointia vallankäytön välineenä esiintyy eri maissa ja eri olosuhteissa eri mittakaavassa. Tätä ongelmaan kielten arvioinnissa on käsitelty esim. Shohamy (1993).

2.3 Yhteenveto

Arviointi on laajimmillaan arvon (*worth*) ja ansion (*merit*) määrittämistä. Ansio viittaa siihen, miten hyvä arviointikohde on eli onko sen puitteissa tehty työ siinä määrin tuloksellista, että käyttäjällä on sille arvoa. Ansio voidaan määrittää rajaamalla sen dimensiot eli arvioinnin kohteet ja standardit eli arvioinnin taso. Käsillä olevassa tutkimuksessa arvot (*worth*) on otettu annettuina, mutta muutoin koko arviointiprosessi on pyritty esittämään mahdollisimman kattavasti eri näkökulmat huomioon ottaen. Koska tavoitteena on selvittää onko palaute annetusta opetuksesta odotusten mukaista, tässä tutkimuksessa korostuu ansion näkökulma.

Ihanteellisen arviointiteorian – jota kuitenkin ei vielä ole kehitetty – tulee kuvata ja perustella, miksi tietyt evaluaatiokäytännöt johtavat tietynlaisiin tuloksiin. Sen tulee ottaa kantaa koko arviointiprosessiin: ohjelmien laatimiseen, tiedon hankkimiseen ja arvojen liittämiseen ohjelman kuvaukseen sekä saadun tiedon käyttöön. Arviointia koskevassa keskustelussa käydään usein rajaa arvioinnin ja arviointitutkimuksen välillä. Cronbachiin viitaten esitetään usein tulkinta, jonka mukaan edellinen on päätös- ja jälkimmäinen johtopäätöskeskkeistä. Suhteessa arvioinnin käyttötarkoituksen huomioon ottamiseen arviointiprosessissa on olemassa useita koulukuntia. Jotkut katsovat Scrivenin näkemyksiin vedoten, että arvioinnin eri puolet eivät tule riittävän hyvin esiin, jos tavoite on ennalta määrätty, kun taas toisten mielestä tarkoituksenmukaisen menetelmän valitseminen edellyttää arviointitavoitteen tuntemista. Tämä tutkimus sijoittuu näiden määritelmien mukaan arvioinnin ja arviointitutkimuksen välimaastoon pyrkiessään tekemään johtopäätöksiä, jotka vaikuttavat myös päätöksentekoon.

Suomessa 1960-luvulla alkanut koulu-uudistus nojasi yhteiskunnallisiin muutoksiin: kaupungistumiseen sekä koulutuksellisen että sukupuolten välisen tasa-arvon vaatimukseen. 1970-luvun öljykriisi ja kohonneet työttömyysluvat johtivat valtionhallinnon jakamaan päätösvaltaa toimijoille. Koulutusohjelmien arvioinnin esiinmarssin taustalla olivat toisaalta havahtuminen siihen, että koulutuksen alueella panos – tuotos -ajattelu ei ollut osoittautunut riittävän hyvin toimivaksi, toisaalta resurssien väheneminen, mikä edellytti sekä tiukempaa talouskuria että myös lisääntyntä läpinäkyvyyttä. Näiden seurauksena syntyivät kansalliset arvioinnit, joista ensimmäisiä oli yhdysvaltalainen NAEP, jota seurasivat useiden Euroopan maiden arviointiohjelmat. Valvovasta ohjauksesta siirryttiin tavoiteohjaukseen. Desentralisointi viitoitti tietä arvioinnille asettamalla tavoitteet, joiden toteutumista oli arvioitava. Opetushallituksen antamat opetussuunnitelman perusteet ovat tällaisia tavoitteita. Niiden arviointi edellyttää arviointia tavoitteiden suunnassa. Tämän tutkimuksen kohteena oleva arviointi on eräs esimerkki tästä.

Vastuullisen arvioinnin takaamiseksi eri järjestöt ovat laatineet jäseniään varten eettisiä ohjeita ja arviointistandardeja. Tunnetuimpia näistä on ”The Program Evaluation Standards”, jotka sisältävät arvioinnin hyödyllisyydelle, sovellettavuudelle, asianmukaisuudelle ja tarkkuudelle asetettuja vaatimuksia.

Näiden standardien huomioon ottamiseen on pyritty arvioinnin laadun varmistamiseksi.

Seuraavassa luvussa kuvataan aikaisempia englannin kieltä koskevia arviointeja sekä esitellään joitakin muissa maissa toteutettuja arviointisovelluksia.

3 ENGLANNIN KIELEN OPPIMISTULOSARVIOINTEJA

Suomessa on järjestetty oppimistulosarviointeja 1960-luvulta saakka. Valtaosaltaan ne ovat kohdistuneet vain osaan kielitaidon osa-alueista. Samoin on niiden harvojen kansainvälisten arviointihankkeiden laita, joihin Suomi on osallistunut. Eri vuosina järjestettyjen arviointien tulokset eivät ole välttämättä vertailukelpoisia mm. siksi, että niissä on käytetty eri kokeita, ja niiden tulokset ilmaistu ratkaisuprosentteina, mikä ei anna riittävään yhdenmukaista ja vertailukelpoista kuvaa koko populaation tasosta. Niistä voi kuitenkin päätellä kielitaidon tason trendin eri aikoina. Kansainvälisten arviointien tulokset puolestaan antavat vertailupohjaa Suomessa järjestettyjen kokeiden tuloksiin. Tuloksia ja trendejä voidaan verrata tämän tutkimuksen luvussa 9 esitettyihin oppimistuloksiin.

Kansalliset arvioinnit korvaavat Suomessa entistä tarkastajajärjestelmää. Niitä toteutetaan eri maissa eri tavoin. Siksi tähän tutkimukseen on sisällytetty esimerkkejä muiden maiden jo vakiintuneista tai kehitteillä olevista kansallisista arvioinneista. Koska arviointimenetelmät ovat jatkuvan kehityksen alaisia, esimerkit eivät ehkä vastaa tämän julkaisun ilmestymisajankohdan tilannetta. Silti ne täyttänevät tarkoituksensa kansallisten arviointien kehittämismalleina.

3.1 Suomen perusopetuksen englannin kielen arviointeja

Olennaisena osana yhtenäiskoulutyypin oppivelvollisuuskoulun toiminnan seuraamista 1960-luvun puolivälin jälkeen oli oppimistulosten systemaattinen seuraaminen ja arviointi englannin kielessä vieraana kielenä. Esimerkiksi vuoden 1969 keväällä järjestettiin luokille 4 ja 6 suullisen kielitaidon kokeet, joiden tulokset kuitenkin kyettiin raportoimaan vasta yli 20 vuotta myöhemmin. Kokeen merkittävä anti oli havainto, jonka mukaan kriteereihin tukeutuen oli mahdollista arvioida produktiivisia kokeita luotettavasti. Sukupuolten väliset erot todettiin vähäisiksi. Tytöt menestyivät kuitenkin molemmilla luokka-asteilla poikia paremmin (Takala 1993:16-18).

Vuoteen 1975 saakka toteutettiin joka vuosi useita koulusaavutusmittauksia, joita käytettiin hyväksi opetussuunnitelmien kehittämisessä (Takala 1982a:25). Esimerkiksi keväällä 1972 varsinaista peruskoulukokeilua suorittavien kuntien 6. ja 9. luokkien oppilaille laadittiin englannin kokeet, joissa mitattiin tasokursseittain (laaja, keski- ja yleiskurssi) tekstin ja puheen ymmärtämistä, sanastoa ja rakenteita. Viimeksi mainituissa taidoissa todettiin puutteita molemmilla luokka-asteilla ja kaikilla tasokursseilla (Havola-Pitkänen 1974).

1970-luvun puolivälissä keskeisen mielenkiinnon kohteeksi Suomessa oli noussut kaikille oppilaille yhteisten tavoitteiden pohtiminen eli ns. ydinaineiksen määrittäminen, koska tasokurssijärjestelmä ei antanut alimman tasokurssin oppilaille mahdollisuutta esimerkiksi lukio-opintoihin. Vuoden 1976 loppupuolella käynnistyi peruskoulun tilannekartoitus 1:n suunnittelu, jossa tavoitteena oli selvittää, missä määrin oppilaat hallitsisivat englannin kielen ydinaineiksen ja perustaidot sellaisina kuin ne oli määritelty ns. VIKKE -oppaassa (1976). Muita tavoitteita olivat mm. seuraavat:

1. Arvioinnin tulee tuottaa valideja ja reliaabeleita ns. oppiaineen yhteisen sisällön ja perustaitojen kokeita sekä osiopankki (ns. koetehtäväpankki).
2. Tutkitaan, missä määrin tavoitteiden saavuttamista selittävät
 - a) opetusta ja luokkahuonetapahtumia kuvaavat tekijät,
 - b) oppilaan sukupuoli, kotitausta, kodin antama tuki ja vuorovaikutus kotona,
 - c) koulun ja luokan koko ja
 - d) tasokurssit.

Arviointi toimeenpantiin 2.-20.4.1979. Tutkimusaineiston keruu edellytti oppilaalta kolmen tunnin käyttöä eri päivinä. Ensimmäinen tunti käytettiin oppilaskyselyyn vastaamiseen, minkä yhteydessä oli tarkoitus opettaa oppilaita optisen vastauslomakkeen käyttöön. Toisen tunnin aikana oppilaat suorittivat puheen ymmärtämisen ja sanaston kokeet, ja kolmannella tunnilla tekstin ymmärtämisen ja rakenteiden kokeet.

Eri osakokeista parhaiten menestyttiin puheen ymmärtämisen kokeessa (ratkaisuprosentti 52). Tekstin ymmärtämisessä menestyttiin jonkin verran heikommin (46 %), ja heikoimmin rakennekokeessa (39 %). Sanaston kokeen tuloksia ei ole raportoitu tässä julkaisussa (Takala 1982a:38). Tarkempaa pohdintaa erityisesti tämän kokeen tekstin ymmärtämisen tehtävien tuloksista on esitetty selvityksessä englannin kielen tasosta Suomessa (Takala 1998).

Tutkimusasetelma oli osittain hierarkkinen, millä tarkoitetaan sitä, että koetehtävät jaetaan useihin koeversioihin ja esitetään vain yksi versio kullekin oppilaalle (Takala 1982a:26-32). Suunnittelussa hyödynnettiin silloin uusia ideoita mm. otantateorian alalla eli monimatriisiotantaa ja testiteorian alalla ns. latenttien piirteiden testiteoriaa, jotka mahdollistivat käytännön mittauksien olennaisen laajentamisen aiempaa paremman tulosten luotettavuuden ja yleistettävyyden saavuttamiseksi (Takala 1982a:22).

Vuonna 1989 Kasvatustieteiden tutkimuslaitoksessa käynnistettiin opetushallinnon kanssa yhteistyössä peruskoulun toiminnan ja tuloksellisuuden arviointihanke "Peruskoulun arviointi 90", jonka aineisto koottiin vuosina 1990-1 (Linnakylä & Saari 1993:1) ja joka englannin kielen osalta kohdistui 9. luokan op-

pilaisiin. Tämä koe koski ensi kerran koko ikäluokkaa¹, josta tosin vain määritetty otos osallistui tutkimukseen. Tarkoitus oli mitata englannin kielen puheen ymmärtämistä (Linnakylä & Saari 1993:3). Myös puhetaidon testaamista pidettiin kiinnostavana, mutta se olisi vaatinut runsaasti resursseja, joten molempia suullisen kielitaidon alueita ei voitu ottaa testattavaksi (Havola & Saari 1993:158).

Suurimmat alueelliset ja tilastollisesti samoin kuin käytännössäkin erittäin merkitsevät erot todettiin silloisen Uudenmaan läänin ja Itä- ja Pohjois-Suomen tulosten välillä. Uudenmaan keskiarvo oli 16/20 pistettä (ratkaisuprosentti 80), Itä- ja Pohjois-Suomen 14/20 pistettä (ratkaisuprosentti 69). Jälkimmäisellä alueella oppilaat siis osasivat noin 86 % siitä, mitä Uudenmaan läänin oppilaat hallitsivat. Kaupunkien ja maaseudun välillä todettiin myös tilastollisesti erittäin merkitsevä ero kaupunkien hyväksi. Tyttöjen ja poikien välinen ero ei sen sijaan osoittautunut tilastollisesti merkitseväksi eikä myöskään pienten ja isojen koulujen välinen ero (Havola & Saari 1993:168). Tässä arvioinnissa saavutettuja tuloksia oli mahdollista verrata vuonna 1971 järjestetyn IEA -tutkimuksen (ks. Kansainvälinen arviointiyhteistyö) kohderyhmien, rinnakkaiskoulujärjestelmän 14-vuotiaiden sekä abiturienttien tuolloin osoittamaan kielitaitoon. Tulokset osoittavat keskimääräisen osaamisen parantuneen huomattavasti tutkimusajankohdasta toiseen. Vuonna 1991 peruskoulun oppilaiden keskiarvo on korkeampi kuin laajan kurssin keskiarvo vuonna 1979 (Havola & Saari 1993:165).

Opetushallituksen vuonna 1993 aloittamassa peruskoulun oppimistulosten arviointihankkeessa toteutettiin peruskoulun 9. luokan oppilaiden englannin kielen arviointiin ja sen kehittämiseen liittyvä tutkimus. Tavoitteena oli saada tietoa kielikokeiden kehittämisen pohjaksi, koulujen itsearviointin toteuttamiseksi ja opetussuunnitelman kehittämiseksi. SUKOL² toimitti kokeen, joka pidettiin keväällä 1994. Kokeessa mitattiin tekstin ja puheen ymmärtämistä, kirjallista tuottamista, sanaston ja rakenteiden hallintaa sekä reagointia. Viimemainittu oli eräänlainen puhekokeen simulaatio, jossa oppilas reagoi nauhalta kuulemiinsa väittämiin, kysymyksiin tai keskustelukatkelmiin kirjoittamalla vastauksen sen sijaan että sanoisi sen.

Kokeeseen osallistui runsaat 2000 oppilasta, ja otannassa oli pyritty huolehtimaan siitä, että mukaan tuli kaupunki- ja maaseutukouluja samoin kuin isoja ja pieniä kouluja. Ruotsinkielisiä kouluja mukana oli neljä. Kokeen maksimipistemäärä oli 100. Suomenkielisten koulujen keskiarvoksi saatiin 63 ja ruotsinkielisten 73 pistettä. Tyttöjen keskiarvo suomenkielisissä kouluissa oli 66 pistettä ja poikien 59, ja koulujen keskiarvo oli 49-80 pistettä. Ruotsinkielisissä kouluissa se asettui välille 70-75. Ruotsinkielisten koulujen määrä oli kuitenkin pieni, joten tulokset eivät ole yleistettävissä.

Alueellisista eroista ilmeni, että pienimpien (alle 5 000 asukasta) kuntien koetulosten keskiarvot olivat 9 prosenttiyksikköä heikommät kuin suurempien (yli 25 000 asukasta). Tyttöjen parempi koemenestys ei kuitenkaan ollut merkitsevässä määrin riippuvainen kunnan koosta: alle 5 000 asukkaan kunnissa keskiarvo oli 63 pistettä ja yli 25 000 asukkaan kunnissa 68 pistettä. Poikien kohdal-

¹ Aikaisemmat kokeet oli järjestetty tasokursseittain.

² Suomen kielenopettajain liitto.

la ero oli merkitsevämpi, poikien pistemäärien keskiarvo oli 53 ja tyttöjen 66 pistettä. Seuraava kuvaus luonnehtii saavutettuja tuloksia: "Etelä-Suomessa isolla paikkakunnalla isoa koulua käyvän tytön menestyminen kielten opiskelussa hyvin on todennäköisempää kuin Itä- ja Pohjois-Suomessa pienellä paikkakunnalla pientä koulua käyvän pojan." (Pasanen 1996, 362).

Viimeinen laajamittainen arviointi peruskoululaisten englannin taidosta ennen vuoden 1999 kansallista arviointia on vuodelta 1995. Vuosikymmenen alkupuolen laman aikana kunnat olivat joutuneet tekemään rajujakin leikkauksia opetustoimen määrärahoista. Monilla paikkakunnilla jouduttiin esimerkiksi lomauttamaan opettajia. Tutkimuksen tarkoituksena oli mm. selvittää, oliko näillä toimenpiteillä yhteyttä englannin oppimistuloksiin. Tutkimuksessa arviointiin sekä puheen että tekstin ymmärtämisen taitoja. Puheen ymmärtämisen kokeessa käytettiin Peruskoulun arviointi 90 -tutkimuksen vuoden 1991 kokeen kolmesta osakokeesta kahta, joita oli jo sitä ennen käytetty kaksi kertaa. Tekstin ymmärtämisen kokeessa tuloksia ei voitu verrata vuosien 1995 ja 1991 välillä, koska peruskoulun arviointitutkimukseen ei vuonna 1991 sisällynyt tekstin ymmärtämisen arviointia. Vertailu kohdistettiin sen sijaan vuoteen 1983, jolloin tutkija Liisa Havola oli arvioinut 9. luokkalaisten tekstin ymmärtämisen taitoja, ja luovutti nyt ennen julkaisemattomia aineistojaan tuon tutkimuksen käyttöön. Takala (1998) on myös kommentoinut vuoden 1983 tuloksia. Saari toteaa, että puheen ymmärtämisen koetulokset olivat osoittaneet myönteistä kehitystä jo vuodesta 1971 lähtien ja että vuosien 1991 ja 1995 välillä ne ovat parantuneet noin seitsemän prosenttia. Tekstin ymmärtämisen taso näytti pysyneen suurin piirtein vuoden 1983 tasolla (Saari 1996:462).

Vuoden 1999 englannin kansallisen arvioinnin kielitaidon eri osa-alueiden ratkaisuprosentit olivat: koko koe 64, puheen ymmärtäminen 58, tekstin ymmärtäminen 70, puhuminen 60, kirjoittaminen 61 ja kieliopilliset rakenteet 65 (Tuokko 2000:5). Taulukossa 9 esitetään vuosien 1993 ja 1999 englannin kansallisten arviointien oppimistulosten vertailun keskeiset johtopäätökset (Tuokko 2004).

TAULUKKO 9 Opetushallituksen kahden viimeisen englannin arvioinnin (1993, 1999) tulosten vertailun keskeiset johtopäätökset. (Lähde: Tuokko 2004:162.)

Vuoden 1993 arviointi	Vuoden 1999 arviointi
Osaamistaso oli tyydyttävä	Osaamistaso oli tyydyttävä
Eri koulujen ja maan eri alueiden välillä oli suuria oppimistulosten eroja	Eri koulujen ja maan eri alueiden välillä oli suuria oppimistulosten eroja.
Itä- ja Pohjois-Suomen alueella kielten oppimistulokset jäivät keskimäärin muita alueita heikommiksi.	Itä- ja Pohjois-Suomen alueella kielten oppimistulokset jäivät keskimäärin muita alueita heikommiksi.
Kaupunkikoulut menestyivät maaseutukouluja paremmin	Kaupunkikoulut menestyivät 2-3 prosenttiyksikköä maaseutukouluja paremmin
Ruotsinkielisten oppilaiden koetulokset olivat keskimäärin kymmenen prosenttiyksikköä suomenkielisten tuloksia paremmat.	Ruotsinkielisten oppilaiden koetulokset olivat keskimäärin kymmenen prosenttiyksikköä suomenkielisten tuloksia paremmat.
Suomen- ja ruotsinkielisen oppilaiden kouluarvosanojen keskiarvo oli kuitenkin sama, 7,5.	Suomen- ja ruotsinkielisen oppilaiden kouluarvosanojen keskiarvo oli kuitenkin sama, 7,5.
Tytöt suoriutuivat kaikissa muissa paitsi sanaston osakokeessa poikia paremmin.	Tytöt suoriutuivat kaikissa osakokeissa poikia paremmin; pojista alimmassa kvartiilissa oli noin 32 %, tyttöjä taas noin 19 %.

Pojat saivat todistusarvosanaan suhteutettuna kokeesta keskimäärin parempia pistemääriä kuin tytöt	Pojat saivat todistusarvosanaan suhteutettuna kokeesta keskimäärin parempia pistemääriä kuin tytöt arvosanaluokissa 7-10.
Pojat jäivät produktiivisessa kielitaidossa suhteellisesti enemmän tyttöjen jälkeen kuin reseptiivisissä taidoissa	Pojat jäivät produktiivisessa kielitaidossa suhteellisesti enemmän tyttöjen jälkeen kuin reseptiivisissä taidoissa.
Kielenopetuksessa olisi kiinnitettävä enemmän huomiota oppilaiden sanaston täsmällisyyteen, sanavaraston kartuttamiseen esimerkiksi kontekstista päättelyn ja sananjohtamisen avulla sekä keskeisten kieliopin seikkojen kuten verbiopin hallintaan.	Kuullun ymmärtämisen ja kirjoittamisen osaluokilla todettiin runsaasti heikkoja suorituksia. Kielenopetuksessa olisi kiinnitettävä enemmän huomiota tekstin ymmärtämisessä päättelyyn ja johtopäätösten tekemiseen sekä tarkkan merkityssisällön hallitsemiseen, puhumisessa sosiolingvistiseen kompetenssiin sekä perussanavaraston (numerot, kuukaudet yms.) omaksumiseen sekä ääntämisen tasoon, kieliopillisissa rakenteissa verbiopin, artikkelien ja pronomien hallintaan, kirjoittamisessa tekstilajien erottamiseen ja sisällön monipuolisuuteen ja kuullun ymmärtämisessä perustaso vaativamman tekstin ymmärtämiseen.

Pasanen 1996,1, 31, 34 53, 50-51, 74-75.

Tuokko 2000, 6, 5, 42, 48, 51, 62, 94, 37, 93, 116-119

Em. johtopäätökset osoittavat, että monissa asioissa, joihin opetuksen tulisi voida vaikuttaa, ei ollut tapahtunut juurikaan parannusta. Tuloksissa ilmeni koulujen sijainnista johtuvia eroja, pojat menestyivät heikommin erityisesti puhumisessa, ymmärtämistaitojen hallinta oli puutteellista ja poikien kouluarvosanat olivat arviointituloksiin verrattuna paremmat kuin tyttöjen.

3.2 Kansainvälinen arviointiyhteistyö

Suomi osallistui vuonna 1971 laajaan IEA:n³ kuuden oppiaineen tutkimukseen, jossa englannin kielen taitoja testattiin kaikilla neljällä kielitaidon osa-alueella (Postletwhwaite 1975:5). Vain tekstin ymmärtämisen koe oli pakollinen kaikille osanottajamaille. Englannin kielen tutkimukseen ottivat osaa Belgia, Chile, Alankomaat, Israel, Ruotsi, Saksan liittotasavalta, Suomi, Thaimaa ja Unkari. Kielen ymmärtämistä testattiin pääasiassa monivalintatehtävin. Tuottamistehävissä oppilaiden tuli itse tuottaa vastaukset (Takala 1998:74; 2004:256).

Suomen 14-vuotiaiden ryhmä oli Thaimaan ohella englannin opiskelun kannalta heterogeenisin. Saksan kielellä oli tuolloin selvästi suurempi osuus oppilaiden ensimmäisenä vieraana kielenä, ja lisäksi vain osa kansalaiskoululaisista opiskeli vierasta kieltä. Koetuloksissa keskikoululaisten ryhmä menestyi keskitasoisesti ja tuottamiskokeissa jopa keskitasoa paremmin, mutta kansalaiskoululaiset suoriutuivat ymmärrettävästi kokeista todella heikosti. Yleisesti ottaen siis tämän ryhmän englannin kielen taito oli kansainvälisesti katsoen keho (Takala 1998:77). Tulokset olivat jossakin määrin parempia kaupunkikouluissa kuin maaseudulla. Tyttöjen tulokset olivat vain vähän parempia kuin poikien. Isojen ja

³ The International Association for the Evaluation of Educational Achievement.

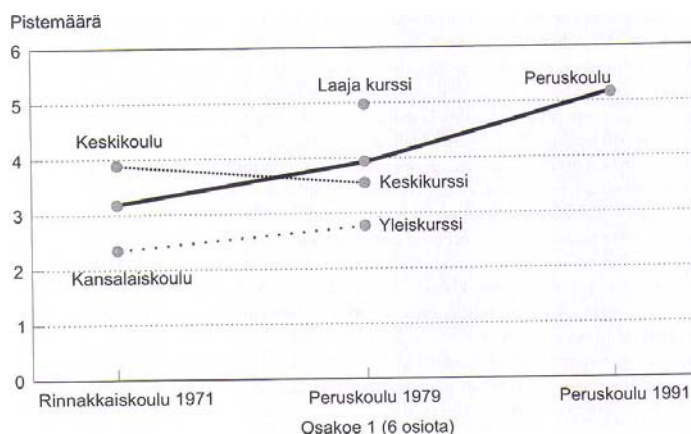
pienien koulujen välillä ei todettu eroja. Sen sijaan vanhempien koulutustaso – jossa tuohon aikaan oli rinnakkaiskoulujärjestelmän ja kansalaiskoulun oppilaiden vanhempien välillä huomattava ero – ja kodin kirjamäärä sekä kirjojen käyttö olivat selvästi yhteydessä puheen ymmärtämisen tasoon (Takala 1998:86). Tätä tulosta Takala tulkitsee osittain regressioanalyysien soveltamisen tuottamaksi, koska siinä annettiin ajallisesti kauimmin vaikuttaneiden tekijöiden (kotitaustan) ensin selittää tuloksia niin paljon kuin mahdollista (Takala 1998: 78-79).

Kuten Suomen kansallisten arviointien yhdessä mainittiin, IEA:n tutkimukseen sisältyneitä tehtäviä käytettiin myöhemmin Suomen vuoden 1979 tekstin ymmärtämisen ja puheen ymmärtämisen arvioinneissa. Samoja tekstejä käytettiin lisäksi vuonna 1983 toimeenpannuissa kokeissa kuitenkin niin, että yksi teksteistä esitettiin 7.-luokkalaisille, yksi 8.-luokkalaisille ja vain yksi 9.-luokkalaisille. Taulukko 10 osoittaa näiden kolmen tekstin ratkaisuprosentit eri vuosina samoin kuin kehityksen, mikä oli tapahtunut tekstin ymmärtämisen taidoissa vuodesta 1971 vuoteen 1983.

TAULUKKO 10 Tekstin ymmärtämisen kokeen tehtävien keskimääräiset ratkaisuprosentit teksteittäin vuosina 1971-83 (Lähde: Takala 1998:84.).

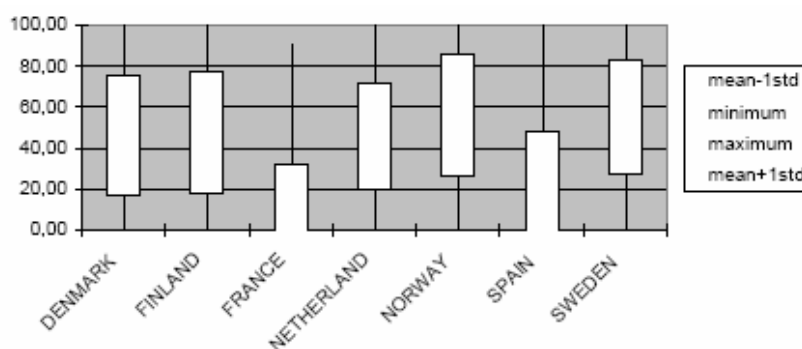
Tehtävä	IEA/71 14v.	1979 9. lk	1983
My brother Robert	57,1	66,9	75,2 (7. lk)
Pamela	43,2	50,2	65,1 (8. lk)
A day out	46,4	59,7	73,4

Joitakin IEA:n puheen ymmärtämisen kokeen tehtäviä käytettiin myös Peruskoulun tilannekartoituksessa 1 vertailukohdan saamiseksi. Kuvio 3 osoittaa IEA:n vuoden 1971 tuloksien lisäksi myös Suomen vuosien 1979 ja 1991 kokeiden tulokset ja taitojen kehittymisen suunnan.

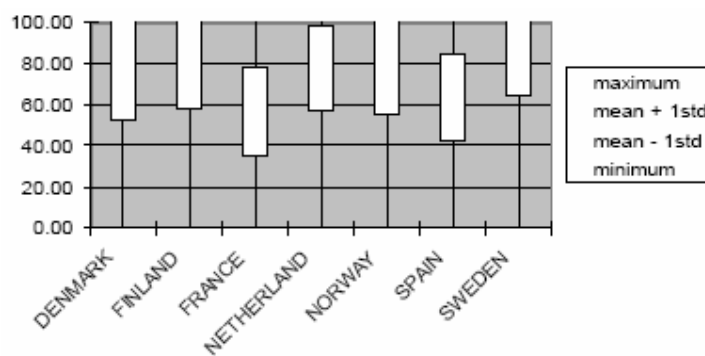


KUVIO 3 Englannin puheen ymmärtäminen eri koulumuodoissa ja oppilasryhmissä vuosina 1971, 1979 ja 1991, keskiarvot. (Lähde: Havola & Saari 1993:165.)

Vuonna 2002 Suomi osallistui kansainväliseen, kahdeksan EU-maan⁴ perusopetuksen päättövaiheen englannin kielen arviointiin, johon osallistui 99 suomenkielistä ja 10 ruotsinkielistä koulua. Oppilasotoksen koko oli 1614, joista 1464 oli suomenkielistä ja 150 ruotsinkielistä. Poikia oli 826 ja tyttöjä 788. Kielitaitoa testattiin puheen ja tekstin ymmärtämisen, sanaston ja rakenteiden sekä kirjallisen tuottamisen osakokein. Koe tai osia siitä oli vuosina 1995-1996 pidetty Espanjassa, Ranskassa ja Ruotsissa sekä myöhemmin myös Hollannissa. Arviointiin sisältyivät oppilas- ja opettajakyselyt. Suomen tulokset on julkaistu Opetushallituksen Arviointi-sarjassa (Tuokko 2003:3). Koko arvioinnin tulokset on julkaistu ruotsiksi Skolverketin raportissa (2004) ja englanniksi Ranskan opetusministeriön raportissa (Bonnét 2004). Kuviot 4-7 osoittavat eri osa-alueilla eri maissa saavutetut tulokset.



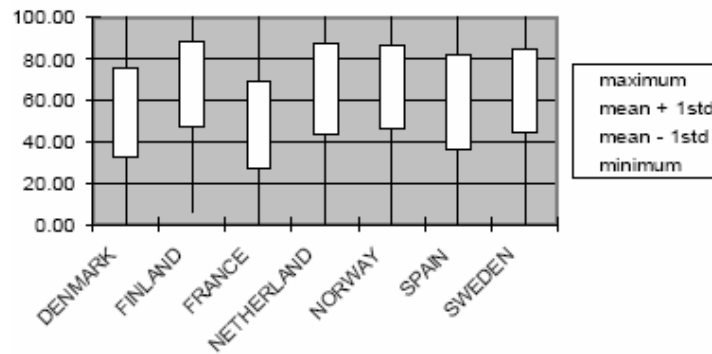
KUVIO 4 Puheen ymmärtämisen kokeen tulokset seitsemässä⁵ Euroopan unionin maassa. (Lähde: Bonnét 2004:69).



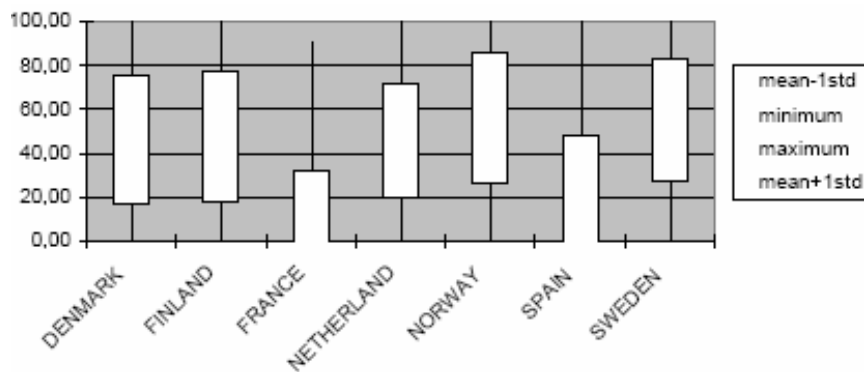
KUVIO 5 Tekstin ymmärtämisen kokeen tulokset seitsemässä Euroopan unionin maassa. (Lähde: Bonnét 2004:71).

⁴ Espanja, Hollanti, Norja, Ranska, Ruotsi, Saksa (osittain), Tanska ja Suomi.

⁵ Saksan tuloksia ei saatu analysoitaviksi raportin käsikirjoituksen jättämiseen mennessä.



KUVIO 6 Kieliopillisen kompetenssin kokeen tulokset seitsemässä Euroopan unionin maassa. (Lähde: Bonnét 2004:70).



KUVIO 7 Kirjoittamisen kokeen tulokset seitsemässä Euroopan unionin maassa. Lähde: (Bonnét 2004:72).

IEA yritti 1990-luvun puolivälissä käynnistää toista vertailevaa kielenoppimistutkimusta, mutta se keskeytyi taloudellisen tuen puuttumiseen (Leimu 2004c:63). Opetushallituksessa käytiin vuosina 1999-2003 myös keskustelua pohjoismaiden yhteisestä englannin kielen arvioinnista, joka kuitenkin kaatui ilmeisesti Tanskan arviointitoiminnan uudelleenjärjestelyihin.

Tämän vuosituhannen kansainvälisiin arviointihankkeisiin PISAan⁶ tai PIRLSiin⁷ ei ole valitettavasti sisällytetty vieraiden kielten arviointeja. Leimu korostaa, että kansainvälisiin vertailututkimuksiin olisi kuitenkin tärkeitä osallistua, koska niillä on merkitystä suomalaisen koulun ja koko yhteiskunnan ominaislaadun kuvan muotoilijoina (2004b:41-42). Suomalaista kieltenopetusta ja sen tuloksia pidetään yleensä korkeatasoisina. Niille tulisi myös antaa tilaisuus kansainväliseen näkyvyyteen.

3.3 Kansalliset arvioinnit muissa Euroopan maissa

Euroopan maista ainakaan Belgiassa, Italiassa, Itävallassa, Kreikassa, Latviassa, Liettuaissa ja Saksan Liittotasavallan Schleswig-Holsteinissa ei järjestetä perus-

⁶ Programme for International Student Assessment

⁷ Progress in International Reading Literacy Study

opetuksen kansallisia arviointeja englannin kielessä. Tsekin tasavallassa on vuodesta 2003 ollut käynnissä projekti päättövaiheen oppilaiden englannin taitojen testaamiseksi. Seitsemän muun Euroopan maan kansallisia arviointikäytänteitä, joiden tiedot perustuvat joulukuun 2005 tilanteeseen, esitellään liitteessä 1.⁸

Kahdessa Euroopan maassa, Ruotsissa ja Hollannissa, on pitkään tehty ansiokasta työstä kansallisten arviointien kehittämiseksi. Näiden maiden käytänteet on tämän vuoksi valittu tässä esimerkkeinä esiteltäviksi.

3.3.1 Ruotsin kansallisen englannin ainekokeen kehittäminen

Ruotsin koululaitoksessa siirryttiin 1990-luvulla opetussuunnitelmauudistuksen yhteydessä normiviitteisestä arviointijärjestelmästä kriteeriviitteiseen sekä poistettiin tasokurssijärjestelmä. Uudistuksen yhteydessä vahvistettiin myös peruskoulun kansallisten kokeiden asemaa päättämällä järjestää niitä 5., 7. ja 9. luokalla (Erickson 1999:202). Englannin kieli oli tullut osaksi ns. standardikoejärjestelmää 1950-luvun lopulla siten, että ensin kokeita järjestettiin 6. luokalla, mutta sittemmin myös 8. luokalla. Vuodesta 1984 englannin kokeiden kehittäminen ja tutkimus siirtyivät Kouluylihallitukselta (*Skolöverstyrelsen*, myöhemmin *Skolverket*) Göteborgin yliopiston Kielipedagogiikan osaston tehtäviksi, tarkemmin tutkija Gudrun Ericksonille (emt. s. 202-203).

Standardikokeet olivat pakollisia vuodesta 1987 lähtien. Niitä laadittiin paitsi molemmille tasokursseille, myös tasokurssiton versio. Koulut järjestivät kokeet keväällä sovitun kolmen viikon aikana itselleen parhaiten sopivana aikana. Satunnainen otos toimitettiin arvosteltavaksi, ja lukukauden lopulla koulut saivat tiedoksi keskimmäisen arvosanan pisterajat sekä kokeisiin liittyviä analyyseja (emt. s. 203). Standardikokeiden tarkoitus oli osoittaa kouluille arvosanan 3 (asteikolla 1-5) taso sekä ilmentää vuoden 1980 opetussuunnitelman kommunikatiivista ja funktionaalista kielitaitokäsitystä. Vuodesta 1994 lähtien standardikokeeseen sisällytettiin pakollinen kirjoittamisen osakoe. Suullinen koe sisällytettiin standardikokeeseen vuonna 1991, ja viimeiset standardikokeet järjestettiin vuonna 1996 (emt. s. 204).

Jo ennen standardikokeiden järjestämisen lopettamista oli tehty runsaasti työtä uuden kansallisen kokeen kehittämiseksi. Lähtökohtana oli opetussuunnitelman tavoitteiden analysointi, koska kokeiden tuli mahdollisimman suurella määrällä heijastaa niitä. Tämä oli välttämätöntä, koska oli päätetty siirtyä normiviitteisestä arvioinnista kriteeriviitteiseen, mikä edellytti sisältöjen tuntemista, jotta koetehtävät olisivat edustava otos niistä. Oli siis siirrytty pois siitä normiviitteisen arvioinnin käsityksestä, että koe määrittää ryhmän tason, näkemykseen, jonka mukaan jokaisen tiedot ja taidot arvioidaan suhteessa tavoitteisiin ja kriteereihin.

⁸ Sähköpostitiedonannot EALTA:n jäseniltä maaliskuussa 2005, tiedot tarkistettu joulukuussa 2005..

Periaatteeksi tuli, että

- kansallisten kokeiden tuli mahdollisimman laajasti heijastaa opetus- ja kurssisuunnitelmia ja että tämän tuli näkyä koespesifikaatioissa⁹,
- oppilaiden taitoja kielitaidon eri osa-alueilla ei testattaisi erikseen vaan integroidusti,
- kokeiden tuli edelleen olla ns. *proficiency* -kokeita, millä tarkoitetaan sitä, että koe ei perustu tiettyyn oppimäärään tai kurssiin,
- kokeet eivät perustu tiettyihin oppimateriaaleihin eikä työtapoihin,
- oppilaiden tulee saada mahdollisuus osoittaa, *mitä he osaavat* eikä *mitä he eivät osaa*,
- testattavien sisältöjen, tehtävien vaikeustason, tehtävä- ja vastaustyyppien tulee olla vaihtelevia,
- pyritään mahdollisimman suureen autenttisuuteen,
- tuottamistehtävissä painotetaan sisältöä ja ymmärrettävyyttä ja tehdään ero *häiritsevien ja vähäisten virheiden* välillä sekä
- tulokset esitetään profiilina, jonka muoto osoittaa oppilaan vahvuudet ja puutteet (emt. s. 212-213).

Vuosina 1994-1995 kehitettiin intensiivisesti 7. luokan diagnostista koetta, jonka piti toisaalta auttaa opettajia diagnosoimaan oppimisvaikeuksia, mutta samalla tuottaa myös prognostista tietoa oppilaiden tiedoista ja taidoista suhteessa 9. luokan oppimistavoitteisiin. Tämän kokeen oli myös tarkoitus olla perustana ainekokeille, joita tulotisiin järjestämään 5. ja 9. luokalla (Äp 5 ja Äp 9). Ensimmäinen 5. luokan ainekoe järjestettiin vuonna 1996. Sen tarkoitus oli auttaa opettajia päättämään, olivatko oppilaat saavuttaneet kyseisen luokka-asteen tavoitteet (emt. s. 214). Ensimmäinen 9. luokan ainekoe pidettiin keväällä 1998 (emt. s. 216). Ainekoe koostui kolmesta osasta:

- A - suullinen vuorovaikutus
- B - reseptiiviset taidot
- C - kirjallinen tuottaminen.

Kokeeseen liittyi seuraavaa opettajille ja oppilaille tarkoitettua tukiaineistoa:

- selvitys kokeen taustasta, toteuttamisesta ja arvostelusta,
- oppilaan ohjeet,
- esimerkkejä oppilaille ehkä vieraista tehtävätyypeistä (suositeltiin käsiteltäväksi hyvissä ajoin ennen koetta).

Skolverketin päätöksen mukaan englannin ainekoe järjestettiin kevätlukukaudella niin, että B- ja C -koe suoritettiin tiettyinä päivinä koko maassa, kun taas A-kokeen saattoi järjestää milloin tahansa 3-4 kuukauden sisällä. Eri osakokeiden tulokset tuli lopulta yhdistää, ja kouluun tuli toimittaa ohjeet siitä, miten painottaminen suoritetaan. Oppilaan tulokset annettiin profiiliarviointina, ja kaavakkeessa oli tilaa oppilaan ja opettajan kommenteille kokeesta ja sen tuloksesta (emt. s. 219-221). Kokeet pidetään salaisina vuoden ajan (emt. s. 216).

⁹ Koe- I. testispesifikaatiot ovat kokeen ja tehtävien sisältökuvauksia.

3.3.2 Kansallinen englannin päättöarviointi Hollannissa

Hollannin kansallinen englannin kielen arviointijärjestelmä on otettu esimerkiksi pitkään jatkuneesta ja ilmeisesti pääpiirteissään hyväksi havaitusta käytännöstä, jossa päättöarvosana koostuu sekä ulkopuolisen tahon että oman opettajan arvioista. Luvun 3.3.2. tiedot on antanut CITOssa arvioinnista vastaava Noud van Zuijlen sähköpostiviestinä 4.3.2005.

Suomen perusopetuksen päättövaihetta vastaa Hollannissa 10. luokka, jolloin oppilaat ovat keskimäärin 16-vuotiaita. Kansallisia arviointeja on järjestetty vuodesta 1970, ja ne järjestää hallituksen nimittämä toisen asteen yleissivistävän koulutuksen kansallinen tutkintolautakunta CEVO. Siihen kuuluu toisen ja kolmannen asteen opettajia, opettajien ammattijärjestön edustajia sekä kirkon edustajia. CEVO vastaa tutkintojen sisällöistä ja taitotasojen asettamisesta. Sen lisäksi myös yksi ministeriön osasto osallistuu tutkintojen järjestämiseen.

Kokeessa on kaksi tasoa, joista ylemmällä testataan tekstin ymmärtämistä ja kirjoittamista, alemmalla tekstin ja puheen ymmärtämistä ja kirjoittamista (hyvin vaatimattomalla tasolla) digitaalisesti. Kokeen kesto on 120 minuuttia, digitaalisen kuitenkin 60 minuuttia. Kaikki kokeet järjestetään koulumuodotain samana päivänä ja kaikki suorittavat saman kokeen. Digitaalinen koe on poikkeus: eri oppilasryhmät suorittavat sen eri päivinä ja siinä on useita variantteja, joissa on ankkuriosioita. Kokeisiin osallistuvat kaikki oppilaat, se ei siis ole otantaperustainen. Kokeen voi uusia kerran, erityistapauksissa kaksi kertaa.

Kansallisessa kokeessa on kaksi osaa. Toinen on keskitetysti järjestetty osa, johon ottavat osaa kaikki, jotka ovat osallistuneet tietyn tyyppiseen opetukseen. Oppilaiden omat opettajat arvostelevat kokeet, mutta toisen, satunnaisotoksella valitun koulun opettajan tulee myös arvostella koe, ja näiden kahden opettajan on päästävä yksimielisyyteen arvostelusta. CEVO varustaa kokeet perusteellisilla arvosteluohjeilla, joita opettajien tulee noudattaa. Tarkastajat tarkastavat pistokokein arvosteluja.

Toisessa, koulukohtaisessa osassa, testataan niitä taitoja, joita keskitetyssä kokeessa ei testata, tavallisesti puheen ymmärtämistä, puhumista ja kirjoittamista, joskus myös kirjallisuutta. Koulut saavat laatia omat kokeensa, mutta ne voivat ostaa myös valmiita kokeita. Opettajat arvostelevat itse kokeensa, joskus tosin myös kollegan kanssa. Oppilaan kokonaistuloksen laskemiseksi koulujen on laskettava keskitetyn ja koulukohtaisen kokeen keskiarvo.

CITO¹⁰ laatii vuosittain kokeita kielitaidon mittaamiseksi. Suosituin myyntiartikkeli on mallivideon sisältävä puheen ymmärtämisen koepaketti. Noin 90 % kouluista hankkii sen. CITO myös analysoi kokeet. Tulokset raportoi CEVO, joka toimittaa kouluille joka vuosi myös kuhunkin kokeeseen sovellettavat standardit. Raportin tuottaminen kestää noin kaksi vuotta. Hollannissa esiintyy pyrkimyksiä saada tutkinnon hallinto joustavammaksi ja ensimmäisiä kokeiluja on harkittu vuodeksi 2006, mutta ne ehkä siirtyvät vuoteen 2007. Kokeita pyritään linkittämään CEFiin, mikä saattaa tuoda muutoksia kokeiden rakenteeseen.

¹⁰ Dutch National Institute for Test Development, www.Cito.com

3.4 Yhteenveto

Suomessa kansallista vieraiden kielten arviointia on harjoitettu suhteellisen tiiviisti peruskoulun siirtymävaiheesta 1960-luvulta alkaen. Tavoitteena on ollut joko opetus suunnitelman kehittäminen tai oppimistulosten selvittäminen; itse asiassa molemmat ovat saman asian eri puolia. Tämän tutkimuksen kohteena ollut arviointi samoin kuin käsillä oleva tutkimus palvelevat samaa tarkoitusta vaikkakin edeltäjiään yksityiskohtaisemmin. Arvioinnin kohteena olivat 1990-luvulle saakka yleensä tekstin ja puheen ymmärtäminen sekä rakenteet ja sanasto, joskus kaikki, joskus vain jokin tai jotkut osa-alueet. Peruskoulun siirtymävaiheen uudistushenkisyyttä osoittaa kuitenkin jo vuonna 1969 kokeiluperuskoulujen 4. ja 5. luokkien oppilaille järjestetty puhumisen koe. Peruskoulun tilannekartoitus 1:een sisältynyt englannin arviointi taas oli merkittävä siksi, että sen suunnittelussa käytettiin monimatriisiotantaa ja tulosten raportoinnissa uutta testiteoriaa, joita on sovellettu tässäkin tutkimuksessa. Vuoden 1993 peruskoulun oppimistulosten arviointihankkeen yhteydessä järjestettyyn englannin arviointiin sisältyi myös kirjoittamisen koe sekä puhekokeen simulaatio. Vuonna 1999 järjestettyyn arviointiin, johon tämä tutkimus kohdistuu, sisältyvät kaikkien kielen osa-alueiden sekä rakenteiden ja sanaston kokeet.

Suomen englannin kielen oppimistulosten kehitystä tarkasteltaessa voidaan todeta, että ne näyttävät parantuneen koko ajan vuoden 1993 arviointiin saakka. Se, että kehitys tuolloin näyttää jossakin määrin pysähtyneen, johtuneen siitä, että lähtötaso oli rinnakkaiskoulujärjestelmästä johtuen alkujaan hyvin matala, kun taas 1990-luvulle päästyä on ehkä saavutettu jonkinlainen "tasankovaihe". Tyttöjen paremmuus näyttää olleen pysyvää, mutta kasvamassa.

Suomen perusopetuksen kieltenopetuksen tuloksellisuudesta kansainvälisessä vertailussa on hyvin vähän tietoa johtuen kansainvälisen arviointitoiminnan hämmästyttävästä – niin voitaneen perustellusti sanoa – vähäisyydestä. Vuonna 1971 osallistuttiin ensimmäiseen ja toistaiseksi myös ainoaan merkittävään kansainväliseen englannin kielen arviointitutkimukseen (IEA:n tutkimus) melko heikoin tuloksin perusasteen osalta. Vuosituhannen vaihteessa järjestettyihin laajoihin IEA- ja PISA tutkimuksiin ei sisältynyt vieraiden kielten arviointia. Vuonna 2002 Suomi osallistui kuitenkin kahdeksan EU-maan englannin arviointiin, jonka tulokset olivat varsin hyvät.

Vieraiden kielten arviointitoiminta on Suomessa ollut jossakin määrin haipuilevaa. Eräänlaisena vastakohtana sille ovat Ruotsin ja Hollannin pitkään jatkuneet ammattimaisesti toteutetut kansalliset arvioinnit ja kokeet.

Euroopassa vieraiden kielten opetusta ja arviointia rakennetaan tai pyritään kasvavassa määrin rakentamaan Eurooppalaisen viitekehyksen periaatteiden mukaisiksi. Suomessa ylioppilastutkinnon A-englannin ja B-ruotsin kokeet on linkitetty alustavasti CEF -asteikkoihin (Takala 2005a ja 2005b), ja Euroopan neuvostossa toivotaan palautetta siellä laadittuihin linkitysohjeisiin. Tämä tutkimus on yksi vastaus näihin odotuksiin.

Muissa Euroopan maissa tapahtuvan kansallisten englannin kielten arviointien kehittämisen esimerkkien jälkeen siirrytään käsittelemään kielitaidon arvioinnin lähtökohtana olevia erilaisia kielitaitokäsityksiä.

4 KIELIT AidON ARVIOINNIN LÄHTÖKOHTIA

Kielikokeen laadinnan lähtökohtana on kielitaidon määrittely, mistä tutkijoilla on eri aikoina ollut erilaisia käsityksiä. Niiden ääripäiksi voidaan katsoa kielellisten yksikköjen testaaminen ja mahdollisimman suuressa määrin autenttisia kielenkäyttötilanteita korostava kommunikatiivinen kielitaitokäsitys. Kielitaidon arviointi voidaan myös kohdistaa prosessiin tai tuotokseen. Näitä käsitellään kielitaidon osa-alueittain.

Erityisesti ymmärtämistehtävien vaikeustason arvioiminen on vaativaa. Pääsyyinä tähän katsotaan olevan sen, että "oikeaan" vastaukseen voidaan päätyä monin eri prosessointitavoin. Korkeampien korrelaatioiden saavuttaminen arvioijien ja empiiristen tulosten välillä edellyttää syvempää ymmärtämisprosesseihin perehtymistä. Koska tällä alueella on tehty arvokasta tutkimusta, johon olisi syytä kiinnittää enemmänkin huomiota, sitä esitellään tässä yhteydessä huolimatta siitä, että käsillä olevassa tutkimuksessa kielitaitoa on arvioitu tuotoksena.

Arvioinnin lähtökohtia esitetään taustaksi tämän tutkimuksen kohteena olevien kokeiden laadinnassa sovelletulle kielitaitokäsitykselle ja sen operationaalistamiselle kokeiksi.

4.1 Testauksen perustana käytettyjä kielitaitokäsityksiä

Kielitaidon käsitteen määrittäminen rajaa sen, mitä koe mittaa. Spolsky (1993) jakaa testauksen historian kolmeen kauteen, traditionaaliseen, moderniin ja postmoderniin sen perusteella, millaisiin testauskäytänteisiin kulloinkin vallinnut kielitaitokäsitys on johtanut. Nykyinen kommunikatiivisuutta, autenttisuutta ja suoritusta korostava lähestymistapa edustaa postmodernia kautta. Koska modernille kaudelle ominainen kielen yksiköiden, komponenttien, testaamisperinne on kuitenkin edelleen laajasti käytössä, ja sitä on tämänkin tutkimuksen kohteena olevassa arvioinnissa sovellettu, esitellään seuraavassa edellä mainittujen kahden viimeisen kauden testauksen kehittämiseen eniten vaikuttaneiden tutkijoiden kielitaitokäsityksiä.

Moderniin kauteen kuuluva strukturalistinen kielitiede painottaa kielellisten yksiköiden eli kieliopillisten rakenteiden, sanaston ja ääntämisen testaamista (*discrete point testing*). Tätä testaustyyppiä tuki 1960-luvulla voimakkaasti edennyt tieteenala, kognitiivisten taitojen mittaamiseen keskittynyt psykometriikka, joka korosti erityisesti reliabeliuden ja samanapysyvyyden (*consistency*) (Heinonen 1961:38) merkitystä tutkimustuloksissa. Robert Lado, strukturalistisen testauskoulukunnan ja näin ollen analyttisten, kielen yksikköjä testaavien kokeiden puolestapuhuja (Davies 1978:150), on ensimmäisenä määritellyt kielitaidon testattavalla tavalla ansiokkaassa teoksessaan "Language Testing" (1961). Kielitaitoon ja samalla testauksen kohteeksi kuuluivat hänen mukaansa kielen elementit (*elements*), kielitaidon osa-alueet (*skills*) ja suoritustaso (*degree of achievement*). Kielitaidon elementit integroituvat puhumisen, kuuntelemisen, lukemisen ja kirjoittamisen osa-alueiksi, jotka ovat erillisten elementtien lailla testattavissa olevia muuttujia.

John B. Carroll on toinen tunnettu teoreetikko, jonka kielitaitomalli, ns. Carrollin taulukko (taulukko 11) heijastaa hänen kiinnostustaan taitojen kognitiivista rakennetta ja tässä tapauksessa kielitaidon rakennetta kohtaan. Hän katsoo kuitenkin, että taulukon yksittäisten ruutujen sisältöjen (*discrete points*) testaamisen lisäksi tulisi käyttää myös integroitua testaamista, jossa kaikki neljä kielitaidon osa-aluetta ovat mukana (Takala 1980:191).

TAULUKKO 11 Carrollin kielitaidon testaamisen luokitusjärjestelmä. (Siteerattu: Bachman & Palmer 1984.)

Taito	Kielen aspekti			
	Fonologia tai ortografia	Morfologia	Syntaksi	Sanasto
Kuullun ymmärtäminen				
Suullinen tuottaminen				
Lukeminen				
Kirjoittaminen				

Carroll oli tunnustettu psykolingvisti ja myös huipputason psykometriikko. Hänen tunnetuimpia kontribuutiotaan kielitaidon alalla on empiiriseen tutkimukseen rakentuva kielellisen lahjakkuuden testi (esim. Carroll 1958) ja Foreign Service Institutun suullisen kielitaidon taitotasoasteikon perusteiden luominen. Teoksessaan "Human Cognitive Abilities", joka on katsaus suureen joukkoon faktorianalyysiin perustuvia tutkimuksia, Carroll (1993) esitti käsitteellisen kuvauksen kielellisen taidon alueen faktoreista, jolla hän osoittaa, että ihmiset ovat yleensä erilaisia – ainakin, jos he ovat eri-ikäisiä, mutta myös saman ikäryhmän puitteissa – kielellisen kehityksen yleistasoltaan. Mutta on olemassa myös kykyjen erikoistumista: jotkut erikoistuvat puhumisen taitoihin, jotkut lukemisen ja kirjoittamisen jne. Faktorianalyysien tuottamien faktoreiden eli taitojen kuvaaminen eri tasoilla ja eri suuntiin haarautuvina "oksina" kielenkäytön aluetta kuvaavassa pylväässä pyrkii osoittamaan tätä taitojen erikoistumista (Carroll 1993:147). Carrollin (1993) mukaan kielellisen kompetenssin

(*linguistic competence*) todellista ilmentymää käytöksessä voidaan kutsua kielelliseksi suoritukseksi (*linguistic performance*) (vrt. Chomsky), mutta siihen vaikuttavat lisäksi ei-kielelliset muuttujat. Kielellisen suorituksen ei-lingvistiset muuttujat saattavat siis hyvinkin estää kielellisen kompetenssin ilmenemisen esimerkiksi mitattavassa muodossa (McNamara 1996:52-53).

Noam Chomsky (1972) pyrki selvittämään erityisesti sitä, mikä kielitieteen varsinainen tutkimuskohde ja sen keskeinen tutkimuskysymys on, mutta hänen arvioinnille tärkein antinsa on kielitaidon teoria, jossa taito jaetaan kompetenssiin ja performanssiin (Chomsky 1972:3; ks. myös Takala & Havola 1983:319; Huhta & Takala 1999:184). Hänen kompetenssikäsitteensä sisältää ajatuksen ihanteellisesta kielenkäyttäjistä, joka omaa synnyntänsä, sisäisen, ei-kielispesifin kieliopin, jonka avulla hän pystyy tuottamaan rajattomasti syntaktisesti moitteettomia rakenteita. Performanssiin sisältyvät kieliopillinen ja pragmaattinen kompetenssi, joista jälkimmäinen toimii jonkinlaisena välittäjänä (kieli)tiedon ja itse puhe-suorituksen välillä (Chomsky 1980:224). Analysoidessaan performanssi-käsitteen käyttöä McNamara (1996:55) huomauttaa kuitenkin jo Hymesin (1972:289), joka kehitti Chomskyn kompetenssi-performanssi-käsitteitä edelleen, kiinnittäneen huomiota Chomskyn tämän käsitteen epäjohdonmukaiseen käyttöön.

Vähitellen tapahtui siirtymistä mikrotietojen eli kielen elementtien testaamisesta makrotietojen eli kielen osa-alueiden testaamisen suuntaan täydentämään ja monipuolistamaan kokeita (McNamara 2000:14; ks. myös Huhta & Takala 1999:183). 1970-luvun alkupuolella tutkijat, erityisesti John Oller, alkoivat puhua myöhemmin integratiivisina tunnettujen testien puolesta. Kielen käyttöön sisältyy hänen mukaansa kaksi tekijää. Toinen on kielen reaaliaikainen prosessointi esimerkiksi aidoissa puhe- ja kuuntelutilanteissa, ja toinen pragmaattinen komponentti (*pragmatic mapping component*), jolla tarkoitetaan tapaa käyttää tietoa kielen muodoista ja systemaattisista piirteistä osoittamaan ymmärtämistä ja helpottamaan tuottamista kielenkäyttöyhteyksissä. Ollerin (1979) mukaan kaikki kielellinen toiminta perustuukin sisäistettyyn ennakoitukielioppiin (*expectancy grammar*), jonka mukaan puhuttu ja kirjoitettu kieli esiintyvät sellaisina peräkkäisinä elementteinä, että tekstiyhteydestä on mahdollista päätellä, mikä elementti on seuraavana vuorossa. Elementeillä Oller (1979) tarkoittaa äänneitä, tavuja, sanoja, sanontoja, lauseita ja kappaleita, jopa pitempiäkin tekstin yksiköitä.

Oller (1979) kiteytti ajatuksensa väittämään, joka on tullut tunnetuksi nimellä "Unitary Competence Hypothesis", joka sisältää kielitaidon testauksen yhteydessä sen ajatuksen, että suoriutuminen pragmaattisissa kokeissa, joiksi hän nimittää suosimiaan kokeita, johtuu samasta kielenkäyttäjän kyvystä kuin mitä tarvitaan suoriutumiseen integratiivisissakin kokeissa eli kyvystä yhdistää kieliopillinen, sanastollinen, leksikaalinen, kontekstuaalinen ja pragmaattinen tieto koesuorituksessa. Oller (1979) katsoi erityisesti cloze -kokeen korvaavan hyvin integratiiviset kokeet. 1970-luvulla alettiinkin käyttää cloze -tehtäviä, joiden väitettiin testaavan ei ainoastaan yleistä kielitaitoa vaan myös tekstin ymmärtämistä (vrt. Mauranen 1986:82-84). Cloze -tehtävien avulla on pyritty testaamaan kon-

tekstiin liittyvien oikeiden kielellisten muotojen käyttöä, sanan semanttisen ulottuvuuden tuntemista ja oikean morfologisen ja syntaktisen vaihtoehdon tuntemista (Finocciaro & Sacks 1983:20). Ne ovat hyvin käytännöllisiä monissa tilanteissa, koska niitä on nopea laatia ja korjata, mutta niiden validius tekstin ymmärtämisen kokeina on kuitenkin kiistanalainen (Alderson 2000:205). On osoittautunut, että tämäkin koetyyppi testaa lähinnä kieliopillista kompetenssia ja sanastoa (McNamara 1996:16). On syytä huomata, että edellä mainittu arvio koskee joka n:n sanan aukottamiseen kohdistuvaa alkuperäistä clozea eikä harkinnanvaraista aukotettua varianttia (*gap-filling*). Viimemainittua on käytetty tämänkin tutkimuksen kohteena olevassa kieliopillisen kompetenssin kokeessa.

Ollerin tutkimukset näyttivät osoittavan, että kielitaito olisi yksi kokonaisuus, Chomskyn taas, että se koostuisi useammasta osasta. (Oller kuitenkin joutui myöntämään tehneensä aineiston käsittelyssä metodologisia virheitä, ja perui väitteensä unitaarisesta kielikompetenssista. Sauli Takala, suullinen tiedonanto 23.11.2006.) Erilaiset vastakkaiset päätelmät johtivat yhä kiihtyvään kielitaidon tutkimukseen 1970-luvulla. Vaikka Oller (1979) puhuu kielen käyttämisestä eikä vain sen osaamisesta, hänen määritelmässään ei kuitenkaan ole mitään mainintaa kommunikaatiosta. Pääero edelliseen testustraditioon on, että integratiivinen testaus painottaa kielen kokonaisvaltaista arviointia, kun taas kielen yksiköiden arviointi korostaa kielen elementtien tuntemusta (Buck 2001:66).

Kun kielen rakenteen ja ymmärtämis- ja tuottamisprosessien tutkimuksen rinnalle nousi suuntaus, jonka mukaan kieli oli osa kommunikaatiota, kommunikoinnin väline, voidaan katsoa postmodernin kauden alkaneen. Tällöin alettiin korostaa kielen käyttötarkoituksen eli funktion sekä kielen autenttisuuden ja sosiaalisen kontekstin merkitystä (Huhta 1993:82-83; Huhta & Takala 1999:184; McNamara 2000:16). Funktionaalisuutta painottavaa kommunikatiivista kielitaitoa luonnehtivat mm. seuraavat ominaisuudet:

- vuorovaikutus,
- ennakoimattomuus
- tilannekonteksti = fyysinen ympäristö (osallistujien rooli ja asema, asenne, muodollisuusaste)
- kielellinen konteksti (tekstin koheesio)
- tarkoitus
- rajoitukset (muisti, melu, väsymys)
- autenttisuus (esim. yksinkertaistettujen tekstien lukeminen)
- kielestä riippumattomat tekijät
- kontekstista toiseen vaihtuvat merkitykset
- jatkuva kommunikoinnin onnistumisen arviointi ja
- kielitaidon osa-alueiden integroitunut käyttö. (Morrow 1979:16; Huhta & Takala 1999:184; ks. myös Huhta 1993:83-84).

Dell Hymesin (1972) katsotaan luoneen kommunikatiivisen kompetenssin käsitteen ja vaikuttaneen sen myötä merkittävästi myös kielen testaukseen. Hän laajensi erityisesti kielen käytön aluetta kiinnittäen huomiota kielenkäyttökontekstiin käyttäen esimerkkinä syntyperäisen kielenkäyttäjän kykyä sovittaa kielenkäyttönsä tilanteen vaatimusten mukaiseksi. Kun Chomsky (1972) liitti kompetenssi-käsitteen pelkästään kieliopin sääntöjen tuntemiseen, Hymes (1972) näki kompetenssin laajempänä käsitteenä, kommunikatiivisena kompetenssina, johon

ei sisälly yksinomaan kieliopillinen kompetenssi (implisiittinen ja eksplisiittinen kieliopin sääntöjen tuntemus) vaan myös konteksti- tai sosiolingvistinen kompetenssi (kielen käytön sääntöjen tunteminen) (Canale & Swain 1980:15). Tämä uusi lähestymistapa korosti itsensä ymmärretyksi tekemistä kiinnittämällä vähemmän huomiota kielelliseen tarkkuuteen. Fokus siirtyi muodosta funktioon. On kuitenkin syytä korostaa, että Hymesin (1972) kommunikatiivisen kompetenssin käsitteeseen ei sisällynyt pelkästään soveltuvuus ja (hyväksytyt) käyttötavat, vaan myös kielitaito chomskyläisessä merkityksessä, jossa kieliopilla yhä on hallitseva osa (Köhlmeyer, 2003:1-2). Kommunikatiivisella performanssilla Hymes (1972) tarkoitti kielen varsinaista käyttöä, kun taas puhujan edellytyksistä käyttää kieltä hän käytti termiä "ability for use". Kommunikatiivisen kompetenssin teoria edusti syvällistä siirtymää psykologisesta näkemyksestä kieleen sisäisenä ilmiönä (hallidaylaiseen) sosiaaliseen näkemykseen, jossa korostuu kielen ulkoinen, sosiaalinen funktio (McNamara 2000:17). Samoilla linjoilla ovat olleet Canale ja Swain (1980), joiden kielitaitokuvaukseen sisältyy kolme tekijää: kielellinen/kieliopillinen kompetenssi, joka sisältyy myös Ladon kielitaidon määrittelmään, sosiolingvistinen kompetenssi ja strateginen kompetenssi, josta esimerkiksi Hymes (1972) käytti syntyperäisen kielenpuhujan menettelytapoja.

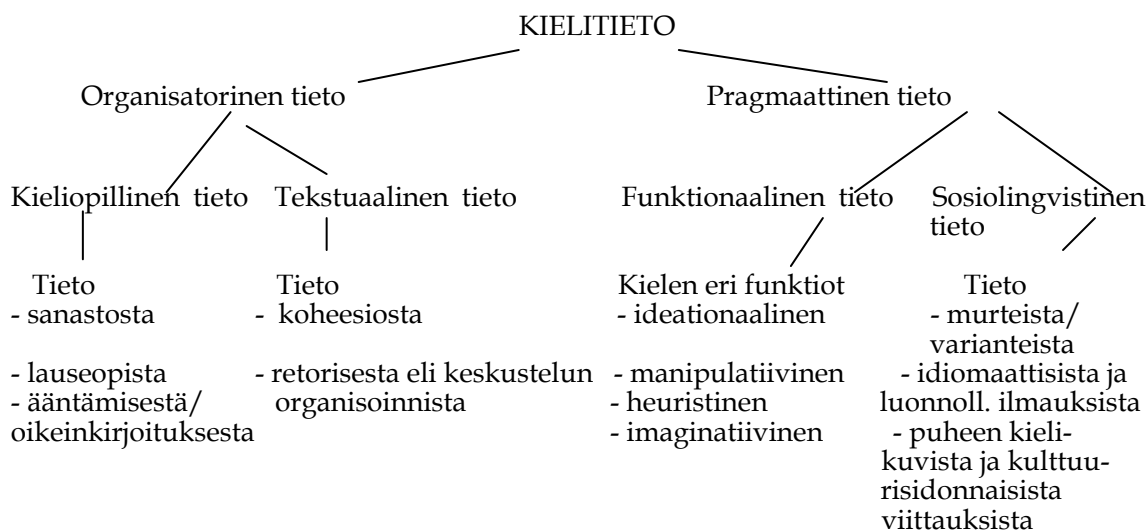
Canalen ja Swainin (1980) kieliopillisen kompetenssin käsitteeseen kuuluvat sanaston, morfologian, lauseopin, semantiikan ja fonologian tuntemus (vrt. Carroll 1961. Siteerattu Bachman & Palmer 1984:35). Sosiolingvistinen kompetenssi koostuu kahdesta sääntöryhmästä: sosiokulttuuriset ja diskurssisäännöt. Edelliset spesifioivat tavat, joilla ilmauksia tuotetaan ja ymmärretään oikein suhteessa tilanteeseen, osallistujien rooleihin ja vuorovaikutuksen normeihin. Diskurssisääntöinä Canale ja Swain (1980) näkevät puhuntojen koheesion eli soveltuvien kieliopillisten linkkien käytön ja koherenssin eli kommunikatiivisten funktioiden sopivat yhdistelmät. Strateginen kompetenssi koostuu verbaalisista ja nonverbaalisista kommunikaatiostrategioista. Niitä on kahta päätyyppiä. Toiset liittyvät ensisijaisesti kieliopilliseen kompetenssiin ja korvaavat riittämättömää kompetenssia (esim. parafrasit). Toiset taas liittyvät enemmän sosiolingvistiseen kompetenssiin (Canale & Swain 1980:28-30). Kielen käyttöpotentiaalin (Hymesin "ability for use") Canale ja Swain katsovat kuuluvan kommunikatiiviseen suoritukseen, performanssiin (Canale & Swain 1980:7).

Kommunikatiivisen kompetenssin ja performanssin välillä tehty ero johti Canalen ja Swainin (1980:34) mukaan siihen päätelmään, että kommunikatiivisessa testauksessa tulee ottaa huomioon paitsi sen selvittäminen, mitä testattava tietää jostakin toisesta kielestä ja kuinka sitä käytetään, myös se, missä määrin testattava pystyy demonstroimaan tätä tietoa kommunikatiivisessa tilanteessa. He toteavat myös, että kielen yksikköjä mittaavat testit voivat olla hyödyllisiä kommunikatiivisessakin lähestymistavassa, koska näin testattava ehkä tulee tietoiseksi kommunikatiivisen kompetenssin elementeistä ja komponenteista. Ne selkeyttävät myös näiden hallinnan arviointia. Tämän tutkimuksen kohteena oleva arviointi lähtee pääpiirteissään myös tästä käsityksestä.

Tähän asti parhaana (Alderson 2000; Buck 2001) pidetyn kielitaidon testauksen käsittelevän teoksen "Language Testing in Practice" (Bachman & Palmer

1996) kielitaidon määritelmän perustana on Bachmanin (1990) esittämä kielitaidon malli. Bachmanin ja Palmerin (1996) malli lähtee siitä ajatuksesta, että jos kielitaidosta halutaan tehdä johtopäätöksiä, se on määriteltävä riittävän tarkasti, jotta se voidaan erottaa muista yksilöllisistä piirteistä, jotka voivat vaikuttaa koesuoritukseen. Mallin tulee tarjota perusta kokeen laadinnalle riippumatta kokeen tarkoituksesta, testattavista ja kohdekielen käyttöalueesta. Kielitaidon määritelmä tarjoaa myös perustan konstruktin (= mittauksen kohteen) eli mitattavan taidon määrittämiselle ja toimii selustana, johon tukeudutaan tehtäessä johtopäätöksiä kielitaidosta (Bachman & Palmer 1996:66).

Bachman ja Palmer (1996) katsovat kielitaitoon (*language ability*) sisältyvän kaksi komponenttia, kielitiedon (*language knowledge*) ja strategisen kompetenssin (*strategic competence*). Kielitieto katsotaan muistissa olevaksi tietoaalueeksi, joka on metakognitiivisten strategioiden avulla hyödynnettävissä tuottaessa ja tulkittaessa diskurssia. Se puolestaan jakaantuu kahteen laajaan kategoriaan, *organisatoriseen* ja *pragmaattiseen* tietoon (Bachman & Palmer 1996:67). Kielelliseen tietoon kuuluvat alueet voidaan esittää Huhdan (1993:89) laatimaana kaaviona (kuvio 8), joka perustuu Bachmanin (1990) malliin. IEA:n kansainvälisessä kirjoitelmatutkimuksessa kehiteltiin 1980-luvun alussa vastaavanlaista kuvausta (Sauli Takala, suullinen tiedonanto 27.6.2006).



KUVIO 8 Kielellisen tiedon alueet (Lähde: Bachman ja Palmer 1996:68.)

Organisaatorinen tieto kattaa kielen formaalisen rakenteen tuntemuksen, ja se jakaantuu kieliopilliseen ja tekstuaaliseen tietoon. Edelliseen liittyy muodollisesti oikeiden puhuntojen tai lauseiden tuottaminen ja ymmärtäminen, jälkimmäiseen laajempien kielen yksiköiden ja niiden välisten yhteyksien tuottaminen ja ymmärtäminen (Bachman & Palmer 1996: 68-69). Pragmaattisen tiedon toinen komponentti, funktionaalinen tieto, auttaa löytämään yhteydet puhuntojen tai lauseiden ja tekstien välillä ja havaitsemaan kielen käyttäjien tarkoitukset. Kuvaukset, luokittelut, selitykset ja surun tai vihan ilmaukset ovat esimerkkejä ideationaalisista funktioista. Manipulatiivisten funktioiden avulla taas vaikutaan ympäröivään maailmaan eli eräällä tavalla manipuloidaan sitä. Näihin

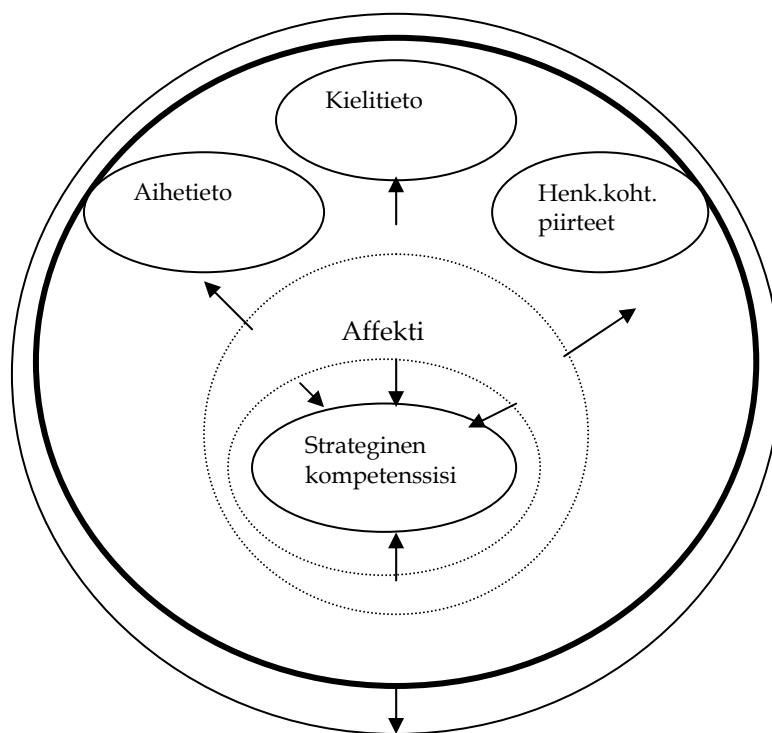
kuuluvat instrumentaaliset funktiot, joiden avulla pyydetään, ehdotetaan, käsketään tekemään jotakin tai varoitetaan tekemästä jotakin, säätelevät (*regulatory*) funktiot, ja interpersonaaliset (ihmisten väliset) funktiot, joiden avulla luodaan suhteita (tervehdykset, kohteliaisuudet, loukkaukset ja anteeksipyyhinnöt). Tieto heuristisista funktioista auttaa lisäämään tietoa ympäröivästä maailmasta, kun taas imaginatiiviset funktiot edesauttavat käyttämään kieltä humoristiin tai esteettisiin tarkoituksiin tai jopa luomaan oman mielikuvitusmaailman (Bachman & Palmer 1996:69-70). Sosiolingvististä tietoa tarvitaan tuottamaan tai tulkitsemaan erilaisiin kielenkäyttötilanteisiin sopivaa kieltä. Se auttaa päättämään, onko vaikkapa murteen käyttö sopivaa, ymmärtämään idiomeja ja kulttuurisia viittauksia sekä kuvaannollista kielenkäyttöä (Bachman & Palmer 1996:70). Lingvististä tietoa käytetään esimerkiksi ymmärrettäessä, että sanonnalla *pull one's socks up* ei ole mitään tekemistä sukkiensa kanssa.

Strateginen kompetenssi on joukko metakognitiivisia komponentteja tai strategioita, joiden voidaan ajatella olevan korkeamman tason kielen käytön toimeenpanevia prosesseja. Kielen käyttöön liittyy edellä käsitellyn kielenkäyttäjän kielitiedon lisäksi hänen tietonsa aiheesta, aihetieto (*topical knowledge*) ja affektiiviset skeemat eli affektit (*affective schemata*). Näiden kaikkien komponenttien onnistunut yhdistäminen mahdollistaa kuhunkin kielenkäyttötilanteeseen sopivan kielen tuottamisen ja tulkinnan. Strategisen kompetenssin hahmottaminen metakognitiivisina prosesseina tarjoaa perustan sellaisten koetehtävien laadinnalle, joissa käytetty prosessi voidaan myös arvioida. Bachman ja Palmer (1996:71) erottavat yleisen kognitiivisen psykologian tapaan kolme metakognitiivisten komponenttien toiminta-aluetta, tavoitteen asettaminen, arviointi ja suunnittelu, joiden sisältö esitetään taulukossa 12.

TAULUKKO 12 Strategisen kompetenssin käyttö. (Lähde: Bachman & Palmer 1996:68.)

Strateginen kompetenssi		
Tavoitteen asettaminen	Arviointi	Suunnittelu
<ul style="list-style-type: none"> - tehtävien tunnistaminen - tehtävän/tehtävien valinta - tehtävän suorittamisesta päättäminen 	<ul style="list-style-type: none"> - tehtävän piirteiden arviointi - oman tiedon (aihe, kielitieto) arviointi - vastauksen oikeellisuuden/sopivuuden arviointi 	<ul style="list-style-type: none"> - aihetiedon ja kielitiedon elementtien valinta - vastauksen muotoilun suunnittelu - lopullisen toimintatavan valitseminen

Jotta koetuloksista voisi tehdä valideja johtopäätöksiä kielitaidosta, on Bachmanin ja Palmerin (1996:61) mukaan osoitettava, millainen kokeessa ja todellisessa elämässä esiintyvien kielenkäyttötilanteiden välinen yhteys on. Tämä yhteys on kyettävä osoittamaan paitsi koetehtävien myös testattavien osalta. Tehtävien piirteiden kuvaukseen on Bachmanin ja Palmerin (1996:47) laatiman viitekehysten mukaan sisällytettävä viisi aspektia: puitteet (*settings*), ohjeet (*rubrics*), syötös, odotettu vastaus sekä syötöksen ja vastauksen välinen suhde. Kirjaamalla viitekehysten edellyttämät yksityiskohdat (emt. s. 49-50) sekä todellisen elämän kielenkäyttötilanteiden että koetehtävien osalta ja vertaamalla näitä toisiinsa, saadaan kestävä pohja kielen kokeen kehittämiseksi ja käytölle.



KUVIO 9 Kielen käytön ja kielikokeessa osoitetun taidon komponentit. (Lähde: Bachman & Palmer 1996:63.)

Testattavia koskevia piirteitä ovat näiden kielitaito, henkilökohtaiset ominaisuudet, aihetieto ja affektiiviset skeemat eli affektit (emt. s. 61), joilla he tarkoittavat aihetiedon affektiivisiä eli emotionaalaisia vastineita (emt. s. 65). Kielen käyttöön liittyy monenlaista prosessointia toisaalta kielen käyttäjien yksilöllisten ominaisuuksien välillä ja toisaalta näiden ominaisuuksien ja kielen todellisessa elämässä tapahtuvan käytön tai koetilanteessa tarvittavan käytön välillä. Näiden prosessien kompleksisuuden vuoksi Bachman ja Palmer (1996) katsovat, että kokeessa testattava kielitaito täytyy määritellä kielen käytön vuorovaikutuksellisesta viitekehystä käsin. Näkemyksensä todellisen elämän kielen käytön ja kielikokeessa osoitetun taidon välillä he ovat esittäneet kuviona 9. Komponentit, jotka sijaitsevat pienemmän, paksummalla painetun ympyrän sisällä, edustavat yksittäisen kielenkäyttäjän ominaisuuksia, kun taas ulompi kehä rajaa niiden kielenkäyttötilanteiden ominaisuudet, joiden kanssa kielen käyttäjä on vuorovaikutuksessa.

Bachmanin ja Palmerin (1996) malli, johon liittyvät kokeen laadinnan kannalta arvokkaat kokeen käyttökelpoisuuden kriteerit, paikkaavat pitkälle aikaisempia puutteita kielitaidon määrittämisessä testauksen näkökulmasta. Mallia kohtaan on suunnattu myös kritiikkiä, erityisesti siksi, että se ei arvostelijoiden mielestä ota riittävästi huomioon todellista interaktiivisuutta kokeen ja sen suorittajan välillä. McNamara (1996) nimeää kommunikatiivisessa testissä huomioon otettavina seikkoina mm. toimijoiden välisen luottamuksen asteen ja henkilökemiat sekä testattavan motivaation ja tunnetilan. Chalhoub-Deville (2003)

korostaa kielenkäyttötilanteen luonnetta sosiaalisena tapahtumana, jossa kieli-taito, kielen käyttäjät ja konteksti liittyvät erottamattomasti yhteen. Bachmanin ja Palmerin malli on kuitenkin toistaiseksi pisimmälle kehitetty ja käyttökelpoi-sin malli (Alderson 2000; Alderson & Banerjee 2002; Alderson 2005).

Douglas (2000) on laatinut Bachmanin ja Palmerin kielitaidon kuvauksen pohjalta kommunikatiivisen kielitaidon komponenttien taksonomian, jossa hän on strategisen kompetenssin komponentteihin lisännyt toteutuksen kontrolloin-ti -kohdan (*control of execution*) (taulukko 13).

TAULUKKO 13 Kommunikatiivisen kielitaidon komponentit. (Lähde: Douglas 2000.)

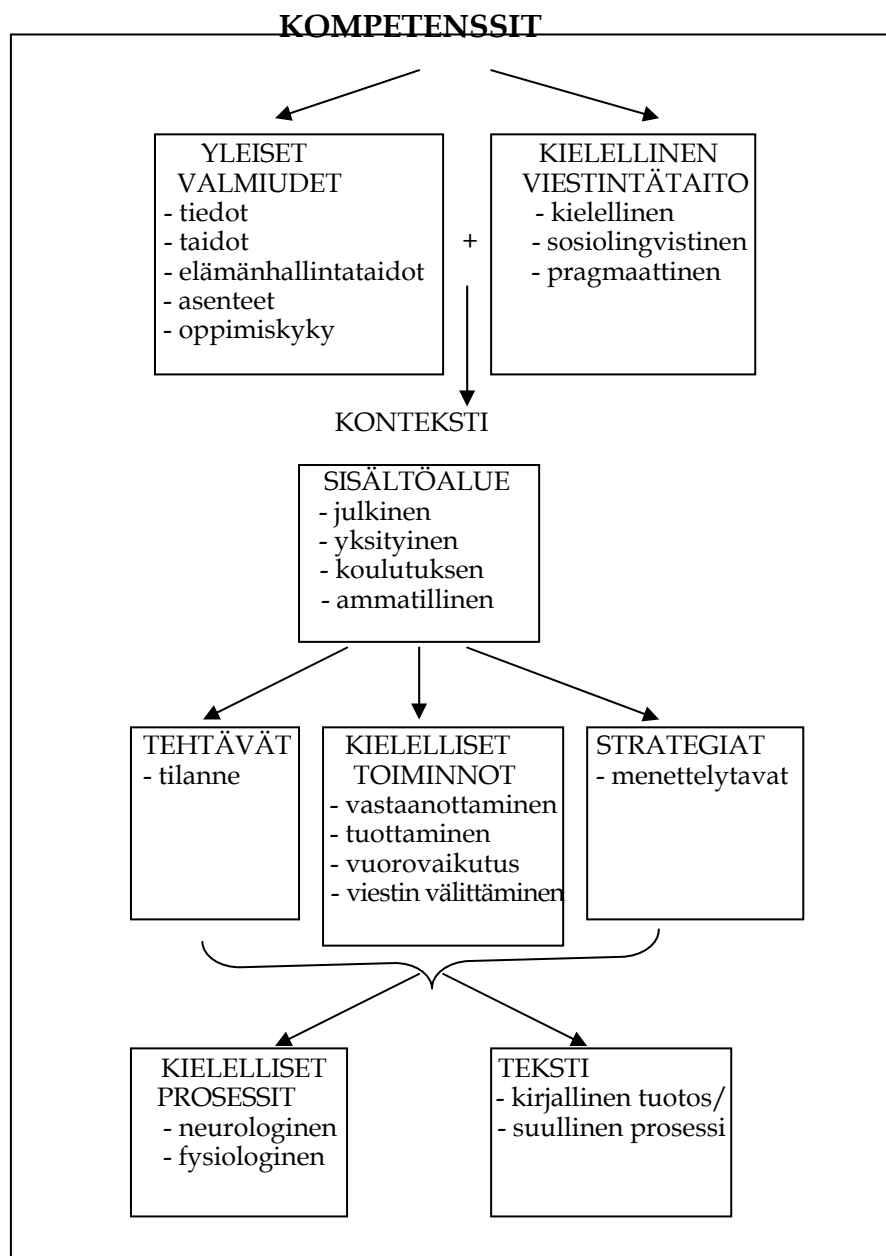
<p>Kielitieto</p> <p>Kieliopillinen tieto</p> <ul style="list-style-type: none"> * Sanaston tuntemus * Morfologian ja syntaksin tuntemus * Fonologian tuntemus <p>Tekstuaalinen tieto</p> <ul style="list-style-type: none"> * Sidosteisuuden hallinta * Retorisen tai keskustelun organisoinnin hallinta <p>Funktionaalinen tieto</p> <ul style="list-style-type: none"> * Ideationaalisten funktioiden tuntemus * Manipulatiivisten funktioiden tuntemus * Heurististen funktioiden tuntemus * Imaginatiivisten funktioiden tuntemus <p>Sosiolingvistinen tieto</p> <ul style="list-style-type: none"> * Murteiden/Varianttien tuntemus * Rekisterien tuntemus * Idiomaattisten ilmausten tuntemus * Kulttuuristen viittausten tuntemus
<p>Strateginen kompetenssi</p> <p><i>Arviointi</i></p> <ul style="list-style-type: none"> * Kommunikatiivisen tilanteen tai koetehtävän arviointi ja soveltuvan diskurssialueen käyttöön ottaminen * Vastauksen kielellisen oikeellisuuden tai soveltuvuuden arvioiminen <p><i>Tavoitteen asettaminen</i></p> <ul style="list-style-type: none"> * Päätöksen tekeminen kommunikatiivisessa tilanteessa käyttäytymisestä <p>Suunnittelu</p> <ul style="list-style-type: none"> * Päätöksen tekeminen siitä, mitä kielitiedon elementtejä tai taustatietoa tarvitaan tavoitteeseen pääsemiseksi. <p><i>Toteutuksen kontrollointi</i></p> <ul style="list-style-type: none"> * Kielitiedon tarpeellisten elementtien käyttöön ottaminen ja järjestäminen suunnitelman toteuttamiseksi

Eurooppalainen viitekehys (2003) tarjoaa Bachmanin ja Palmerin (1996) malliin läheisesti liittyvän kuvauksen kielitaitoon sisältyvistä elementeistä pyrkimällä osoittamaan kielen käytön monia erilaisia ulottuvuuksia ("vaakasuora ulottu-vuus") sekä niissä tarvittavia strategioita portaittain vaikeutuvien asteikkojen avulla ("pystysuora ulottuvuus")¹ laajaa eurooppalaista asiantuntemusta mo-nipuolisesti hyödyntäen. Se käyttää kielitaidosta nimitystä kielenkäyttäjän ja kielenoppijan kyvyt, taidot ja valmiudet, joista englanninkielisessä alkuteokses-

¹ De Jong (2004) on käyttänyt näiden sijasta termejä "kvalitatiivinen" ja "kvantitatiivi-nen" ulottuvuus.

sa Common European Framework of Reference for Languages: Learning, Teaching, Assessment (2001) käytetään nimitystä *the user's/learner's competences*. Nämä jaetaan yleisiin valmiuksiin (*general competences*) ja kielelliseen viestintätaitoon (*communicative language competences*).

Eurooppalainen viitekehys (2003:28) laajentaa kielitaidon käsitettä kielen käytön suuntaan osoittaen, millaisia ulottuvuuksia kielelliseen toimimiseen yhteiskunnan jäsenenä liittyy. Näitä ulottuvuuksia ja niiden suhdetta toisiinsa kuvataan kuviossa 10.



KUVIO 10 Kielen käytön ulottuvuudet. (Muokattu taulukoksi Eurooppalaisen viitekehysten mukaan. 2003:28-29.)

Kompetenssi on kaiken kielellisen toiminnan perusta. Jotta sen merkitys kielellisen toiminnan jokaisessa vaiheessa korostuisi, kompetenssi -sanaa ei ole sijoitettu erilliseen laatikkoon. Näitä kompetensseja eli kielenkäyttäjän kykyjä, taito-

ja ja valmiuksia ryhmitellään neljäksi sisältö- eli elämänalueeksi (*domain*) (emt. s. 35). Niistä jokaisella esiintyy erilaisia tehtäviä, tilanteita, joista kielen avulla suoriudutaan käyttämällä erilaisia strategioita. Niistä valitaan soveltuvimmin tai soveltuvimmat, kun tehtävä on hahmotettu ja edellytetty kielellisen toiminnan muoto tiedostettu. Kielellisen prosessoinnin kautta päästään tekstiin, joka tarkoittaa mitä tahansa kirjallista tai suullista tiettyyn sisältöalueeseen liittyvää prosessia tai tuotosta (emt. s. 29).

Eurooppalaisen viitekehysten (2003) konkreettisin anti kielitaidon arvioinnille ovat sen lukuisat taitotasosteikot, joiden sisältöjä yhdistämällä saadaan räätälöityä moniin testaustilanteisiin soveltuvia asteikkoja. Kielitaidon kokonaiskuvauksia esitetään Eurooppalaisen viitekehysten (2003) taulukossa *Yleiset taitotasot: Kielitaidon laajuus* (emt. s. 157). Liitteenä 2 olevassa koontataulukossa kielellistä viestintätaitoa lähestytään yksityiskohtaisemmin kuvaamalla sen pääkomponenttien sisällöt ja luettelemalla niiden arviointiin laaditut taitotasosteikot, joihin viitataan Kategoriat -sarakeessa. Niiden avulla testattavia sisältöjä esitetään Esimerkkejä -sarakeessa. Sosiolingvististen tietojen ja taitojen elementtien sijoittaminen asteikon tai asteikkojen eri tasoille ei laatijoiden mukaan onnistunut siksi, että esimerkiksi rekisterien erottamista ei voi alkeiskielitaidon tasolla edellyttää (emt. s. 170-171). Siksi tämän pääkomponentin kohdalla Kategoria -sarakeen merkintä A1-B1 tarkoittaa, että otsikon toimintoja voidaan edellyttää näillä taitotasoilla, kun taas B2 viittaa siihen, että testattavan tulee olla vähintään tällä tasolla, jotta kyseisiä asioita voidaan ottaa arvioinnissa huomioon. Liitteessä 2 käytetyt englanninkieliset termit ovat alkukielisessä teoksessa käytettyjä ja ne on sijoitettu taulukkoon oikeiden vastineiden löytämiseksi.

Eurooppalainen viitekehys (2003) ottaa esille myös arvioinnin ja opetus-suunnitelman välisen suhteen todeten, että mitä tahansa liitteen 2 pääkomponenttia saatetaan painottaa opetussuunnitelmassa, mikä on otettava huomioon koetta laadittaessa (emt. 188). Viitekehys ryhmittelee tavoitetyypit seuraavasti:

1. Yleisten valmiuksien kehittäminen eli huomion kiinnittäminen deklaratiiiviseen tietoon, taitoihin ja taitotietoon, persoonallisuuden piirteisiin tai asenteisiin.
2. Kielellisen viestintätaidon laajentaminen eli jokin taulukon 12 pääkomponenteista.
3. Parempi suoriutuminen yhdestä tai useammasta kielellisestä toiminnosta, joita ovat vastaanottaminen, tuottaminen, vuorovaikutus tai viestin merkityksen välittäminen.
4. Parhaaseen mahdolliseen tulokseen pääseminen jollakin elämänalueella (julkinen, ammatillinen, koulutuksen tai yksityinen). (Emt. s. 188-191)

On kuitenkin huomattava, että vaikka jokin edellisistä olisi korostetusti opiskelun lähtökohta, se ei sulje pois muiden tavoitteiden ainakin osittaista huomioon ottamista.

Kielitaitoa voidaan tarkastella myös kielellisten toimintojen ja viestintästrategioiden näkökulmasta, ja viitekehys sisältääkin runsaasti niiden arviointiin tarkoitettuja taitotasosteikkoja, jotka on koottu liitteeseen 3. Sivunumerot viittaavat Eurooppalaiseen viitekehykseen (2003).

Eurooppalainen viitekehys suosittelee liitteen 3 asteikkoja käytettäväksi laadittaessa koetehtäviä kommunikatiivista arviointia varten, mutta korostaa

samalla mahdollisimman erilaisten tehtävien sisällyttämistä kokeeseen (emt. 2003:244). Samalla niiden käyttäminen tuo läpinäkyvyyttä arviointeihin ja auttaa suorittamaan kansainvälisiä vertailuja (emt. 2003:248).

4.2 Prosessi ja tuotos kielitaidon arvioinnin kohteina

4.2.1 Ymmärtäminen

Ymmärtäminen voidaan tulkita sekä prosessiksi että tuotokseksi. Ymmärtämisprosessilla tarkoitetaan sitä, mitä ymmärtämisellä yleensäkin: vastaanottajan ja tekstin välistä vuorovaikutusta (Alderson 2000:3)². Tämän prosessin tuote, tuotos, taas on se, mitä tekstistä on ymmärretty. Aikaisemmin tutkimus kohdistui tuotteeseen, josta kiinnostus siirtyi prosessiin. Viime aikoina ollaan oltu taas palaamassa tuotoksen tutkimiseen johtuen prosessin selvittämisen ongelmallisuudesta (Alderson 2000:5).

Ymmärtämisen tulkitseminen tuotokseksi lähtee siitä, että lopputulos on yleensä samantapainen edeltävästä prosessista riippumatta. Niin ymmärtämisen teorioissa kuin jokapäiväisessä puheessakin on tavallista erottaa ymmärtämisen tasoja, joita kuvataan määrittämällä, mitä asioita tekstistä testattava ymmärtää. Tällaisia tasoja ovat esimerkiksi eksplisiittisesti (*literal meaning*) tai implisiittisesti (*inferred meaning*) ilmaistun asian ymmärtäminen, kriittinen sekä arvioiva ymmärtäminen. Joskus yksinkertaiseen, esimerkiksi yksittäisten sanojen, ymmärtämiseen viitataan mikroprosesseina (paikallinen ymmärtäminen) ja lauseiden merkityksen ymmärtämiseen makroprosesseina (globaali ymmärtäminen).

Kun tiettyjä tasoja pidetään 'vaikeampina', 'syvempää' ymmärtämistä osoittavina, ymmärtämisen käsitteisiin alkaa liittyä yhä enemmän hierarkiaa. Se on myös johtanut käsitykseen, jonka mukaan 'alemmat' taidot opitaan ensin ja että ne ovat edellytyksinä 'korkeammille' (Alderson2000:8).

Puheen ymmärtäminen prosessina

Faerch ja Kasper (1986) olettavat äidinkielen puheen mallin koostuvan seuraavista sisällöistä.

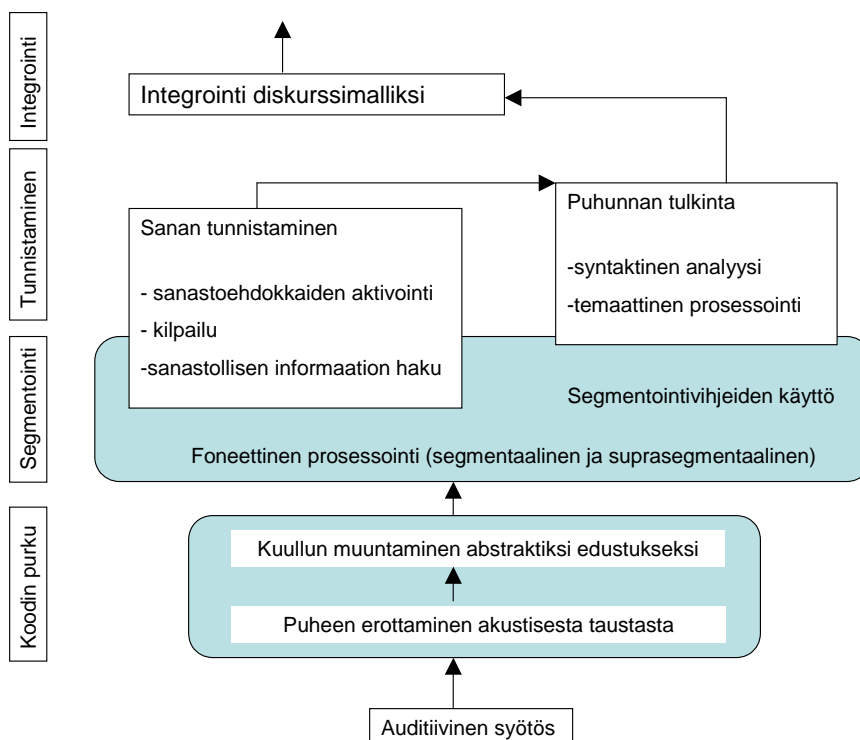
- Ymmärtämisprosessit perustuvat kolmentyyppiseen informaatioon: kielellinen syötös, vastaanottajan kielellinen ja sosiokulttuurinen tieto, kontekstuaalinen informaatio.
- Syötöksen ja vastaanottajan tiedon yhteensovittaminen voi alkaa joko syötöksestä tai tiedosta. Molemmissa tapauksissa kontekstuaalista tietoa käytetään tukemaan ymmärtämistä.
- Eri ymmärtämistehtävät kohdentuvat jompaankumpaan kahdesta prosessoinnin suunnasta.
- On pikemminkin sääntö kuin poikkeus, että syötöksen ja tiedon välinen yhteensovitus on epätäydellinen, mikä johtaa "aukkoihin" eli ymmärtäminen on tyypillisesti enemmän osittaista kuin kokonaisvaltaista.

² Alderson käsittelee tässä vain tekstin ymmärtämistä.

Cutler ja Clifton (1999) ovat esittäneet Willem Leveltin (1989, 1999) tunnettua puheen tuottamisen mallia mukaillen puheen ymmärtämisen prosessimallin. Mallin mukaan kuulijaa voidaan luonnehtia välineeksi, joka muuntaa akustisen syötöksen merkitykseksi. Kuvio 11 kuvaa kuulijan toiminnan alkaen siitä, kun kuultu syötös saavuttaa korvan. Koodin purkuvaiheessa kuulija erottaa ensin puheen muusta samaan aikaan mahdollisesti kuullusta. Tämän jälkeen hän muuntaa sen abstraktimmaksi edustukseksi, jonka puitteissa esimerkiksi tietty äänne yhdistetään jossakin aikaisemmassa yhteydessä jonkin muun puhujan eri nopeudella tuottamana kohdatuksi äänneeksi (Cutler & Clifton 1999:122).

Seuraavassa vaiheessa (jatkuva) signaali segmentoidaan tämän signaalin komponenteiksi. Itse asiassa ei ole kysymys eri vaiheesta, mitä osoittaa laatikkojen päällekkäisyys, vaan segmentointi johtuu suurelta osin prosessoinnin toiminnoista, jotka liittyvät sanojen tunnistamiseen ja puhunnan tulkintaan. Kuuntelija käyttää kuitenkin myös osia puhutusta syötöksestä määrittääkseen sanojen ja lauseiden rajat. Puhunnan tulkinta on tämän jälkeen suhteutettava diskurssikontekstiin, oletuksiin kuulijasta ja hänen tiedoistaan sekä ottaen huomioon monenlaiset sosiolingvistiset tekijät. Näihin ei kuvio kuitenkaan varsinaisesti ota kantaa (Cutler & Clifton 1999:122).

Runsaan kahdenkymmenen vuoden tutkimus puhuttujen sanojen tunnistamisesta on johtanut sellaisten uusien mallien tuottamiseen, joissa tunnistaminen on keskeinen mekanismi keskenään kilpailevien sanaehdotusten hyväksymisessä oikeiksi vaihtoehdoiksi. Sanastosta saatavan tiedon lisäksi on tarkasteltava niitä syntaktisia yhteyksiä ja temaattisia rooleja, joissa sana voi esiintyä tietyn semanttisen rakenteen sisällä sekä tulkittava tämän prosessoinnin tulos. Saattaa olla, että ihmisen omaaman lauseiden prosessointijärjestelmän erityispiirteet voivat ratkaista tämän peräkkäisen ja luonteeltaan ainutkertaisen akustisen syötöksen mukanaan tuomat ongelmat. Prosodinen informaatio saattaa tosin myös rajoittaa ymmärtämisprosessia. Edelleen selvitetäviksi jäävät prosessoinnin eri komponenttien välisen vuorovaikutuksen luonne ja laajuus samoin kuin se, perustuvatko kaikki tämän prosessointijärjestelmän aspektit yleisiin taustalla oleviin periaatteisiin vai edellyttävätkö järjestelmän eri aspektit erilaisia toimintaperiaatteita ja erilaisia sanaston edustuksia (Cutler & Clifton 1999:154-5).



KUVIO 11 Kuulijan auditiivisen syötöksen prosessointi. (Lähde: Cutler & Clifton 1999:124.)

Tekstin ymmärtäminen prosessina

Lukijan suhdetta ymmärtämisprosessiin kuvaa kaksi erilaista lähestymistapaa: *top-down* ja *bottom-up*. *Top-down* -prosessointi korostaa skeemoja, vastaanottajan tietojen sisällyttämistä ymmärtämiseen ja sitä, että ymmärtämisen ei tarvitse tapahtua tietyssä järjestyksessä eikä edetä pienemmistä yksiköistä suurempiin. *Bottom-up* -lähestymistapa katsoo ymmärtämisen lähtevän koodin (foneemit/kirjaimet) purkamisesta: tunnistaen akustiset tai ortografiset ärsykkeet ja muuntaen ne äänneiksi, tunnistaen sanat ja lopuksi päätyen merkityskokonaisuuksiin (Alderson 2000, Buck 2001, Hellekjaer 2005). Yhä yleisempi kanta on se, että ymmärtäminen jakaantuu kahteen komponenttiin, koodin purkuun ja ymmärtämiseen, joista jälkimmäiseen kuuluvat lauseiden prosessointi, lauseiden ymmärtäminen diskurssissa, diskurssirakenteen kokoaminen ja aikaisemman tiedon integroiminen uuteen (Alderson 2000:12; Buck 2001:52). Ymmärtäminen voidaan myös jakaa kahteen ryhmään kiinnittäen huomiota prosessointitapaan: a) assimilointiin, jossa prosessoidaan diskurssia lineaarisena sekvenssinä ja b) diskriminointiin, jossa valikoivan prosessin kautta abstrahoidaan pääkohdat ja arvioidaan tekstin osien keskinäistä tärkeyttä.

Lukemista pidetään usein reaktiona merkityksiin, joita ei kuitenkaan aina pystytä ymmärtämään, vaikka ne näyttäisivätkin tekstissä selviltä. Widdowsonin (1978) mielestä luullaan, että syynä siihen on tekstikoodin purkamisen vaikeus. Hän katsoo kuitenkin, että kyse on pikemminkin merkityksen tuottamisen epätarkkuudesta. Teksti itsessään ei sisällä merkitystä eikä merkityksen tuottaminen ole pelkästään tekstin muotoilemista. Se on samalla ohjeiden sisäl-

lyttämistä tekstiin siitä, mistä oman tietonsa ja kokemuksensa käsitemaailmasta koodin purkajan eli lukijan tulee hakea koodin laatijan eli kirjoittajan tarkoituksia. Widdowson (1978:169) ei siis pidä lukemista reaktiona tekstiin, vaan tekstin välittämänä vuorovaikutuksena kirjoittajan ja lukijan välillä.

Harste (1985), jota Takala (1992) on referoinut, on esittänyt kolme näkemystä tekstin ymmärtämisestä. Ensimmäinen on ns. informaation siirtämisprosessi, jonka lähtökohtana on se, että merkitys on tekstissä ja lukeminen on prosessi, jossa merkitys siirtyy tekstistä lukijalle. Avaintekijöinä tässä prosessissa ovat lukutaidon eri osataidot. Hyvät lukijat siis siirtävät tekstistä enemmän informaatiota kuin huonot lukijat ja lukemisen onnistumisen arvioinnin kriteerinä on siirretyn informaation määrä (Takala 1992:71).

Informaationäkemyistä, joka lienee yleisin näkemys, pidetään kuitenkin liian suppeana tekstin ymmärtämisprosessien selittämiseksi. Sitä on pyritty täydentämään vuorovaikutusnäkemyksellä, jonka mukaan merkitys on sekä tekstissä että lukijan päässä. Lukeminen on prosessi, jossa nämä kaksi merkitysjärjestelmää ovat vuorovaikutuksessa keskenään. Avaintekijöitä tässä prosessissa ovat lukemistaidot ja tekstin rakenne, joten hyvien lukijoiden vuorovaikutus tekstien kanssa on erilainen kuin heikkojen lukijoiden. Lukemisen onnistumisen kriteerinä on se, mitä strategioita hyvä lukija olisi käyttänyt tekstin lukemiseen kyseisessä tilanteessa (Takala 1992:72).

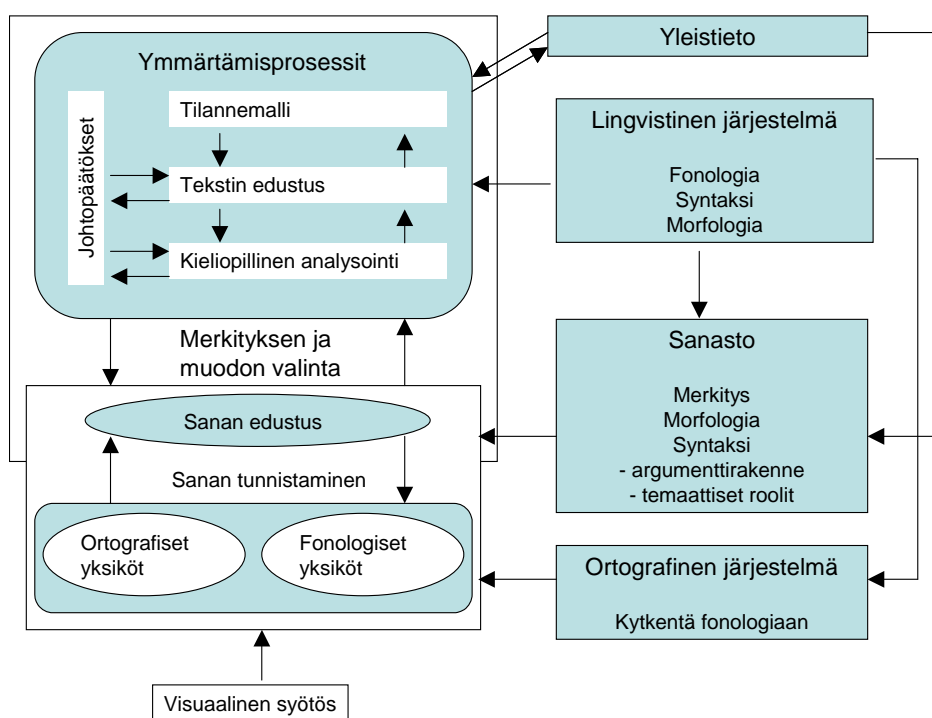
Vuorovaikutusnäkemysten puutteita on puolestaan pyritty korjaamaan ns. transaktionäkemyksessä. Sen keskeiset oletukset ovat:

- Merkitys on suhteellinen ja se vaihtelee jonkin verran lukijasta riippuen. Lukija myös tulkitsee merkitystä eri lailla eri aikoina ja eri tilanteissa.
- Lukeminen on tulkintaprosessi. Oletetaan, että eri lukijat antavat tekstille erilaisia tulkintoja. Samaan kieli- ja tulkintayhteisöön kuuluvat jäsenet (lukijat) tulkitsevat sisältöjä samansuuntaisesti.
- Tulkintaprosessissa avaintekijöitä ovat kulttuuri- ja historiallis-yhteiskunnallinen konteksti, kokemustausta ja se, mitä lukija on aikaisemmin lukenut.
- Hyvät lukijat oppivat lukemastaan. Mitä paremmin onnistutaan lukemaan, sitä enemmän opitaan (Takala 1992:73).

Vuosikymmeniä tekstin ymmärtämistä systemaattisesti tutkinut Walter Kintsch (1998) on kehittänyt ymmärtämisen konstruktiivis-integratiivisen prosessimallin (CI), joka on merkinnyt tuntuvaa painopisteen siirtymistä skeemapohjaisesta ymmärtämisprosessin kontrollijärjestelmästä *bottom-up* -pohjaiseen järjestelmään, jossa kontekstivaikutukset hahmotetaan rajoitusten täyttymisenä. Tämä on tehnyt ensi kertaa mahdolliseksi ottaa ymmärtämisprosessissa huomioon sekä tieto ja muisti. Malli pyrkii kuvaamaan prosessin, jossa vastaanotettava kieli muunnetaan henkiseksi malliksi (*mental model*) vastaanottajan mielessä. Se olettaa tämän prosessin sisältävän kaksi vaihetta, konstruoimis- ja integroimisvaiheen. Tekstien henkisillä malleilla on kaksi lähdettä: teksti itse ja se, mitä vastaanottajat tuovat siihen omien tarpeidensa mukaan omista kokemuksistaan ja tiedostaan. Kintsch (1998) katsoo, että on usein hyödyllistä tehdä ero tekstin mallin ja ns. tilannemallin (*situation model*) välillä, jolla hän tarkoittaa vastaanottajan panosta ymmärtämiseen. Kintsch (1998) korostaa myös sitä, että tekstistä oppiminen on monella tavalla erilaista kuin tekstin muistaminen. Jälkimmäinen

(esim. parafraasin tai tiivistelmän tuottaminen) ei välttämättä vaadi tekstin ymmärtämistä, mutta edellinen, kyky käyttää tekstiä produktiivisesti uudessa ympäristössä, saattaa edellyttää hyvinkin syvällistä ymmärtämistä (Kintsch 1998:290). Kintschin mallia ei tietyvästi ole kuitenkaan ainakaan toistaiseksi käytetty ymmärtämisen arvioinnissa.

Perfetti (1999), samoin kuin Cutler ja Cliftonkin (1999), on käyttänyt hyväkseen Leveltin (1999) puheen tuottamisen prosessoinnin mallia ja laatinut sen pohjalta lukijan suorittaman prosessoinnin kaavion (kuvio 12). Se esittää ne informaation lähteet, joita lukijan oletetaan käyttävän ymmärtäessään kirjoitettua kieltä. Tämän informaation käyttöä tulee tulkita kahdesta näkökulmasta, kirjoitusjärjestelmän, joka luo pohjan lukemisen yksiköille, ja niiden kognitiivisten prosessien, joiden tuloksena pystytään lukemaan (Perfetti 1999:169).



KUVIO 12 Kirjoitetun tekstin prosessointi. (Lähde: Perfetti 1999:170.)

Kaavio kuvaa lukemisen yleisiä komponentteja. Nuolet osoittavat informaation tai vaikutuksen suunnan. Kielellinen tieto ohjaa fonologiaa, morfologiaa ja syntaksia, joita käytetään sanan lukemisessa ja lauseen ymmärtämisessä. Yleistieto (ei-kielellinen) ohjaa sanastoa ja ymmärtämisprosessia. Sanojen tunnistaminen esitetään prosessina, joka synnyttää fonologis-ortografisia kokonaisuuksia (Perfetti 1999:169).

Kirjoitusjärjestelmä on erittäin tärkeä selitettäessä lukemisprosessia. Koska se tarjoaa yksiköt, jotka liittyvät yhteen tai useampaan kielen tasoon – foneemiin, tavuihin, morfeihin, sanastoon – se vaikuttaa kirjoitettujen sanojen tunnistamisprosessiin. Seuraava tarkastelukulma on kirjoitusjärjestelmän yksikköjen muuntaminen henkiseksi edustuksiksi eli painettujen viestien ymmärtämiseksi. Tämä koostuu kahdesta kielen tärkeästä komponentista: sanojen tunnistamisesta

ja kielen prosessointimekanismeista, jotka yhdistävät sanat viesteiksi. Nämä mekaniemit tuottavat kontekstiin sopivia sanojen merkityksiä, jäsentelivät sanajonot sanoiksi ja auttavat tekemään johtopäätöksiä lauseen tarjoamasta informaatiosta, minkä avulla luodaan täydellisempiä, koko tekstin edustuksia. Nämä eivät kuitenkaan ole pelkästään kielellisten prosessien tulosta, kuten kuvio 12 osoittaa, vaan niitä on täydennettävä monilla yleisillä ja erityisillä tietolähteillä. Selvitettäväksi jää vielä, kuinka nämä eri komponentit toimivat yhdessä, esimerkiksi ovatko toiset tietolähteet ehkä toisia tärkeämpiä. Näitä toivotaan saatavan selville kognitiivisin ja neurokognitiivisin tutkimuksin (Perfetti 1999:167-198).

Perfettin (1999) malli, joka äidinkielen kontekstistaan huolimatta soveltuu myös vieraiden kielen lukemisen hahmottamiseen, osoittaa havainnollisesti, kuinka monitahoinen ilmiö tekstin ymmärtäminen on, ja auttaa ymmärtämään, miksi lukemisessa voi joillakin olla suuriakin vaikeuksia. Perfettin (1999) mallia ei ole tiettävästi sovellettu, ainakaan vielä, kielitaidon arviointikontekstissa.

Puheen ymmärtäminen tuotoksena

Buck (2001:32) esittää, että kokeen takaistusvaikutuksen kannalta olisi parempi keskittyä korostamaan tietoa, taitoja ja kykyä prosessin sijasta, koska nämä ovat ominaisia todellisen elämän kuuntelutilanteille. Monet teoreetikot ovat yrittäneet kuvata näitä kuunteluprosessin taustalla olevia taitoja taksonomioiden avulla. Useimmat niistä ovat kuitenkin tutkijoiden olettamuksia ymmärtämisprosessiin sisältyvistä taidoista eivätkä perustu empiirisiin tutkimuksiin. Yleisimmin kuuntelu jaetaan kaksivaiheiseksi prosessiksi, jossa ensin erotetaan kielellinen perustieto ja sitten informaatiota käytetään kommunikatiiviseen tarkoitukseen (Rivers 1972, Carroll 1972, Clark & Clark 1977, Buck 2001).

Eräs ensimmäisistä taksonomioista on Valetten (1977) laatima, ja se on kiinnostava siksi, että se on laadittu testaamisen pohjaksi ja että se seuraa ilmeisen selvästi Bloomin portaittain vaikeutuvien kognitiivisten taitojen kuvausta. Taidot osoittavat hänen mukaansa lisääntyvää puheen ymmärtämisen sujuvuutta (Buck 2001:55). Taulukon 14 Prosessit -sarake on peräisin Bloomin (1956) oppimistavoiteluokittelusta ja Taidot -sarake Valetten (1977) luokittelusta.

TAULUKKO 14 Puheen ymmärtämisen taitojen kognitiivinen luokittelu Valetten mukaan (Bloom 1956; Valette 1977).

Prosessit	Taidot	Taitojen kuvaus
Ymmärtäminen	Mekaaniset taidot	Kuuntelija ymmärtää muistin avulla, ei prosessoimalla, esim. havaitessaan eroja äänteiden välillä ja erottamalla ne.
Soveltaminen	Kielentuntemus	Kuuntelija osoittaa tämän osoittamalla tuntevansa faktoja, sääntöjä jne. esim. vastaamalla yksinkertaisiin kysymyksiin.
Analyysi	Siirtovaikutus	Kuuntelija käyttää tätä tietoa uusissa tilanteissa esim. hankkimalla tietoa yhdestä lähteestä ja käyttämällä sitä toisessa.
Synteesi	Kommunikaatio	Kuuntelija käyttää kieltä luonnollisen kommunikaation välineenä esim. ymmärtäessään viestin, joka sisältää uutta informaatiota.
Arviointi	Kritiikki	Opiskelija analysoi tai arvioi kieltä suhteessa sen vaikuttavuuteen, soveltavuuteen, tyyliin, sävyyn jne.

Valetten lähetyksistä oli kognitiivinen. Buckin (2001) mukaan ensimmäisiä kommunikatiivisen lähestymistavan soveltajia oli Aitken (1978. Siteerattu Buck 2001:31), joka kytki taidot laajempaan kommunikatiiviseen tilanteeseen (Buck 2001:53-54). Weir (1993) kokosi Aitkenin (1978) esittämää kattavamman taksonomian kommunikatiivisen kuuntelun osatiedoista, vaikkakaan hän ei kutsunut sitä taksonomiaksi vaan kuuntelukokeiden edellyttämien operaatioiden tarkistuslistaksi. Nämä eivät siis kuitenkaan ole tutkimusperustaisia taksonomioita. Taulukossa 15 esitetään heidän taitotaksonomiensa rinnakkain siten, että kursiivilla osoitetaan taksonomioiden yhtäläisyydet. Tämän tutkimuksen kohteena olevassa arvioinnissa puheen ymmärtämisen tehtävät luokiteltiin Weirin tarkistuslistojen pohjalta (luku 8).

TAULUKKO 15 Puheen ymmärtämisen taitojen kommunikatiivinen luokittelu. (Muokattu taulukoksi Buckin 2001:54-55 mukaan.)

Aitken (1978)	Weir (1993)
	Suora ymmärtäminen
- sanaston ymmärtäminen ja kyky arvata vieraiden tai epäselvien sanojen merkitys lauseyhteydestä	- keskeisen asian ymmärtäminen
- puhutulle kielelle ominaisten syntaktisten mallien ja morfologisten muotojen ymmärtäminen sekä sen diskurssimallien noudattaminen	- pääajatuksen /-ajatusten tai tärkeän tiedon ymmärtäminen ja sen/niiden erottaminen täydentävistä yksityiskohdista tai esimerkeistä
- painotettujen ja painottamattomien sanojen sekä intonaation ja muiden puheelle ominaisten äänenkäyttöön liittyvien vihjeiden ymmärtäminen	- yksityiskohtien ymmärtäminen, johon kuuluu tärkeiden yksityiskohtien muistaminen
- <i>puhujan tarkoituksen ymmärtäminen</i>	- <i>puhujan asenteen tai aikeen määrittäminen kuulijaa tai aihetta kohtaan</i>
	Johtopäätösten tekeminen
- oikeiden johtopäätösten tekeminen sosiaalisesta tilanteesta, puhujan tarkoituksesta tai kontekstista yleensä	- johtopäätösten tekeminen
- <i>puhujan asenne kuulijaa ja keskustelun aihetta kohtaan</i>	- <i>puheen / puhunnan suhteuttaminen sosiaaliseen tai tilannekontekstiin</i>
- puhujan viestin välittämiseen käyttämien retoristen keinojen identifioiminen	- puhumisen kommunikaatiotehtävän tunnistaminen
	- vieraiden sanastollisten yksiköiden merkityksen päättelemine kontekstista
	<i>Puhujan tarkoituksen ymmärtämistä avustavat keinot</i>
	- fonologisten piirteiden ymmärtäminen
	- kieliopillisten käsitteiden kuten vertailun, syyn, tuloksen, asteen jne. ymmärtäminen
	- päälauseen rakenteiden tai ajatusyksikköjen ymmärtäminen
	- koheesion, erityisesti viittausten ymmärtäminen
	- sanastollisen koheesion, erityisesti sanaperheiden ja kollokaatioiden ymmärtäminen
	- sanaston ymmärtäminen

Weir (1993) toteaa myös, että muistiinpanojen tekeminen kuullun pohjalta edellyttää olennaisten avainkohtien havaitsemista ja niiden valitsemisen taitoa pu-

heesta (Weir 1993:98-99). Richards (1985) puolestaan korostaa kuuntelemisen tarkoitusta todeten, että se vaihtelee sen mukaan, onko kyseessä kuuntelemisen liittyminen sosiaaliseen toimintaan, esimerkiksi keskustelemiseen, informaation hankkiminen vai jokin muu syy. Hän katsoo myös, että kuhunkin näistä sisältyvät omat mikrotaitonsa. Richardson koonnut 33 -kohtaisen taksonomian keskustelussa tarvittaviksi ja 18 -kohtaisen akateemisessa (luentojen) kuuntelussa tarvittaviksi taidoiksi (Richards 1985:198-199). Myös Buck (2001) lähestyy puheen ymmärtämistä tekstin tarkoituksen näkökulmasta, mutta hän on koonnut sitä varten esimerkeiksi taulukot, joissa luetellaan tehtäviä, jotka hänen mukaansa testaavat eksplisiittisesti ja implisiittisesti ilmaistujen asioiden ymmärtämistä (emt. s. 135-136). Hughes (1989) puolestaan lähtee siitä, mitä kuuntelijan tulisi osata tehdä tietynlaisissa puhumistilanteissa. Hän määrittelee taidot operaatioiksi kuten Weirkin (1993) ja toteaa, että eräät operaatioista ovat globaaleja, nimittäin pääasian ymmärtäminen, argumentin kehittelyn seuraaminen ja puhujan asenteen tunnistaminen. Muut operaatiot hän jakaa sen perusteella, liittyykö puhuminen enemmän informaation hankkimiseen/saamiseen vai vuorovaikutukseen (taulukko 16). Kielitaitoa arvioitaessa kunkin operaation osalta tulisi myös ilmaista onko kyseessä implisiittinen vai eksplisiittinen ymmärtäminen. Tämän tutkimuksen kohteena olevassa arvioinnissa sovellettiin pääosin Weirin (1993), mutta myös Riversin (1972) ja Hughesin (1989) esittämiä näkemyksiä.

TAULUKKO 16 Kuuntelijan taidot. (Lähde: Hughes (1989:161-162).)

Informaation hankkimisen operaatioita	Vuorovaikutteisia operaatioita
<ul style="list-style-type: none"> • faktatiedon hankkiminen • ohjeiden noudattaminen • kysymysten ymmärtäminen tietojen antamiseksi • tarpeen ilmausten ymmärtäminen • avun pyytämisen ymmärtäminen • luvan pyytämisen ymmärtäminen • anteeksipyyntöjen ymmärtäminen • tapahtumasarjojen (kerronnan) seuraaminen • mielipiteiden tunnistaminen ja ymmärtäminen • mielipiteiden perusteluketjun seuraaminen • vertailun ymmärtäminen • ehdotusten tunnistaminen ja ymmärtäminen • kommenttien tunnistaminen ja ymmärtäminen • puolustelujen ja pahoittelujen ymmärtäminen • preferenssien tunnistaminen ja ymmärtäminen • valitusten tunnistaminen ja ymmärtäminen • spekulatioiden tunnistaminen ja ymmärtäminen. 	<ul style="list-style-type: none"> • tervehdysten ja esittäytymisten ymmärtäminen • yksimielisyyden ilmausten ymmärtäminen • erimielisyyden ilmausten ymmärtäminen • puhujan tarkoituksen ymmärtäminen • epävarmuuden ilmaisemisen ymmärtäminen • lisäselvityspyyntöjen ymmärtäminen • lisäselvityspyyntöjen tunnistaminen • mielipiteen kysymisen tunnistaminen • ymmärtämisen ilmauksien tunnistaminen • ymmärtämättä jäämisen ilmaukset • puhujan oman tai toisen tuotoksen korjaamisen tunnistaminen ja ymmärtäminen • puheen tai kommentin muotoilun tunnistaminen ja ymmärtäminen • perustelujen tai tuen tunnistaminen • vahvistuskysymysten tunnistaminen • taivutteluyritysten tunnistaminen.

Englannin arvioinnissa Hughesin (1989) esittämiä yksityiskohtaisia operaatiota tarkasteltiin kielenkäyttötilanteista (tehtävät) ja aihepiireistä käsin (luku 8).

Hughes (1989) ehdottaa myös testattavaksi alemman tason kuuntelutaitoja kuten vokaalien ja konsonanttien erottamista ja lausepainon aiheuttamia merkityseroja ainakin diagnostisessa tarkoituksessa. Hughes (1989) pitää myös tärkeänä tekstien määrittelemistä mahdollisimman tarkasti tekstityypeittäin (monologi, dialogi, useita osallistujia; keskustelu, kuulutus, puhe tai luento, ohjeiden tai suunnan antaminen jne.), tekstimuodoittain (kuvailu, selitys, argumentointi, ohje, kertonta) sekä pituuden suhteen (minuutit ja sekunnit)³. Näin tehtiin tässä tutkimuksessa arvioitaessa ymmärtämisen tehtäviä Grid -taulukoiden avulla (luku 8).

Tekstin ymmärtäminen tuotoksena

Tekstin, samoin kuin puheenkin ymmärtämisen arvioinnissa, yhdeksi epävarmuustekijäksi nähdään se, että empiirisesti ei ole voitu todeta, mitä taitoja ymmärtämiseen hallintaan sisältyy. Epätietoisuudesta huolimatta on arvioinnin validiuden kannalta syytä testata niin monia taitoja kuin mahdollista (Hughes 1989:136-137). Näin meneteltiin myös tämän tutkimuksen kohteena olevassa arvioinnissa. Huolimatta siitä, että on joukko tekstin ymmärtämisen teoreetikkoja ja teorioita, jotka eivät hyväksy ymmärtämisen jakamista taitoihin, kehitetyt taksonomiat antavat käytännössä hyvää pohjaa arvioinnille. Davis (1965. Siteerattu Alderson 2000:9-10) erottaa kahdeksan ymmärtämisen alataittoa, Munby (1978. Siteerattu Alderson 2000:10-11) kahdeksantoista. Jälkimmäinen, mikrotaitojen taksonomia, on ollut hyvin vaikutusvoimainen sekä oppimateriaalien että kokeiden laadinnassa (Alderson 2000:9). Taulukossa 17 esitetään molempien ehdotukset tekstin ymmärtämistaidoiksi.

TAULUKKO 17 Tekstin ymmärtämisen taidot. (Muokattu taulukoksi Aldersonin 2000:9-10 mukaan.)

Davis (1968)	Munby (1978)
- sanojen merkityksen muistaminen	- kielen skriptin tunnistaminen
- sanapäätely tekemällä johtopäätökset kontekstista	- vieraiden sanojen ja niiden käytön merkityksen päätely
- kysymyksiin vastaaminen silloin, kun vastaukset ovat ilmeisiä tai parafraasin muodossa	- selvästi ilmaistun informaation ymmärtäminen
- ajatusten yhdistäminen	- selvästi ilmaisemattoman informaation ymmärtäminen
- johtopäätösten tekeminen sisällöstä	- käsitteellisen merkityksen ymmärtäminen
- kirjoittajan tarkoituksen, asenteen, sävyn ja mielialan tunnistaminen	- lauseiden kommunikatiivisen arvon ymmärtäminen
- kirjoittajan tekniikan identifiointi	- lauseen sisäisten suhteiden ymmärtäminen
- tekstin rakenteen seuraaminen	- tekstin osien välisten suhteiden ymmärtäminen käyttämällä hyväksi sanastollisia koheesiokeinoja

TAULUKKO 17 jatkuu

³ Grideissä (ks. s. 158) on lähdetty tästä periaatteesta.

TAULUKKO 17 jatkuu

	- tekstin osien välisten suhteiden ymmärtäminen käyttämällä hyväksi kieliopillisia koheesiokeinoja
	- tekstin tulkitseminen tarkkailemalla sitä ulkopuolelta
	- diskurssi-indikaattorien tunnistaminen
	- diskurssin sisältämän informaation pääkohtien havaitseminen
	- pääajatuksen erottaminen sitä tukevista yksityiskohdista
	- tekstin tai ajatuksen tiivistäminen sisällyttämällä siihen tarpeelliset yksityiskohdat
	- tarpeellisten kohtien valitseminen tekstistä
	- perusviittaustaitojen käyttäminen
	- silmäilyluku
	- kysytyn tiedon paikallistaminen ja diagrammien sisältämän tiedon muuntaminen "luettavaksi"

Weirin (1993) mukaan lukemisessa on kolme operaatiota:

- a) silmäily: tekstin nopea lukeminen
- b) lukeminen huolellisesti pääasioiden ja tärkeiden yksityiskohtien ymmärtämiseksi
- c) spesifimpien lingvististen aputaitojen tunteminen: kieliopilliset käsitteet (syy, tulos, tarkoitus, vertailu), virkkeiden ja lauseiden syntaktinen rakenne, diskurssi-indikaattorit (*discourse markers*), sanastollinen ja / tai kieliopillinen koheesio ja sanasto (Weir 1993:73).

Weir (1993) katsoo kuitenkin viimeisen operaation sisältyvän myös kahteen ensimmäiseen, mutta myöntää, että ei tiedetä, missä määrin ne ovat tarpeellisia tai korvattavissa, ja että sitä olisi luultavasti vaikea määrittää. Tekstin ymmärtämisen tehtävien luokitteluun sovellettiin tässä tutkimuksessa Weirin (1993) jakoa operaatioihin (luku 8).

Hughes (1989) kutsuu lukijan lukiessaan käyttämiä taitoja operaatioiksi, ja on koonnut kokeenlaatijoiden avuksi tekstin ymmärtämisen tarkistuslistat, joita hän kuitenkin varottaa pitämästä kattavina. Hughes (1989:138) jakaa taidot nopean ja tarkan lukemisen sekä johtopäätösten teon toiminnoiksi taulukossa 18 esitetyllä tavalla.

TAULUKKO 18 Lukemisen toiminnot. (Lähde: Hughes 1989:138-139.)

Toiminto	Testattavalta edellytettävät taidot
Nopea lukeminen	
Silmäily	Testattava osaa <ul style="list-style-type: none"> - löytää pääajatuksat ja diskurssiaiheen nopeasti ja tehokkaasti - havaita nopeasti tekstin rakenteen - päättää tekstin tai sen osan soveltuvuuden omiin tarpeisiinsa
Haku	Testattava löytää nopeasti tietoa ennalta tiedossa olevasta aiheesta
Selailu	Testattava löytää nopeasti <ul style="list-style-type: none"> - tietyt sanat tai fraasit - kuviot, prosenttiluvut, - luettelon tietyt kohdat, - lähteiden tai viitteiden olennaiset sanat
Tarkka lukeminen	Testattava osaa <ul style="list-style-type: none"> - tunnistaa pronominiin viittaukset - tunnistaa diskurssimerkit - tulkita komplekseja lauseita - tulkita topiikkilauseita - hahmottaa tekstin loogisen jäsentelyn - seurata argumentin kehittelyä - erottaa väittämät esimerkeistä - yksilöidä selkeästi ilmaistut pääajatuksat - yksilöidä implisiittisesti ilmaistut pääajatuksat - tunnistaa kirjoittajan aikomuksen - tunnistaa kirjoittajan asenteet ja tunteet - tunnistaa tekstin aiotun vastaanottajan - tunnistaa, millaisesta tekstistä on kyse (pääkirjoitus, päiväkirja ym.) - erottaa faktat mielipiteestä - erottaa olettamukset faktoista - erottaa faktat huhuista
Johtopäätösten tekeminen	Testattava osaa <ul style="list-style-type: none"> - päätellä kontekstista tuntemattoman sanan merkityksen - tehdä päätelmiä tekstin sisällöstä pystyäkseen vastaamaan sanoilla Kuka? Milloin? Mitä? alkaviin kysymyksiin - tehdä päätelmiä tekstin sisällöstä pystyäkseen vastaamaan motivaatiota, syytä, seurausta ja kykyä koskeviin sanoilla Miksi? ja Kuinka? alkaviin kysymyksiin - tekemään pragmaattisia (tekstiin sisältymättömän aineksen pohjalta tehtäviä) päätelmiä.

Kuten edellä mainitut tutkijat, myös Eurooppalainen viitekehys (2003) toteaa lukijoilla olevan erilaisia tavoitteita lukiessaan. Näitä lukemisen tarkoituksia ovat esim.

- keskeisen sisällön ymmärtäminen (yleiskäsitys),
- valikoiva ymmärtäminen (tietyn informaation selville saaminen),
- yksityiskohtainen ymmärtäminen, ja
- ymmärtäminen johtopäätösten tekemistä varten. (Eurooppalainen viitekehys 2003:105)

Cambridgen yliopiston kommunikatiivisen kielitaidon testi CCSE⁴ näkee kielitaidon hierarkkisesti rakentuvana yhtenäisenä kokonaisuutena. Samaa tekstiä käytetään eri tasoille opiskelijoille erilaisin tehtävin. Huomiota on kiinnitetty siihen, millaisia tekstejä todellisessa elämässä luettaisiin (Alderson 2000:296-297). CCSE:n käsikirjassa erotetaan neljä taitotasoa sen mukaan, minkä asteista kielitaitoa lukemistehtävä edellyttää. Taidon asteet luokitellaan seuraavasti:

- kompleksisuus	tekstistä saatavan tiedon kompleksisuus
- laajuus	kuinka monia tekstin aiheita käsitellään
- nopeus	miten nopeasti tekstiä käsitellään
- joustavuus	lukemisen tyylin sopeuttaminen tehtävään
- riippumattomuus	riippumattomuus annetuista ohjeista ja sanakirjan käytöstä

Näistä kompleksisuus, laajuus ja riippumattomuus näyttäisivät Aldersonin (2000) mielestä liittyvän tekstin kielellisiin piirteisiin, informaatioon, jonka kieli välittää ja käyttäjän kasvaneeseen kielitaitoon. Joustavuuden hän ei kuitenkaan katso liittyvän kielitaitoon. Toisin kuin muut testit, CCSE lähtee siitä, että minkä tasoiset lukijat tahansa voivat käsitellä monia tekstityyppejä. Edistymistä ei näin ollen voikaan määritellä sen perusteella, millaista tekstiä kokeen suorittaja kykenee prosessoimaan, vaan nimenomaan siitä, millaista tietoa hän pystyy tekstistä hankkimaan (Alderson 2000:300-301). Yleensä on oletettu, että oppimisen alkuvaiheessa olevat opiskelijat tarvitsevat helpompia tekstejä, ja siksi ns. portaittain vaikeutuvat tekstit ovat olleet suosittuja. CCSE on kuitenkin osoittanut, että heikotkin lukijat voivat ymmärtää jotakin vaikeista teksteistä. Tämä johtopäätös sotii esim. ACTFL:ssa käytettyjä portaittain vaikeutuvien tekstien käyttöä vastaan (Alderson 2000:279-280). Toisaalta tätä ei voi pitää yksiselitteisesti selviönä, koska lukijanhan tulee tuntea huomattavan suuri määrä tekstin sisällöstä ymmärtääkseen tekstin ja toisaalta nuorille lapsille ei voi tarjota luettavaksi heidän kognitiivisen tasonsa ylittäviä tekstejä.

4.2.2 Tuottaminen

Huolimatta siitä, että kielen ymmärtämistä ja tuottamista käsitellään usein eräänlaisina toistensa vastakohtina, niillä on kuitenkin myös paljon yhteistä. Lukijan tekstin ymmärtämistä helpottaa kirjoittamisen käytänteiden tunteminen. Kirjoittamista onkin kuvattu viestintätapahtumaksi lukijan ja kirjoittajan välillä (Vähäpassi 1986/2004, 1987a). Suullisessa vuorovaikutustilanteessa taas sekä puhuminen että puheen ymmärtäminen ovat usein yhtä keskeisessä asemassa. Esimerkiksi Hughes (1989), käsitellessään kielitaidon osa-alueiden arviointia, katsoo, että suullisen kielitaidon opettamisen tavoitteena on onnistuneen vuorovaikutuksen kehittäminen ja että siihen kuuluvat sekä ymmärtäminen että tuottaminen (Hughes 1989:113).

Clark (1996) korostaa kielen vuorovaikutuksellista luonnetta ja on esittänyt seuraavat kuusi väitettä kielen käytöstä. Vaikka ne liittyvät puhumiseen, ne soveltuvat myös kirjoittamisen kuvaamiseen.

⁴ Certificate in Communicative Skills in English.

1. Kieltä perimmäisessä merkityksessään käytetään sosiaalisiin tarkoituksiin.
2. Kieli on yhteistoiminnan laji.
3. Kieleen liittyy aina puhujan tarkoitus ja puhuteltavan tarve tulla ymmärretyksi.
4. Kielen käytön perustilanne on kasvokkain tapahtuva keskustelu.
5. Kielen käyttö koostuu monesti useammasta kuin yhdestä toiminnan kerroksesta.
6. Kielen käyttö on sekä kognitiivista että sosiaalista.

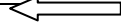





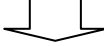
Vaikka puhetta ja kirjoitusta käytetään yleensä erilaisissa yhteyksissä, erilaisista syistä ja eri kommunikatiivisiin tarkoituksiin, puhuttu ja kirjoitettu diskurssi perustuvat monille yhteisille lingvistisille resursseille ja niitä voidaan käyttää monissa tapauksissa kohtaamaan samat lingvistiset tavoitteet. Puhutun ja kirjoitetun kielen väliset erot väliset erot liittyvät olennaisesti viestintätilanteeseen ja tuottamisprosessiin (Takala 1982b, Tiittula 1993:64). Taulukossa 19 puhumista ja kirjoittamista on vertailtu Bygaten (1987), Tiittulan (1993), Brownin (2001), Weiglen (2002) ja Luoman (2004) esittämien näkökantojen pohjalta.

TAULUKKO 19 Suullisen ja kirjallisen tuottamisen vertailua.

Puhuminen	Kirjoittaminen
Ajatusyksiköt ovat lyhyempiä	Ajatusyksiköt ovat tiiviimpiä
Lauseet liittyvät toisiinsa rinnasteisten konjunktioiden, esim. <i>and, or, but</i> avulla	Runsaammin sivu- ja alisteisia lauseita tiiviimmän ja runsaamman tiedon välittämiseksi.
Epäröintitaukoja, täytesanoja, toistoa	Selkeät virkkeen ja lauseen aloitukset ja lopetukset.
Käytetään puhekieltä, murretta, slangia. Alempi muodollisuuden taso.	Käytetään muodollista ja korrektaa, ehkä konservatiivistakin standardikieltä. Yleensä korkeampi muodollisuuden taso.
Sanasto yksinkertaista; täsmälliset ilmaukset saatetaan korvata ilmauksilla kuten <i>this one, the round one, put jne.</i>	Sanasto täsmällisempää, eri rekisteriä edustavaa.
Persoonallisuutta, tunnetta, epätarkkuutta, yleistyksiä.	Persoonatonta, tarkkaa.
Prosessoidaan reaaliajassa, katoavaa	Voidaan lukea yhä uudelleen, pysyvää.
Lyhyt aika sanottavan suunnittelemiseen, muotoilemiseen ja tuottamiseen.	(Ehkä paljonkin) pitempi suunnittelu-aika.
Reaaliaikainen esitys antaa mahdollisuuden väärinkäsitysten oikaisemiseen.	Etäisyys kirjoittajan ja lukijan välillä edellyttää tuotokselta sidosteista ja jäsentynyttä ilmaisua.
Vähäisempi redundanssi.	Runsaampi redundanssi.

Puhuminen prosessina

Bygate (1987) näkee puhumisen puhujan liittyvänä prosessina, jossa on kolme vaihetta: suunnittelu, valinta ja tuotos (kuvio 13). Näiden pääotsikoiden alle hän sijoittaa tieto- ja strategiakomponentit. Hän tekee myös eron käsitteiden tieto (*knowledge*) ja taito (*skill*) välillä siten, että tieto tekee puheen mahdolliseksi, kun taas taito on se aktiivinen komponentti, jota tarvitaan silloin, kun puhuja on aktiivisesti mukana vuorovaikutuksessa (ks. myös Luoma 2004:104).

Tieto		Taito
Suunnittelu Rutiinien tuntemus: - informatiivinen - interaktiivinen Diskurssin tilan tuntemus	 	Viestin suunnittelu: - informaation suunnittelu - vuorovaikutuksen suunnittelu Hallintataidot: - keskustelun kulku - vuorojen vaihto
Valinta - sanasto - fraasit - kielioppi	 	Merkityksestä neuvottelemine - selkeyttämistaidot - proseduraaliset taidot
Tuotos - tuottamiskeinot		Tuottamistaidot: - helpottaminen - kompensointi
Kieliopilliset säännöt Ääntämissäännöt		Tarkkuuden hallinta
		Ilmaisu 

KUVIO 13 Puheen prosessointi. (Lähde: Bygate 1987:50.)

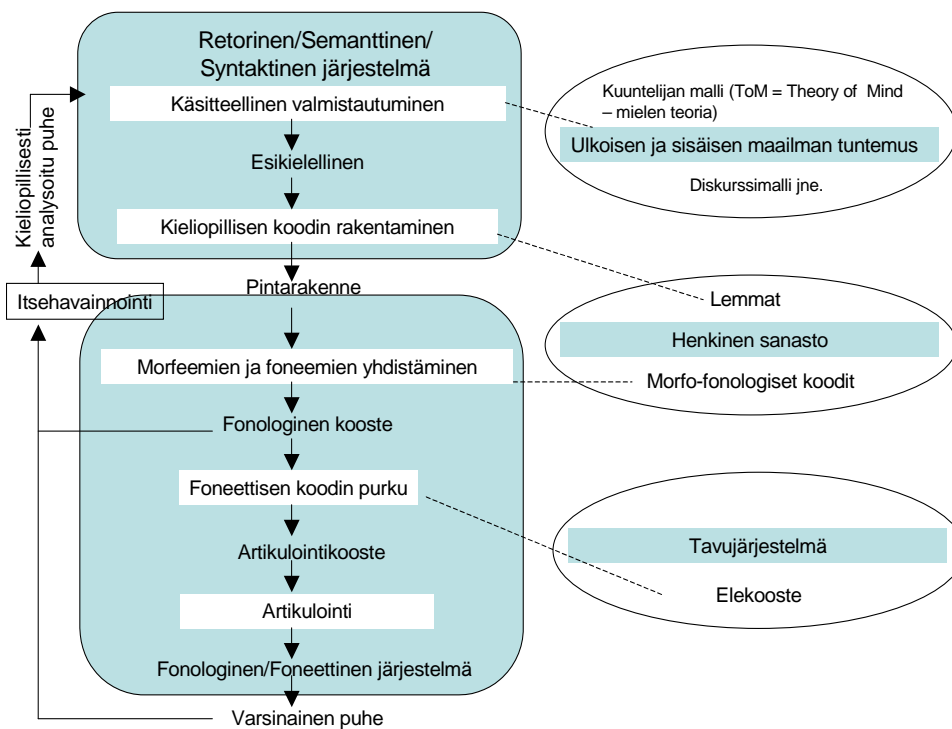
Bygaten (1987) kuviossa on kaksi erityispiirrettä. Ensinnäkin taidot ovat riippuvaisia jostakin tilanteeseen sopivasta tietovarastosta ja niihin liittyy tiettyjen tarkoitusten ilmaisemisen hallinta sovittuja konventioita käyttäen. Taitoja käyttämällä muodostuu samalla rutiineja, jotka varastoituvat tiedoksi. Toinen kuvion piirre on se, että taidot ovat myös riippuvaisia toisistaan. Niihin liittyy samanaikaisia päätöksentekoprosesseja tehtäessä päätöksiä viestin välittämisestä, sen muotoilusta ja tuottamisesta. Suulliseen tuottamiseen liittyy toisin sanoen kyky käsitellä kaikkia osataitoja kaavion alusta loppuun (Bygate 1987:49).

Bygate (1987) lähtee siitä, että keskustelijoiden tulee tuntea informaatio- ja vuorovaikutusrutiinit ja rakentaa jatkuvasti mielessään kuvaa käynnissä olevasta keskustelusta, jotta suunnittelu olisi mahdollista vuorovaikutteisessa puhetilanteessa. Informaatiiorutiinit ovat usein toistuvia informaatiiorakenteita kuten tarinoita, kuvauksia tai vertailuja. Vuorovaikutusrutiinit ovat tyypillisiä vuoroteturakenteita, joiden puhujat tietävät kuuluvan tietyn tyyppisiin tilanteisiin, kuten palveluun, puhelinkeskusteluihin tai oppitunteihin. Taidot, joita keskustelijat tarvitsevat käyttääkseen tätä tietoa, ovat viestin suunnittelutaidot, joissa rutiinien taustalla oleva tieto mahdollistaa sen, että keskustelijat pystyvät ennustamaan, miten keskustelu etenee. Näin he voivat suunnitella ennalta oman osuutensa ja vuorovaikutuksen hallintataidot, joihin sisältyvät aiheen etenemisen ja puheenvuorojen vaihdon hallinta. Valintavaiheessa oppijat käyttävät tietojansa sanastosta, fraaseista ja kieliopista valitakseen, miten sanovat sen, mitä haluavat sanoa. Nämä taidot liittyvät merkityksestä neuvottelemiseen. Taito selkeyttää sanottavansa auttaa puhujaa valitsemaan ilmauksensa sen valossa,

mitä hän arvioi kuulijan tietävän, kun taas proseduraaliset taidot auttavat häntä varmistamaan sen, että ymmärtämistä on tapahtunut, esimerkiksi käyttäen korostusta, toistoa, tarkennuspyyntöjä jne. Suunnittelu- ja valintatoimintoja voidaan kutsua vuorovaikutustaidoiksi, koska ne vaikuttavat siihen, kuinka puhujat suhtautuvat toisiinsa keskustelussa (Luoma 2004:104).

Tuottamistoiminnot liittyvät suulliselle tuottamiselle tyypilliseen aikaan sitoutumiseen. Tässä tarvittavat tiedot ovat artikulointi ja puhujan tieto kielioppi- ja ääntämissäännöistä. Tuottamiseen liittyvät taidot ovat helpottaminen (*facilitation*) ja korvaaminen (*compensation*). Puhujat voivat helpottaa puheensa tuottamista yksinkertaistamalla rakennetta, käyttämällä ellipsiä, sanontoja ja idiomeja, täyte- ja epäröintikeinoja. Kompensaatiotaidot muistuttavat helpotustaitoja, mutta niitä käytetään ymmärtämisen selkeyttämiseksi. Kompensaatiotaitoihin kuuluvat joustava sanontojen, korjaamisen, uudelleen muotoilun, laajentamisen tai vähentämisen avulla toteutetun toiston ja epäröinnin käyttö (Luoma 2004:105)

Leveltin (1989) urauurtava kuvaus puheen prosessoinnista pohjautuu käsitykseen ihmisen puheen tuottamisen keinoista (kuvio 14). Semanttisyntaktinen järjestelmä liittyy ilmaistavaksi aiotun käsitteellistämisen leksikon yksiköiden (*lemma*) lineaariseen pintarakenteeseen. Fonologis-foneettisen järjestelmän tehtävä on valmistella artikulointimalli, jonka puhekumppani pystyy tulkitsemaan tuon pintarakenteen ja sen takana olevan käsitteellistyneen ilmaukseksi. Tästä Levelt (1999) käyttää nimitystä *articulatory score* (artikulointikooste) (Levelt 1999:86).



KUVIO 14 Puhujan prosessointikaavio. (Lähde: Levelt 1999:87)

Kuviota tulkitaan ylhäältä käsin. Käsitteellisessä valmistautumisessa puhuja luo viestin, joka on edellä esitetyn kaltainen käsitteellinen rakenne. Tässä käytetään hyväksi sosiaalista kompetenssia ottaen huomioon puhekumppanin kanssa yhteinen taustatieto (ToM = Theory of Mind = mielen teoria), joka auttaa rakentamaan komplekseja tietorakenteita sosiaalisesta ympäristöstämme. Tällöin kiinnitetään puhekumppanin huomio siihen, mikä kulloinkin on uutta tai relevanttia. Tähän päästään kokeilemalla erilaisia lähestymistapoja tietolähteisiin, jotka on kuviossa ilmaistu ellipseinä. Viesti on viime kädessä käsiterakenne, joka koostuu sanastollisista käsitteistä, joita kielenä edustavat sanat. Kieliopillisen koodin rakentamisessa viestin sanastolliset käsitteet aktivoivat vastaavat syntaktiset sanat (*lemma*) mielen sanastossa. Puhuja siis käyttää sanastollis-syntaktista informaatiota rakentaakseen sopivan syntaktisen mallin, pintarakenteen, ja ensimmäisen järjestelmän prosessointi päättyy tähän (Levelt 1999:27-8).

Morfeemien ja foneemien yhdistäminen tarkoittaa sitä, että heti kun lemma valitaan, sen muodon koodi aktivoituu. Tällöin puhuja pääsee käsiksi sen morfologiseen ja fonologiseen rakenteeseen, joka on perusainesta fonologisten sanojen rakentamiseen. Foneettisen koodin purku sisältää ajatuksen siitä, että fonologinen tulos laukaisee artikuloitikoosteen, joka johtaa artikulointiin. Kun tämä yhdistyy puhe-elinten toimintaan, päädytään lopputuotteeseen, varsinaiseen puheeseen. Itsehavainnoinnilla tarkoitetaan sitä, että samoin kuin kuunnellessamme tarkkailemme puhuessakin omaa sisäistä ja ulkoista puhettamme ja jos tuotettavassa puheessa syntyy ongelma, puhe voidaan keskeyttää ja korjata sitä (Levelt 1999:86-8).

Puhuminen tuotoksena

Eräät tutkijat erottavat neljäntyyppistä tiedon hankintaan tai välittämiseen liittyvää puhetta: kuvailu, ohjeiden antaminen, tarinan kerronta ja mielipiteen ilmaisu/perustelu. Jaon taustalla on näkemys siitä, että kussakin näistä käytetään omanlaisiaan rutiineja järjestettäessä informaatiota helpottamaan ymmärtämistä (Luoma 2004:32). Samaa perustelua käyttäen Bygate (1987) päätyi hienojakoisempaan jaotteluun (taulukko 20). Puhujan kielenkäyttö on erilaista kussakin näistä kategorioista. Jos joku on hyvä kuvailemaan, ei hän välttämättä ole taitava myös vertailemaan asioita, kertomaan tarinaa tai perustelevaan mielipiteitään.

TAULUKKO 20 Informaation järjestämISRutiinit. (Lähde: Bygate 1987:24.)

Selittävät rutiinit (<i>Expository routines</i>)	Arvioivat / Tulkitsevat rutiinit (<i>Evaluative routines</i>)
Kuvailu	Selittäminen
Kerronta	Perusteleva
Ohjeiden antaminen	Ennustaminen
Vertaileminen	Päätöksen tekeminen
	Etusijan antaminen

Toinen lähestymistapa suullisen kielitaidon arviointiin on tarkkailla, mitä testattavat sanovat heille annetuissa tilanteissa (Luoma 2004:33). Eurooppalainen

viitekehys (2003) jakaa nämä funktionaaliset kompetenssit kahteen kategoriaan, *makro- ja mikrofunktioihin* (taulukko 21). Edelliset ovat vähintään muutaman virkkeen pituisia esityksiä, kun taas jälkimmäiset koostuvat lyhyistä ilmauksista, jotka usein liittyvät vain puheenvuoron vaihtamiseen (Eurooppalainen viitekehys 2003:177).

TAULUKKO 21 Makro- ja mikrofunktiot. (Lähde: Eurooppalainen viitekehys 2003:177.)

Makrofunktiot	Mikrofunktiot
- kuvaus	- asiatietojen ilmaiseminen ja pyytäminen
- kertomus	- asenteiden ilmaiseminen ja tunnistaminen
- kommentointi (huomautukset ja selitykset)	- suostuttelu
- erittelevä asiateksti	- seurustelu
- tulkinta	- diskurssin jäsentäminen
- selitys	- kommunikaation korjaukset
- havainnollistus	
- opastaminen	
- argumentointi (argumentoiva, pohtiva teksti)	
- suostuttelu	

Eurooppalainen viitekehys (2003) sisällyttää pragmaattisiin taitoihin diskurssi- ja funktionaaliset kompetenssit, joista suoriutumisen onnistuneisuus arvioidaan kuusiportaisella taitotasosteikolla (A1-C2). Diskurssikompetenssia arvioitaessa huomiota kiinnitetään seuraaviin tuotoksen ominaisuuksiin: olosuhteiden mukainen jousto, puheenvuoron vaihtaminen, teeman kehittäminen (kertomuksessa), koherenssi ja koheesio (Eurooppalainen viitekehys 2003:173-6). Testattavan onnistumisen perusteet funktionaalisia taitoja arvioitaessa ovat sujuvuus (ks. kuitenkin luku 6) ja ilmaisun tarkkuus (emt. s. 180:28).

Fulcherin mukaan puhumisen konstruktin (= mitattava, arvioitava piirre) voidaan kuvailla taulukon 22 mukaisesta viitekehuksesta käsin (vrt. taulukko 13, s. 68). Sijoittamalla kokeeseen arvioinnin kannalta tarkoituksenmukaisia piirteitä testaavia komponentteja testattavan tuotosta voidaan ohjata haluttuun suuntaan.

TAULUKKO 22 Puhumisen konstruktin määrittäminen. (Lähde: Fulcher 2003:48.)

Kielellinen kompetenssi
Fonologia
- ääntäminen
- painotus
- intonaatio
Tarkkuus
- syntaksi
- sanasto
- sidosteisuus
Sujuvuus
- epäröinti
- toisto
- sopivamman sanan valitseminen
- sopivamman rakenteen valitseminen
- koheesio

TAULUKKO 22 jatkuu

TAULUKKO 22 jatkuu
Strateginen kapasiteetti Saavutusstrategiat - liiallinen yleistäminen - parafraasi - sanan keksiminen - uudelleen muotoilu - koodin vaihto - ei-kielelliset strategiat Välttämistästrategiat - muotojen välttäminen - funktioiden välttäminen
Tekstuaalinen tieto Puheen rakenne - puheen vuorottelu - vierusparit (<i>adjacency pairs</i>) - aloitukset/lopetukset
Pragmaattinen tieto - soveltuvuus - implikaatiot - itsensä ilmaiseminen
Sosiolingvistinen tieto - tilanteeseen liittyvä - aiheeseen liittyvä - kulttuuriin liittyvä

Kirjoittaminen prosessina

Kirjallisen tuottamisen edellytyksiä ovat viestin/tekstin luominen ja sen muotoileminen (kognitiiviset ja kielelliset taidot) sekä kirjoittaminen (manuaaliset taidot) (Eurooppalainen viitekehys 2003:131). Vieraalla kielellä kirjoittamiseen tarvitaan ainakin jonkinlaista tietoa kieliopista ja sanastosta. Näiden taso määrittää kirjoittamistehtävästä suoriutumisen. On todettu, että vieraan kielen taito määriteltynä toisen kielen lingvististen elementtien hallintana, ja taito kirjoittaa, ovat eri taitoja. Kuitenkin tuottaessaan vierasta kieltä kirjoittajat soveltavat monia samoja kirjoittamisprosesseja kuin äidinkielessäänkin (Weigle 2002:23). Näin ollen on perusteltua esittää eräitä äidinkielen kirjoittamisen prosessointimalleja. Vaikka kognitiivisen prosessin kuvaus jääkin monilta osin puutteelliseksi, sen analysointi palvelee testattavien taitojen määrittämistä.

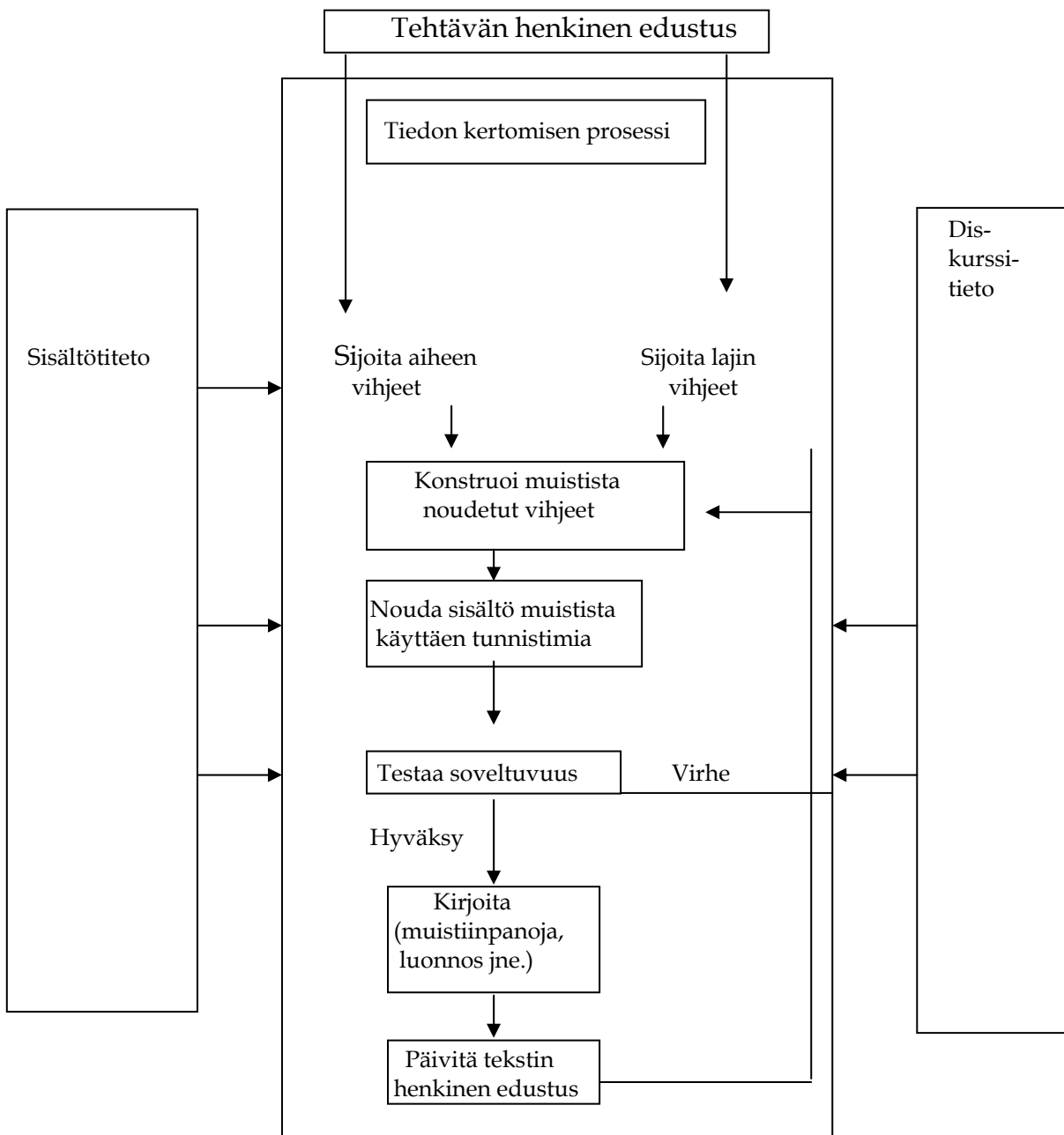
Kun puhutaan kommunikatiivisesta kirjoittamisesta, kirjoittamistilanne koostuu joukosta parametrejä, jotka vaikuttavat kirjoittamisprosessiin. Näitä ovat kirjoittaja – vastaanottaja, palaute, tarkoitus, sisältö, tehtävän laajuus ja monimuotoisuus sekä tehtävän tuttuus. Kehittyneen kirjoittamisprosessin oletetaan koostuvan kolmesta pääprosessista: suunnittelusta, suunnitelmien muuntamisesta vastaanottokelpoiseksi, luettavaksi kieleksi sekä muokkaamisesta (Takala 1982b:229). Takala (1982b) referoi myös Bereiteriä, jonka käsityksen mukaan kirjoittamisen taidot integroituvat hierarkkisesti seuraavasti:

1. Assosiativinen kirjoittaminen (muistuttaa transkriboitua puhetta, tajunnan virtaa, on ekspressiivistä).
2. Performatiivinen (edellinen taso kytkeytyy tähän siten, että nyt osataan ja havaitaan tyypillisiä konventioita).

3. Kommunikatiivinen (performatiivinen taito yhdistyy tässä sosiaaliseen kognitioon, jossa kirjoittaja jo pyrkii vaikuttamaan lukijaan).
4. Integroiva (ottaa huomioon kirjoittajan oman tuotoksensa lukijana, joka muokkaa sitä kunnes on tyytyväinen tuotteeseensa ja toivoo sen tyydyttävän muitakin.)
5. Episteeminen (jossa reflektointi yhdistyy edelliseen, jolloin tuotos ei ole vain ajatellun tuote, vaan olennainen osa ajattelua) (Takala 1982b:226).

Vieraalla kielellä kirjoittaminen ei kuitenkaan aina ole ”kehittynyttä”. Bereiter ja Scardamalia (1987. Siteerattu Weigle 2002) ovat kirjoittamisen prosessimallissaan lähteneetkin siitä, että kirjoittamista on kahdenlaista. Toinen on rupattelun luonteista kirjoittamista, jota Bereiter ja Scardamalia kutsuvat tiedon kertomiseksi (*knowledge telling*). Toinen on tiedon muuntamista (*knowledge transformation*), joka edellyttää selvästi suurempaa ponnistelua, runsaasti harjoitusta ja hyvää kielitaitoa sekä tietoa ja joukkoa taitoja. Tiedon kertomiseen liittyy tuskin lainkaan suunnittelua tai korjailua (Weigle 2002:31) ja siihen pystyy suhteellisen vähäiselläkin kielitaidolla. Tottumattoman kirjoittajan suurin ongelma on viestin tai tekstin vastaanottajan välinen fyysinen etäisyys. Lukijalla ei ole mahdollisuutta pyytää toistoa tai selitystä kuten puhetilanteessa, ja näin ollen kirjoittajan epäonnistuminen jossakin tai joissakin kerrontaprosessin kohdissa saattaa jättää viestin vajavaiseksi tai johtaa väärinkäsitykseen.

Kuviossa 15 esitetään tiedon kertomisprosessi. Sen mukaan kirjoittaja käyttää kirjoittamistehtävän henkistä edustusta noutaakseen sisältö- eli aihetietoa ja tehtävän vaatiman diskurssityypin skeeman, joka voi olla esimerkiksi mielipidekirjoitus tai kuvaus. Tämän jälkeen arvioidaan näiden vihjeiden avulla saatujen ehdotusten soveltuvuus tekstiin. Jos ne hyväksytään, ne kirjoitetaan muistiin. Prosessi toistetaan niin moneen kertaan kuin muistissa riittää soveltuvaa lisättävää sisältöä.

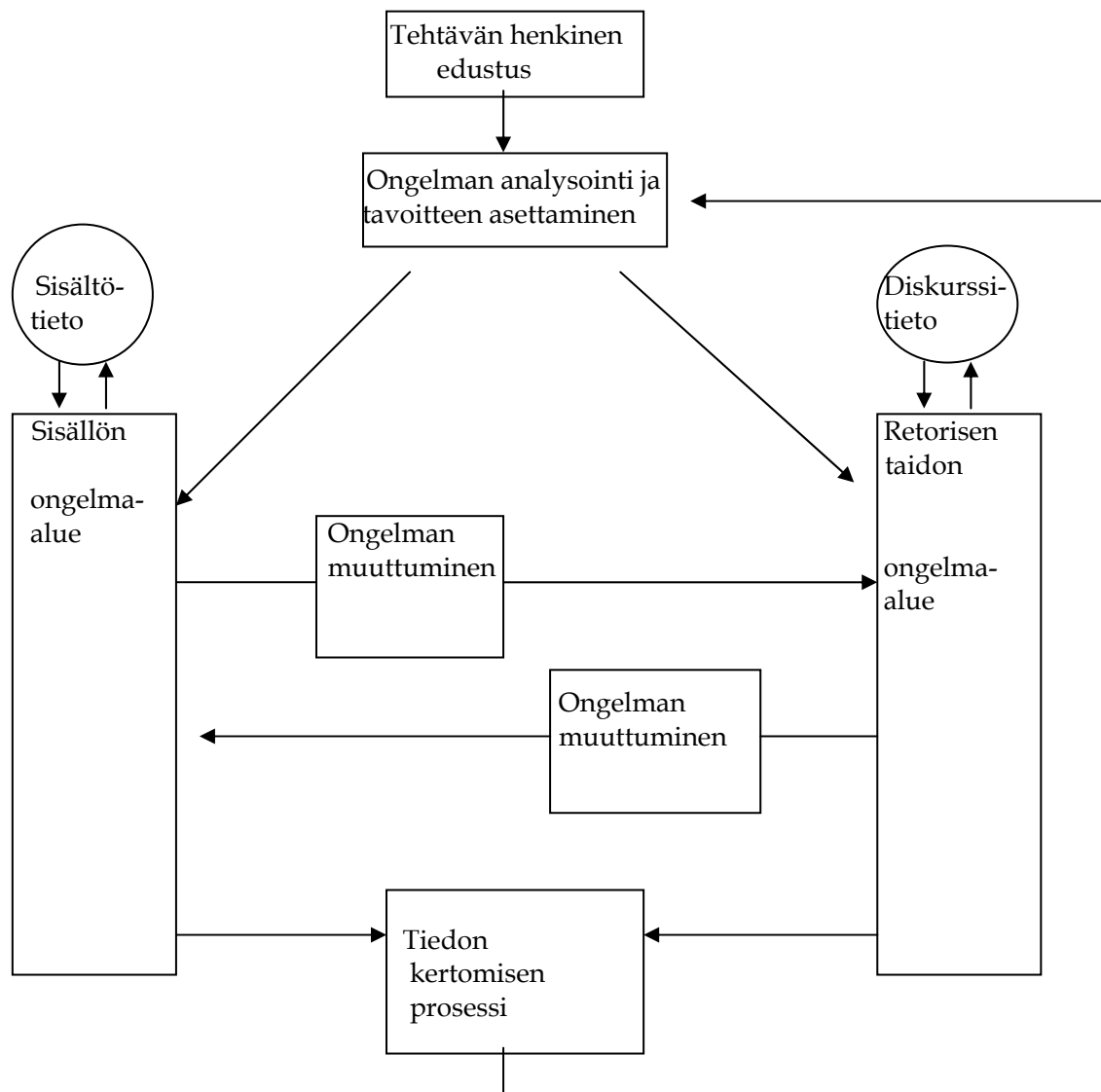


KUVIO 15 Tiedon kertomismallin rakenne. (Lähde: Bereiter & Scardamalia 1987. Siteerattu Weigle 2002:33.)

Kehittyynyttä kirjoitustaitoa kuvaavat Bereiterin (1980. Siteerattu Takala 1982b: 226) mukaan seuraavat kuusi tietojärjestelmää tai taitoa:

1. Kirjoitetun kielen tuottamisen sujuvuus.
2. Ajatusten tuottamisen sujuvuus.
3. Kirjoittamisen konventioiden hallinta.
4. Sosiaalinen kognitio, joka ilmenee taitona ottaa lukija huomioon.
5. Kirjallisten piirteiden tunnistaminen ja arvostaminen
6. Reflektointi.

Mitä korkeammalla tasolla nämä taidot ovat, sitä paremmin kirjoittaja onnistuu tiedon muuntamisessa, mikä edellyttää kirjoittamisen käyttämistä luomaan kirjoittamisprosessin kautta uutta tietoa (Weigle 2002:32-33). Kuviossa 16 esitetään Bereiterin ja Scardamalian (1987) tiedon muuntamisen malli.



KUVIO 16 Tiedon muuntamismallin rakenne. (Lähde: Bereiter & Scardamalia. Siteerattu Weigle 2002:34.)

Tässä mallissa ensimmäinen vaihe on ongelman analysointi ja tavoitteen asettaminen, minkä jälkeen siirrytään sekä sisältöä että esittämistapaa koskevien ongelmien ratkaisemiseen. Ratkaisun löytyminen sisältöalueella saattaa johtaa uuden ongelman syntymiseen toisella ongelma-alueella. Kaavion laatijat ovatkin kuvanneet prosessia kaksisuuntaiseksi vuorovaikutukseksi tietosisällön ja tekstin muodon välillä. Näiden kahden ongelman ratkaisut muodostavat tiedon kertomisen syötöksen, jolloin varsinainen kirjoittaminen tapahtuu (Weigle 2002:34).

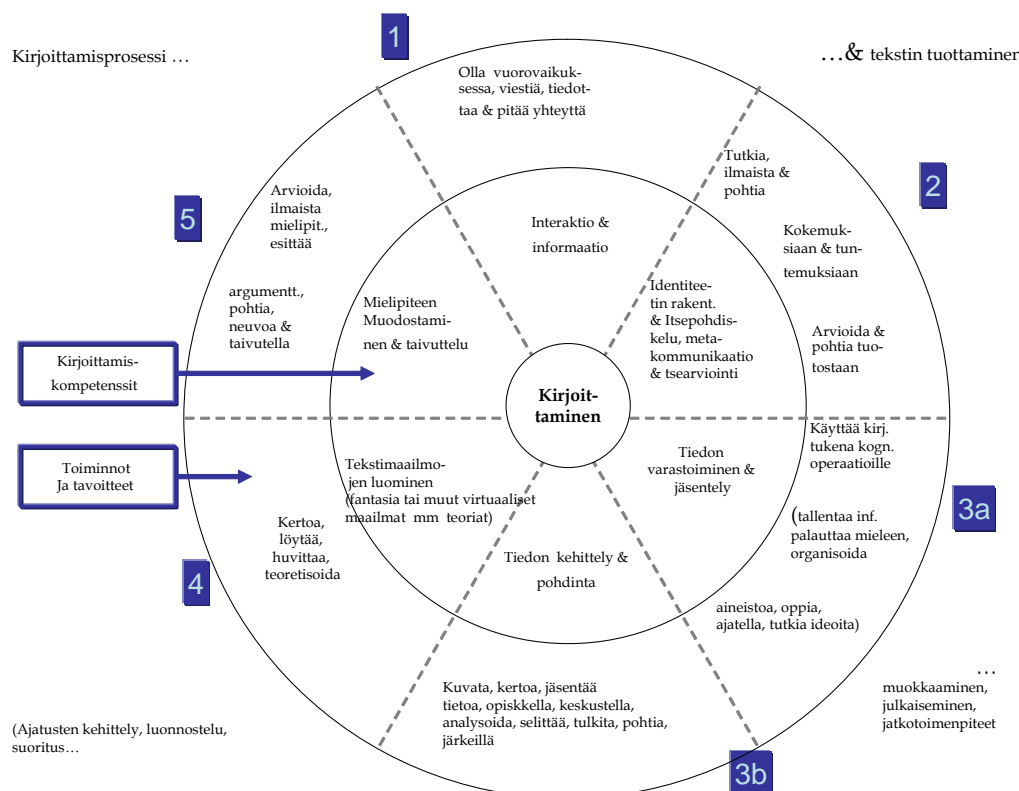
Kognitiivinen prosessointi		Toistaminen	Jäsentäminen/Uudelleen jäsentäminen		Keksiminen/Kehittäminen
Hallitseva intentio/ tarkoituserä	Ensisijainen sisältö Ensisijainen vastaanottaja	Tosiasiat Ideat	Tapahtumat	Asiat, tosiasiat, henkiset tilat, ideat	Ideat, henkiset tilat, vaihtoehtoiset maailmat
Oppia (metakielinen)	I T S E	Kopiointi Sanelusta kirjoittaminen	Kuullun (tai luetun) tarinan uudelleen kertominen	Muistiinpanot Yhteenvedo Tiivistelmä Referaatti Mukaelma	Kommentit kirjan marginaaliin Metaforat Analogiat
Välittää tunteita (emotiivinen)	ITSE MUUT	Tajunnanvirta	Persoonalliset kertomukset	Kuvaukset	Pohtiva kirjoittaminen
			Omakohtaiset päiväkirjat	Kirjeet	- omakohtaiset esseet
Informoida (viittaava)	M U U T	Sitaatti Lomakkeet	Kertovat raportit Uutiset Tiedotteet Ohjeet Sähkeet Kiertokirjeet	Ohjeet Kuvaukset Tekniset kuv. Elämäkerrat Tiet. raportit Raportit	Asiakaskeskeinen kirjoittaminen - määritelmät - akat. esseet - artikkelit - kirja-arvostelut - kommentit
Vakuuttaa/ Vaikuttaa (konatiivinen)	M U U T	Auktoriteetti/asiantuntijasitaatit	Hakemukset	Suosituks Ohjekirjeet	Argumentoiva/vaikuttava kirj. - pääkirjoitukset - kriitt. esseet - artikkelit
Viihdyttää, ilahduttaa, miellyttää (poeettinen)	M U U T	Runo- ja proosasitaatit	Tarinan kirjoittaminen, kun loppu annettu Tarinan lopun kirjoittaminen tarinan uudelleen kirjoittaminen	Kirjalliset muotokuvat tai sketsit Pakinat	Viihdyttävä kirjoittaminen - parodiat - loppusointuiset runot, riimit
		Dokumentoiva esitys	Toteava esitys		Tulkitseva esitys

Perinteisiä kirjallisia tyyliä ja muotoja voidaan käyttää yhden tai useamman tässä mainitun tarkoituserän mukaan

KUVIO 17 Kirjallisen esityksen yleinen malli. (Vähäpassi 1987:11)

Vähäpassi (1987a)⁵ on tutkinut koulussa opetettavan kirjoittamisen tavoitteita ja yhdistänyt siinä vaadittavat kognitiiviset prosessit erilaisiin kirjoittamisen tyyppeihin (kuvio 17). Hän erottaa kolme kognitiivisen prosessoinnin perustasoa. Vähiten vaativa on toistaminen, sellaisen tiedon tuottaminen, joka on jo aikaisemmin määritelty kielellisesti, esim. sanelu tai lomakkeen täyttäminen. Seuraava taso, uudelleenjäsentäminen, sisältää kirjoittajalle tutun aineksen järjestämistä, esim. raportin kirjoittaminen. Kolmas ja vaativin taso on keksiminen, kehittäminen, ajatusten tuottaminen. Kirjoittamisen tarkoituksia Vähäpassi erottaa viisi: oppiminen, tunteiden välittäminen, informoiminen, vakuuttaminen tai suostuttelu ja viihdyttäminen tai ilahduttaminen (Vähäpassi 1987a:11). Kaikkia näitä tarkoituksia voi toteuttaa erikseen toisistaan riippumatta, mutta kielellinen vaikeustaso eli kognitiivinen prosessointi sen sijaan vaihtelee riippuen kirjoittajan kielitaidosta ja annetusta kontekstista (Weigle 2002:10). Mallin pääulottuvuuksien risteyskohtiin, ruutuihin, on sijoitettu erilaisia tekstilajeja, mikä ei kuitenkaan tarkoita sitä, että tietty tekstilaji voitaisiin sijoittaa vain yhteen lokeroon. Vähäpassi (1987b:11) ottaa esimerkiksi tiivistelmän, jota voidaan käyttää sekä tiedonvälittämiseen että oppimiseen, jopa taivutteluun.

Kuviossa 18 esitetään kirjoittamisprosessi lähtien ajatusten kehittelystä ja tekstin luonnostelusta tavoitteiden suunnassa edeten käytettävissä olevien kirjoittamiskompetenssien avulla tekstin tuottamiseen, sen muokkaamiseen ja muihin jatkotoimenpiteisiin.



KUVIO 18 Kirjoittamisprosessi. (Lähde: Evensen ym. 2004.)

⁵ Myös Weigle (2002:8-9) ja Leimu (2004c:233) ovat lainanneet Vähäpassin kaaviota.

Kirjoittaminen tuotoksena

Suullisen tuottamisen yhteydessä esiteltiin tiedon hankintaan tai välittämiseen liittyvän puheen jaotteluperiaatteita (taulukko 20) samoin kuin Eurooppalaisen viitekehäyksen funktionaalisten taitojen makro- ja mikrofunktiot sisältöineen (taulukko 21). Sama jaottelu on sovellettavissa kirjalliseen tuottamiseen ottaen luonnollisesti huomioon tuottamiskanavasta johtuvat erityispiirteet.

Graben ja Kaplanin (1996) näkemys kielitaidosta, joka perustuu Hymesin (1972), Canalen ja Swainin (1980) ja Bachmanin (1990) teoksiin, jakaa kielitaidon kolmeen tyyppiin, lingvistiseen tietoon, diskurssitietoon ja sociolingvistiseen tietoon. Lingvistiseen tietoon kuuluu tieto kielen perusrakenteista ja sociolingvistiseen tieto niistä tavoista, joilla kieltä käytetään soveltuvalla tavalla erilaisissa sosiaalisissa ympäristöissä. Diskurssitiedolla tarkoitetaan tietoa tavoista, joilla sidosteista tekstiä rakennetaan. Grabe ja Kaplan (1996) määrittävät kirjoittamisen tehtäväympäristön suhteessa kirjoitustapahtuman osallistujiin, puitteisiin, tehtävään, tekstiin ja aiheeseen, ja he ovat koonneet yksityiskohtaisen luettelon niistä kielitaidon komponenteista, jotka liittyvät erityisesti kirjoittamiseen (taulukko 23). Tällainen taksonomia on hyödyllinen kirjoittamisen arvioinnille esimerkiksi silloin, kun pyritään selvittämään, mitä komponentteja hyvien kirjoittajien tuotoksissa on ja mitkä taas ovat tyypillisiä heikoille kirjoittajille.

TAULUKKO 23 Kirjoittamisen taksonomia (Grabe & Kaplan 1996:220-1. Siteerattu: Weigle 2002:30-31.)

I *Kieliopillinen tieto*

- A. Kirjoitetun koodin tuntemus
 1. Ortografia
 2. Oikeinkirjoitus
 3. Välimerkit
 4. Muotoilukonventiot
- B. Fonologinen ja morfologinen tieto
 1. Äänteiden ja kirjainten vastaavuus
 2. Tavut (rytmi)
 3. Morfeemien rakenne
- C. Sanasto
 1. Osallistujien väliseen kielenkäyttöön liittyvät sanat ja sanasto
 2. Akateemiseen ja pedagogiseen kielenkäyttöön liittyvät sanat ja sanonnat
 3. Muodolliset ja tekniset sanat ja sanonnat
 4. Aiheeseen liittyvät sanat ja sanonnat
 5. Ei-kirjaimellinen kielenkäyttö ja metaforat
- D. Lauseoppiin ja rakenteeseen liittyvä tieto
 1. Peruslauseenrakenteet
 2. Yleisesti hyväksytyt muodolliset kirjoittamisen rakenteet (sopiva tyyli)
 3. Kielikuvat ja kuvakieli
 4. Metaforat ja vertaukset
- E. Tietoisuus eri kielten välisistä eroista
- F. Tietoisuus omista vieraan kielen taidoista ja niiden rekisterien hallinnasta

II *Diskurssitieto*

- A. Tieto lauseiden sisäisistä ja välisistä merkintäkeinoista (koheesio, syntaktinen parallelismi)
- B. Tieto tekstin informaatorakenteesta (topiikki/ kommentti, vanha/uusi tieto, teema/reema, vierus- l. likiparit)
- C. Tieto lauseiden välisistä semanttisista suhteista.
- D. Tieto pääasioiden tunnistamistaidoista
- E. Tieto tekstilajin rakenteesta ja rajoituksista

- F. Tieto tekstin jäsenkäytännöistä
- G. Tieto päätelmien teosta (inferenssi, elaborointi)
- H. Tieto eri kielten diskurssin rakentamisen piirteiden eroista
- I. Tietoisuus omista diskurssitaidoista eri kielissä

III Sosiolingvoinen tieto

- A. Kirjoitetun kielen funktionaaliset käytöt
- B. Gricen (1975) maksimien⁶ soveltaminen ja tietoinen rikkominen
- C. Rekisteri ja tilanneparametrit
 1. Kirjoittajan ikä
 2. Kirjoittajan käyttämä kieli
 3. Käytetyn kielen kielitaito
 4. Lukijoiden huomioon ottaminen
 5. Kirjoittajan ja kohteen välinen status (valta/kohteliaisuus)
 6. Muodollisuuden aste (kunnioitus/yhteenkuuluvuus)
 7. Etäisyyden aste (ulkopuolisuus/osallisuus)
 8. Vuorovaikutuksen aihe
 9. Kirjoittamisen välineet (kynä, tietokone, sanelu, pikakirjoitus)
 10. Viestin välityskeino (yksi sivu/ kirja/ääneen lukeminen/painettu)
- D. Tietoisuus kielten ja kulttuurin välisistä sosiolingvivistisistä eroista
- E. Tietoisuus rekisterien ja tilanneparametrien rooleista.

Kirjoittamisen osa-alueella korostuvat kielellisten tietojen ja taitojen (sanastolliset, kieliopilliset, semanttiset sekä oikeinkirjoitustaidot) (emt. s. 156) lisäksi pragmaattiset tiedot ja taidot, joiden mukaisesti viestejä/tekstejä

- rakennetaan, jäsenetään ja yhdistetään (diskurssikompetenssi)
- käytetään viestintätarkoituksiin (funktionaalinen kompetenssi)
- järjestetään vuorovaikutus- ja asiointiskeemojen mukaisesti (skeemakompetenssi) (Eurooppalainen viitekehys 2003:173).

Tekstilaji voidaan määritellä kirjoittamisen aiotun muodon ja funktion suhteen. Muodolla tarkoitetaan sellaisia tuotoksia kuin kirje, raportti tai essee. Funktiota voidaan tarkastella kommunikatiivisten funktioiden suhteen, kuten esimerkiksi anteeksi pyytämisen tai diskurssimoodin suhteen kuten kertominen, kuvailu, selittäminen tai argumentointi (Weigle 2002:96). Eurooppalainen viitekehys esittää samanlaisia muotoon liittyviä esimerkkejä (Eurooppalainen viitekehys 2003:96).

4.3 Kieliopillinen kompetenssi

Kieliopin asema kielitaidon osa-alueena kielitaitoa arvioitaessa on kiistanalainen. Bachman ja Palmer (1996) pitävät sitä kielitaidon elementtinä, ei osa-alueena. Alderson (1993, 2000) lähtee siitä, että erilliset kielioppitehtävät kokeen osana ovat tarpeettomia, koska sama tieto voidaan saada tekstin ymmärtämisen kokeella. Hughes (1989) perustelee niiden mukaan ottamista sillä, että kieliopil-

⁶ Gricen maksimin (periaatteen) mukaan vuorovaikutus on yhteistoimintaa, jossa kumpikin osapuoli tekee parhaansa kommunikaation onnistumiseksi (*maxim of cooperation*). Muita hänen maksimejaan ovat *maxim of quantity/quality/relation/manner*. (Buck 2001:24)

linen kompetenssi on yleensä yksi oppimistavoitteiden osa. Koska virheettömyys on hyvän kielitaidon tärkeä edellytys ja se osoitetaan selkeimmin kieliopillisen kompetenssin tasona, lienee paikallaan sijoittaa sen teoreettinen kuvaus kielitaidon osa-alueiden yhteyteen. Kieliopin käsitettä on myös määritelty monin eri tavoin. Laajimmat käsite-erot ovat lauseoppikeskeisen ja kommunikatiivisen näkökulman välillä. Uutta, kiinnostavaa tietoa kieliopillisen kompetenssin käsitteestä on esittänyt Purpura (2004).

4.3.1 Lauseoppikeskeinen kielioppikäsitys

Lauseoppikeskeiseen kielioppikäsitukseen kuuluu mm. formaalinen kielioppi, joka on systemaattinen tapa selittää ja ennustaa ihannepuhujan tai kuulijan tietoa kielestä (*knowledge of language*) (Purpura 2004:6). Tätä ryhmää edustavat traditionaalisen kieliopin, strukturaalisen eli deskriptiivisen kielitieteen ja transformatiivis-generatiivisen kieliopin käsitykset. Formaalinen kielioppi analysoi pääasiassa lauseiden ja virkkeiden rakennetta jättäen näiden muotojen merkityksen ja niiden käytön lauseyhteydessä esimerkiksi semantiikan ja pragmatiikan vastuulle.

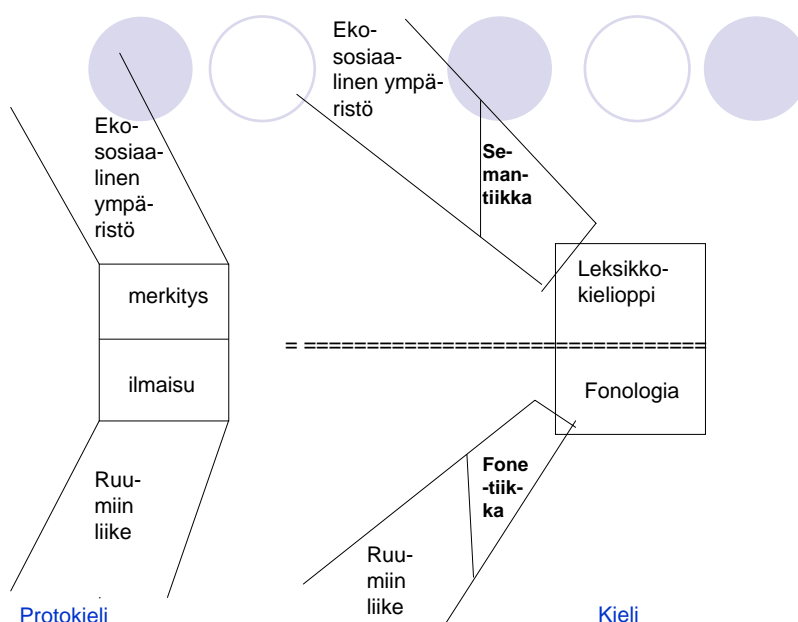
Traditionaalisen kieliopin ansiona pidetään sitä, että se ottaa huomioon myös kieliopillisten muotojen käytön. Tämä on toisaalta puute, sillä monien poikkeuksien vuoksi kieltä on ollut vaikea sijoittaa mihinkään viitekehykseen, mikä taas estää tai ainakin vaikeuttaa sen yleistämistä muihin kieliin (Purpura 2004:8). Toisin kuin traditionaalinen kielioppi, strukturaalinen kielitiede ei perustu sääntöihin, vaan pyrkii kuvaamaan kieltä kohdentamalla huomion tiukasti kieliopilliseen muotoon. Tämä lähestymistapa jättää huomiotta kieliopin semanttiset aspektit eikä ohjaa juurikaan kieliopillisten muotojen käyttöä kontekstissa (Purpura 2004:9). Transformatiivis-generatiivinen kielioppi lähtee siitä, että jokaisen kielellisen järjestelmän taustalla olevat ominaisuudet voidaan paljastaa yksinkertaisen lausetason analyysin avulla. Tätä kielioppikäsitystä kehitti edelleen Chomsky universaalikieliopissaan, jonka mukaan kielen prinssiipit ovat valmiina mielessä, joten niitä ei tarvitse oppia eikä opettaa. Hänen mukaansa lausekerakenteet muodostetaan kaikissa kielissä samoin: lausekkeen ydin on pääsana, jolla voi olla kaksi erilaista määrettä, pääsanaan kiinteästi liittyvä komplementti, jonka kanssa pääsana muodostaa konstituentin, ja tarkenne, joka määrää tätä konstituenttia ja muodostaa sen kanssa uuden konstituentin. Pääsana heijastaa syntaktiset ominaisuutensa koko lausekkeen alueelle. Kielet puolestaan eroavat pääsanana, tarkenteen ja komplementin järjestyksen suhteen (Jaakkola 1997:39).

4.3.2 Kommunikatiivinen kielioppikäsitys

Kommunikatiivisessa kielioppikäsityksessä yhdistyvät kieliopillinen muoto ja merkitys. Tätä on kuvattu esittämällä, että puheakti sisältää sekä kirjaimellisen että aiotun merkityksen. Kommunikaatioperustainen perspektiivi näkee kieliopin kielellisten normien, preferenssien ja odotusten järjestelmänä, jotka yksilö ottaa käyttöön välittääkseen niitä pragmaattisia merkityksiä, jotka ovat sopivia, hyväksyttäviä ja luonnollisia kussakin tilanteessa. Tässä kontekstissa kielioppi

merkitsee jotakuinkin samaa kuin itse kieli eikä tarkoita vain muotoa, vaan myös mielekkyyttä ja pragmaattista sopivuutta, hyväksyttävyyttä ja luonnollisuutta (Purpura 2004:16).

Kommunikatiivisen kielioppitaidon testaamisen kannalta kahden tutkijan kielioppiteoriat, Austinin (1962) puheaktiteoria ja Hallidayn (2003) systeemifunktionaalinen kielioppiteoria, ovat olleet merkityksellisiä. Puheaktiteoria sisältää ajatuksen siitä, että puhunto koostuu kolmesta toisiinsa liittyvästä puheaktista. Ensimmäinen on merkityksellisen proposition tuottaminen (*locutionary act*), joka välittää puhunnon kirjaimellisen merkityksen (*locutionary meaning*). Toinen viestii puhujan aikomuksesta tietyssä kontekstissa (*illocutionary act*), joka välittää puhujan puhunnon aiotun merkityksen. Kolmanneksi puhunto aiheuttaa vaikutuksen puhekumppaniin, mistä Austin (1962) käyttää termiä *perlocutionary act* (Purpura 2004:17-18). Toisen, systeemifunktionaalisen kielioppiteorian mukaan kommunikaatio ilmenee kielessä merkityksen kautta, jossa kielioppi on vain yksi osa (Purpura 2004:21). Hallidayn kielioppikäsitys korostaa näkemystä, jossa muoto ja merkitys nähdään rinnakkaisina ja kieltä käytettäessä tehdään valintoja samanaikaisesti eri osajärjestelmistä (Jaakkola 1997:21). Halliday on kehittänyt oman teorian kielen kehitymisestä ja kieliopin ja sanaston (leksikkokielioppi) asemasta siinä (kuvio 19). Vaikka kuvaus koskeekin lapsen kielen kehitystä, sen leksikkokielioppi -käsite voidaan rinnastaa (aloittelevan) vieraan kielen puhujan kieliopillisen kompetenssin ja sanaston prosessointiin kieltä tuotettaessa.



KUVIO 19 Protokielen ja kielen suhde ekososiaaliseen ja fyysiseen ympäristöön. (Halliday 2003:13)

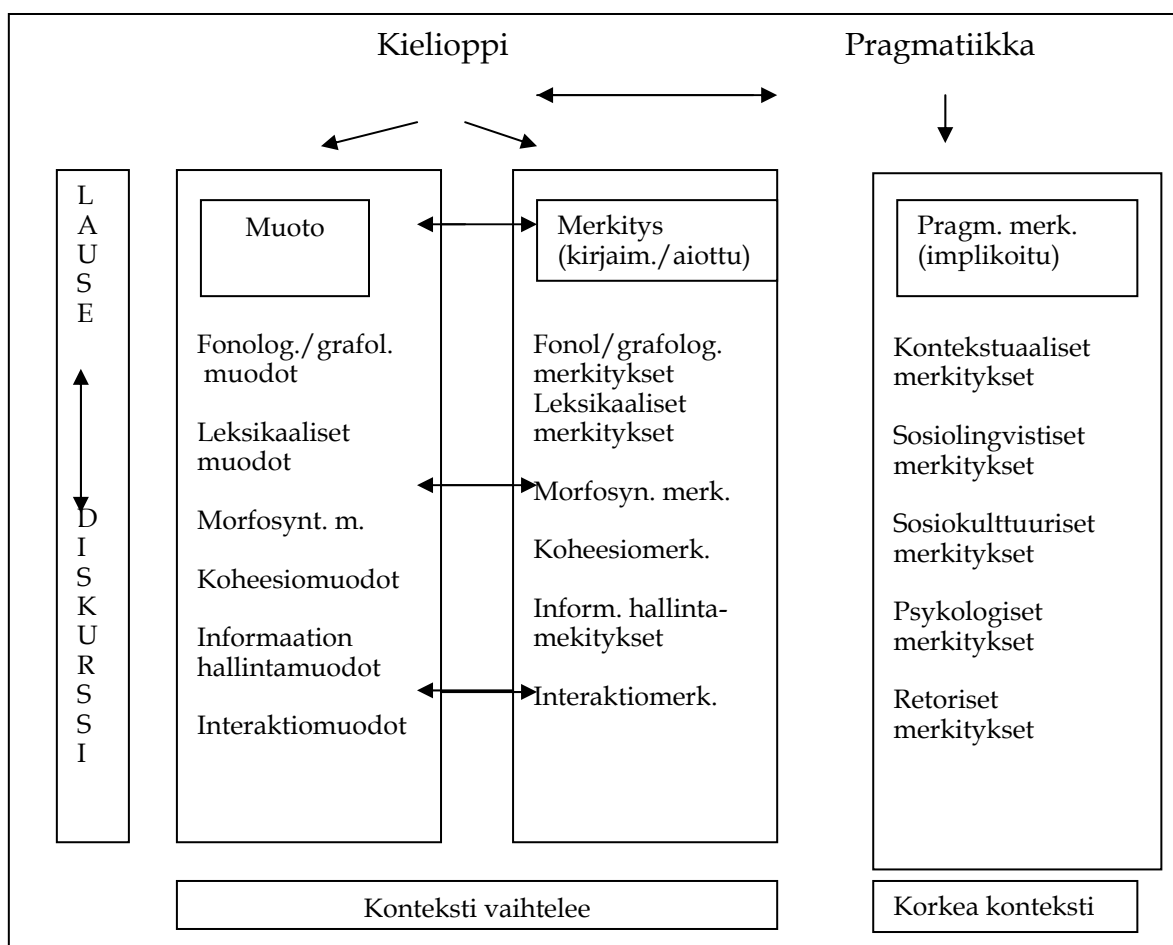
Kuvio 19 kuvaa sitä, miten lapsi kehittää ensin puhekielen, joka on yhdistelmä eleitä ja ääntä, mutta jossa ei vielä esiinny sanoja. Tässä kielessä ikään kuin yh-

distyvät semantiikka ja fonologia, joiden välillä ei kuitenkaan ole kieliopillista tasoa. Merkit ovat erillään, kieli ei ole vielä osittunut (*stratified*) niin, että merkkejä voitaisiin käyttää kerroksittain. Protokielen kehittyessä kieleksi syntyy tila, jossa merkitykset järjestäytyvät keskinäisten suhteiden abstraktiksi verkostoksi, ja jota Halliday kutsuu leksikkokieliopiksi (*lexicogrammar*).

Aluksi tämä kielioppi palvelee eräänlaisena välikielenä, jolloin alkuperäisen merkin kaksi osaa venyvät merkitsemään monia asioita. Seuraavassa vaiheessa merkin pystyy purkamaan varioitaviksi merkityskomponenteiksi niin, että artikulaatiomuoto saattaa merkitä yhtä asiaa, merkkien järjestys toista ja prosodinen väritys kolmatta. Tämän jälkeen onkin mahdollista järjestää merkitykset funktionaaliseen järjestykseen niin, että jokainen puhunto sisältää valikoiman kaikesta edellä esitetystä, mikä vuorostaan mahdollistaa samanaikaisen reflektoinnin ja toiminnan. Viime kädessä se merkitsee sitä, että merkitys pystytään luomaan dialogin – ja myöhemmin myös monologin – kuluessa. Alkuperäinen sisältömerkki on nyt mahdollista stratifioida (osittaa), ja samanlainen stratifiointi tapahtuu samanaikaisesti ilmaisuosassa. Ääni korvaa eleen, ja tämä jakaantuu samalla tavoin kahtia: tilaksi (fonologia), jossa ääni ja leksikkokielioppi järjestyvät, ja puheeksi artikuloitavana ja kuultavana prosessissa, joka tapahtuu äänielimissä (fonetiikka) (Halliday 2003:112-15).

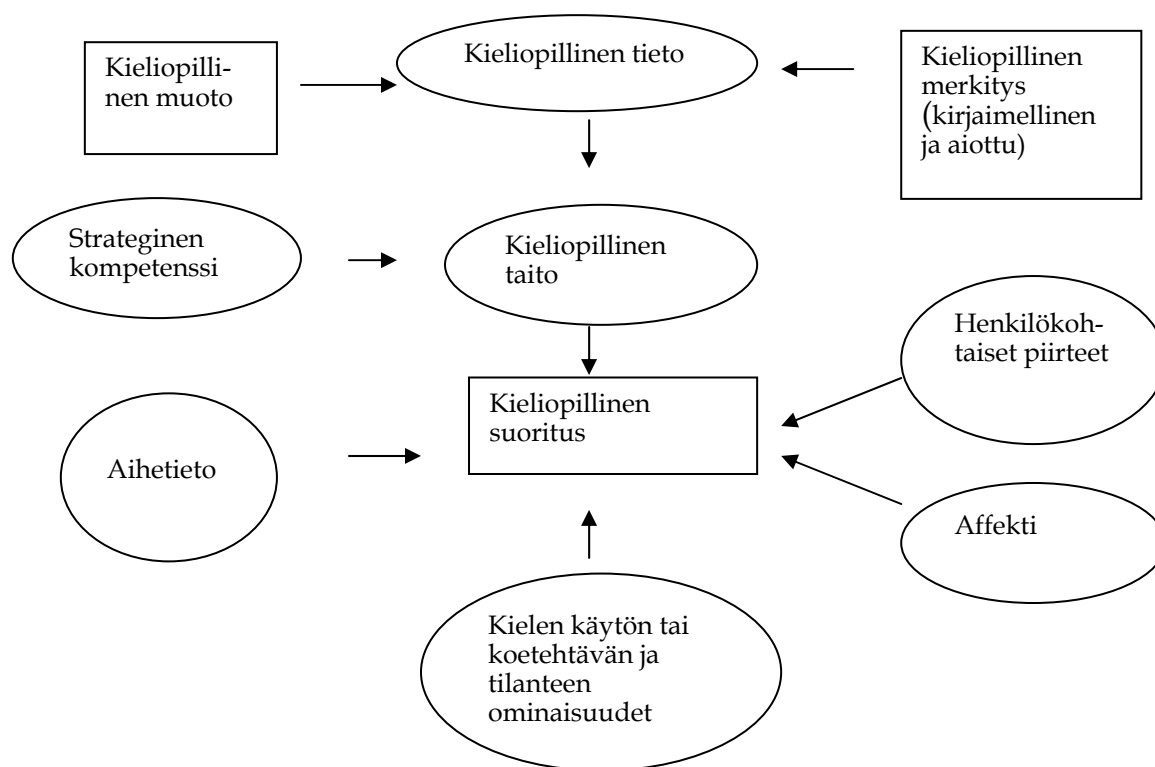
Kieliopin asema kielen käytössä

Purpura on esittänyt teoreettisen mallin (kuvio 20) kielioppikäsitteistään, joka jakaantuu muotoon ja merkitykseen sekä pragmatiikkaan, joiden hän katsoo yhdessä muodostavan kielen (2004:62). Jos kirjaimellinen merkitys selviää suoraan kieliopillisesta muodosta, puhutaan matalasta kontekstista. Mitä enemmän kontekstin apua viestin ymmärtäminen edellyttää, sitä korkeampi katsotaan kontekstin olevan. Arviointitarkoituksia varten Purpura on tähän peruskuvioon lisännyt tarkennuksia (Purpura 2004:78, 91) sekä esimerkkejä käytännön sovelluksista (emt. s. 107).



KUVIO 20 Kieliopin ja pragmatiikan teoreettinen määritelmä. (Lähde: Purpura 2004:62.)

Kuvio 21 kuvaa Purpuran (2004) malliin perustuvaa näkemystä kieliopillisen tiedon (*grammatical knowledge*), kieliopillisen taidon (*grammatical ability*) ja kieliopillisen suorituksen (*grammatical performance*) välisestä suhteesta. Kieliopillisella tiedolla Purpura (2004) tarkoittaa pitkäaikaismuistissa käytettävissä olevia tietorakenteita, ja kieliopillisella taidolla edellä mainittujen resurssien käyttöön ottamista. Strategiseen kompetenssiin hän katsoo kuuluvan paitsi Bachmanin ja Palmerin (1996) määrittelemät metakognitiiviset strategiat myös kognitiiviset strategiat (esimerkiksi assosiaatioiden muodostaminen, selventäminen). Kielioppitaitoon siis sisältyvät sekä kielioppitieto että strategiset kompetenssit. Arvioitavat muuttujat on ilmaistu suunnikkaina ja teoreettiset konstruktit soikioina.



KUVIO 21 Kielioppitiedon, kieliopillisen taidon ja kielellisen suorituksen välinen suhde. (Lähde: Purpura 2004:88.)

Eurooppalaisessa viitekehyksessä (2003) kielellisen viestintätaidon yhden pääkomponentin, kielelliset tiedot ja taidot, eräs kategoria on kieliopilliset tiedot ja taidot, josta käytetään myös nimitystä kieliopillinen kompetenssi. Se määrittelee kieliopin periaatteina, joiden mukaan kielen elementeistä muodostetaan mielekkäitä lauseita. Kieliopillisen kompetenssin osoitus taas on kyky ymmärtää tällaisia lauseita sekä muodostaa itse lauseita näiden periaatteiden mukaisesti (emt. s. 160). Viitekehys ei sitoudu mihinkään malliin tai teoriaan. Se katsoo kuitenkin, että kieliopillisen järjestelmän kuvausta varten on määriteltävä elementit, kategoriat, luokat, rakenteet⁷, kuvailevat prosessit ja suhteet (emt. s. 160-161).

4.4 Sanastollinen kompetenssi

Sanastoa ei lueta varsinaisesti kielitaidon osa-alueeksi, vaan yhdeksi sen elementiksi tai komponentiksi (ks. esim. Bachman & Palmer (1996:75) kuten kielioppikin. Viestinnän tehokkuuden kannalta laaja ja tarkka sanavarasto kuitenkin on välttämätön ja näin ollen tärkeä arvioinnin kohde. Osa kielen ilmiöistä sijoittuu sanaston ja kieliopin väliin. Kielessä on myös suuri määrä vakioilma-

⁷ Rakenne -sanaa ei tässä käytetä kielioppi- tai kieliopillinen kompetenssi -termin synonyymina kuten usein suomalaisissa kielitaidon kuvauksissa.

uksia, jotka koostuvat joukosta sanoja ja joita tulee ymmärtää ja osata itse käyttää viestinnässä. Joskus ne ovat hyvinkin olennaisia viestin välittymiselle, kun taas niiden puute tai virheellinen käyttö paljastaa ongelmat idiomaattisen kielen hallinnassa (Takala 1997:107). Sanastokokeissa menestyminen korreloi tekstin ymmärtämisen kokeissa menestymisen kanssa (Takala 1984:18; Alderson 2000:35 ja Read 2000:115), ja sitä voidaan pitää jopa tärkeämpänä kuin kielioppia (Takala 1984:19). Read (2000) toteaa, että opiskelijat joutuvat tavan takaa huomaamaan puutteita sanavarastossaan ja saattavat peräti samaistaa sanaston kartuttamisen kielen opiskeluun (Read 2000:1).

Sanaston hallinnan merkitys näyttää kohonneen kommunikatiivisen kielenopetuksen myötä (Read 2000). Tämä on havaittavissa myös Suomen englannin kielen oppimateriaaleissa lisääntyneinä sanastotehtävinä. Kun opetus yleensä kulkee testauksen edellä, on ehkä odotettavissa, että sanaston opettamisen monipuolistamiseen aletaan kiinnittää entistä enemmän huomiota ja että sanasto otetaan laajempien arviointien osaksi.

4.4.1 Sanan ´osaaminen`

Sanaston arvioinnin pohjaksi tulisi määritellä, mikä ´sana` on. Käsite on kuitenkin vaikeaselkoinen (Takala 1984:6-7; Read 2000:16-17). Ortografisesti sen määrittely on yksinkertaista, mutta asiaa voidaan lähestyä myös kieliopillisesti ja semanttisesti, jolloin määrittely vaikeutuu (Takala 1984:6). Sanojen muodon määrittämiseen sisältyy tieto siitä, miten sana kirjoitetaan tai äännetään sekä mistä sana alkaa ja mihin se päättyy. Siihen kuuluu myös sanan morfologisten piirteiden (taivutus, sananjohto tai yhdyssanojen muodostamisen mahdollisuus) sekä sanan paradigmaattiset suhteet toisiin sanoihin. Semanttisen aspektin kanssa ollaan tekemisissä silloin, kun sanan merkitystä ei tunneta riittävän tarkasti tai kun ei tiedetä, millaisten sanojen kanssa kyseistä sanaa käytetään. Semanttiseen tietoon kuuluvat myös merkitys- ja vivahde-erojen tunteminen samoin kuin sanan tyyliarvo (Takala 1984:56-57). Eurooppalainen viitekehys (2003:157) määrittää sanastollisen kompetenssin kielen sanaston ja kieliopin elementeistä koostuvan sanavaraston tuntemisena ja käytön hallintana. Elementit esimerkkeineen esitetään taulukossa 24.

TAULUKKO 24 Sanastolliset tiedot ja taidot (sanastollinen kompetenssi). (Muokattu taulukoksi Eurooppalaisen viitekehyksen 2003:157-158 ja CEFin 2001:110-111 pohjalta.)

Sanastoelementit	a) vakiintuneet sanonnat	* lauseparret	- kielen funktioiden suorat ilmenemät	<i>How do you do?</i>
			- sananlaskut	
			arkaaiset ilmaukset	<i>Be off with you!</i>
		* idiomaattiset sanonnat		<i>He kicked the bucket.</i>
		* vahvistimet		<i>He drove hell for leather.</i>
		* kiinteät lausemallit		<i>Please, may I have...?</i>
		* muut kiinteät sanonnat		<i>to put up with; in front of</i>
		* kiinteät kollokaatiot		<i>make a mistake</i>
	b) yksittäiset sanat		- erilaisia merkityksä omaavat sanat	<i>tank; stand</i>
			- sulkeiset sanaryhmät	viikonpäivät, kuukaudet, painot, mitat jne.
		- kieliopillisia tai semanttisia tarkoituksia varten muitakin sanaryhmiä		
Kieliopilliset elementit	* artikkelit			<i>a, the</i>
	* määrää ilmaisevat sanat			<i>some, all, many jne.</i>
	* demonstratiivipronominit			<i>this, that, these, those</i>
	* persoonapronominit			<i>I, we, s/he, they, me jne.</i>
	* interrogatiivipronominit			<i>who, where, how jne.</i>
	* relatiivipronominit			<i>who, which, that jne.</i>
	* possessiivipronominit			<i>my, your, his, its jne.</i>
	* prepositiot			<i>in, at, by, with, of jne.</i>
	* apuverbit			<i>be, do, have; modaaliapuverbit</i>
* konjunktiot			<i>and, but if, although jne.</i>	

Miten sanan "osaamisen" voi määrittää? Yksi määrittelytapa on tehdä ero reseptiivisen ja produktiivisen taidon välillä. Sanan voi tunnistaa jonkun muun käyttämänä, mutta sitä ei välttämättä pysty käyttämään itse. Sana voi osoittautua hankalaksi määrittellä tai kuvata, sanalle voi olla vaikeata löytää äidinkielistä vastinetta tai synonyymia/antonyymia tai sen kollokaatio on epäselvä (Takala 1984:55).

Nation (1990) on koonnut analyttisen taulukon siitä, mihin kysymyksiin oppijan/testattavan on ikään kuin vastattava osoittaakseen tuntevansa sanan merkityksen eli 'osatakseen sanan' (taulukko 25).

TAULUKKO 25 Sanan tuntemisen komponentit. (Lähde: Nation 1990:26.)

Muoto	Resept/ produkt.	Kysymys
Puhuttu	R	Miltä sana kuulostaa?
	P	Miten sana äännetään?
Kirjoitettu	R	Miltä sana näyttää?
	P	Miten sana kirjoitetaan ?
<i>Sijainti</i>		
Kielipiilliset rakennemallit	R	Missä rakennemalleissa sana esiintyy?
	P	Missä rakennemalleissa sanaa tulee käyttää?
Kollokaatiot	R	Mitä tai minkä tyyppisiä sanoja voi odottaa esiintyvän tämän sanan edellä tai jäljessä?
	P	Mitä tai minkä tyyppisiä sanoja tämän sanan kanssa tulee käyttää?
<i>Funktio</i>		
Frekvenssi	R	Kuinka yleinen sana on?
	P	Kuinka usein sanaa tulisi käyttää?
Sopivuus	R	Missä yhteydessä tämä sana voi esiintyä?
	P	Missä yhteydessä tätä sanaa voi käyttää?
<i>Merkitys</i>		
Käsite	R	Mitä sana tarkoittaa?
	P	Mitä sanaa tulisi käyttää tämän merkityksen ilmaisemiseen?
Assosiaatiot	R	Mitä muita sanoja tämä sana tuo mieleen?
	P	Mitä muita sanoja tämän sanan sijasta voisi käyttää?

4.5 Yhteenveto

Kielitaidon arviointi alkaa kielitaidon käsitteen määrittämisestä. Eri aikoina siihen on katsottu kuuluvan eri asioita. Kaikilla teorioilla on edelleenkin paikkansa testauskäytännöissä, minkä käsityksen mukaisesti tämän tutkimuksen kohteena olevassa arvioinnissakin on toimittu. Nykyään käytetyimpiä kielitaidon malleja ovat Bachmanin ja Palmerin arviointia silmällä pitäen laatima malli sekä siihen läheisesti liittyvä Eurooppalaisen viitekehyksen kielitaitokäsitys, jonka arviointia tuetaan lukuisilla yksityiskohtaisilla taitotasokuvauksilla. Tämän tutkimuksen kohteena olevassa arvioinnissa tukeuduttiin Bachmanin ja Palmerin kielitaitokäsitykseen, ja käsillä olevaa tutkimusta varten tehtäviä ja tuotoksia arvioitiin Eurooppalaisen viitekehyksen kulloinkin tarkoituksenmukaisiksi katsottuja taitotasosteikkoja käyttäen (luku 9).

Kielitaitoa arvioitaessa taidon määrittämisessä on myös päätettävä, arvioidaanko ymmärtämis- ja tuottamisprosessia vai prosessoinnin tulosta eli tuotosta, tuotetta. Kohde on tässäkin tapauksessa vaihdellut ajasta ja tutkijasta riippuen. Ymmärtämisprosessin määrittäminen on ollut dualistista: *top-down* tai *bottom-up*. Nyt lähdetään yleensä siitä, että ymmärtäminen alkaa koodin purkamisesta ja päättyy aikaisemman tiedon integroitumisesta uuteen. Prosessi on osoittautunut käytännössä varsin vaativaksi arviointikohteeksi.

Ymmärtämisprosessin katsotaan perustuvan kolmenlaiseen informaatioon: kielelliseen syötökseen (jonka testattava voi tulkita eri tavoin), kielelliseen

ja sosiokulttuuriseen tietoon (joka on käytännöllisesti katsoen jokaisella omanlaisensa) sekä kontekstuaaliseen informaatioon (jonka ymmärtämisen määrä riippuu testattavan luku/kuuntelutaidoista). Näin ollen yleensä arvioidaan ymmärtämistä tuotoksena, joka on testattavan (oletettu) ymmärtämisen taso, jonka hän osoittaa suorittamalla (oletettavasti) vaikeustasoltaan erilaisia tehtäviä. Tuotosnäkökulma taas on tuottanut erilaisia kognitiivisia ja kommunikatiivisia taksonomioita, joilla on pyritty luokittelemaan ymmärtämisen taitoja. Eräät tutkijat jakavat ne kahteen ryhmään, suoraan ja päättelyn kautta ymmärtämiseen, joista jälkimmäiseen liittyvien katsotaan osoittavan korkeampaa kielitaidon tasoa. Toiset lähtevät tekstin tarkoituksista, joista kukin edellyttää omia taitojaan. Kolmas koulukunta luettelee tilanteita, joissa kuulijan tulee kyetä osoittamaan kielitaitonsa. Tuotoksen arvioinnin perustana on arvioitavien taitojen sisältö ja vaikeustaso.

Puhuminen prosessina vaatii tavallisesti samanaikaista toimintaa sekä puhujalta että kuuntelijalta, ja sisältää suunnittelua, valintaa ja itse tuotoksen. Prosessointia on myös kuvattu kaaviolla, jossa puhuja luo käsitteellisessä valmistelussa viestin, käsiterakenteen, jota kielenä edustavat sanat. Kieliopillisessa koodauksessa nämä aktivoivat vastaavat syntaktiset sanat. Tällöin puhuja pääsee käsiksi muodon koodin morfologiseen ja fonologiseen rakenteeseen, joka laukaisee artikulatorisen eleen, joka johtaa artikulatoriseen tulokseen. Tämän yhdistyessä puhe-elinten toimintaan päädytään itse puheeseen. Tuotoksena puhetta arvioitaessa voidaan lähtökohtana pitää viestinnällisiä tehtäviä kuten kuvailu, ohjeiden antaminen, tarinan kerronta ja mielipiteen ilmaisu/perustelu, joiden yhteydessä käytetään tietynlaisia rutiineja. Näitä voidaan kutsua makro-funktioiksi. Mikro-funktioita puolestaan ovat lyhyet ilmaukset, jotka voivat olla sellaisinaan arvioinnin kohteina.

Kirjoittaminen voidaan määrittää sekä kognitiivisesta että sosiaalisesta näkökulmasta, sitä voidaan tulkita sekä prosessina että tuotoksena ja sitä voidaan arvioida joko lingvistiseltä tai kommunikatiiviselta kannalta. Nämä molemmat aspektit voidaan myös yhdistää. Kirjoittamisen kognitiivisessa prosessoinnissa voidaan erottaa kolme perustasoa: toistaminen, jäsentäminen ja kehittäminen sekä ajatusten tuottaminen, jotka voidaan yhdistää kirjoittamisen erilaisiin tarkoituksiin. Tällöin kirjoittajan kielitaito luonnollisesti määrittää tuotoksen kielellisen tason. Kommunikatiivinen prosessointi taas lähtee vuorovaikutuksesta, viestinnästä, ja etenee tuotoksen itsearvioinnin, uudelleenorganisoinnin ja analysoinnin kautta joko mielikuvi- tai vakavien teoreettisten pohdiskelujen maailmojen tulkittamiseen. Kirjoittamista tuotoksena tarkasteltaessa taksonomioita jaottelevat tuotoksen komponentit. Kehittyneessä tuottamisessa kiinnitetään kuitenkin suurempaa huomiota kieliopilliseen kompetenssiin ja tuotoksen rakenteeseen liittyvään tietoon. Mitä korkeampaan tasoon pyritään, sitä suurempia vaatimuksia asetetaan myös sanastolliselle kompetenssille. Kieliopillinen kompetenssi voidaan myös asettaa ymmärtämis- ja tuottamisprosessien lähtökohdaksi.

Kielitaidon määritelmän ja sen ilmenemisen pohdiskelusta siirrytään itse kokeen laadinnassa huomioon otettaviin asioihin.

5 KIELITAIDON KOKEEN LAADINTA JA ARVIOINNIN TOIMEENPANO

Arvioinnin tarkoitus on löytää ne korjaamista vaativat kohdat, joihin myöhemmin voidaan etsiä parannuskeinoja. Näin ollen esimerkiksi kokeesta saatavien tulosten tulisi liittyä jonkin taidon määrittelyyn kokeen ulkopuolella. Tässä arvioinnissa käytetyn Bachmanin ja Palmerin kokeen laadinnan mallin mukaan koetehtävien tulee vastata todellisen elämän kielenkäyttötilanteita sekä olla käyttökelpoisia. Tätä ominaisuutta he arvioivat kuudella kriteerillä. Kun halutaan sanoa jotakin testattavan yleisestä taitotasosta, tarvitaan viitekehys selittämään sitä, mitä taidolla tarkoitetaan ja miten se ilmenee kyseisessä koesuorituksessa.

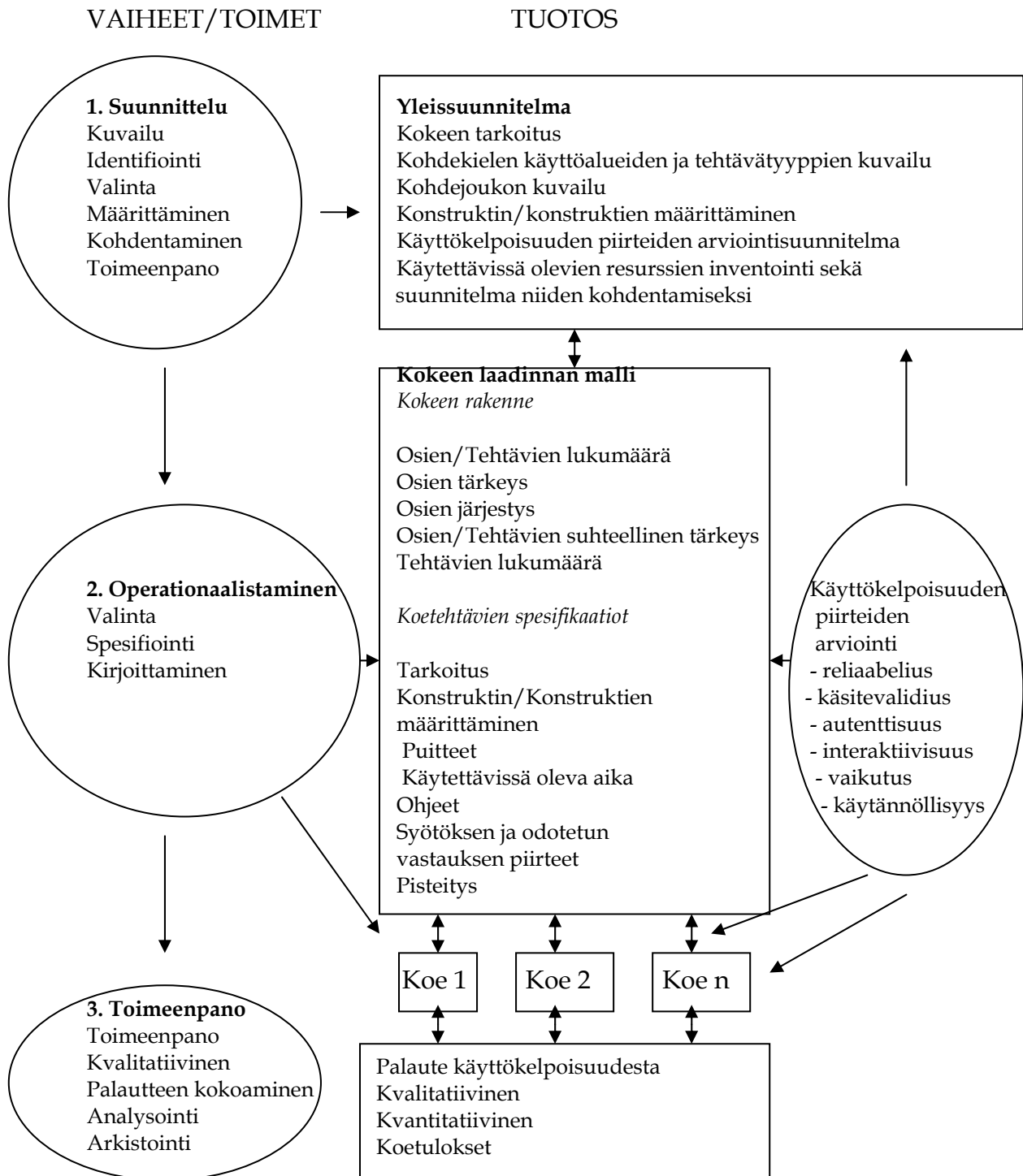
Jokaisessa arvioinnissa on tehtävää menetelmää ja käytettäviä asteikkoja koskevia ratkaisuja. Luvussa 2 todettiin erilaisten arviointien edellyttävään erilaisia menetelmiä. Mitä merkityksellisempi arviointi on, sitä perusteellisemmin menetelmän valintaan on paneuduttava. Esimerkiksi tämän tutkimuksen kohteena olevassa arvioinnissa päädyttiin aikanaan lyhyiden vapaiden vastausten polytomiseen pisteitykseen, joka käsillä olevaa tutkimusta varten muutettiin dikotomiseksi.

Tässä luvussa käsitellään normi- ja kriteerivoitteisen arvioinnin eroja erityisesti siksi, että jälkimmäinen soveltuu erinomaisen hyvin juuri opetussuunnitelmaan pohjautuvien oppimistulosten esittämiseen, jollaisesta tässä tutkimuksessa on kysymys.

5.1 Kokeen laadinnan malli

Arviointi- ja kielen testauksen prosessien kuvauksia löytyy useista lähteistä (Lindquist 1961, Norris 1990, Alderson 1992, Rea-Dickins & Germaine 1992, Davidson & Lynch 1993, Nitko 2001, Lynch 1996, 2003). Mm. Alderson; 67-68 (2000) ja Purpura (2004) toteavat kuitenkin Bachmanin ja Palmerin (1996) laatiman kielen käyttöä ja koetehtäviä analysoivan viitekehyksen edustavan uusinta ja selkeintä ajattelua kielitaidon testaamisessa. Kritiikkiäkin on siihen kuitenkin kohdennettu (ks. s. 68).

Bachman ja Palmer (1996:86) ovat esittäneet mallin, jossa kokeen laadinta jaetaan kolmeen vaiheeseen: suunnittelu, operationaalistaminen ja toimeenpano. Kuviossa 22 kuvataan em. mallin mukaiset kokeen laadinnan vaiheet. Vasemmalla olevissa ympyröissä esitetään prosessit, joiden kautta päästään kokeen laadinnan konkreettiseen toteuttamiseen, jonka komponentit esitetään oikealla puolella olevissa laatikoissa.



KUVIO 22 Kokeen laadinta. (Lähde: Bachman & Palmer 1996:87.)

5.1.1 Suunnitteluvaihe

Suunnitteluvaiheessa kokeen tarkoituksen kuvaaminen tarkoittaa niiden johtopäätösten kirjaamista, joita koetulosten perusteella aiotaan tehdä. Käyttökelpoisuuden arviointisuunnitelmaa tarvitaan kokeen laadinnan kaikissa vaiheissa. Se alkaa komponenttien mahdollisimman tasapuolisesta sisällyttämisestä kokeeseen ja päättyy koevastausten kautta palautteen kokoamiseen. Tähän liittyy myös arviointimenetelmän ja -tapojen valinta, joka tämän tutkimuksen kohteena olevassa alkuperäisessä arvioinnissa oli osaksi määräytynyt arviointitoimeksiannon kautta. Näitä teemoja käsitellään jäljempänä tässä luvussa ja niiden käytäntöön soveltamista luvussa 8.

Kohdekielen käyttöalueen (*target language use = TLU domain*) kartoittamisen eli niiden tilanteiden kuvauksen, joissa kieltä todellisessa elämässä käytettäisiin, on tarkoitus luoda pohjaa varsinaisten koetehtävien sisällölle ja samalla niiden autenttisuuden ja niihin sisältyvien prosessointitapojen arvioinnille. Testattavien kuvaus tarjoaa mahdollisuuden johtopäätösten tekemiseen kokeen vaikutuksista. Kun kokeen tarkoitus ja kohdejoukko ovat selvillä, on määriteltävä konstruktin, mittauksen kohde, sillä koe on käyttökelpoinen ja validi vain, jos se mittaa oikeata konstruktia (Buck 2001:1). Buck puhuu kokeen validiudesta, vaikka nykykäsityksen mukaan olisi mieluummin puhuttava pistemäärien tulkintojen ja käyttötapojen validiudesta. Eräät lukemisen teoriat sisältävät esimerkiksi ajatuksen siitä, että lukemiseen sisältyy monia eri konstruktteja (silmäily, selailu jne.) ja että konstruktit eroavat toisistaan (Alderson ym. 1995:17). Ne ovat siis kielitaitoteorian osa, jota täsmentämällä kielenkäytön suuntaan päästään koetehtävään tai osioon, joka mittaa testattavaa taitoa kohteena olevan konstruktin tai sen osan tuottamisessa. Tätä menettelyä sovellettiin ymmärtämistehtävien valinnassa ja sitä käsitellään luvussa 8. Luoma (2001) muistuttaa kuitenkin, että testauksen ollessa kyseessä koetehtävien ja testattavien odotetun käyttäytymisen tarkentaminen ja konkretisointi lisääntyvät jatkumona alkaen käsitteistä ja siirtyen konstruktien kautta niiden operationaalistamiseen (Luoma 2001:4).

Konstruktin voi vaihtoehtoisesti määrittää niiden tehtävien kautta, jotka testattava pystyy suorittamaan. Näin menetellään erityisesti silloin, kun halutaan tietää, millaisista kielenkäyttötilanteista testattavat selviytyvät. Tähän prosessiin sisältyy kaksi vaihetta. Ensinnä on oltava sopiva lista kohdekielisistä tehtävistä, jotka testattavan pitäisi hallita. Toiseksi on tärkeää tuottaa joukko koetehtäviä, jotka jollakin tavalla jäljittelevät todellisia kielenkäyttötilanteita (Buck 2001:106; Luoma 2004:42-43). Taustalla oleva teoreettinen käsite on autenttisuus, jota Bachman (1990) on todennut olevan kahdenlaista: tilanteen autenttisuus (*situational authenticity*) eli missä määrin tehtävät sisältävät kohdekielen käyttötilanteiden piirteitä, ja prosessoinnin autenttisuus (*interactional authenticity*) eli missä määrin koetehtävät edellyttävät samoja taitoja kuin kohdekielen käyttötehtävät. Näihin kiinnitettiin huomiota erityisesti puhumisen kokeiden tehtävien laadinnassa ja osittain myös kirjoittamistehtävien yhteydessä. Tätä kommentoidaan luvussa 8.

Kompetenssin ja tehtävän väliseen interaktioon perustuvan konstruktin määrittämisellä saadaan käyttöön molempien lähestymistapojen edut ja voidaan olettaa, että toisen puutteet korvaavat toisen heikkoudet. Tämän tekemiseksi saattaa riittää se, että laaditaan lista tarvittavista taidoista ja lista tärkeimmistä kohdekielen käyttötehtävistä ja kehitetään sitten koetehtäviä, jotka edellyttävät molempia (Buck 2001:108-109). Näin meneteltiin osittain puhumisen ja kirjoittamisen tehtävien yhteydessä tutkimuksen kohteena olevan arvioinnin tehtäviä laadittaessa. Toisaalta voidaan tehdä tarveanalyysi, jonka avulla määritetään tehtävä, joka on tarpeen tietyssä kohdekielen tilanteessa, ja käytetään sitä pohjana konstruktin määrittämiselle ja kokeen suunnittelulle (Buck 2001:111).

Konstruktit voidaan myös määrittää kolmesta pääviitekehyksestä käsin: lingvistisesti orientoitunut, kommunikatiivisesti orientoitunut ja tilannepohjainen. Kaikki kolme lähestymistapaa voidaan myös yhdistää. Esimerkiksi kirjoittamisen kokeessa, joka on tarkoitettu heikommille oppilaille, voidaan päättää kiinnittää huomiota pääasiassa lingvistiseen ja tekstuaaliseen tietoon tai kieliopin ja sanaston hallintaan pikemminkin kuin kielen toimivuuteen kommunikatiivisissa tilanteissa. Jos taas arvioinnin kohteena on kieltä paremmin hallitseville testattaville kohdennettu suostuttelevan tekstin kirjoittaminen, on painotettava funktionaalista ja sosiolingvististä tietoa. Edellä mainitut asiat pyrittiin ottamaan huomioon tämän tutkimuksen kohteena olevaa koetta laadittaessa käyttämällä tuotosten arvioinnissa useita eri asteikkoja (luku 8). Myös aihetiedon rooli on määriteltävä, erityisesti silloin, kun testattavien yleistiedon tasosta ei ole varmuutta, koska määritelmän mukaan kirjoittamisen täytyy aina käsitellä jotakin (Weigle 2002:79-80). Tärkeintä kuitenkin lienee, että koe mittaa tavoitteiden suuntaisesti.

Purpuran (2004 s. 99-100) mukaan kieliopillisen muodon tunteminen liittyy fonologiaan, sanastoon, morfosyntaksiin, koheesioon, informaation hallintaan ja vuorovaikutukseen. Kuvio 21 sivulla 104 esittää kielioppitiedon komponentit, joiden määrittelystä pääsee sellaisiin kieliopin kohtiin, joita voi käyttää kunkin komponentin hallinnan konstruktin pohjana. Kommunikatiivisesti orientoituneiden konstruktien käsitteet määritellään ensin yleisellä tasolla, esimerkiksi kertominen tai mielipiteen ilmaiseminen, minkä jälkeen täsmennetään puitteet, jotka rajaavat ja määrittävät onnistuneen suorituksen. Tilannepohjainen konstrukt kuuluu edellä käsiteltyyn tehtäväpohjaiseen lähestymistapaan, ja suunnittelun lähtökohtana ovat testattavan kielelliset tarpeet kyseisessä tilanteessa (Luoma 2004:162-163).

Konstruktin operationaalistaminen saattaa olla riittämätön kahdesta syystä: jotakin sellaista, jonka pitäisi olla mukana, on saattanut jäädä pois tai siihen on lisätty jotakin, mikä ei sinne kuulu. Messick (1994) käyttää termiä *construct-underrepresentation* (konstruktin aliedustus) viittaamaan edelliseen ja *construct-irrelevant* (konstruktin kannalta asiaankuulumaton) kuvaamaan jälkimmäistä. Molemmat vaikuttavat kokeen validiuteen. (Buck 2001, McNamara 2000). Se, että tässä tutkimuksessa arvioitavan oppilaat olivat opiskelleet perusopetuksen opetussuunnitelman perusteiden mukaisen oppimäärän, takaa jossakin määrin samanlaiset taidot, vaikka yksilöiden välisiä eroja muodostuu oppivelvalli-

suuskoulussa monistakin syistä. Jos konstruktissa on jotakin liikaa⁸, on selvitetävä, aiheuttaako se varianssia. Usein kokeessa joudutaan käyttämään muitakin taitoja kuin ne, jotka ovat juuri mittauksen kohteena, mutta jos testattavien kyseiset taidot ovat samanlaisia, varianssia ei esiinny. Testattavien suoriutuminen kokeesta voi vaihdella tehtävätyypistä riippuen. Jos kokeeseen sisällytetään erilaisia tehtävätyyppejä, vältetään se vaara, että jokin tehtävätyyppi vaikuttaisi liikaa koetulokseen (Erickson 1999:220; Buck 2001:100-101). Kokeessa käytettyjen eri tehtävätyyppien suhteellista osuutta käsitellään luvussa 8. Mitattavat konstruktit luovat pohjan operationaalistamisvaiheen koetehtävien laatimiselle ja tukevat myös koetuloksista tehtävien tulkintojen konstruktivalidiutta.

5.1.2 Operationaalistaminen

Operationaalistaminen yhtenä kokeen laadinnan pääkomponenttina on vaihe, jossa siirrytään yleissuunnitelmasta varsinaisen kokeen laadintaan. Jokaisella kokeen järjestäjällä on jokin käsitys siitä, millaista kielitaitoa testataan, miten arviointi toteutetaan ja millaisia arviointikriteerejä käytetään. Näiden ajatusten esittämistä kirjallisessa muodossa kutsutaan koe- eli testispesifikaatioiksi. Ne sisältävät laatijan määritelmän arvioitavista konstruktista ja yksityiskohtaiset kuvaukset tehtävistä ja arviointikriteereistä (Popham 1978, Takala 1980, Davidson & Lynch 1993, Alderson ym. 1995, McNamara 2000, Lynch 2003, Luoma 2004). Ne pidetään usein salaisina, jotta uusia versioita voitaisiin kehittää (Buck 2001:197). Tämä johtuu luultavasti "tuotesalaisuuden" säilyttämisyrittämisestä. Mm. Luoma on pohtinut syitä tähän (Luoma 2001:148).

Koespesifikaatioiden laatimisen yhteydessä tehdään tärkeimmät yksittäisiä koetehtäviä koskevat päätökset. Niihin kirjataan perusteet sille, miksi arviointi kohdistuu tiettyihin konstruktihin ja kuinka tehtävät ja kriteerit operationaalistavat ne (Luoma 2004:113). Bachman ja Palmer (1996) perustelevat niiden merkitystä useilla laadunvarmennuksen näkökulmilla. Ne

- auttavat rinnakkaisten, samoja ominaisuuksia arvioivien kokeiden laadinnassa,
- toimivat keinona kokeenlaatijan tavoitteiden selvittämisessä,
- tarjoavat mahdollisuuden verrata lopputulosta spesifikaatioihin ja
- ovat keino arvioida kokeen autenttisuutta (1996:177).

Koska tutkijat ovat yhä epävarmoja siitä, mitkä muuttujat vaikuttavat käsitevalidiuteen, spesifikaatioiden tulisi mielellään olla mahdollisimman yksityiskohtaisia (Alderson 2000:124).

Alderson (2000) suosittelee, että testispesifikaatioihin sisällytettäisiin ainakin suurin osa seuraavista kohdista:

⁸ Vrt. Töttö 2005

1. Kokeen tarkoitus
2. Testattavan tiedot (ikä, sukupuoli, kielitaidon taso)
3. Kokeen taso (suhteessa testattavien taitoon)
4. Koekonstruktio
5. Kuvaus sopivasta kielikurssista tai oppikirjasta
6. Kokeen osien lukumäärä
7. Kuhunkin osaan käytettävä aika
8. Kunkin osan painotus
9. Kohdekielen tilanne/tilanteet
10. Tekstilajit
11. Tekstien pituus
12. Tekstien vaikeustaso
13. Testattavat taidot
14. Kielen elementit (rakenteet/sanasto/funktiot)
15. Tehtävätyypit
16. Osien lukumäärä ja painotus
17. Testaustavat
18. Ohjeet, esimerkit, arviointikriteerit
19. Pisteitysohjeet
20. Kuvaus eri tasoista tyypillisistä tuotoksista
21. Kuvaus siitä, mitä eri tasoiset testattavat osaavat tehdä kyseisellä kielellä todellisessa elämässä
22. Esimerkkejä kokeista
23. Esimerkkejä testattavien suorituksista (Alderson 2000:169-70; Ks. myös Alderson, Clapham & Wall 1995: 9-38).

Edellisestä, erittäin käyttökelpoiseksi osoittautuneesta luettelosta voidaan alkuperäisessä, tämän tutkimuksen kohteena olevassa arvioinnissa katsoa otetun huomioon jokseenkin kattavasti kohdat 1-18 (luku 8). Kohtien 19-23 toteuttaminen jäi puutteelliseksi syistä, joita esitetään luvussa 11.

Koespesifikaatiot hyödyttävät muitakin osapuolia kuin kokeenlaatijoita. Ne voivat olla myös erilaisia erilaisille kohdeyleisöille (mm. Weigle 2002:83). On kuitenkin muistettava, että testaus on väistämättä kompromissi ihanteen ja käytännön välillä yritettäessä yhdistää testisuoritus kohdekielen käyttöalueeseen (Alderson 2000:154).

Paitsi yhtä kokeen laadinnan vaihetta operationaalistamisella tarkoitetaan konstruktin täsmentämistä, mikä johtaa kokeen laadinnan päämäärään, koeosien kehittämiseen. Näin ollen pyritään suunnittelemaan kokeita, jotka arvioivat tavoitetaitoja joko erikseen tai integroidusti. Arvioitiinpa taitoja millä tavalla tahansa, tehtävät muodostavat osan teoreettisesta konstruktista, mutta ne operationaalistetaan eri tavoilla (Alderson 2000:118). Koespesifikaatioiden tulisi ilmaista yksityiskohtaisesti, millaisia näiden erilaisten tehtävien tulisi olla, sillä kukin yksittäinen osio saattaa operationaalistaa vain osan konstruktia, mutta yhdessä osien tulisi edustaa koko konstruktia. Kokeella eli siitä tehtävillä johtopäätöksillä, tulkinnoilla, on käsitevalidiutta samassa suhteessa kuin missä tämä tavoite toteutuu (Buck 2001:116). Tärkein tekijä tehtävien suunnittelussa on konstruktin liittyvä tieto, jonka pisteet antavat, toisin sanoen, mitä pisteiden käyttäjän tulee tietää pisteiden saajan kielitaidosta. Tämä sanelee sen, millaisia taitoja tehtävien tulisi saada testattava osoittamaan (Luoma 2004:29).

Suomessa on testispesifikaatiota laadittu yleisiin kielitutkintoihin, ns. AMKKIA -projektiin, valtionhallinnon kielitutkintoihin ja inkerinsuomalaisten paluumuuttajien kielitutkintoon. Tämän tutkimuksen pohjana olevaan arviointiin niitä siis laadittiin jossakin määrin. Nyt myös ylioppilastutkinnon kielikokeisiin niitä ollaan laatimassa (Sauli Takala, sähköpostitiedonanto 15.10.2006).

5.1.3 Toimeenpaneminen

Kolmas kokeen laadinnan vaihe on kokeen toimeenpaneminen. Se kattaa varsinaisen kokeen järjestämisen ensin esikokeiluna tai mahdollisesti pilottina⁹. Sen aikana kootaan palautetta testattavien kielitaidosta, koejärjestelyistä ja itse kokeesta. Varsinaisessa kokeessa kerätään tietoa kokeen toimeenpanosta, mm. ajankäytöstä ja ohjeista sekä testattavien koevastauksista (Bachman & Palmer 2004:245). Arvioinnin toimeenpanoon kuuluu lisäksi tulosten analysointi, mikä perusteella tehdään johtopäätöksiä arvioinnin kohteena olleesta kielitaidosta.

Huolimatta siitä, että käytettäväksi ehdotettu viitekehys osoittaa kokeen laadinnan olevan lineaarista, se ei sitä kuitenkaan ole, kuten alkuperäinen arviointi (luku 8) osoittaa. Työn edetessä joudutaan esimerkiksi resurssien (henkilöstö, materiaali, aika) allokointia ehkä muuttamaan useaankin kertaan (Bachman & Palmer 2004:86-9).

5.1.4 Kokeen käyttökelpoisuuden piirteiden arviointi

Bachmanin ja Palmerin (1996) kielitaidon arvioinnin lähestymistapa perustuu seuraavien ehtojen toteutumiseen:

- Kielellinen suoritus vastaa todellisen elämän kielenkäytön vaatimuksia ja
- Kokeen käyttökelpoisuus, jonka komponentit (reliaabelius, käsitevalidius, autenttisuus, vuorovaikutteisuus, vaikutus ja käytännöllisyys) on määriteltävä selvästi ja yksiselitteisesti. Näiden komponenttien laatua ja määrää on tarkkailtava niin kokeen suunnittelu-, laadinta- kuin käyttövaiheessakin. Tämä osoitetaan kuvion 22 oikeanpuoleisessa ympyrässä.

Ensimmäisen periaatteen toteutumiseksi koetta suunniteltaessa on kiinnitettävä huomiota siihen, että sekä kielenkäyttötilanteiden ja koetehtävien ominaisuudet että todellisen elämän kielenkäyttäjien ja testattavien ominaisuudet vastaavat toisiaan. Näiden ominaisuuksien perusteella tehtävät johtopäätökset voidaan yleistää koskemaan kyseisiä kielen käyttöalueita (*domains*). Testattavien tärkein ominaisuus on kielitaito, koska siitä on tarkoitus tehdä johtopäätöksiä. Muita ovat aihetieto ja affektiiviset skeemat, affektit, jotka myös vaikuttavat voimakkaasti sekä todellisen elämän kielenkäyttöön että koesuoritukseen ja joiden huolellisella huomioon ottamisella voidaan myös merkittävästi helpottaa testattavien koesuoritusten onnistumista. Näitä teemoja käsitellään luvuissa 6 ja 8.

⁹ Esikokeilu = Pre-testing, kokeiden tai niiden osien testaaminen ennakoita. Pilot-ti(kokeilu) = Pilot testing, projektin, jonkin kokonaisuuden toimivuuden testaaminen.

Kokeen käyttökelpoisuuden komponenttien sisällöt Bachman ja Palmer (1996) määrittelevät seuraavasti. Reliaabeliuden voidaan katsoa olevan pisteiden johdonmukaisuuden funktio kahden kokeen ja niiden osioiden välillä. Ihannetapauksessa testattavan tulisi siis voida osallistua kumpaan tahansa kahdesta rinnakkaiseksi arvioidusta kokeesta ja saada sama tulos. Myös ryhmän paremmuusjärjestyksen tulisi olla sama, jotta kokeen voi katsoa täyttävän reliaabeliuden vaatimukset ja olevan luotettava mitattavaksi aiotun taidon indikaattori. On syytä todeta, että edellä esitetty on reliaabeliuden käsitteellinen määritelmä. Sen empiirinen operationaalistaminen tapahtuu lähes aina vain yhden kokeen sisäisen laadukkuuden perusteella (Bachman & Palmer 1996:19-20). On myös yhä korostettava, että kyse on pistemäärien, ei kokeen reliaabeliudesta (mm. Thompson 2003).

Käsitevalidius liittyy kokeen tulosten perusteella tehtäviin tulkintoihin (ei itse kokeeseen, kuten usein esitetään; mm. Messick 1988, Luoma 2001). Kun tuloksia käsitellään testattavan kielitaidon indikaattoreina, on kyettävä perustelemaan tulosten perusteella tehtävät tulkinnat. Tämän vuoksi on oltava todisteita siitä, että tulokset heijastavat juuri sellaista kielitaitoa, jota on haluttu mitata. Konstrukti on siis sen taidon määritelmä, joka muodostaa tietyn kokeen tai tehtävän perustan siitä saatavan tuloksen tulkitsemiselle. Käsitevalidius liittyy myös siihen yleistettävyyssalveeseen, josta tulosten tulkinnalla voidaan tehdä yleistyksiä (Bachman & Palmer 1996:20).

Kokeen tarkoitus vaikuttaa sekä reliaabeliuden että validiuden tasoon. Tärkeissä kokeissa edellytetään korkeampaa reliaabeliutta ja validiutta kuin vähemmän tärkeissä. Kapea-alainen konstrukti samoin kuin usea suhteellisen yhtenäinen konstrukti ovat omiaan kohottamaan reliaabeliutta. Korkea käsitevalidius taas varmistetaan mahdollisimman korkealaatuisella kokeenlaadintaprosessilla ja sen osoittaminen keräämällä mahdollisimman paljon aineistoa sen tueksi. Tämän tutkimuksen yhteydessä tätä laadunvarmistusta kuvataan luvussa 8.

Autenttisuus voidaan määritellä tietyn kielikokeen tehtävien ominaisuuksien ja todellisen elämän kielenkäyttötehtävän välisen vastaavuuden määränä. Autenttisuuden tason osoittamiseksi Bachman ja Palmer (1996) suosittelevat kohdekielen käyttöalueen piirteiden kirjaamista koko kokeenlaadinnan ajan. Tehtävät voivat kuitenkin olla suhteellisen hyödyllisiä aiottuihin tarkoituksiinsa, vaikka niiden autenttisuuden tai prosessoinnin tarpeen määrä olisikin vähäinen. Kvalitatiivista tukea sekä autenttisuudelle että interaktiivisuudelle voidaan hankkia myös esikokeilun yhteydessä pyytämällä testattavia arvioimaan näiden piirteiden esiintymistä. Tutkimuksen kohteen kaltainen laaja kansallinen koe ei kuitenkaan mahdollista tätä. Testattavan kannalta koe on hyödyllinen, jos koetehtävien aiheet ja tehtävätyypit ovat asianmukaisia, todellisen elämän tarpeisiin liittyviä. Tällöin myös hänen affektiivinen reaktionsa koetta kohtaan on todennäköisesti positiivinen, ja se auttaa häntä osoittamaan kykynsä parhaalla mahdollisella tavalla (Bachman & Palmer 1996:23-24).

Interaktiivisuus viittaa tässä yhteydessä kielen prosessointiin. Sillä tarkoitetaan sitä, missä määrin ja millä tavoin testattava käyttää tehtävän suorittamiseen yksilöllisiä ominaisuuksiaan, joita tässä tapauksessa ovat kielitaito (=

kielitieto ja strateginen kompetenssi), aihetieto ja affektiiviset skeemat. Jotta siis kielitaidosta voitaisiin tehdä päätelmiä, kokeen on testattava sekä kielitietoa että strategista kompetenssia. Interaktiivisuus on tärkeä linkki tehtävän käsittevalidiuden osoittamiseen. Alin hyväksyttävä taso löytyy niistä prosessointitapaa osoittavista piirteistä, jotka sisältyvät konstruktiin (Bachman & Palmer 1996:25-6).

Vaikutuksen määrä arvioidaan suhteessa vaikutettaviin, joita voivat olla asianosaiset, koulutusjärjestelmä, yhteiskunta, testattavat, opetus tai opettajat. Näissä tapauksissa vaikutus ilmaistaan tavallisesti niinä negatiivisina ilmiöinä, jotka tulisi minimoida. Eniten keskustelua on herättänyt kokeiden takaistusvaikutus (*washback*), erityisesti silloin, kun se kohdistuu kielitaitoon prosessina eli kielen opetukseen ja oppimiseen. Bachman ja Palmer (1996) ehdottavatkin, että arvioinnin yhteydessä

- listattaisiin mahdollisimman yksityiskohtaisesti kokeen käyttötarkoitus,
- kirjattaisiin sekä positiiviset että negatiiviset seuraukset kokeen käyttämisestä tällä tavoin,
- kirjattaisiin tulokset niiden toivottavuuden ja epätoivottavuuden mukaisessa järjestyksessä
- koottaisiin tietoa siitä, miten todennäköisesti näitä erilaisia tuloksia ilmenee (Bachman & Palmer 1996:30-33).

Viimeinen kokeen käyttökelpoisuuden ehto on käytännöllisyys, joka eroaa edellisistä siinä, että se liittyy kokeen toimeenpanoon eikä tulosten tulkintaan kuten edelliset, ja sen määrä riippuu siitä, miten käytettävissä olevat resurssit vastaavat koespesifikaatioiden asettamia vaatimuksia. (Bachman & Palmer 1996:35-36).

Kokeen laadintaan tarvittavat resurssit ovat kolmenlaisia: ihmiset, joilla on tarvittavat taidot, materiaaliset resurssit ja riittävästi aikaa (Bachman & Palmer 1996:37; Buck 2001:101). Resurssit ovat tärkeitä (Weigle 2002:81), mutta kokeenlaatijoiden täytyy useinkin toimia olemassa olevissa olosuhteissa, vaikka heiltä puuttuisikin resursseja tehdä työnsä niin hyvin kuin mahdollista (Buck 2001:101). Näihin viitataan luvussa 8.

5.2 Arviointimenetelmät

Arviointimenetelmän valintaan vaikuttaa paitsi arvioinnin tarkoitus myös sen kohderyhmä. Tällä perusteella voidaan vieraan kielen arvioinnissa erottaa erilaisia menetelmiä (taulukot 26 ja 27).

TAULUKKO 26 Arviointimenetelmiä. (Lähde: Takala1997:91-92.)

Diagnostinen arviointi	Formatiivinen arviointi	Summatiivinen arviointi
Pyritään saamaan selville testattavan lähtötilanne tai oppimisvaikeuksien syyt.	Selvitetään, onko opetettu omaksuttu.	Kartoitetaan oppimistuloksia oppimäärän lopussa yleensä arvosanan antamiseksi.
Saavutusmittaus	Taitotason arviointi	
Arvioidaan tietyn kurssin tai oppimäärän tavoitteiden saavuttamista.	Arvioidaan määriteltyjen taitotasojen saavuttamista oppimäärän tavoitteista riippumatta.	
Standardikokeet	Ei-standardisoidut kokeet	
Käytetään laajoissa arvioinneissa, joissa on ollut mahdollisuus laajaan kokeiden esitestaukseen.	Standardikokeiden vastakohta, jota Suomen arvioinneissa käytetyt kokeet yleensä edustavat.	
Sisäinen arviointi	Ulkoisen arviointi	
Oppilaitoksen oman henkilökunnan suorittama arviointi, jonka muoto voi olla - itsearviointi - toveriarviointi - testaajan/opettajan suorittama arviointi	Oppilaitoksen ulkopuolisen tahon suorittama arviointi	
High stakes -arviointi	Low-stakes -arviointi	
Arvioinnit, joissa menestyminen on testattavalle erittäin tärkeää, esim. Suomen ylioppilastutkinto.	Arviointi, jonka tulokset eivät esimerkiksi estä testattavan jatko-opintoja tai työpaikan saantia.	
Tutkintoarviointi	Ohjelma-arviointi	
Kaikki suorittavat saman kokeen ja saavat todistuksen suorituksestaan	Perustuu otokseen sekä testattavien että tehtävien osalta. Todistusta ei anneta. Voi olla myös kansainvälinen.	
Normiviitteinen arviointi	Kriteeriviitteinen arviointi	
Testattavan saama arvosana riippuu siitä, kuinka hyvin hän on suoriutunut suhteessa muihin osallistujiin.	Testattavien suorituksia verrataan taitotason kriteereihin ja saatujen arvosanojen jakaantuminen on täysin riippuvainen siihen osallistuvien osaamistasosta.	

Kun kokeen tarkoituksena on opitun arviointi suhteessa ohjelman (= opetus suunnitelman) tavoitteisiin eli opetustavoitteiden hallinta kurssin tai ohjelman päättyessä, kokeen tulisi olla taulukon mukaan kriteeriviitteinen. Jokaisen tehtävän tulisi siis mitata opetussuunnitelmassa hallittavaksi esitettyjä tavoitteita tasolla *Osa* / *Ei osaa*.

TAULUKKO 27 Koetyypin valinnan ja arviointitarkoituksen välinen suhde. (Lähde: Brown 1995:118.)

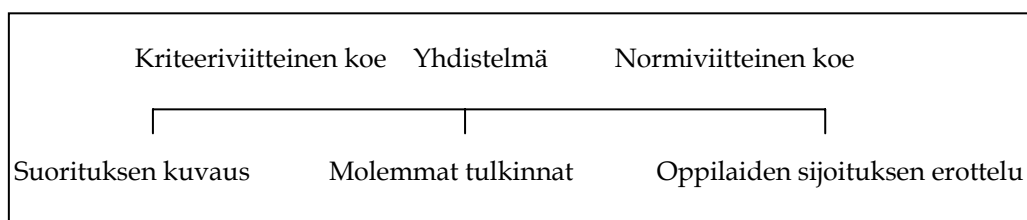
Johtopäätöksen / Kokeen tarkoitus				
Kokeen ominaisuudet	Yleinen kielitaito	Opiskeluryhmään sijoittaminen	Tavoitteiden hallinta	Diagnosointi
1. Kokeen tuotettaman tiedon yksityiskohtaisuus	Hyvin yleinen	Yleinen	Yksityiskohtainen	Hyvin yksityiskohtainen
2. Fokus	Yleiset ohjelmaan pääsyn edellyttämät taidot	Kurssilla käsiteltävät asiat	Kurssin tai opinto-ohjelman määritellyt tavoitteet	Kurssin tai opinto-ohjelman määritellyt tavoitteet
3. Arvioinnin/päätöksenteon tarkoitus	Opiskelijan tason vertaaminen muihin	Opiskelijan tasolle sopivan opiskeluryhmän löytäminen	Opitun arviointi suhteessa ohjelman tavoitteisiin	Opitun arviointi suhteessa ohjelman tavoitteisiin
4. Vertailun tyyppi	Vertailu muihin oppilaitoksiin	Vertailu ohjelmassa esille tuleviin opiskeltaviin asioihin	Opitun arviointi suhteessa ohjelman tavoitteisiin	Opitun arviointi suhteessa ohjelman tavoitteisiin
5. Milloin pidetään	Alku- tai loppumittauksena	Kurssin alussa	Kurssin tai ohjelman päättyessä	Kurssin tai ohjelman alussa tai kuluessa
6. Pisteiden tulkinta	Pistemäärien jakauma	Pistemäärien jakauma	Opetustavoitteiden hallinta	Opetustavoitteiden hallinta
7. Kokeen tyyppi	Normiviitteinen	Normiviitteinen	Kriteeriviitteinen	Kriteeriviitteinen

Koetta kutsutaan normiviitteiseksi, jos se mittaa suoritusta suhteessa testattavan tehtävien hallintaan tietyssä ryhmässä. Koe on kriteeriviitteinen, jos sen tavoitteena on mitata suoritusta suhteessa selkeästi määritellyä ja rajattua kielenkäyttöaluetta edustavien tehtävien hallintaan (Linn & Gronlund 1995:16). Taulukossa 27 tämän arvioinnin tyyppi määritellään kriteeriviitteiseksi. Kriteerimittauksen painopiste oli alkujaan Yhdysvalloissa. Takala (1980) mainitsee jo Thorndiken vuonna 1913 osoittaneen selvästi suhteellisen ja absoluuttisen mittaamisen välisen eron. Suomessa ainakin Vahervuo (1958) tutki jo varhain vastaavia 'A- ja S-systeemien' välisiä eroja. Robert Glaserin sanotaan varsinaisesti tuoneen esille käsitteet normiviitteinen ja kriteeriviitteinen mittaaminen (Popham 1978). Sen taustalla oli ohjelmoidun opetuksen tulosten arviointi, jossa ajatuksena oli, että tehokkaan opetuksen tuloksena asia(t) opitaan ja tulokset ovat hyviä. Tukena näille ajatuksille olivat samoihin aikoihin syntyneet käyttäytymistavoitteiden määrittelyt (Mager 1970). Popham itse on käsitellyt seikkaperäisesti kriteeriviitteisten kokeiden osionlaadintaa ja oppilaiden suoritusten kuvausta. Hän myös katsoi, että kriteeriviitteisen kokeen testispesifikaatioiden tulisi sisältää seuraavat viisi komponenttia:

1. Kokeen mittaaman käyttäytymisen yleinen kuvaus.
2. Mittaustapaa havainnollistava esimerkkiosio.
3. Hallittavan ärsykeaineiston ominaisuuksien kuvaus ja rajaus.
4. Edellytettyjen vastauksien ominaisuuksien kuvaus (joko valittavien vaihtoehtojen ominaisuuksien kuvaus tai tuotettavan vastauksen arviointikriteerien määrittely).
5. Joissakin tapauksissa lisätäsmennyksiä erillisessä liitteessä (Takala 1980).

Normiviitteistä arviointia luonnehtii esimerkiksi se, että jonkun sanotaan olevan parempi kuin 80 % luokasta. Tällöin tiedetään henkilön sijoitus suhteessa tuohon viiteryhmään, mutta ei sitä, mitä hän osaa tai ei osaa tehdä. Kriteeriviitteinen tulkinta ilmaisee arvioidun suorituksen määritelyjen konstruktien ja niistä johdettujen tehtävien perusteella, jotka henkilö, joka on saavuttanut tietyn pistemäärän, osaa suorittaa (Takala 1980, Nitko 2001, Lynch 2003, Luoma 2004). Sen taustalla on viime aikoina kaikenlaiseen julkiseen toimintaan liitetty avoimuuden (läpinäkyvyyden) ja tilivelvollisuuden vaatimus. Opetuksen halutaan olevan tehokasta ja oppimistulosten osoittavan, missä määrin asetetut tavoitteet on saavutettu. Popham luonnehtii kriteeriviitteistä arviointia toteamalla, että jos testattava hallitsee 85 % tehtävistä, hänen tuloksensa voidaan yleistää koskemaan kaiken sen alueen hallintaa, josta koetehtävät olivat otos (Popham 1999:99). Kriteeriviitteiseen arviointiin soveltuvan kokeen laadinta edellyttää todellista tai tuotettavissa olevaa osioiden perusjoukkoa, joiden on määritelty testaavan tiettyä asiaa. Tästä perusjoukosta otetaan satunnaisotos tai ositettu satunnaisotanta, jonka suoritustaso on yleistettävissä kyseisen asian osaamiseen (Takala 1980:8). Tarkkojen johtopäätösten tekeminen kriteerikokeen perusteella edellyttää siis tarkasti määriteltyjä kielenkäyttöalueita. Koetta tarkastelemalla ei voi päätellä, kumpaan kategoriaan koe kuuluu. Ero tulee esiin paitsi siinä, millaisia johtopäätöksiä kokeesta tehdään myös esimerkiksi siinä, miten se on esikokeilun jälkeen koottu. Jos kokeesta ensimmäisen osioanalyysin jälkeen poistetaan ”liian” helpot osiot ja pyritään keskivaikeisiin, päästään sellaiseen pistejakaumaan, että testattavien tehokas sijoittaminen eri tasoille on mahdollista. Jos taas kokeeseen jätetään kaikki ne tehtävät, jotka on laadittu tuottamaan tietoa opetustavoitteista, ollaan tekemisissä kriteeriviitteisen arvioinnin kanssa. Jos tehtävät osoittautuvat helpoiksi, tavoitteet on asetettu matalalle ja päinvastoin (Linn & Gronlund 1995:17).

Eri arviointitavat eivät välttämättä ole toisensa pois sulkevia. Jos normiviitteistä arviointia käytettäessä havaitaan puutteita testattavan taidoissa, niitä voidaan selvittää kriteeriviitteisellä (Nitko 2001:13). Näiden kahden mittaustavan yhdistämistä, niiden pitämistä mittausjatkumon ääripäinä, kuvataan kuviossa 23.



KUVIO 23 Kriteeri- ja normiviitteisen arvioinnin jatkumo. (Lähde: Linn & Gronlund 1995:17.)

Tehokas kriteerimittaaminen edellyttää ohjelma-arvioinnissa siis

- mitattavan alueen tarkkaa kuvaamista,
- mittarin tarkkaa kuvausta, joka sisältää sekä ärsyke- että reaktio-osan erittelyn,
- vastausten pisteitysjärjestelmän,
- validit tehtävät,

- tehtävät, jotka ovat edustava tai ositettu satunnaisotos mitattavalta alueelta,
- edustavaa oppilasotosta ja
- oppilaiden ja osioiden otostamista eli matriisiotantaa (Takala 1980:109).

5.3 Arviointiasteikon valinta

Produktiivisen kielitaidon arvioinnin kannalta keskeisiä kysymyksiä ovat arviointiasteikkojen valinta ja se, miten arvioijat niitä käyttävät. Weigle (2002) toteaa kirjoittamisesta, että kirjoitustehtävän pisteet eivät ole tulosta testattavan ja testin välisestä vuorovaikutuksesta, vaan vuorovaikutuksen osapuolia ovat testattavan lisäksi tehtävän ohjeet tai virike, tuloksena oleva teksti, arvioijat ja arviointiasteikko (Weigle 2002:108)¹⁰. Sama määritelmä soveltuu myös puhumisen arviointiin.

Arviointiasteikkojen kolme päätyyppiä ovat kohdistettu piirrearviointi l. retorinen arviointi (*primary trait*) (Takala 1987:31-32) sekä holistiset ja analyttiset (Weigle 2002:109) asteikot. Arviointityypit voidaan ryhmitellä myös sen perusteella, millaista palautetta testattava saa kielitaidostaan. Seuraavaa Huhdan (1990:2) jakoa suullisen kielitaidon palautteisiin voidaan soveltaa myös kirjoittamiseen. Testattava saa

- kielitaidon kokonais- tai yleisarvion,
- vain osataitojen erilliset arviot tai
- edellisten yhdistelmän.

Taulukkoon 28 on ryhmitelty produktiivisen kielitaidon suoritusten pisteittämisvaihtoehtoja. Mitään tyyppiä ei sellaisenaan voi pitää muita parempana, vaan käyttökelpoisuus riippuu mm. kokeen tavoitteista, arviointiin käytettävissä olevasta ajasta sekä arvioija- ja taloudellisista resursseista (Bachman ja Palmer 1996).

¹⁰ Tässä yhteydessä Weigle viittaa myös Hamp-Lyonsiin, Kenyoniin ja McNamaraan (1996).

TAULUKKO 28 Produktiivisen kielitaidon suoritusten pisteittämisvaihtoehdot. (Muokattu taulukoksi Huhdan 1990:2 mukaan.)

Yksi arvosana			Usea arvosana	
<i>Suora</i> (kokonaisvaltaisesti)	<i>Epäsuora</i>		Piirteet arvioidaan erikseen (analyttisesti), kokonaisarvioita ei anneta	- tasokuvausten avulla - ilman tasokuvauksia
Ilman tasokuvauksia	Yksittäisistä kriteereistä kokoamalla (tasokuvausten avulla tai ilman)	- yhtä tai useampaa kriteeriä painottaen		
Tasokuvausten avulla		- kutakin kriteeriä yhtä paljon painottaen	- vaikutelman mukaan - mekaanisesti (summa tai keskiarvo)	Piirteet arvioidaan sekä osina että kokonaisuutena
	Yhdistämällä kokonaisvaltainen ja yksittäisten kriteerien arviot			

Arviointitapaa kutsutaan holistiseksi, jos testattavalle annetaan kokonaisarvio suorituksesta. Ellei tasokuvauksia käytetä, kyseessä on impressionistinen arviointi, jota aikaisemmin käytettiin runsaasti (Weigle 2002:112). Tällöin arvioinnin tueksi annetaan usein maamerkkisuorituksia, joihin arvioijat voivat tukeutua. Tasokuvauksia käytettäessä arviointi tarkentuu, koska ne ilmaisevat, mihin piirteisiin on tarkoitus kiinnittää huomiota, vaikka niitä ei arvioida erikseen. Esimerkkejä holistisista arviointiasteikoista ovat TOEFL¹¹- kirjoittamisen kokeen (Weigle 2002:112), YKI¹² -ja ACTFL/OPI¹³-suulliset kokeet (Luoma 2004:67) sekä CEFin¹⁴ Yleisen suullisen ja kirjallisen kielitaidon arviointitaulukot Overall Oral Production (2001:58) ja Overall Written Production (2001:61). Yksi asteikkojen ongelma on se, että asteikoissa heikkoja suorituksia kuvataan usein negatiivisesti: "ei osaa" sen sijaan että todettaisiin "osaa". Holistisessa arvioinnissa voidaan kuitenkin kiinnittää huomiota suorituksen ansioihin puutteiden sijasta (Weigle 2002:112). Laatimalla tarkat pisteitysohjeet voidaan arvioijat saada kiinnittämään huomiota juuri niihin asioihin, joita halutaan mitata. Holistisen arvioinnin heikkous on se, että yksi arvosana/pistemäärä ei anna riittävää tietoa suorituksen puutteista. Tätä kuitenkin tarvittaisiin erityisesti vieraan kielen oppimisen yhteydessä, ja esimerkiksi diagnostisessa testauksessa tulisikin pyrkiä tarkempaan kuvaukseen. Kaikkea asteikkojen käyttämistä koskeva ongelma on arvioijien erilainen asteikkojen tulkinta, ellei koulutukseen ole

¹¹ Test of English as a Foreign Language.

¹² Yleiset kielitutkinnot.

¹³ American Council on the Teaching of Foreign Languages/Oral Proficiency Interview

¹⁴ Common European Framework for Languages.

kiinnitetty riittävää huomiota (Weigle 2002:114). Tätä korostetaan myös tämän tutkimuksen johtopäätöksissä.

Jos testattavalle annetaan produktiivisesta suorituksesta usea arvostus/pistemäärä, kyse on analyttisestä arvioinnista. Tällöin perustana ovat kriteerit, piirteet, joiden toteutumista arvioidaan eri taitotasoilla. Kirjoittamisen kokeen kriteereinä käytetään yleisesti sisältöä, organisointia, sidosteisuutta, rekisteriä, sanastoa ja rakenteita (Weigle 2002:114), kun taas suullisessa kokeessa kiinnitetään usein huomiota laajuuteen, tarkkuuteen, sujuvuuteen, vuorovaikutukseen ja koherenssiin (CEF 2001:72-73). Analyttinen arviointi antaa testattavan taidoista näin ollen hollistista selvästi yksityiskohtaisempaa tietoa, mikä puolestaan on hyödyksi varsinkin silloin, kun testattavan taidot eri kriteerien suhteen ovat eri tasoilla. Testattavalle annettavat pistemäärät tai nykyaikaisempi profiilikuvaus (vrt. Ruotsi, luku 3) osoittavat kielitaidon puutteet ja ansiot. Se on myös reliabelimpi, koska osioiden, tässä tapauksessa kriteerien, lisääminen lisää myös reliabeliutta (Weigle 2002:120). Toisaalta voidaan huomauttaa, että reliabeliutta voi vaarantaa arvioijan rajallinen kyky erotella ja arvioida luotettavasti useita eri piirteitä ("tasainen rivi").

Tunnetuimpia kirjoittamisen arvioinnissa käytettäviä analyttisiä asteikkoja on neliportainen asteikko, jossa käytetään viittä kriteeriä, joita kaikkia painotetaan eri tavoin siten, että korkein, sisällöstä annettava maksimipistemäärä on 30 ja alin, mekaanisista taidoista annettava, vain 5 pistettä (Jacobs ym. 1981. Siteerattu Weigle 2002:116). Toinen tunnettu asteikko on Weirin TEEP¹⁵ -asteikko (1988). Tässä suoritukset arvioidaan kiinnittäen huomiota seitsemään kriteeriin, jotka kukin arvioidaan asteikolla 0-3, josta muodostuu testattavan kokonaispistemäärä (Weigle 2002:115). Toisaalta jos halutaan tehdä ratkaisuja/päätelmiä jokaisen kriteerin osalta, niiden kunkin arvioinnin tulee olla reliabelia, mikä on kova vaatimus. Suullisen tuottamisen analyttisistä asteikoista tällä hetkellä tunnetuimpia lienevät Eurooppalaisen viitekehysten lukuisat asteikot sekä ACTFL/OPI, jota Saleva (1997) käytti tutkimuksessaan vertailukohteena kielistudiossa suoritettavaan kokeeseen. Esimerkki suullisen kielitaidon lingvistisiä piirteitä arvioivasta analyttisestä asteikosta on CEFin puhumisen laadullisia piirteitä kuvaileva taulukko 8 (CEF 2001:28-9). Jos arviointikriteereihin kuuluu esimerkiksi vuorovaikutus, tukena voidaan käyttää taulukkoa Overall Spoken Interaction (CEF 2001:74) ja strategioita arvioitaessa esimerkiksi taulukkoa Turn Taking (CEF 2001:124). Näin meneteltiin käsillä olevassa tutkimuksessa (luku 7).

Tarkin produktiivisten taitojen arviointiasteikko on kohdistettu piirrearviointi, jossa asteikot laaditaan tiettyä tehtävää silmällä pitäen. Tällöin pisteitsohjeisiin sisältyvät

- itse kirjoitustehtävä,
- maininta tehtävän ensisijaisesta retorisisestä piirteestä (diskurssityyppi),
- ohjeet siitä, miten tehtävänannon kunkin osan odotetaan vaikuttavan testattavien suoritukseen,
- tehtävän ja arvioitavan piirteen välisen suhteen arvioiminen,
- taitotason arvioinnin kuvaus,
- mallikirjoitelmia, jotka on arvioitu tiettyjen pistemäärien arvoisiksi

¹⁵ Test of English for Educational Purposes.

- malliaineille annettujen pisteiden perustelut (Lloyd-Jones 1977; Takala 1987; Weigle 2002)

Kohdistetun piirrearvioinnin tunnetuin sovellus on USAn koulujen testausohjelma NAEP¹⁶, jonka kehitti Richard Lloyd-Jones vuonna 1974 kirjoitelmien arvioinnin tutkimusta varten (Takala 1987:31). Asteikon laatiminen on kuitenkin hyvin työlästä ja aikaa vievää, vaikka sen perusteella saadaankin tarkkaa tietoa osaamisesta. Näin ollen sitä on käytetty vähän esimerkiksi vieraan kielen testaamisessa.

Reseptiivistä kielitaitoa samoin kuin kieliopillista kompetenssia arvioidaan laatimalla etukäteen vastausten pisteityssäännöt. Tällaista testausta on totuttu pitämään luotettavana. Produktiivisen kielitaidon kokeissa arviointi taas jää arvioijan harkittavaksi, ja tällaisiin tuloksiin on suhtauduttu pitkään epäilevästi. Produktiivisen kielitaidon arvioinnissa todistustaakka jää kokeen laatijan sijasta arvioijalle, minkä oikeastaan tulisi olla täysin luonnollista. Kun taitotasot määritellään tarkasti ja arvioijat koulutetaan tulkitsemaan kuvaimet sisällöltään samanlaisina, tulokset ovat hyvinkin luotettavia (Takala:1997:99; Weigle 2002:59).

Arviointiasteikkojen laatimista vaikeuttaa se, että kielitaidon kehittymisprosessista ei ole riittävästi tietoa (Huhta 1990:6). Kielitaitoa arvioidaan konstruktien kautta, jotka taas perustuvat kielitaidon piirteiden ja tasojen kuvauksiin. Reseptiivisistä taidoista ei ole kuitenkaan kyetty toistaiseksi laatimaan riittäviä tasokuvauksia, koska luotettavaa tietoa kielen ymmärtämisen prosessoinnista ja kehittymisestä ei ole. Suuressa osassa kokeita osoituksena testattavan taidoista saadaan vastauksia testaajan laatimiin kysymyksiin, mutta ei sen sijaan testattavan omakohtaista ymmärtämisestä osoittavaa tuotosta, prosessoinnista puhumattakaan (ks. esim. Takala 1997:103; Alderson 2000:150; Buck 2001:99). Testattavalle tarjotaan eri tasoisia ja eri tyyppisiä tekstejä, joihin kohdistuvat tehtävät testaavat tiettyjä oletettuja taitoja. Reseptiivisten kokeiden arviointi suoritetaan siis tavallaan etukäteen. Jos tulos on tavoitteen mukainen, tehtävän suorittaminen osoittaa kyseisen taidon hallintaa. Esimerkiksi Euroopalaisen viitekehyksen (2003) tekstin ja puheen ymmärtämisen "Osaa"-asteikot edustavat pikemminkin käyttäytymistaksonomioita kuin kielen kehittymisen teoriaa (Grid 2003:3), mikä johti puheen ja tekstin ymmärtämisen kokeiden laatimista tukevan Gridin laatijat etsimään ratkaisuja tekstien ja niihin liittyvien osioiden vaikeustason määrittämiseen.

Lisäongelmia aiheuttavat kuvainten optimaalisen pituuden ja taitotasojen lukumäärän määrittäminen. Niiden tulisi olla mahdollisimman lyhyitä, jotta ne olisivat käyttäjäystävällisiä (Luoma 2004:60). Kuvauksen niukkuus saattaa kuitenkin johtaa siihen, että arvioijat tulkitsevat niitä eri tavalla (Huhta 1990:8). Arviointitavasta riippuen käytetään yleensä 5-6 taitotasoa, joilla arvioidaan 3-7 piirrettä (Weigle 2002; Huhta 1990; Luoma 2004). Mitä vähemmän tasoa käytetään, sitä yksimielisempiä arvioijat luonnollisesti ovat. Kuitenkin esimerkiksi Pollit (1990. Siteerattu Weigle 2002) huomauttaa, että ei ole realistista odottaa arvioijien pystyvän erottamaan kirjoittamisen kokeessa edes viittä tasoa. Myös

¹⁶ National Assessment of Educational Progress.

arvioitavien piirteiden määrä saattaa arvioijasta riippuen muodostua ongelmalliseksi. Näin tapahtui esimerkiksi kansallisen englannin arvioinnin suullista osaa sensoroitaessa, jolloin yksi sensoreista katsoi voivansa arvioida vain kolmea piirrettä tavoitteeksi asetettujen viiden sijasta. CEFin mukaan kokemus on osoittanut, että yli 4 tai 5 kategorian soveltaminen alkaa aiheuttaa kognitiivista ylikuormitusta ja että 7 kategoriaa on ehdoton yläraja (CEF 2001:193).

Tasojen optimaalisesta määrästä on kuitenkin erilaisia mielipiteitä. Tunnettu arvioinnin asiantuntija Robert Ebel (1969) esitti varhain, että arviointi on luotettavampaa ja informaatiota menetetään vähemmän, kun tasoja on pikemminkin useita kuin vähän. Tätä hän pyrkii osoittamaan taulukolla 29, jossa kuvataan mittausten reliabeliutta suhteessa arviointikategorioiden lukumäärään. Jos kaksi rinnakkaismittausta korreloi .95, ja jos mittaustulokset luokitellaan 15-portaisella asteikolla, arvostelupisteparien korrelaatio luokkien keskikohdalla olisi .94, 10-portaisella asteikolla .92, 5-portaisella .85 ja 2-portaisella .63 (Ebel 1969:220). Mitä vähemmän on arviointiluokkia ja mitä luotettavampi on alkupe-
räinen arviointi, sitä enemmän reliabelius kärsii käytettäessä harvoja arviointiluokkia. Ebelin (1969) mukaan 10-15 arviointiluokkaa takaavat riittävän luotettavuuden.

TAULUKKO 29 Reliabeliuden vähenemisen suhde arviointiasteikon porrastuksen vähenemiseen. (Lähde: Ebel 1969.)

Arviointikategorioiden lukumäärä	Mittausten reliabelius				
	<i>Maksimaalinen</i>	.95	.90	.80	.70
15	.94	.89	.79	.69	.49
10	.92	.87	.78	.68	.48
5	.85	.80	.71	.62	.45
2	.63	.60	.53	.47	.33

Arviointitavan valinta ei siis ole itsestäänselvyys, ja siksi Weigle (2002:120-121) ehdottaa päätöksenteon tueksi Bachmanin ja Palmerin (1996) kokeen käyttökelpoisuuden viitekehystä. Taulukossa 30 vertaillaan holistisia ja National Assessment of Educational Progressin analyyttisiä asteikkoja suhteessa kokeen käyttökelpoisuuden kuuteen piirteeseen. Kutakin piirrettä painotetaan kuitenkin arvioinnin tarkoituksesta ja sen suorittamismahdollisuuksista käsin. Mitä vähemmän resursseja on käytettävissä, sitä yksinkertaisempaa arviointitapaa tulee käyttää.

TAULUKKO 30 Holististen ja analyyttisten asteikkojen vertailu suhteessa kokeen käytökelpoisuuden ominaisuuksiin. (Lähde: Weigle 2002:121.)

Ominaisuus	Holistinen asteikko	Analyyttinen asteikko
Reliaabelius	Alempi kuin analyyttisessä, mutta silti hyväksyttävä.	Korkeampi kuin holistisessa.
Käsitevalidius	Perustuu siihen, että kaikki kirjoitustaidon aspektit kehittyvät samalla nopeudella ja ne voidaan siksi arvioida yhdellä arvosanalla. Tulokset korreloivat kielitaitoon kuulumattomiin aspekteihin kuten tekstin pituuteen ja käsialaan.	Soveltuvat vieraan kielen arviointiin paremmin, koska kirjoittamisen taidon aspektit kehittyvät eri nopeudella
Käytännöllisyys	Suhteellisen nopea ja helppo.	Aikaa vievä, kallis.
Vaikutus	Yksi arvosana saattaa jättää piiloon kirjoittajan epätasaisen taitoprofiilin ja johtaa harhaan esimerkiksi sopivan kurssin valinnassa.	Useampien asteikkojen käyttäminen antaa hyödyllistä diagnostista informaatiota opetusta silmällä pitäen. Käyttökelpoisempi arvioijakoulutuksen kannalta
Autenttisuus	Holistinen lukeminen on luonnollisempi prosessi kuin analyyttinen lukeminen	Arvioijat saattavat lukea holistisesti ja soveltaa kuitenkin analyyttisiä pisteitä holistisen vaikutelman arviointiin
Vuorovaikutteisuus*	Ei relevantti.	Ei relevantti.

* Bachmanin ja Palmerin määrittämä vuorovaikutteisuus tarkoittaa testattavan ja kokeen välistä vuorovaikutusta. Tähän saattaa vaikuttaa myös arviointiasteikko, jos testattava tietää, kuinka hänen suorituksensa arvioidaan.

5.4 Yhteenveto

Tunnetuimpia ja käytetyimpiä kielitaidon arvioinnin käytännön toteuttamisen kuvauksia on Bachmanin ja Palmerin laatima. Siinä on pyritty sellaiseen kielitaidon kuvaukseen, että taitokomponentit olisivat mitattavissa. Se asettaa myös selkeät laatuvaatimukset kokeen laadinnalle. Koetehtävien tulee ensinnäkin vastata mahdollisimman suurella määrällä todellisen elämän kielenkäyttötilanteita, vaikka tähän vaatimukseen vastaamisessa on käytettävä myös tervettä järkeä. Toiseksi kokeen tulee olla käyttökelpoinen, ja tämä ominaisuus määritellään suhteessa kokeen reliaabeliuteen, konstruktivalidiuteen, autenttisuuteen, interaktiivisuuteen eli tapaan, jolla kielen prosessointi otetaan huomioon kokeessa sekä vaikutukseen ja käytännöllisyyteen. Nämä komponentit eivät kuitenkaan esiinny juuri koskaan kokeessa yhtä vahvoina, vaan niiden keskinäinen painotus ratkaistaan mm. kokeen tarkoituksen ja sen laadintaan käytettävissä olevien resurssien perusteella.

Arvioinnin järjestäminen koostuu kolmesta vaiheesta. Suunnitteluvaiheessa kirjataan kokeen tarkoitus ja operationaalistamisvaiheessa täsmennetään kokeen rakenne. Kokeen toimeenpanemiseen liittyy kokeiden laadinnan jälkeen esikokeilu, jota seuraa palautteen analysointi ja mahdollisesti havaittujen puutteiden korjaaminen. Kokeen järjestämisen jälkeen toistetaan samat toimenpiteet

ja dokumentoidaan mahdollisimman tarkasti koko arviointiprosessi.

Arviointimenetelmät voidaan ryhmitellä sekä arvioinnin tarkoituksen että kohderyhmän perusteella. Kokeen ominaisuuksia ja sen soveltuvuutta tiettyyn tarkoitukseen voidaan tarkastella kokeen tuottaman tiedon yksityiskohtaisuuden, kokeen fokuksen, arvioinnin tarkoituksen, ja vertailukohtan kannalta. Tiedon tarpeesta riippuen valitaan joko normi- tai kriteeriviihteinen arviointi. Edellinen antaa tietoa pistemäärän sijoittumisesta tietylle arviointiasteikolle, jälkimmäinen mahdollisimman tarkoin määrätyn opiskelualueen hallinnasta. Ymmärtämis- ja kieliopillisen kompetenssin kokeissa johtopäätösten tekeminen kokeiden tuloksista edellyttää pisteiden tulkitsemista. Pistemäärän selville saaminen taas edellyttää pisteitystä, päätöksen tekemistä sitä, mikä kunkin osion numeerinen arvo on. Produktiivisen kielitaidon kokeet taas arvioidaan yleensä taitotasoin joko kriteerein tai ilman niitä. Asteikko valitaan kokeen tarkoituksen ja käytettävissä olevien resurssien mukaan. Valintaa tai yhdistelmää tehtäessä asteikon käyttökelpoisuus on kuitenkin arvioitava kokeen käyttökelpoisuuden kriteerien perusteella.

Produktiivisen kielitaidon kokeiden arviointia pidetään usein epäluotettavampana kuin reseptiivisten kokeiden arviointia, jossa kunkin osion pisteet on päätetty etukäteen. Kuitenkin on todettu, että jos taitotasokuvaukset on huolellisesti laadittu, koulutettujen arvioijien välinen korrelaatio ja näin ollen arvioinnin luotettavuus on hyvinkin korkea.

Taitotasojen optimaalisesta määrästä ei vallitse yksimielisyyttä. CEF katsoo 4-5 tason olevan arvioijan kannalta lähes maksimi, kun taas esimerkiksi Ebel lähtee siitä, että pistemäärien taitorasorajoja määriteltäessä tulisi erottaa 15 tasoa, mikä parantaisi merkittävästi arvioinnin luotettavuutta.

Seuraavassa luvussa käsitellään operationaalistamisen tavoitteita eli kielitaidon eri osa-alueiden arviointiin soveltuvien tehtävätyyppien valinnan perusteluita, jotka on pyritty mahdollisimman hyvin ottamaan huomioon tämän tutkimuksen kohteena olevan arvioinnin kokeiden sisällön valinnassa. Niitä kuvaillaan luvuissa 8.

6 KIELITAIDON OSA-ALUEIDEN ARVIOINTI

Kielitaito jaetaan perinteisesti neljään osa-alueeseen, puheen ja tekstin ymmärtämiseen sekä puhumiseen ja kirjoittamiseen. Niitä testaavien tehtävien lisäksi kokeisiin sisällytetään usein myös kieliopillisen kompetenssin, (kieliopillisten rakenteiden) ja sanaston tehtäviä. Sen vuoksi niitä esitellään tässä luvussa. Vuoden 1994 opetussuunnitelman perusteiden tueksi laaditussa Perusopetuksen päättöarvioinnin kriteerit -oppaassa (OPH 1999) sovelletaan em. osa-alueajattelua, samoin Eurooppalaisessa viitekehyksessä ja Suomen vuoden 2004 opetussuunnitelman perusteissa. Viimemainittuun on lisätty myös kieliopillisen kompetenssin (kieliopillisten rakenteiden) tavoitteita. Myös tutkimuksen kohteena olevan arvioinnin tehtävät noudattavat tätä jakoa.

Kielitaidon eri osa-alueiden arviointiperiaatteiden käsittelyn on tarkoitus perustella valintoja niiksi kokeiksi, joita on kuvattu luvussa 8, jossa esitetään tämän tutkimuksen kohteena oleva arviointi.

6.1 Ymmärtäminen

Kielitaidon osa-alueet voidaan jakaa kahteen ryhmään tuottamis- ja vastaanottamiskanavan perusteella (kuulo, näkö) ja kielenkäyttömuodon perusteella (produktiivinen tai reseptiivinen). Esimerkiksi Bachman ja Palmer (1996) katsovat kuitenkin, että nämä taidot eivät ole kielitaidon osia, vaan tietyssä kontekstissa ilmeneviä kykyjä käyttää kieltä (Bachman & Palmer 1996:75). On myös jossakin määrin kiistanalaista, liittyykö eri osa-alueisiin erilaisia kognitiivisia mekanismeja vai ovatko ne vain erilaisia sosiokulttuurisesti väritettyjä yleisemmän kielitaidon ilmentymiä (Weigle 2002: 14).

Ymmärtämisen kokeen suunnittelun klassinen lähestymistapa on koe-konstruktin määrittäminen valitun kielitaidon teorian pohjalta ja kokeen laadinta käyttämällä valittuja tekstejä ja tehtäviä, joita testattavien halutaan ymmärtävän. Kokeen avulla tehdään johtopäätöksiä ymmärtämisestä, mikä tyypillisesti ilmaistaan pisteinä (Alderson 2000:138). Aikaisemmin ei kiinnitetty paljon huomiota tekstin ja testikysymysten väliseen suhteeseen, vaan etsittiin tekstejä, joista sai ns. hyviä kysymyksiä. Nykyään päätetään ensin, mitä taitoja halutaan

testata, valitaan sopiva teksti ja päätellään, mikä osa tekstistä vaatisi kyseisen kohdetaidon käyttöä (Alderson 2000:255). Tärkeätä on myös yrittää tehdä kysymyksiä, joihin saatavien vastauksien vuoksi lukijat ylipäättään lukisivat/kuuntelisivat kyseisen tekstin. Tällä taas tuetaan myös kokeen autenttisuuden vaatimuksen täyttymistä (Alderson 2000:160).

Koska kilpailevat teoriat eivät ole toisensa täysin poissulkevia, testien laatijat voivat käyttää osia eri teorioista kokeen kehittämiseen (Alderson 2000:137). Näin on menetelty myös tämän tutkimuksen kohteena olevassa arvioinnissa, jonka teoriapohja, perusopetuksen opetussuunnitelman perusteet 1994, antaa kommunikatiivisen näkemyksen puitteissa varsin vapaat kädet konstruktien laatimiseen. Teorian jokaista komponenttia ei ole mielekästä testata missään kokeessa, vaan tavallisesti keskitytään niihin osiin, jotka ovat olennaisia testaus-tarkoitukselle. Ne määritellään sitten sekä teoreettisesti että operationaalisesti tähän tarkoitukseen soveltuvilla tavoilla. Mikään yksittäinen koe ei voi mitata kaikkea. (Alderson 2000:203).

Ymmärtämisen konstruktia määritellään myös tulkitsemalla taitotasoa-asteikkojen kuvaimia, jotka jakavat taidon sen mukaan, miten kielitaidon uskotaan kehittyvän. Tunnetuimpia uusia ei-kaupallisia asteikkoja ovat Eurooppalaisen viitekehityksen asteikot. Reseptiivisten taitojen kuvaaminen asteikoilla ja niin muodoin arvioiminen on kuitenkin vaikeata siksi, että tehtävien tulisi saada testattava käyttämään kokeessa niitä taitoja, joita halutaan mitata ja lisäksi osoittamaan käyttäneensä niitä. Ongelmana on, että varsinaisesti ei tiedetä, mitä taitoja ymmärtämiseen sisältyy, kuten luvuissa 4 ja 5 todettiin, ja onko käytetty koe todella mitannut juuri näitä taitoja (Hughes 1989; Alderson 2000). Muuttujia, jotka kielitaidon lisäksi vaikuttavat ymmärtämiseen, ovat mm. seuraavat vastaanottajan taidot, tiedot ja ominaisuudet:

- taustatieto
- kielen hallinta
- tekstityypin tuntemus
- sisältö
- organisatoriset piirteet
- pragmaattiset piirteet
- vastaanottamistaidot
- kuuntelemisen/lukemisen tarkoitus ja
- vastaanottajan motivaatio (Chalhoub-Deville ym. 1997; Alderson 2000; Buck 2001).

6.1.1 Kysymysten vaikeusasteen hierarkia

Ymmärtämisen tasoa testataan yleensä laatimalla kysymyksiä (ks. taulukko 31), joihin annetuista vastauksista tehdään johtopäätöksiä testattavan kielitaidon tasosta. Bachman ja Palmer (2004) määrittelevät lukemistehtävän ja vastauksen suhteen laajaksi tai kapeaksi riippuen sen edellyttämän prosessoinnin määrästä. Jos tehtävässä kysytään tekstin pääasiaa, on koko teksti prosessoitava, jolloin kyseessä on laaja suhde, kun taas jotakin yksityiskohtaa kysyttäessä suhde on kapea (Alderson 2000:163). Tekstuaalisesti yksiselitteiset (*textually explicit*) kysymykset ovat sellaisia, joiden vastaus löytyy samasta virkkeestä, kun taas tekstuaalisesti epäsuorat vaativat sitä, että lukija yhdistää tietoa yli virkkeiden.

Widdowson (1978) jakaa tekstin sisällöstä tehtävät kysymykset kahteen ryhmään. Toiset kiinnittävät huomion tiettyihin lauseisiin/virkkeisiin ja lukija vastaa osoittamalla, miten kysymyksen merkityssisältö liittyy kyseisten lauseiden merkityssisältöihin eli: Mitä tekstin lauseissa sanotaan siitä, mitä kysymyksessä kysytään? Toisen ryhmän kysymyksiin vastatessaan lukijan edellytetään tekävän johtopäätöksiä kontekstista ja käsittelevän tekstikatkelmia diskurssina eikä pelkkänä joukkona lauseita. Ne auttavat lukijaa havaitsemaan, kuinka lauseita käytetään diskurssissa.

Pearson ja Johnson (1978) erottavat tunnetussa taksonomiassaan kolmenlaisia kysymyksiä, jotka vaihtelevat vaikeusasteeltaan. Tekstuaalisesti eksplisiittisiin kysymyksiin löytyy vastaus yhdestä lauseesta, kun taas tekstuaalisesti implisiittiset kysymykset edellyttävät vastaajien yhdistävän eri lauseissa annettua tietoa. Skriptiperustaiset kysymykset vaativat oikean vastauksen tuottamiseksi vastaajalta myös taustatietoja. Tiettyyn kysymystyyppiin vastaamisen edellyttämien kognitiivisten ja affektiivisten piirteiden laajuuden määrittämiseksi kysymyksiä pyritään systematisoimaan laatimalla ns. kysymystypologioita. Candlin ym. (1974. Siteerattu Rost 1990) jakavat kysymykset viiteen luokkaan:

- Kirjaimellista tietoa testaavat kysymykset, jotka koskevat faktatietoa, sanojen merkityksiä ja ajatusten tunnistamista.
- Uudelleenorganisointia edellyttävät kysymykset, jotka edellyttävät synteesiä tai luokittelua tai informaation uudelleen järjestämistä.
- Johtopäätöstä edellyttävät kysymykset, jotka koskevat yhteyksiä, oletuksia ja ennustuksia.
- Arvioivat kysymykset, jotka perustuvat vastaajan etukäteistietoon, mikä mahdollistaa arviot validiudesta, tarkkuudesta, hyväksyttävyydestä tai arvosta.
- Arvottavat kysymykset, jotka koskevat kielen arvostusta, merkitystä, kommunikatiivista arvoa ja tyylin piirteitä.

Myös Gerot (1987. Siteerattu Rost 1990) on luokitellut kysymysten vaikeustasoa suhteessa vaadittuun vastaukseen.

- Replikatiivinen: jäljittelee tekstiä sana sanalta.
- Kaiku: vastaus toistaa tekstin sisällön, vaikka voi poiketa siitä sanastollisesti tai kieliopillisesti.
- Synteesi: lukijan/kuulijan täytyy yhdistää ja sulauttaa. toisiinsa joukko identifioitavissa olevia informaation osia.
- Epäsuora: testattavan tulee päätellä fakta tai ajatus, joka seuraa jostakin tekstissä eksplisiittisesti ilmaistusta.
- Oletus: testattavan tulee päätellä fakta tai ajatus, joka seuraa jostakin tekstissä implisiittisesti ilmaistusta.

Taulukossa 31 esitetään kahden kysymyshierarkioita syvemmin analysoineen tutkijan kokoamia typologioita. Rost (1990) on varustanut ryhmittelynsä myös varsinaisin kysymyksin, joiden avulla toivotun tyyppisiä vastauksia oletetaan saatavan. Kummassakaan typologiassa ei esitetä vaikeustason vertikaalista jatkumoa. Barrettin (1974) osuudessa on horisontaalinen jatkumo siten, että hän jakaa tekstit ensin `helpompiin` ja `vaikeampiin` ja määrittää kysymyksen tavoitteen tästä lähtökohdasta käsin. Kummankin tutkijan typologioissa Joka toinen sarake on tummennettu ainoastaan lukemisen helpottamiseksi.

TAULUKKO 31 Kysymysten kognitiivinen vaikeustason jatkumo (Lähteet: Rost:1990; Barrett 1974.)

Rost	Barrett	
<i>A. Määrittely</i>	Eksplisiittisesti ilmaistun informaation kirjaimellinen tunnistaminen tai muistaminen. Seuraavien asioiden tunnistaminen.	Johtopäätös, tekstin kirjaimellisen sisällön synteesi, henkilökohtainen tieto, intuitio ja mielikuva pohjana otaksumille tai oletuksille.
- Määrittele ____. - Mitä ____ merkitsee tässä? - Mitä ____ merkitsee? - Mitä termejä voi käyttää ____:stä?	<i>Yksityiskohta</i>	<i>Asiaa tukevat yksityiskohtat</i>
	Paikallistaa/Identifioida/Muistaa henkilöiden nimet, tarinan tapahtuma-aika, puitteet tai tarinassa kerrottu tapahtuma.	Paikallistaa/Identifioida/Muistaa kirjoittajan/puhujan tekstikatkelmaan lisäämät tiedot, jotka tekevät siitä informatiivisemman, kiinnostavamman ja vetoavamman.
	<i>Pääajatus</i>	
	Selvästi ilmaistu toteamus, joka on kappaleen tai pitemmän tekstikappaleen pääajatus.	Yleinen merkittävyys, teema tai opetus, jota ei ole ilmaistu selkeästi tekstissä.
<i>B. Esimerkki</i>	<i>Tapahtumajärjestys</i>	
- Anna esimerkki ____:stä. - Osaatko nimetä yhden ____?	Paikallistaa/Identifioida/Muistaa tekstikatkelmassa selkeästi ilmaistujen toimintojen tai tapahtumien järjestys	Otaksuman tekeminen siitä, mitä olisi tapahtunut kahden selkeästi ilmaistun toiminnan tai tapahtuman välillä; oletusten tekeminen siitä, mitä tapahtuisi seuraavaksi tai miten tarina olisi alkanut, ellei se olisi alkanut siitä, mistä alkoi.
	<i>Vertailu</i>	
<i>C. Tärkeys</i>	Paikallistaa/Identifioida/Muistaa kirjoittajan/puhujan selkeästi vertailemien henkilöiden, historian tapahtumien tai paikkojen samankaltaisuudet tai erilaisuudet.	Tehdä johtopäätöksiä henkilöiden, aikojen tai paikkojen samankaltaisuudesta tai erilaisuudesta.
- Mikä on ____:n tärkeys? - Miksi ____ on tärkeä? - Mikä on ____:n merkitys x:lle? - Mikä on ____:n rooli? - Kuka oli ____?	<i>Syy- ja vaikutussuhteet</i>	
	Paikallistaa/Identifioida/Muistaa syyt selkeästi tekstikatkelmassa ilmaistuihin sattumuksiin, tapahtumiin tai henkilöiden toimintaan.	Tehdä johtopäätöksiä henkilöiden motiiveista ja heidän vuorovaikutuksestaan toisten kanssa ja eri aikoina ja paikoissa; otaksuman tekeminen siitä, mikä sai kirjoittajan/puhujan sisällyttämään tekstiin tietyt ajatukset, sanat, luonnehdinnat ja toiminnot.
	<i>Luonteenpiirteet</i>	
	Paikallistaa/Identifioida/Muistaa toteamukset luonteesta, jotka auttavat osoittamaan, millainen henkilö oli kirjoittajan/puhujan kertomana.	Tehdä oletuksia henkilöiden luonteesta tekstikatkelmassa selkeästi ilmaistujen vihjeiden pohjalta henkilöiden luonteesta.
<i>D. Fyysinen mallikuva</i>		<i>Lopputuloksen ennustaminen</i>
		Tehdä tekstikatkelman alun lukemisen jälkeen otaksumia tekstin lopusta.
- Kuvaa, miltä ____:n on täytynyt näyttää. - Kuvaa _____.		<i>Kuvaannollisen kielen merkityksen ymmärtäminen</i>
		Kirjoittajan/Puhujan käyttämän kuvaannollisen kielen ilmaisemisen kirjaimellisesti.

<i>E. Funktio / Tarkoitus</i>		<i>Arviointi</i>
- Mikä on ___:n funktio / tarkoitus. - Mihin ___ käytetään?		Arvioiden tekeminen.
<i>F. Järjestyksen ja paikan merkitykset</i>		
- Mitä tapahtui sen jälkeen, kun...? - Mitä tapahtui ___:lle? - Kuvaa, mitä ___ teki?		
<i>G. Samankaltaisuus ja erilaisuus</i>		
- Mikä on ___:n ja ___:n ero? - Vertaile ___ ja ___. - Miten / Miltä piirteiltään ___ eroaa ___:sta? - Miksi ___ on erilainen kuin ___?		
<i>H. Luokittelu</i>		
- Kuinka monta lajia ___ löytyi? - Mitkä ___:n ryhmät soveltuisivat ___:än?		
<i>I. Syy – seuraus</i>		
- Mikä aiheutti ___:n? - Mitkä tekijät vaikuttivat...?		
<i>J. Väittämän tutuus</i>		
- Mitä perusteluja ___ vaatii ___:lta? - Mitä perusteluja ___ antaa ___:lle? - Selitä, miksi...		
<i>K. Arvioiva ajttelu</i>		
- Minkä luulet olevat ratkaisun ...? - Anna toinen selitys ___:lle. - Oletko tätä mieltä?		

Kokeessa ymmärtämistehtävä määritellään usein osioiden tai tehtävien odotetun vastauksen tyyppin mukaan. Bachman ja Palmer (1996) erottavat kolme päätyyppiä:

- vastauksen valitseminen (*selected response*), joka on yleensä joko monivalinta tai oikein/väärin-väittäjä
- lyhyen vastauksen tuottaminen (*limited production response*), mikä yleensä edellyttää yhden tai muutaman sanan vastausta ja
- laajan vastauksen tuottaminen (*extended production response*), jolloin vastaus on pitempi kuin yksi virke, kahdesta virkkeestä käytännöllisesti katsoen vapaaseen kirjoitelmaan (Alderson 2000:156).

Taulukossa 32 kuvataan ymmärtämisen arvioinnin eri tehtävätyyppien ansioita ja puutteita Hughesin (1989), Aldersonin (2000) ja Purpuran (2004) näkemyksiä yhdistäen.

TAULUKKO 32 Ymmärtämisen kokeen tehtävätyyppien ansioita ja puutteita.

Tehtävätyyppi	Kuvaus	Ansiot	Puutteet
Cloze/Aukko-täydennys	Tekstistä poistetaan joka n:s sana/kokeen laatijan päättämät sanat, jotka oppilas/testattava täydentää	- nopea suorittaa ja korjata	- mittaa muutakin kuin tekstin ymmärtämistä
Monivalinta	Tekstistä tehdään kysymyksiä tai väittämiä ja annetaan kuhunkin 3-4 kohde- tai äidinkielistä vastausvaihtoehtoa	- nopea suorittaa ja korjata - mittaa hyvin, jos distraktorit onnistuneita	- tekniikka opittavissa - arvaamalla voi saada 25-30 % oikein - laadinta vaatii aikaa ja taitoa - järkeviä vaihtoehtoja on vaikea laatia (3 usein luonnollinen maksimi)
O/V-valinta	Tekstiin liittyvä kohde- tai äidinkielen väittä- mä on arvioitava oikeaksi tai vääräksi luetun tekstin perusteella	- nopea suorittaa ja korjata - vaikeusaste helposti varioitavissa	- arvaamismahdollisuus 50 %
Virkkeiden järjestäminen	Joukko virkkeitä on järjestettävä loogiseksi kertomukseksi.	- mittaa hyvin - nopea suorittaa ja korjata - vaikeusaste helposti varioitavissa	- laatijan tarkistettava huolellisesti, että vain yksi järjestys on mahdollinen - selvitettävä kaikki vaihtoehdot
Tekstien editointi	Virheiden korjaaminen tekstistä.	- korjattujen virheiden laadusta voidaan tehdä päätelmiä testattavan lukutaidon luonteesta	- oppijat eivät usein ole tottuneet koetyyppiin (eivät usein osaa tai halua editoida omaakaan tekstiä)
Lyhyet vastaukset	Testattava vastaa muutamalla sanalla tekstistä äidin- tai kohdekielellä tehtyihin kysymyksiin.	- jos vastausavain on kattava, mittaa hyvin tekstin ymmärtämistä	- edellyttää sensoreita - edellyttää mielellään esikokeilua, koska kattavan vastausavaimen laatiminen on vaikeaa.
Pitkät vastaukset	Testattava osoittaa hallitsevansa kielen käyttötilanteita (suullinen esitys, kirjoitelma)	- jos alue on hyvin määritetty, antaa hyvän kuvan testattavan taidoista	- edellyttää koulutettuja sensoreita - edellyttää tilanteeseen soveltuvien arviointias- teikkojen käyttöä

Vaikka taulukoissa esitetyt käsitykset ovat vailla vankkaa evidenssipohjaa, ne auttavat kuitenkin ymmärtämisen kokeiden laadintaa sekä sopivan tehtävätyypin valintaa käytettävissä olevien resurssien (esim. aika) puitteissa.

6.1.2 Puheen ymmärtämisen arviointi

Buck (2001:32-52) katsoo, että puheen ymmärtämisen kokeissa olisi syytä keskittyä arvioimaan nimenomaan niitä puheen ymmärtämisen taidon piirteitä, joista tekstin ymmärtämisen tai rakennekokeen avulla ei saada tietoa. Tällaisia ominaisuuksia ovat esimerkiksi fonologinen modifiointi (assimilaatio, elisio, intruusio) vahvat/heikot muodot, prosodiset piirteet (paino, intonaatio, lausepaino), puheen nopeus ja diskurssirakenne. Tehtävissä vaikeuksia aiheuttavat

runtas prosessoinnin tarve, tiedon löytyminen useista kohdista, muun kuin tarkan sisältökohdan muistaminen, faktan erottaminen mielipiteestä, epäoleelliset yksityiskohdat ja vastauksen tuottaminen välittömästi (Buck 2001:149-151). Buck (2001) on koonnut seuraavan taulukon (taulukko 33) puheen ymmärtämisen kokeen tekstin helpottamista edesauttavista tekijöistä.

TAULUKKO 33 Puheen ymmärtämisen kokeen helpottamiskeinoja. (Lähde: Buck 2001:149-51.)

Kielelliset piirteet	<ul style="list-style-type: none"> - Hitaampi puhenopeus. - Pidemmät tauot. - Tuttu ääntämys. - Luonnolliset intonaatiomallit. - Tavalliset sanat. - Yksinkertaiset kieliopilliset rakenteet. - Toisiaan seuraavat lauseet sisäkkäin sijoitettujen sijasta. - Yksinkertaiset pronomini viittaukset.
Selkeys	<ul style="list-style-type: none"> - Selkeästi ilmaistut ajatukset. - Runtas redundanssi.
Jäsentely	<ul style="list-style-type: none"> - Lineaarisesti tai aikajärjestyksessä etenevät tekstit. - Pääkohtien ilmaiseminen selkeästi ennen esimerkkien esittämistä.
Sisältö	<ul style="list-style-type: none"> - Tuttu aihe - Harvoja asioita. - Vähän toimijoita tai kohteita. - Tekstin elementtien suhde kiinteä. - Konkreettinen sisältö.
Konteksti	<ul style="list-style-type: none"> - Visuaalinen tai muu tuki

De Jong (1987) kehottaa tarkastelemaan puheen ymmärtämisen tehtäviä neljästä näkökulmasta: kuunneltavan tekstin, tehtävien, vastauksien sekä pisteityksen perspektiivistä. Taulukossa 34 esitetään hänen ryhmittelynsä sekä se, mitkä ovat kunkin tehtävän osan kriittisiä kohtia.

TAULUKKO 34 Puheen ymmärtämisen tehtävissä huomioon otettavia asioita. (Lähde: De Jong 1987.)

Tehtävän osa	Avainkysymys	Toimenpide	Huomioon otettavaa
Kuunneltava teksti	Millaista kieltä testattavan tulisi pystyä kuuntelemaan?	- konstruktin/kielen käyttöalueen täsmennys	- kokeeseen sisällytetyt otokset kielen käyttöalueista - vaikeusaste, joka vaihtelee pääasiassa akustisten ja kielellisten tekijöiden välillä: melu, äänen laatu, puhunnan nopeus, sanasto, aiheen abstraktiotaso
Tehtävä	Mitä testattavan tulisi ymmärtää tästä tietyistä kielen käyttöalueista?	- kuuntelun käytön / tarkoituksen / funktion typologia	- kuuntelutehtävään liittyvät kysymykset tai muut virikkeet - vaikeusaste riippuu kuunneltavasta tekstistä, itse ymmärtämistehtävästä sekä näiden kahden välisestä suhteesta - osioiden vaikeusastetta voi varioida käyttämällä erilaisia ymmärtämistehtäviä - kysymystypologiat
Vastaukset	Kuinka testattavan odotetaan osoittavan ymmärtämisen?	- vastauksen tyyppi	- testattavalta edellytettävät reaktiot - kuunneltavan tekstin, ymmärtämistehtävän ja edellytetyn vastauksen sekä erikseen että yhdessä määrittelemä vaikeustaso - jos kuunneltava teksti ja ymmärtämistehtävät ovat vaikeustasoltaan samanlaisia, osion vaikeutta voi varioida antamalla vaikeustasoltaan erilaisia vastauksia edellyttäviä tehtäviä - tehtävän edellyttämän vastauksen operationaalistaminen on puheen ymmärtämisen kokeiden merkittävä virhelähde a) avotehtävät: tuottamiskyvyn vaikutus (mitattavan taidon kannalta irrelevanttia varianssia aiheuttava) b) monivalinta: vaihtoehtojen ymmärtäminen
Pisteitys	Mitä pisteitysmenetelmiä aiotaan noudattaa?	- arviointiasteikot	- koska pisteet johdetaan vastauksista, asteikon valinta on tärkeä

6.1.3 Tekstin ymmärtämisen arviointi

Jokseenkin yksimielisiä ollaan siitä, että vieraan kielen lukijoiden täytyy ylittää kielellinen kynnyks ennen kuin he voivat ruveta hyödyntämään äidinkielessä saavuttamaansa lukutaitoa. Tällä kynnyksellä on myös yhteyttä lukijan taustatietoihin ja luettavaan tekstiin siten, että joissakin tiettyä aihetta käsittelevissä teksteissä kynnyks on matalampi ja tarvitaan vähemmän kielitaitoa, joissakin korkeampi, mikä edellyttää enemmän kielitaitoa (Alderson 2000:112). Erään tutkimuksen mukaan (Bernhardt & Kamil 1995) äidinkielen taito selitti 20 %

vieraalla kielellä lukemisen varianssista, kun taas toisen tutkimuksen mukaan vieraan kielen taito selitti siitä 30% (Chun & Plass 1997). Freedle ja Kostin (1993) tutkivat TOEFLin lukemisen kokeen tehtävien vaikeustasoa ja havaitsivat, että seitsemän tehtävien piirrekategoriaa selitti 58 % tehtävien vaikeuden varianssista. Nämä olivat: sanastollinen päällekkäisyys tekstin ja ratkaisun välillä, lauseiden pituus, tehtävän pituus, tekstikatkelman pituus, kappaleen pituus, retorinen organisaatio, kieltomuotojen ja kappaleiden välisten viittausten käyttö.

Tekstin ymmärtämisen koe on sekä luettavan tekstin että tehtävän vaikeuden funktio eli helposta tekstistä voidaan tehdä vaikea kysymys ja päinvas-toin (Alderson 2000:85; ks. myös CCSE -koe, luku 4). Lukemisen tehtävien sanotaan vaihtelevan vaikeusasteeltaan riippuen tekstistä, tehtävistä, lukijasta ja kaikkien niiden vuorovaikutuksesta. Tekstit vaihtelevat sisällöltään (abstraktio-tasoltaan, informaatiotiheydeltään, aiheiltaan, tekstimuodoltaan tai tyybiltään, kontekstiltään ja kulttuurisilta konventioiltaan) ja kirjoittajan tyylin suhteen (sanasto ja rakenteet, koheesio ja koherenssi ja redundanssin käyttö) (Alderson 2000:127).

Tekstin valinta on tärkeä, jotta voitaisiin ymmärtää lukemisen luonne ja se, että tekstilajilla ja aiheella saattaa olla merkittävä vaikutus sekä lukemisen tuloksiin että itse prosessiin sekä siihen, miten muut muuttujat, erityisesti luki-jan motivaatio ja taustatiedot välittyvät tehtävän suorittamisen kautta (Alder-son 2000:255). Bachman ja Palmer (2004) eivät anna ohjeita aiheen valitsemisek-si, vaikka he tosin huomauttavat, että koevalidiuden käyttökelpoisuus-dimensio on riippuvainen siitä, kuinka lähellä todellisen elämän tilannetta esi-merkiksi jokin lukemisen tehtävä on. Heidän näkemyksensä mukaan kielitaitoa voidaan arvioida vain tietyissä tilanteissa, joiden tuloksia ei voi yleistää (Alder-son 2000:159). TOEFL-testin kokeissa käytetään useita lyhyitä tekstikatkelmia laajemman aihevalikoiman kattamiseksi, mikä vähentää liian vähäisen aiheva-likoiman käytöstä aiheutuvaa harhaisuutta (Alderson 2000:109). Joka tapauk-sessa joudutaan toteamaan, että on vaikeata arvioida testattavien kykyä riip-pumatta tekstin luonteesta ja lukutehtävästä (Alderson 2000:304). Jos lukemi-seen liittyy joukko erilaisia komponentteja, nämä komponentit ovat eri lukijoilla eri tavoin kehittyneet (Alderson 2000:93). Jos eri testattavat vastaavat eri tavoin tehtävään ja saavat kuitenkin tehtävän oikein, on todellinen ongelma määrittää sitä, mitä tehtävä ja koe testaavat (Alderson 2000:97). Tämän vuoksi ymmärtä-misprosessin tutkiminen ja tunteminen olisi tärkeätä. Tarkka tieto ei ole aina kovin välttämätöntä johtopäätösten tekemiseksi kielitaidosta, vaan useimmiten voidaan tyytyä siihen, että vastaajat ovat voineet päätyä vastauksiinsa eri ta-voin. Jos halutaan määrittellä tarkasti, mitä tehtävä ja koe testaavat, joudutta-neen joko rajaamaan tilanne hyvin tiukasti ja yrittämään sulkea ei-halutut piir-teet pois tai järjestämään yksilötestausta käyttäen syvälle diagnosoivaa menet-telyä, mm. ääneen ajattelua. Tavallisessa testauksessa jouduttaneen toimimaan "keskivertoajattelulla": teksti on luonteeltaan x ja siihen liittyvät kysymykset on laadittu testaamaan taitoja a, b ja c. Näin ollen kokeen voidaan väittää "lähtö-kohtaisesti" mittaavan z:aa.

Lukemistehtävän muoto voi olla kielellinen tai ei-kielellinen tai niiden yhdistelmä, ja siihen saattaa liittyä kuvia, kuvioita, kaavioita ym. Nämä voivat kuitenkin johtaa erilaiseen tekstin prosessointiin, mikä tulisi ottaa huomioon vertailtaessa tuloksia (Alderson 2000:152). Samanlainen vaikutus tuloksiin voi olla myös tekstin pituudella, joka voi vaihdella yksittäisistä sanoista, joita on käytetty käskyinä tai kieltoina, pitkiin tieteellisiin artikkeleihin (Alderson 2000:153).

Widdowson (1978) esittää seuraavia tapoja tekstin ymmärtämisen testaamiseksi:

1. Kysymykset voivat edellyttää usean lauseen ymmärtämistä.
2. Luetun tekstin väliin voidaan liittää sen tulkintaa koskevia väittämiä, joihin voidaan liittää esimerkiksi perustelujen esittämiskehotus.
3. Tekstistä erotetaan lauseita sellaisenaan tai hieman muokattuina ja korvataan jotkut elementit anaforisilla elementeillä (pronomineilla) tai yleisillä nominilausekkeilla (*the event, this fact*). Kokeeseen osallistujan tulee korvata nämä tekstistä löytyvillä asianmukaisilla sisältösanoilla.
4. Kontekstuaalista synonyymiikkaa voitaisiin käyttää pyytämällä lukijaa kirjoittamaan uudelleen esitetyt lauseet korvaamalla osa sanoista muilla, tekstistä löytyvillä sanoilla.

Toinen tapa luokitella koekysymykset on erottaa paikallinen ja globaali ymmärtäminen. (Alderson 2000:87). Kysymysten kieli, niiden sanamuoto ja osioissa ja vaihtoehtoissa käytettävän sanaston frekvenssi ovat kaikki tärkeitä. Sen informaation taso, jota kysymys vaatii – olipa se paikallinen tai globaali – ja kysymyksen ja vaaditun tekstissä olevan informaation välinen suhde on ratkaisevan tärkeä (Alderson 2000:113). Paitsi lukutehtävän tarjoamaa informaatiota myös testattavan omat taustatiedot tulisi pystyä ennakoimaan. Kuten aikaisemmin on todettu, kaikki kielen ymmärtäminen edellyttää lukijan tekevän johtopäätöksi ja oletuksia ja käyttävän taustatietojaan. Vastauksen ja ymmärtämistehtävän suhteen sanotaan olevan suora, jos haettu informaatio sisältyy lukutehtävään ja epäsuora, jos muitakin tietoja tarvitaan (Alderson 2000:163). On tärkeätä pohtia, mitä eri tekniikoilla pystytään arvioimaan sekä mitä niillä tyypillisesti arvioidaan (Alderson 2000:202). Toinen kysymys, johon koetta laadittaessa tulisi pystyä vastaamaan, on, pitäisikö vieraan kielen lukemisen kokeen mitata kielitaitoa enemmän kuin lukutaitoa, lukutaitoa enemmän kuin älykkyyttä tai mitä tahnassa muuta konstruktia, joka saattaisi liittyä kokeeseen (Alderson 2000:112). Vastauksella on yhtä paljon tekemistä oikeudenmukaisuuden kuin reliabeliudenkin kanssa: jos sanotaan että mitataan lukutaitoa, sitä on mitattava (Alderson 2000:112). Tässä tapauksessa on tyydyttävä ymmärtämisprosessin tuntemuksen puutteessa luottamaan siihen, että on mitattu kielitaitoa. Testaajien tulee myös selkeyttääkseen tulkintaa kyetä kuvailemaan käyttäjille, mitä korkeita pisteitä saavat testattavat osaavat tehdä ja mitä ne, jotka saavat alhaisia pisteitä, eivät osaa (Alderson 2000:271). Tämä on juuri tässä tutkimuksessa sovelletun kriteerimittauksen periaate.

6.2 Tuottaminen

Tuottamistaidon arvioinnissa on tarkoitus tehdä johtopäätöksiä puhumisen tai kirjoittamisen taidon sisällöstä ja tasosta. Tätä varten on laadittava tehtäviä, jotka ovat edustava otos niistä tilanteista, joissa tarvittavan kielitaidon testattavan oletetaan hallitsevan. Tehtävien avulla testattavien taidot tulisi saada mahdollisimman monipuolisesti esiin, ja tässä yhteydessä tulisi myös pystyä identifioimaan ne tehtävän komponentit, jotka eivät liity kyseiseen taitoon (Weigle 2002:57). Pisteityksessä tulee pitää huolta luotettavasta arvioinnista, joka on suhteessa asetettuihin tavoitteisiin (Hughes 1989:83).

Tuottamistaitoja arvioidaan kriteerein, jotka esitetään portaittain vaikeutuvina asteikkoina. Yleisimmin käytettyjä kriteereitä ovat laajuus (*range*), irheettömyys/tarkkuus (*accuracy*), sujuvuus (*fluency*), vuorovaikutus (*interaction*) ja sidosteisuus (*coherence*). Eurooppalainen viitekehys (2003:52-53) käyttää näitä kuvatesaan puhutun kielen laadullisia piirteitä. Niitä käytetään myös, ehkä vuorovaikutteisuuksiin lukuun ottamatta, arvioitaessa kirjoitettua tekstiä. Laajuus on lähinnä sanavarastoon ja tarkkuus kielipiilliseen kompetenssiin liittyvä kriteeri. Sujuvuus -kriteeri liitetään yleensä suulliseen tuottamiseen. Kirjallisessa tuottamisessa sitä vastaisi ehkä sidosteisuus. Vuorovaikutus on strategisten edellytysten kriteeri, joka liittyy myös selvimmän suulliseen tuottamiseen.

Kaikkien kriteerien sisällöstä ei suinkaan vallitse yksimielisyyttä, ja koska tavoitteetkin saattavat olla erilaisia, laajoihin, kaupallisiin kokeisiin onkin yleensä laadittu omat kriteerit ja asteikot. Sujuvuus -kriteeri on ehkä kiistellyin (Hildén 2000; Saleva 1997). Tärkeintä on, että asteikon laatija määrittää kriteerit niin, että tulkitsijan käsitys ei poikkea merkittävästi laatijan käsityksestä. Taulukossa 35 esitetään Eurooppalaisen viitekehysten A2 -tason edellä käsiteltyjen kriteerien sisällöt, joka on keskimäärin se taso, jolle opetus suunnitelmien perusteissa 2004 edellytetään päästävän.

TAULUKKO 35 Suullisen tuottamisen laadulliset kriteerit A2 -tasolla. (Lähde: Eurooppalainen viitekehys 2003:52-3.)

Laajuus	Virheettömyys	Sujuvuus	Vuorovaikutus	Koherenssi
Osaa joitakin lyhyitä, ulkoa opeteltuja ilmauksia, keskeistä sanastoa ja kaikkien tavallisimpia lauserakenteita. Osaa käyttää niitä yksinkertaisesti ja rajallisesti arkipäivän tilanteissa.	Osaa käyttää oikein joitakin yksinkertaisia rakenteita, mutta tekee kuitenkin jatkuvasti virheitä perusasioissa.	Pystyy esittämään sanottavansa ymmärrettävästi hyvin lyhyissä puheenvuoroissa, vaikka tauot, väärät aloitukset ja uudelleenmuotoilut ovat selvästi kuultavissa.	Pystyy vastaamaan kysymyksiin ja reagoimaan yksinkertaisiin lausumiin. Osaa osoittaa, milloin ymmärtää, mutta ymmärtää harvoin niin paljon, että pystyy ylläpitämään keskustelua.	Osaa yhdistää sanaryhmiä lauseiksi tavanomaisilla sidanoilla kuten "ja", "mutta", "koska".

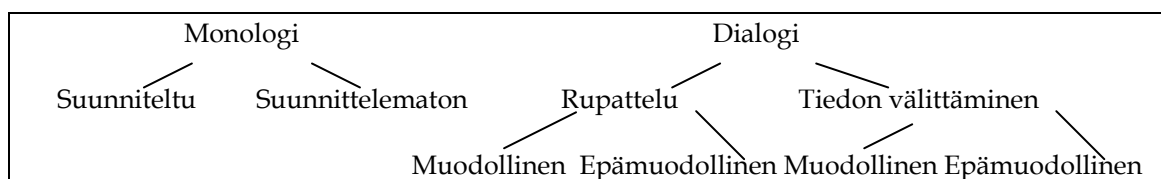
Eurooppalainen viitekehys tarjoaa useita asteikkoja, joita voi käyttää arvioita- vasta kielenkäyttötilanteesta riippuen. Jos esimerkiksi halutaan kiinnittää huomiota vuorovaikutukseen, voidaan käyttää kaikkia tai jo(i)takin taulukossa 38 mainituista asteikoista. Näin arvioinnin voi räätälöidä vastaamaan parhaalla mahdollisella tavalla tavoitteita. Taulukossa 36 esitetään esimerkkejä tästä.

TAULUKKO 36 Taitotasoasteikkojen valitseminen kriteerien arviointiin. Lähde: Eurooppalainen viitekehys 2003: 266.)

Kriteerit	Käytettävät asteikot
Laajuus	Yleinen laajuus
	Sanaston laajuus
Tarkkuus	Kieliopillinen virheettömyys
	Sanaston hallinta
Sujuvuus	Sujuvuus
Vuorovaikutus	Kokonaisvaikutelma vuorovaikutuksesta
	Puheenvuoron ottaminen
	Yhteistyö
Sidosteisuus	Sidosteisuus

6.2.1 Puhumisen arviointi

Puhuminen toimintana, jossa pyritään tiettyyn tulokseen, sisältää ajatuksen jatkumosta, jonka kaksi ääripäätä ovat rupattelu eli interaktionaalinen kieli, jonka päätarkoitus on vuorovaikutus (Brown 2001:251, Buck 2001:13) ja tiedon hankinta tai välittäminen eli transaktionaalinen kielenkäyttö (Brown 2001:251; Buck 2001:13). Rupattelutaitoon liittyvät puhujan persoonallisuus ja sosiaaliset taidot, ja usein keskustellaankin siitä, onko näitä syytä arvioida. Joissakin arvioinnissa small talk -taito voi kuitenkin olla arvioinnin kohde erityisesti silloin, kun näitä asioita on juuri käsitelty opetuksessa. Olisi tärkeätä, että osallistujat tietäisivät, millaiseen kielitaitoon heidän tulisi pyrkiä suoriutuakseen testistä hyvin (Luoma 2004:23). Kohteliaisuutta on myös vaikea arvioida, koska se on interpersonaalinen ja sosiaalinen, ja kokeeseen osallistujien sosiaalinen suhde on keinotekoinen. Yksi ratkaisu voisi olla arvioida kohteliaisuutta vain väljästi esimerkiksi kolmella tasolla: sopiva, jotakuinkin sopiva, kyseenalainen / huonompi (Luoma 2004:27). Koekriteereissä voisi harkita oppijan palkitsemista niiden hallitsemisesta (Luoma 2004:18). Myös eri kategorioihin kuuluvia kielenkäyttörutiineja on hyödyllistä testata, koska ne saattavat antaa erilaista tietoa puhujan taidoista. Tämä jaottelu kattaa kuitenkin vain keskusteluun liittyvän suullisen tuottamisen, mutta ei sen sijaan erilaisia suullisia esityksiä, kuten puheita, luentoja, uutislähetyksiä ym. Kuviossa 24 esitetään Nunanin (1991) laajempi ja yksityiskoh- taisempi käsitys suullisesta kielitaidosta.



KUVIO 24 Suullisen kielitaidon tyypit. (Lähde: Nunan 1991.)

Fulcher (2003) katsoo suullisen kokeen laadinnan koostuvan yleensä seuraavien asioiden kirjaamisesta:

1. Kokeen laadinnan yleinen kuvaus eli luonnehdinta siitä, millaista puhumista testataan (konstrukti).
2. Tehtävänanto, joka voi sisältää mm. seuraavaa tietoa:
 - a) Puhumistehtävän kuvauksen: kuulija(kunta), tarkoitus/viestintä-funktio, genre/tekstilaji, sisältöä koskevat mahdolliset lähteet.
 - b) Puheen sisältöä mahdollisesti tukeva teksti-, kuva- tai muu materiaali.
 - c) Tehtävänannon kieltä koskeva kuvaus.
 - d) Vastaamistapaa koskevat ohjeet.
3. Puhumistuotosta koskeva kuvaus siitä, miten puhumistehtävän suorittaja tuottaa suorituksensa, mitä hänen tulee tehdä täyttääkseen tehtävänannon
4. Arvosteluperusteet.
5. Mallisuoritus (ainakin yksi tai mieluummin useita eritasoisia)
6. Testaustilanteeseen soveltuvien tehtävätyyppien malleja, tietoa niiden ansioista ja puutteista.

Tavallisin tapa järjestää puhekokeita on arvioida kokelaita yksi kerrallaan, usein haastattelun muodossa niin, että haastattelija on opettaja tai ulkopuolinen arvioitsija (esim. ACTFL). Se vaatii kuitenkin runsaasti niin ajallisia kuin taloudellisiakin resursseja, joten yhä enemmän on siirrytty käyttämään parityötä, jolloin testattavat hoitavat keskenään koko testaustilanteen arvioitsijan toimiesä vain taustalla varmistamassa, että testattavat tuottavat arviointiin riittävän määrän kieltä. Ohjeiden ja tehtävien selkeydelle asetetaan tällaisissa tehtävissä korkeat vaatimukset, jotta ne helpottavat keskustelua, ja että testattavat tietävät, mitä heiltä odotetaan hyvien tuloksien saamiseksi. Väistämätön tämän arviointitavan ongelma on kuitenkin se, että testattavan puheeseen vaikuttaa toisen osallistujan persoonallisuus, kommunikointityyli ja mahdollisesti myös kielitaidon taso (Luoma 2004:37). Näin ollen kaikki testattavat eivät ehkä saa tilaisuutta osoittaa kaikkea osaamistaan (Weir 1993:55-56).

Hughes (1989) jakaa puhumisen tehtävätyypit kolmeen ryhmään: haastattelu, vuorovaikutus muiden testattavien kanssa ja reagointi kuultuun tai nähtyyn. Haastattelu voi olla vapaamuotoinen tai kontrolloitu kuten esimerkiksi ACTFL, joka etenee vaiheittain lämmittelystä (*warm up*) varsinaiseen haastatteluun, jossa haetaan testattavan tasoa (*probing*), ja päättyy haastattelijan lopetukseen (*wind down*). Mitä vapaamuotoisempi haastattelu on, sitä suurempia vaatimuksia se asettaa tuotoksen pisteittämiselle. Hughes (1989) lukee tähän ryhmään kuuluviksi myös roolileikin, jossa testattavalle annetaan tilanne, josta hänen tulee selviytyä, tulkkaustehtävän, kuvan tai kuvasarjan sisällön/tapahtumien kertomisen ja ääneenlukemistehtävän. Vuorovaikutustehtäväryhmään kuuluvat sellaiset tehtävätyypit, joissa testattavat toimivat keskenään arvioija(ie)n seurattavana sivusta, kuten roolileikki tai keskustelu. Kummankin ohjauksen määrä voi vaihdella, mutta arviointiin pätee sama kuin haastatteluunkin: mitä ohjatumpi, sitä helpompi tuotos on arvioida (Hughes 1989:119-122).

Tiedonhankintatilanteissa tärkeintä on saada viesti perille ja vakuuttua siitä, että vastaanottaja on ymmärtänyt viestin. Yhteisen perustan löytäminen, informaation antaminen sopivissa annoksissa, looginen eteneminen, kysymykset, toistot ja ymmärtämisen tarkistukset auttavat puhujia pääsemään tähän tavoitteeseen.

seen. Näitä piirteitä tulisi siksi esiintyä testattavan suorituksessa silloin kun kyseessä on tiedonhankintatehtävä. Niiden avulla pystytään myös perustelemaan, miksi joku suoriutuu tehtävästä paremmin kuin joku toinen (Luoma 2004:23). Koetta suunniteltaessa onkin yleistä kohdistaa huomio puheen informatiiviseen aspektiin, koska se rajaa sisältöalueet ja kielenkäytön tyypit. Sosiaalista puhetta on syytä sisällyttää kaikenlaisiin tehtäviin, mutta säätää myös arviointikriteereitä ottamaan tämä huomioon, koska se todennäköisesti vaikuttaa arvioijan vaikutelmaan siitä, kuinka hyvin testattava suorittaa tehtävät (Luoma 2004:32).

Kolmannen ryhmän tehtävien hyväksi lähteiksi Hughes (1989) mainitsee ARELSin¹ suullisen englannin kielen tutkinnon sekä ETSin² englannin kielen puhekokeen mallitehtävät. Esimerkkeiksi niistä hän antaa tilannekuvauksen (*described situations*), jonka perusteella on reagoitava, irrallisiin huomautuksiin vastaamisen (*remarks in isolation to respond to*) ja simuloidun keskustelun, jossa toisen osapuolen vuorosanat kuuluvat nauhalta ja jonka aiheesta testattava on saanut edeltä käsin tietoa (Hughes 1989:122-124; vrt. Saleva 1997:182).

Koska koetehtävien avulla on tarkoitus tehdä johtopäätöksiä mitattavaksi aiotuista kielitaidon piirteistä, vastausten ennakoitavuuden punnitseminen tehtäviä laadittaessa ja valittaessa on tärkeää. Taulukossa 37 tehtävätyypit on jaettu kolmeen ryhmään tätä kriteeriä käyttäen. Ryhmittely perustuu Luoman (2004) esittämään jaotteluun. Strukturoidut tehtävät rajaavat jokseenkin tarkasti sen, mitä puhujan odotetaan vastaavan (Luoma 2004:48). Tällaisia ovat esimerkiksi tervehtimisessä, kiittämisessä ja anteeksipyyttämisessä käytettävät fraasit. Puolistrukturoidut tehtävät edellyttävät reagointia erilaisissa tilanteissa, joissa sosiaaliseen kielenkäyttöön liittyvien fraasien lisäksi tarvitaan omakohtaista tuottamista (Luoma 2004:49). Avoimien tehtävien suorittaminen antaa puhujalle eniten mahdollisuuksia omakohtaiselle kielenkäytölle (Luoma 2004:48). Taulukon 37 tehtävätyyppien kokoamiseen ja luonnehdintaan on käytetty Luoman (2004) lisäksi Bygaten (1987), Underhillin (1987) ja Weirin (1993) pohdiskeluja. Taulukko osoittaa myös, että piirre, joka toisessa tehtävässä on etu, voi toisessa olla jonkinasteinen haitta. On siis järkevää miettiä, milloin ja missä määrin esimerkiksi omaehtoisen kielen tuottamisen laajuus on tärkeämpää kuin selkeä vertailtavuus.

¹ Association of Recognised English Language Schools

² Educational Testing Services

TAULUKKO 37 Suullisen tuottamisen tehtävätyyppien vertailua

Tehtävätyyppi	Tehtävä	Vahvuuksia	Heikkouksia
Strukturoitu	<ul style="list-style-type: none"> - lukeminen - kuullun tai luetun perusteella tekstin täydentäminen - korjaaminen - muuntaminen - ääneen lukeminen - tekstin toistaminen - lyhyen fraasin tuottaminen 	<ul style="list-style-type: none"> - vertailtavuus, koska vastausvaihtoehdot / arvioinnin kohteet ovat etukäteen määriteltävissä / rajattavissa - jos ohjeissa edellytetään selitystä, oma kohtainen tuottaminen lisääntyy - osoittaa kontekstin ymmärtämistä - testaa myös kieliopillista tietoa 	<ul style="list-style-type: none"> - autenttisuuden vähäisyys - omakohtaisen tuottamisen vähyys - vertailtavuuden takaamiseksi määriteltävä hyväksyttävät vastaukset etukäteen, mikä vaatii resursseja
Puolistrukturoitu	<ul style="list-style-type: none"> - henkilökohtaisiin kysymyksiin vastausminen (nimi, ikä jne.) - tilanteessa reagoiminen (kiittäminen, anteeksipyyntö, valitukset) - tiivistelmä kuullun / luetun perusteella etukäteen annettujen vihjeiden pohjalta 	<ul style="list-style-type: none"> - vertailtavuus, koska vastausvaihtoehdot / arvioinnin kohteet ovat etukäteen määriteltävissä / rajattavissa - osoittaa sujuvuuden asteen - useat tehtävät testavat keskeistä kielitaitoa - tehtävästä riippuen melko korkea autenttisuuden aste - autenttisuuden määrä 	<ul style="list-style-type: none"> - omakohtaisen tuottamisen määrä - ulkoa opettelu korostuu erityisesti, jos koheen sisältö tunnetaan
Avoin	<ul style="list-style-type: none"> - roolileikki, simulaatio - kuullun / luetun perusteella koetilanteessa tehtyihin kysymyksiin vastaaminen - kuvailu, kerronta, ohjeet, vertailu, selitys, perustelu, ennustus tai päätöksenteko annetun aiheen, kuvan tai kuvasarjan pohjalta 	<ul style="list-style-type: none"> - omaehtoisen tuottamisen määrä ja laatu - autenttisuuden määrä ja laatu 	<ul style="list-style-type: none"> - vertailtavuuden takaamiseksi ohjeiden laatiminen vaatii resursseja - mitä avoimempi tehtävä on, sitä enemmän ulkopuolinen arvioitsija on tarpeen - edellyttää puheen / tekstin ymmärtämistä - parityössä osapuolet tai toinen heistä ei tunne roolia "omakseen".

Puhetehtävät vaativat usein pitkiä ja yksityiskohtaisia ohjeita ja vihjeitä, joten niissä käytettävä kieli on tarkoin harkittava, jotta testattavat ymmärtäisivät tehtävän tarkoitettulla tavalla. Itse tehtävien tulee olla riittävän inspiroivia, jotta ne innostaisivat puhumiseen, riittävän strukturoituja, että ne tuottaisivat todella sellaista puhetta, jota halutaan, ja riittävän odottamattomia, että testattavat eivät voi harjoitella niiden suorittamista näillä tietyillä aiheilla ja tehtävillä. Niissä tulisi myös varautua siihen, että arvioijalla on mahdollisuus johdatella tai auttaa testattavaa tarvittaessa (Luoma 2004:53).

Koska kansallisessa arvioinnissa haluttiin testata myös nimenomaan oppilaiden puhetaitoa, edellä esitetyt näkökohdat olivat voimakkaasti esillä tehtäviä

suunniteltaessa ja niistä oli tässä työssä suurta apua. Luvussa 8 käy ilmi, millaisiin ratkaisuihin päädyttiin.

6.2.2 Kirjoittamisen arviointi

Kirjoitettu kieli on aikaisemmin usein nähty puhumiseen verrattuna toissijaisena. Esimerkiksi de Saussure (1916) totesi, että (puhuttu) kieli ja kirjoittaminen ovat kaksi erilaista merkkijärjestelmää ja että kirjoittamisen ainoa olemassaolon syy on edustaa (puhuttua) kieltä (Takala 1982).

Kirjallinen tuottaminen on toimintaa, joka tapahtuu kontekstissa, jonka tavoitteena on saavuttaa tietty tarkoitus ja joka on sopivalla tavalla muotoiltu aiotulle kuulijakunnalle (Hamp-Lyons & Kroll, 1997. Siteerattu Weigle 2002:19). Kansainvälisen kirjoitelmatutkimuksen (IEA) yhteydessä kirjoittaminen, jota tässäkin tutkimuksessa tarkastellaan omana kielikoetyyppinä, määriteltiin sekä sosiaalisesta että kognitiivisesta näkökulmasta seuraavasti:

- a) Se on viestintätapahtuma kirjoittajan ja lukijan välillä. Tällä viestintätapahtumalla on tietty aihe. Sillä on myös tarkoitus niin kuin suurimmalla osalla ihmisen toiminnoista.
- b) Se on kognitiivinen prosessi, joiden avulla kirjoittaja esittää näkemyksensä tai pohdintansa joko ulkoisesti havaitusta ja/tai sisäisesti koetusta todellisuudesta (Vähäpassi 1987c:9).

Kuten testaamisessa yleensäkin, kirjoittamisen tehtävänannon lähtökohdan tulisi olla kirjoittamisen tarkoitus. Kun testauksen tavoitteet voivat kuitenkin olla erilaisia (kieliopin tuntemus tai kommunikatiiviset taidot), on tärkeä päättää, mitä kirjoitustaito on kokeen kannalta. Taitoa kirjoittaa pidetään nykyään yleisesti eri asiana kuin kielitaitoa määriteltynä lingvististen elementtien hallintana, vaikkakaan eivät aivan irrallaan toisistaan olekaan (Weigle 2002:35). Weigle (2002) katsoo kuitenkin, että kaikenkattavan kielitaidon määritelmän sijaan saattaisi olla hyödyllisempää pohtia sitä, minkälainen kirjoittaminen olisi todennäköisesti tarpeen eri henkilöille erilaisissa tilanteissa (2002:3-4). Hughes (1989/2004:84) antaa esimerkiksi yleisen kirjoitustaidon kokeen sisällön määrittämisestä osan CCSEn³ ensimmäisen tason kokeen testispesifikaatioista (taulukko 38).

³ Cambridge Certificates in Communicative Skills in English

TAULUKKO 38 Esimerkki yleisen kirjoittamisen taidon testispesifikaatioista. (CCSE; Lähde Hughes 1989:84.)

Operaatiot (l. funktiot)	Ilmaiseminen	kiitokset, vaatimukset, mielipiteet, kommentit, asenteet, vahvistus, anteeksipyyntö, puute/tarve, tieto, valitus, syyt, perustelut
	Ohjaaminen	tilaaminen, ohjeiden antaminen, taivuttelu, neuvominen, varoittaminen
	Kuvaaminen	toiminta, tapahtumat, kohteet, ihmiset, prosessit
	Tuottaminen	tieto, ohjeet, palvelut, selvennys, apu, lupa
	Kertominen	tapahtumien järjestys
	Raportointi	kuvaus, kommentti, päätökset
Tekstityyppi		muoto, kirje (henkilökohtainen/liikeasia), viesti, faksi-ilmoitus, postikortti, resepti, raportti, ohjeet
Tekstin vastaanottaja		määrittämätön, vaikka vastaanottaja on muulla tavoin määritetty kirjoittajalle
Aihe		määrittämätön, vaikka joissakin kokeissa tehtävät liittyvät kokeen muiden osien aiheisiin
Murre		määrittämätön
Tekstin pituus		määrittämätön

Yleisintä kirjoittamisen tehtävää, jossa testattavat oikeastaan tuottavat esimerkin kirjoittamistaidostaan, kutsutaan usein suoraksi (*direct*) kirjoittamisen kokeeksi. Hamp-Lyons (1991:5) edellyttää tällaiselta tuotokselta viittä piirrettä. Testattavien tulee kirjoittaa ainakin sadan sanan tuotos, heille annetaan virike mutta heillä on vapaus tiettyssä määrin poiketa siitä, ja kunkin tuotoksen lukee ainakin yksi, mieluummin kaksi tai kolme koulutettua arvioijaa. Arviot perustuvat yhteisiin arviointiohjeisiin ja ne ilmaistaan numeroina tai numeroina ja sanallisina arvioina. White (1985) katsoo, että korkean reliaabeliuden takaamiseksi laajamittaisissa arvioinneissa edellytetään

- yksiselitteiset, kriteerein varustetut arviointiohjeet,
- mallikirjoitelmat,
- kaksi, tarvittaessa kolme, arvioijaa,
- kontrolloidut olosuhteet, so. samanaikainen arviointi,
- tarkkailijat (*table leaders*), jotka varmistavat, että asteikkojen käyttö ei vääristy.

Useimmille tällaisille kokeille on ominaista, että ne on kirjoitettava tietyn ajan kuluessa (yleensä 30-120 min) ja että aiheita ei ilmoiteta testattaville etukäteen (Weigle 2002:58-59). Tämän tutkimuksen kohteena olevassa arvioinnissa kirjoittamistehtävät edustivat tätä tyyppiä. Termiä "suora" Weigle pitää hieman harhaanjohtavana, koska Messickin (1994) mukaan mikä tahansa koe on parhaimmillaankin epäsuora indikaattori kielitaidon taustalla olevasta kyvystä. Weigle käyttääkin nimitystä *timed impromptu writing* (Weigle 2002:59). Takala on luonnehtinut tällaista tyypillistä kirjoitustehtävää luonnostelukompetenssin mittariksi (Sauli Takala, suullinen tiedonanto 27.4.2005).

Tyypillisessä kirjoittamisen kokeessa annetaan vain 3-4 aihetta, usein vähemmänkin. Esimerkiksi TOEFLin kirjoittamisen kokeessa annetaan vain yksi aihe, jolloin kokeen käytännöllisyys ja luotettavuus paranevat (Weigle 2002:147). Useita eri aiheita käytettäessä tehtävän konstruktia on vaikeampi määrittää, eli koska samaan pistemäärään voi päätyä kirjoittamalla eri aiheista, jää epäselväksi,

mitä koe mittaa (Weigle 2002:153). Käytännössä voi kuitenkin olla vaikeaa ellei mahdotonta tietää tätä tarkasti ilman työläitä menetelmiä. Voi myös kysyä, onko niin tarkka tieto itse asiassa tarpeen. Bachman ja Palmer (1996) antavat kolme neuvoa kirjoittamisen kokeen ohjeiden laatijoille: ohjeiden pitäisi olla niin yksinkertaisia, että kokelaat ymmärtävät ne, niin lyhyitä, että niiden lukeminen ei vie liikaa koeaikaa ja niin yksinkertaisia, että kokelaat tietävät tarkasti, mistä heidän odotetaan kirjoittavan. Tehtävien täsmennysten määrän samoin kuin niiden sanamuotojenkin vaihtelu voi myös vaikuttaa suoritukseen (Weigle 2002:63). Kokeen suorittajien täytyy voida tulkita tehtävänanto mahdollisimman samalla tavalla, jotta tulosten vertailukelpoisuus säilyisi (Weigle 2002:61). Kaikilla tulisi olla yhtä hyvät mahdollisuudet kirjoittaa kaikista aiheista. On siis tiedettävä, mitä tarjotaan ja mitä tulisi välttää (Weigle 2002:91).

Silloin kun kirjoittajat saavat valita eri tehtävien välillä, kokeen laatijan tulee tietää, missä määrin tehtäviä voi valita niin, että silti saadaan vertailukelpoisia tuotoksia. Jos tehtävät taas tuottavat erilaisia tuloksia, on hyödyllistä tietää, missä erot ilmenevät. Tehtävien erilaisuus voi liittyä tekstien kielipillisiin, sanastollisiin tai retorisiin eroihin. Ne voivat myös liittyä arvioijien käyttämiin erilaisiin kriteerien tulkintoihin erilaisia tehtäviä arvioitaessa (Weigle 2002:62). Taulukossa 39 esitetään joukko dimensioita, joiden suhteen kirjoittamisen kokeen tehtävät yleensä vaihtelevat.

TAULUKKO 39 Suorien kirjoittamistehtävien dimensiot. (Lähteet: Purves ym. 1984:397-8 ja Hale ym. 1996. Siteerattu Weigle 2002:63.)

Dimensio	Esimerkkejä
Aihepiiri	itse, perhe, koulu, tekniikka jne.
Virike	teksti, useita tekstejä, kuvio, taulukko
Tekstilaji	kirjoitelma, kirje, tuttavallinen viesti, ilmoitus
Retorinen tehtävä	kertominen, kuvaus, selitys, väite
Eesityksen malli	prosessi, vertailu/vastakohta, syy/seuraus, luokittelu, määrittäminen
Kognitiiviset vaatimukset ⁴	tuottaa faktoja/ajatuksia, järjestää/järjestää uudelleen tietoa, soveltaa/analysoida/tehdä synteesi/arvioida
Vastaanottajan määrittäminen	itse, opettaja, luokkatoverit, yleisö
Roolin määrittäminen	itse/tarkkailija, joku muu/oletetut henkilöt
Sävyn/Tyylin määrittäminen	tuttavallinen, muodollinen
Pituus	vähemmän kuin ½ sivua, ½-1 sivua, 2-5 sivua
Aika	vähemmän kuin ½ h, ½-1 h, 1-2 h
Ohjaus	kysymys/toteamus, implisiittinen/eksplisiittinen, kontekstin määrä
Aiheiden määrä	valittavissa / ei valittavissa
Tuottamistapa	käsin / tietokoneella
Arviointikriteerit	ensisijaisesti sisältö ja organisointi; ensisijaisesti kielellinen tarkkuus; määrittelemätön

Kirjoittamisvirikkeinä voidaan käyttää mm. lainauksia, lähdetekstejä tai kuvia. Jos halutaan yksiselitteisesti testata kielitaitoa ja sulkea pois asiatieto, kuva voi olla hyvä lähde. Jos kuva kuitenkin on moniselitteinen, kirjoittamistuotoksen

⁴ Perustuvat Bloomin (1956) oppimistavoitteiden taksonomiaan.

vertailtavuus kärsii. Lähdetekstin käyttäminen virikkeenä takaa yhteisen informaation kaikille kokelaille, vaikka se toisaalta voi olla ongelmallinenkin esimerkiksi silloin, jos joku kokelas ymmärtää tekstin tai sen osan väärin. Vieraalla kielellä kirjoitettaessa ongelmaksi voi muodostua myös lähdetekstin osien kopiointi. Jos taas halutaan arvioida, hallitsevatko testattavat tietyn tyyllilajin tai kykenevätkö he kokoamaan argumentin, joka ottaa huomioon taustatiedot ja kohdeyleisön (mikä edellyttää sekä strategista kompetenssia että kielitaitoa), lainaus voi olla hyvä lähde (Weigle 2002:94-95).

CEF (2001:61) tarjoaa seuraavia tehtävätyyppejä kirjallisen tuottamisen tehtäviksi:

1. lomakkeiden ja kyselykaavakkeiden täyttäminen,
2. artikkelien kirjoittaminen aikakausi- ja sanomalehtiin ja tiedotteisiin,
3. postereiden laatiminen,
4. raporttien, muistioiden ym. laatiminen,
5. muistiinpanojen tekeminen,
6. muistiinpanojen tekeminen sanelusta,
7. luova / mielikuvitukseen perustuva kirjoittaminen ja
8. henkilökohtaisten tai liikekirjeiden kirjoittaminen.

CEFissä (2001:82) luetellaan tehtävätyyppejä, jotka luokitellaan vuorovaikutteiseksi kirjoittamiseksi:

- muistiinpanojen, muistioiden ym. toimittaminen toiselle/toisille, kun puhuminen on mahdotonta
- kirjeet, faksit, sähköposti jne.
- sopimusteksteistä, raporteista ym. neuvottelemine muotoilemalla uudelleen luonnoksia, toimittamalla täydennyksiä, oikoluvun suorittaminen jne.
- osallistuminen on/off -line tietokonevälitteisiin konferensseihin.

6.3 Kieliopillinen kompetenssi

Purpura (2004) on koonnut esimerkkiluettelon kieliopin testaamiseen soveltuvista tehtävätyypeistä (taulukko 40) jaotellen ne sen mukaan, missä määrin tehtävä edellyttää tuottamista. Näiden kaltaisia kieliopin testaamiseen soveltuvia esimerkkejä samoin kuin pisteitystä esittelevät myös mm. Takala (1970), Kärkäinen ja Takala (1977), Hughes (1989:174-8) ja Juurakko (2002).

TAULUKKO 40 Esimerkkejä kieliopillista kompetenssia mittaavista tehtävistä. (Lähde: Purpura 2004:144-145.)

Vastausten valitseminen	Lyhyt tuottaminen	Laaja tuottaminen
Monivalinta	Aukkotäydennystehtävät	Tiivistelmät, esseet
Oikein/Väärin	Cloze-tehtävät	Dialogit, haastattelut
Yhteensovittaminen	Lyhyet avovastaukset	Roolileikit, simulaatiot
Erottelu	Sanelutehtävät	Tarinat, raportit
Sanalistat	Tiedon välittäminen	Tiedon täydentäminen
Kieliopillisen oikeellisuuden tunnistaminen	Tiedon täydentäminen	Ongelmanratkaisu
Kieliopillisen muodon käyttö-tarkoituksen tunnistaminen	Keskustelun täydentäminen	Päätöksenteko

Purpura (2004) esittää myös esimerkkejä useiden kansainvälisesti tunnettujen kielikokeiden kielioppitehtävistä tai kieliopin huomioon ottavan kielenkäytön osakokeista. Taulukko 41 on esimerkki Cambridgen yliopiston *First Certificate of English* -kokeen sisällöstä, jossa on käytetty edellisen taulukon 40 kahden ensimmäisen sarakkeen tehtävätyyppejä.

TAULUKKO 41 Kielen käytön kokeen yleiskatsaus. (Lähde: Purpura 2004:184.)

Osa	Kielitiedon (alue(et))	Tehtävätyyppi	Ohje ja odotettu vastaus	Osioiden lkm	Pisteitys
1.	Sanastollinen merkitys	Monivalinta-cloze	Tekstikatkelma. 15 aukkoa. 4 vaihtoehtoa	15	Oikein/Väärin. Yksi oikeellisuuden kriteeri.
2.	- Morfosyntaktinen muoto ja merkitys. - Sanastollinen muoto	Avocloze	Tekstikatkelma. 15 yhden sanan aukkoa	15	- " -
3.	- Morfosyntaktinen muoto ja merkitys - Sanastollinen muoto	Avainsanan muunnos (Parafraasin tuottaminen)	Avainsanaa käyttäen täydennetään lauseen aukko (2-5 sanaa) merkitsemään samaa kuin mallilause.	10	0-2 pistettä. Kaksi oikeellisuuden kriteeriä.
4.	- Morfosyntaktinen muoto ja merkitys	Virheen tunnistaminen ja korjaaminen	Lauseessa esiintyvä virhe korjataan tai todetaan lauseen oikeellisuus.	15	Oikein/Väärin. Yksi oikeellisuuden kriteeri.
5.	Sanastollinen muoto	Sananmuodostus	10 avainsanaa, joita on käytettävä oikeissa muodoissaan tekstikatkelman aukkojen täydentämiseen	10	Oikein/Väärin. Yksi oikeellisuuden kriteeri

Suullista ja kirjallista tuotosta eli taulukon 40 sarakkeen 'Laaja tuottaminen' tehtävätyyppejä arvioidaan yleensä kriteerein ja taitotaso. Taulukko 42 on esimerkki analyttisestä asteikosta, jonka kriteerit kuvaavat kieliopillista kompetenssia sekä muodon että merkityksen oikeellisuuden näkökulmasta.

TAULUKKO 42 Kielopillisen taidon analyttinen arviointiasteikko. (Lähde: Purpura 2004:121.)

<i>Taso</i>	<i>Muoto</i>	<i>Merkitys</i>
Täydellinen (5)	<ul style="list-style-type: none"> - Edustaa täydellistä tehtävänannon edellyttämää kielipillisten muotojen skaalaa, jossa esiintyy sanastollisia (subst. + subst. -muotoja), morfosyntaktisia (akt. + pass. imperfekti, <i>when</i>-lauseet) ja sidosteisuuden muotoja (kronologia, tulos). - Osoittaa näiden muotojen täydellistä tarkkuutta; vähäisiä satunnaisia virheitä voi esiintyä, mutta ne eivät missään vaiheessa hämää merkitystä. 	<ul style="list-style-type: none"> - Viesti välittyy täydellisesti ja selvästi. - Osoittaa laajaa ja tehtävänantoon sopivaa sanastoa ja sen täsmällistä käyttöä.
Laaja (4)	<ul style="list-style-type: none"> - Edustaa laajaa tehtävänannon edellyttämää kielipillisten muotojen skaalaa, jossa esiintyy sanastollisia, morfosyntaktisia ja sidosteisuuden muotoja. - Osoittaa näiden muotojen hyvää tarkkuutta; joitakin virheitä voi esiintyä, mutta ne eivät missään vaiheessa hämää merkitystä. 	<ul style="list-style-type: none"> - Viesti välittyy yleensä hyvin ja selkeästi. - Osoittaa laajaa ja tehtävänantoon sopivaa sanastoa. Sanavalinnassa saattaa esiintyä joitakin virheitä. - Viesti välittyy riittävällä tarkkuudella joistakin epäselvyyksistä huolimatta.
Keskinkertainen (3)	<ul style="list-style-type: none"> - Edustaa keskitasoista tehtävänannon edellyttämää kielipillisten muotojen skaalaa, jossa esiintyy sanastollisia, morfosyntaktisia ja sidosteisuuden muotoja. - Osoittaa näiden muotojen tyydyttävää tarkkuutta. Muodoissa esiintyy joitakin virheitä, mutta ne hämäävät harvoin merkityksen. 	<ul style="list-style-type: none"> - Viesti välittyy riittävällä tarkkuudella joistakin epäselvyyksistä huolimatta. - Ongelmia tehtävänantoon sopivan sanaston valinnassa, joka voi olla epätäydellistä tai epätarkkaa.
Rajoittunut (2)	<ul style="list-style-type: none"> - Edustaa rajoittunutta tehtävänannon edellyttämää kielipillisten muotojen skaalaa, jossa esiintyy sanastollisia, morfosyntaktisia ja sidosteisuuden muotoja. - Useissa keskeisissä muodoissa virheitä, jotka hämäävät joskus merkityksen. 	<ul style="list-style-type: none"> - Viesti on joskus epäselvä. - Monia ongelmia tehtävänantoon sopivan sanaston valinnassa, joka voi olla epätäydellistä tai epätarkkaa.
Puutteellinen (1)	<ul style="list-style-type: none"> - Edustaa hyvin rajallista tehtävänannon edellyttämää kielipillisten muotojen skaalaa - Vakavia muotovirheitä, jotka hämäävät usein merkityksen. - Ei riitä arvioinnin suorittamiseen. 	<ul style="list-style-type: none"> - Viesti välittyy juuri ja juuri tai ei lainkaan. - Osoittaa vain hyvin yksinkertaisen sanaston hallintaa. Puutteita sanaston valinnassa.

Eurooppalainen viitekehys tarjoaa myös asteikon kielipillisen virheettömyyden arvioimiseen suosittaen sen käyttöä yhdessä yleisen kielitaidon laajuutta kuvaavan asteikon kanssa (Eurooppalainen viitekehys 2003:161). Erilaisilla kielipillinäkömyksillä on arvonsa, ja ne voivat varsin hyvin täydentää toisiaan arviointitarkoituksissa (Purpura 2004:24). Suhtautumisessa kielipillisen kompetenssin testaamiseen erillisenä osakokeena tai tehtävinä voidaan karkeasti erottaa kaksi koulukuntaa. Toinen katsoo, että koska testaamisen kohde on kielitaito eivätkä sen osat, jos kielitaidossa edes on erotettavissa osia, kielipillikoikeita ei tarvita (ks. Alderson 1993:12; 2000: 98). Lienee pelätty, että erillisten elementtien testaaminen johtaa negatiiviseen takaistusvaikutukseen eli että (taas) aletaan keskittyä kielipillien opettamiseen. Toiseksi koulukunnaksi voidaan nähdä ainakin ne kielentestaaajat ja kielentestausorganisaatiot, jotka yhä sisällyt-

tävät kokeisiinsa erillisiä kieliopillisen kompetenssin osakokeita. Syitä tähän on useita. Nämä kokeet voidaan laatia nopeasti ja luotettavasti korjattaviksi ja niiden takaistusvaikutus on tarkkuutta ja täsmällisyyttä korostava. Toisin kuin kokonaiskielitaitoa testaaviin ymmärtämisen ja tuottamisen kokeisiin, kieliopipikokeeseen voidaan sijoittaa selvästi useampia osioita, olkoonkin, että monimatriisimenetelmällä näitäkin taitoja voidaan arvioida laaja-alaisemmin. Jälkimmäistä koulukuntaa edustaa tämän tutkimuksen kohteena oleva arviointi.

6.4 Sanastollinen kompetenssi

Testauksen tutkijat viittaavat toistuvasti "Bachmanin ja Palmerin -malliin" kielikokeiden laadinnan perustana sijoittaen tähän yleiseen kielitaidon testaamisen malliin eri kielitaidon osa-alueiden ominaisuuksia. Read (2000) katsoo kuitenkin, että em. tutkijat ovat jättäneet mallissaan sanaston liian vähälle huomiolle sisällyttäessään sen "vain" osaksi kieliopillista tietoa. Hän toteaa, että sanaston hallinta on tärkeä elementti useissa muissakin mallin kategorioissa, ja nostaa esille erityisesti sosiolingvistisen tiedon (Read 2000:56).

Kokeen tarkoitus luo puitteet sanaston arviointitavoille. Opetussuunnitelmaan perustuvassa ohjelma-arvioinnissa sen kohteet voidaan periaatteessa määrittää suhteessa oppimistavoitteisiin. Tätä vaikeuttaa kuitenkin opiskeltavan sanaston määrittelemättömyys. Read (2000) ohjaa frekvenssisanakirjojen käyttöön huomauttaen samalla, että on tärkeämpää osata suuri osa frekventistä sanastosta kuin suuri määrä sanoja yleensä (Read 2000:87). Sanaston arviointia Read (2000) toteaa suoritettun mm. seuraaviin kysymyksiin vastaamiseksi:

- Kuinka laaja ja syvä (= monipuolinen, erikoistunut) testattavan sanavarasto on?
- Kuinka tehokkaita erilaiset järjestelmälliset sanaston oppimistavat ovat?
- Miten oheisoppimista tapahtuu lukemisen ja kuuntelemisen yhteydessä?
- Pystyvätkö ja kuinka testattavat päättelämään tuntemattomien sanojen merkitykset tekstiyhteydestä?⁵
- Miten testattavat menettelevät korvatessaan puutteita sanavarastossaan? (Read 2000:151).

Sanaston testaamisessa katsotaan olevan kaksi ääripäätä: kielen yksikköjen (*discrete point*) testaus, jonka tavoitteena on selvittää, tietääkö testattava jonkin yksittäisen sanan merkityksen ja/tai osaako hän käyttää sitä, ja testattavan sanan sijoittamisen lause- tai tekstiyhteyteen, jolloin se toimii vuorovaikutuksessa kielitaidon muiden komponenttien kanssa. Read (2000) ehdottaa näille näkemyksille kolmea ulottuvuutta (kuvio 25).

⁵ Tätä on Suomessa tutkinut Vaurio (1998).

<p style="text-align: center;">Erillinen</p> <p>Sanaston tuntemuksen tai käytön mittaaminen itsenäisenä konstruktina.</p> <p style="text-align: center;">Valikoiva</p> <p>Mittaaminen, jossa tietyt sanat ovat arvioinnin kohteina.</p> <p style="text-align: center;">Kontekstista riippumaton</p> <p>Mittaaminen, jossa testattava voi tuottaa odotetun vastauksen viittaamatta mihinkään kontekstiin.</p>	<p style="text-align: center;">Osana laajempaa kokonaisuutta</p> <p>Sellaisen sanaston mittaaminen, joka muodostaa osan jonkin toisen laajemman konstruktin arviointia.</p> <p style="text-align: center;">Kattava</p> <p>Mittaaminen, joka ottaa huomioon materiaalin (tekstin/ puheen ymmärtämisen) tai testattavan vastauksen (kirjallinen ja suullinen tuottaminen) koko sanaston sisällön.</p> <p style="text-align: center;">Kontekstiriippuvainen</p> <p>Mittaaminen, jossa arvioidaan testattavan kyky ottaa huomioon kontekstin tarjoama informaatio ja tuottaa odotettu vastaus.</p>
---	---

KUVIO 25 Sanaston arvioinnin dimensiot. (Lähde: Read 2000:9.)

Tämän tutkimuksen kohteena olevan arvioinnin sanastotehtävät voidaan luokitella dimensiolla valikoiva – kontekstiriippuvainen. Eurooppalainen viitekehys (2003) tarjoaa sanaston arviointiin kaksi asteikkoa, sanaston laajuus ja sanaston hallinta (Eurooppalainen viitekehys 2003:159). Sanan merkityksen päättelemiseenkin on laadittu asteikko: Vastaanottamisstrategiat: vihjeiden tunnistaminen ja johtopäätösten teko puhutussa ja kirjoitetussa tekstissä (Eurooppalainen viitekehys 2003:110).

Konstrukteja määritettäessä törmätään myös käsitpareihin reseptiivinen – produktiivinen ja passiivinen – aktiivinen sanavarasto. Hallinnan osoittamista kuvataan sanoilla tunnistaa – muistaa ja ymmärtää – käyttää. Ei ole kuitenkaan olemassa yksimielisesti hyväksyttyä teoriaa siitä, miten em. käsitteet tulisi määrittellä konstruktin perustaksi eli missä esimerkiksi kulkisi raja sen välillä, milloin sana muuttuu tunnistettavasta muistettavaksi (Read 2000:153-5).

Perinteisissä sanastokokeissa on käytetty erityisesti seuraavia tehtävätyyppejä:

- erilaiset monivalintatehtävät,
- sanojen yhdistäminen synonyymeihin tai määritelmiin,
- käänös,
- tarkistuslista (kyllä/ei -koe, jossa testattavat ilmaiset, tuntevatko sanan vai eivät) (Read 2000:87).

Lause- tai tekstiyhteyttä edellyttäviä tehtävätyyppejä ovat

- tutun sanan merkityksen tunnistaminen,
- sanan merkityksen päätteleminen kontekstin avulla,
- täydennys (ilman vihjeitä) ja
- useiden sanojen ja erilaisten muotojen täydentäminen tekstikatkelmaan (Read 2000:163)

Sanaston testaaminen kontekstissa on aikaa vievää eritoten puheen ymmärtämisen ollessa kyseessä. Jos halutaan saada tietoa sanaston laajuudesta, on taloudellisempaa käyttää kielen yksiköiden testaustyypppejä. Koetta, joka on dimensioiltaan osana laajempaa kokonaisuutta kattava ja kontekstiriippuvainen,

on paikallaan käyttää silloin, kun sanaston hallinnasta halutaan tietää jotakin sellaista, jonka konteksti määrittää eli sen avulla tehdään johtopäätöksiä (Read 2000:161).

6.5 Yhteenveto

Ymmärtämisen kokeen laadinnan lähtökohtana on arvioitava taito, jonka testaamiseksi on siis löydettävä sopivat tekstit ja tehtävätyypit, joista kaikilla on omat ansionsa ja puutteensa. Tehtävästä suoriutumiseen vaikuttavat kuitenkin monet muutkin muuttujat. Erityisesti on huomattava kysymystyyppien yhteys toivotun arviointitiedon tuottamiseen. Ymmärtämisen arvioinnin avainkysymykset kohdistuvat tekstiin (millaista kieltä tulisi pystyä ymmärtämään), tehtävään (mitä testattavan pitäisi ko. alueesta ymmärtää), vastaukseen (miten ymmärtäminen osoitetaan) ja pisteitysmenetelmiin. Mahdollisimman suuri määrä luotettavaa tietoa saadaan laajalla teksti- ja tehtävätyypivalikoimalla sekä tiedostamalla, mitkä tekijät vaikuttavat tekstin ymmärtämiseen vaikeuttavasti tai helpottavasti. Puheen ymmärtämistä tulisi pyrkiä testaamaan nimenomaan sellaisissa kielenkäyttötilanteissa, joissa kielitaitoa ei voida kirjallisella kokeella osoittaa. Tekstin ymmärtämisen testaamisessa korostetaan mielekkäiden, kiinnostavien tekstien valintaa.

Tuottamiskokeen laadintaan sisältyvät konstrukti(e)n ja tehtäväannon laadinta, mahdollisen oheisaineiston hankinta/laadinta, odotetun puhe-suorituksen kuvaus ja arvosteluperusteet. Laatijan käytössä olisi suotavaa olla tehtävätyyppien malleja ja kokeen arvostelijalla eritasoisia mallisuorituksia. Konstruktiivisen kuvauksen tulisi kattaa vaadittu kielellinen kompetenssi, strateginen kapasiteetti sekä tekstuaalinen, pragmaattinen ja sosiolingvistinen tieto. Suullisessa kokeessa keskustelukumppani vaikuttaa eri tavoin testattaviin. Vastauksen ennakoitavuuden punnitseminen on tehtäviä laadittaessa ja valittaessa tämän vuoksi tärkeää, ja siksi tehtävätyyppien vahvuuksien ja heikkouksien tunteminen auttaa arvioinnin suuntaamisessa. Kirjoittamista arvioitaessa tulee pohtia erityisesti virikkeiden (kuvat, kuvat, lähdetekstit, konteksti) vaikutusta suoritukseen yleensä ja tietyn tyyppisiin testattaviin erityisesti. Arvioinnin luotettavuuden varmentamiseksi tulee pyrkiä arvioimaan etukäteen eri aiheista kirjoitettujen tuotosten vertailukelpoisuus.

Arviointitarkoituksiin kieliopillinen kompetenssi on määriteltävät suhteessa siihen, halutaanko tutkia testattavan tietoja rajatussa kontekstissa vai sitä, kuinka hyvin ja missä laajuudessa hän osaa niitä itse käyttää. Kieliopillisen kompetenssin hallinnan testaamiseen soveltuvat siis useimmat yleisesti käytetyt tehtävätyypit edellyttäen, että tehtävät ja arvosteluohjeet on täsmennetty arviointitavoitteiden suuntaisiksi.

Sanaston arvioinnissa huomiota voidaan kiinnittää testattavan sanavaraston laajuuteen ja syvyyteen, siihen, millä tavoin sanoja on opittu, sanapäätelykykyyn tai sanavaraston puutteiden kompensointitapoihin Koetta laadittaessa

on tarpeen kiinnittää huomiota käytännöllisyyteen. Jos halutaan tehdä johtopäätöksiä sanaston laajuudesta, saattaa olla järkevää testata kielen yksiköitä. Jos taas tavoitteena on selvittää sanojen toimiminen vuorovaikutuksessa muiden kielen komponenttien kanssa, on paikallaan testata sanastoa teksti- tai lauseyhteydessä.

Edellä on käsitelty useita kielitaidon eri osa-alueiden arviointiin liittyviä periaatteita ja niiden sovelluksia. Luvussa 8 käsitellään niiden soveltamista arviointivälineiden laadinnassa ja arvioinnin toimeenpanemisessa. Seuraavassa luvussa 7 kuvataan tutkimuksessa käytettyjä metodeja ja tutkimusaineisto.

7 TUTKIMUSAINEISTO JA -METODI

Tässä luvussa taustoitetaan syitä alkuperäisen arvioinnin toimeenpanoon vuonna 1999 ja myöhempää tarvetta analysoida uudelleen tuolloin saavutetut oppimistulokset. Tutkimuskysymysten ja aineiston esittelyn jälkeen kuvataan taitotasojen asettamisen teoreettinen tausta sekä klassisen ja uuden testiteorian sovellusalueet. Seuraavaksi esitetään asiantuntijoiden suorittamat arviointitoimenpiteet, jotka olivat välttämättömiä aineiston tilastollisen käsittelyn pohjaksi. Analyysit kuvataan vain siltä osin, kun ne liittyvät aineiston osoittamiseen IRT -analyyseihin soveltuvaksi. Taitotasojen konkreettiseen asettamiseen liittyvät analyysit esitetään tuloluovussa 9. Näin havainnollistuu se, miksi erilaisia korjaustoimenpiteitä tarvittiin ja miten niihin päädyttiin. Lopuksi kuvataan Euroopan neuvoston kielipolitiikan jaoksen hyväksyttävältä taitotasojen määrittämiseltä edellyttämä arviointiprosessin kuvaus

7.1 Tutkimuksen tausta

Yhtenä syynä englannin kielen ja sen jälkeen välittömästi toisen kotimaisen ja muiden A-kielenä opiskeltavien kielten kansallisten arviointien järjestämiseen vuosina 1999-2002 oli todennäköisesti odotettavissa olevan uuden opetussuunnitelman perusteiden valmistelu. Se tuli tuntuvassa määrin perustumaan toistakymmentä vuotta valmisteilla olleeseen ja vuonna 2001 englanniksi ilmestyneeseen vieraiden kielten oppimista, opetusta ja arviointia käsittelevään Euroopan kieltenopetuksen ja kielitutkintojen läpinäkyvyyden lisäämiseksi tarkoitetun teoksen Common European Framework of Reference for Languages: learning, teaching, assessment (2001) kielitaitokäsitykseen ja erityisesti sen taitotasojärjestelmään. Koska tarkoitus oli muokata niiden pohjalta Suomen kielenopetusjärjestelmään soveltuva asteikko, englannin kielen kansallisen arvioinnin avulla oli mahdollista saada tietoa oppilaiden kielitaidon tasosta.

Englannin kielen vuoden 1999 kansallista arviointia toimeenpanemaan valittiin projektipäälliköksi tämän tutkimuksen tekijä. Arviointia taustoitettaessa todettiin, että yhteys aikaisempaan tutkimukseen oli työläs muodostaa, koska

arviointeja varten oli yleensä laadittu eri kokeet, joiden tulosten välistä luotettavaa vertailua ei voitu suorittaa. Myös kontaktit kansainväliseen vieraiden kielten arviointityöhön puuttuivat, mikä vaikeutti mm. perehtymistä silloin ajankohtaiseen Eurooppalaisen viitekehyksen kehittämistyöhön. Ongelmalliseksi osoittautui myös tuottamiskokeiden arviointi, koska "virallisia" työkaluja arviointiin ei ollut. Sekä arviointiprosessi että tuottamiskokeiden kriteeristö ja arviointiasteikko näytti olevan tarpeen suunnitella. Edellinen kyettiin hahmottelemaan ja kirjaamaan, jälkimmäistä sen sijaan ei. Tuottamiskokeiden arviointiin tarkoitettut "viralliset" kriteerit ja taitotasosteikot (A2001) laadittiin vasta vuosina 2001 ja 2002 järjestettyihin toisen kotimaisen kielen ja ranskan, saksan ja venäjän kansallisiin arviointeihin.

Kiinnostava tilaisuus tämän tutkimuksen tekemiselle tarjoutui, kun osoitautui mahdolliseksi verrata eri taitoja mittaavien koeosioden vaikeustasoa toisiinsa kokeen tyyppistä riippumatta. Kun englannin kielen kansallisen arvioinnin tulokset raportoitiin vuonna 2000, todettiin, että puheen ymmärtämisen kokeen ratkaisuprosentti oli alempi kuin muiden kielitaidon osa-alueiden. Klassisen testiteorian analyysien perusteella ei kuitenkaan voitu selvittää, oliko koe vaikeampi kuin muut osakokeet vai oliko oppilaiden puheen ymmärtämisen taito heikompi kuin muilla osa-alueilla. IRT -tekniikan käyttö analyyseja tehtäessä mahdollistaa tämän asian tutkimisen. Koska englannin ja muiden vieraiden kielten produktiivisen kielitaidon tuotokset oli arvioitu eri kriteereitä käyttäen, niidenkään tulokset eivät olleet vertailukelpoisia.

Jotta yhteys aikaisempaan tutkimukseen voitaisiin muodostaa ainakin vuosien 1999, 2001 ja 2002 ja ehkä seuraavienkin arviointien välille, oli tarpeen toisaalta määrittää eri asteikkojen, CEF, OPS 2004 ja A2001 taitotasojen katkaisukohdat ja linkittää em. arviointien oppimistulokset näihin taitotasoihin. Uudet tilastolliset menetelmät mahdollistavat myös tämän. Koska linkittämismenetely edellyttää arviointiprosessin luotettavaa kuvausta, oli mahdollista kirjata myös se.

7.2 Tutkimuskysymykset ja tutkimusaineisto

Edellä mainittujen pohdintojen jälkeen syntyi päätös seuraavien asioiden selvittämiseksi:

- Vuoden 1999 englannin arviointiprosessin tarkka kuvaaminen.
- Oppimistulosten ilmaiseminen IRT -analyysien avulla.
- Taitotasosteikkojen vastaavuuden määrittäminen
- reseptiivisten kokeiden osalta CEF- ja OPS 2004-taitotasosteikkojen välillä ja
- produktiivisten kokeiden osalta CEF, A2001- ja OPS 2004 -taito-tasosteikkojen välillä.
- Oppimistulosten ilmaiseminen em. taitotasolla.

Tavoitteet kiteytyivät seuraaviksi viideksi tutkimuskysymykseksi.

1. Millaisia reseptiivisen kielitaidon (puheen ja tekstin ymmärtämisen sekä kieliopillisen kompetenssin) kokeiden osiot ovat klassisen ja uuden testiteorian menetelmillä analysoituina?
2. Miten CEF- ja OPS 2004 -taitotasoasteikot vastaavat toisiaan reseptiivistä kielitaitoa arvioitaessa?
3. Mikä perusopetuksen päättövaiheen oppilaiden reseptiivinen kielitaito on CEF- ja OPS 2004 -taitotasoasteikoilla ilmaistuna?
4. Miten CEF-, OPS 2004- ja A2001 -taitotasoasteikot vastaavat toisiaan produktiivista kielitaitoa arvioitaessa?
5. Mikä perusopetuksen päättövaiheen oppilaiden produktiivinen kielitaito on CEF-, OPS 2004- ja A2001 -taitotasoasteikoilla ilmaistuna?

Varsinainen tutkimusaineisto koostuu kolmesta osasta:

- a) Englannin arviointiin vuonna 1998 laaditut kokeet:
 - kolme puheen ja kolme tekstin ymmärtämisen versiota, joista edellisiin sisältyi myös kieliopillista kompetenssia ja sanastoa mittaavia tehtäviä ja jälkimmäisiin kolme kirjoitelman aihetta kuhunkin ja
 - kuusi puheen tuottamisen tehtävävihkoa, joissa jokaisessa osat A toiselle ja B toiselle parityöskentelyn osapuolelle.
 Kokeiden tarkempi sisältö kuvataan luvussa 8.
 - b) perusopetuksen 9. luokan oppilaiden (n = 4854) ymmärtämisen ja kieliopillisen kompetenssin kokeiden osiot (n = 111)
 - c) asiantuntijoiden (n = 2) konsensusarviot em. tehtävien ja osioiden vaikeustasosta sekä otokset puhumisen (n = 52) ja kirjoittamisen (n = 50) kokeen osaotoksista, jotka arvioitiin CEF- ja OPS 2004- A2001 -taitotasoasteikoilla
 - d) asiantuntijoiden (n = 2)⁶ arviot puhumisen (n = 286) ja kirjoittamisen (n = 539) kokeiden tuotoksista⁷, jotka arvioitiin A 2001-taitotasoasteikolla.
- Jotta olisi mahdollista linkittää vuoden 1999 englannin kansallisen arvioiden koetulokset sekä CEF-, OPS 2004- että A2001 -taitotasoasteikkoihin ja myös verrata reseptiivisten kokeiden vaikeustasoja toisiinsa, tuli
- 1) yhdistää (a) asiantuntijoiden arviot reseptiivisen kielitaidon kokeiden tehtävien ja osioiden vaikeustasosta sekä puhe- ja kirjoittamistuotosten tasosta ja (b) empiirinen tieto tehtävien todellisesta vaikeustasosta sekä
 - 2) määrittellä taitotasorajat kansallisen kokeen summapistemääräjakautumassa.

Kokeiden vaikeustason määrittämiseksi ja taitotasojen linkittämiseksi toisiinsa tarvittavat toimenpiteet olivat:

1. Puheen ja tekstin ymmärtämisen kokeiden tekstien ja osioiden vaikeustason arvioiminen CEF- ja OPS 2004-taitotasoasteikkoja käyttäen.
2. Kieliopillisen kompetenssin kokeen osioiden vaikeustason arvioiminen.
3. Puhumisen kokeen suppeahkon, mutta riittäväksi arvioidun, oppilasotoksen (n = 54) koesuoritusten arviointi.
4. Kirjoittamisen kokeen suppeahkon oppilasotoksen (n = 50) tuotosten arviointi A2001-, CEF-, CEF -käsikirjan ja OPS 2004-taitotasoasteikkoja käyttäen.
5. Arviointitiedostojen luominen.
6. Aineistojen tilastollinen käsittely.
7. Tulosten tulkinta.

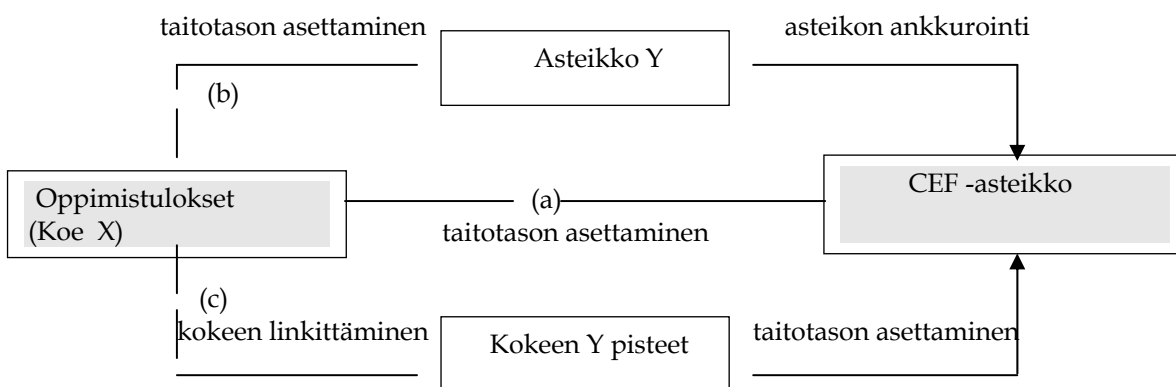
⁶ Toinen asiantuntijoista oli sama kummallakin kerralla.

⁷ Kaikki c)-kohtaan sisältyneet kirjoitelmat eivät sisältyneet tähän otokseen.

7.3 Taitotasojen asettamisen teorettinen tausta

Kielitutkintojen tai -kokeiden linkittämisellä tarkoitetaan eri taitotasoasteikkoja käyttäen arvioitujen tutkinto- tai arviointitulosten välisen vastaavuuden määrittämistä. Vastaavuuden saavuttamiseksi on määritettävä tietty minimipistemäärä, standardi, joka on taitotason katkaisukohta (*cut-off point*) (Figueras ym. 2005:269-70). Linkittäminen voidaan suorittaa jollakin seuraavista kuvion 26 osoittamista tavoista eli linkittämällä tulokset

- suoraan CEF -asteikkoon (esim. produktiivisen kielitaidon tulokset) (a),
- johonkin muuhun asteikkoon, joka puolestaan on jo linkitetty CEF -asteikkoon (b),
- sellaisen kokeen tuloksiin, joka on jo linkitetty CEF -asteikkoon (c).



KUVIO 26 Linkittämisen prosessi. (Lähde: Kaftandjieva 2004:1.)

Tässä tutkimuksessa sovellettiin linkittämismenettelyjä *a* ja *b*. Kuten kuvio 26 osoittaa, taitotasojen asettaminen (*standard setting*) on välttämätön toimenpide olipa kyseessä mikä tahansa linkittämistapa. Termi viittaa kriteeriviitteiseen arviointiin olennaisesti liittyvään ns. absoluuttiseen päätöksentekoprosessiin, jolla pyritään luokittelemaan oppimistulokset rajalliselle määrälle peräkkäisiä, yleensä portaittain vaikeutuvia suoritustasoja (Kaftandjieva & Takala 2002; Lynch 2003; Kaftandjieva 2004; Takala 2005a). Suuri määrä koepisteitä siis ikään kuin puristetaan joukolle taitotasoa, joita esimerkiksi CEF -asteikossa on kuusi, OPS 2004:ssä 10 ja A2001:ssä samoin kuusi.

Taitotasojen asettamisen menetelmiä on hyvinkin 50 (Kaftandjieva 2004:11). Ne edellyttävät asiantuntija-arvioita siitä, mikä taidon, kyvyn tai tiedon riittävä taso on, sekä siitä, että arvioitavan kokeen pisterajat vastaavat sovittua taitotasoa (Lynch 2003:99). Tämän vuoksi arvioijien valinta ja koulutus ovat keskeisiä luotettavien taitotasojen asettamisessa (Kaftandjieva 2004:16). Eräs tunnetuimpia taitotasojen asettamisen menetelmiä on Angoff -menetelmä, joka perustuu testattavien tuntemiseen. Siinä asiantuntijajaneeli (9-11 henkilöä) (Kaftandjieva 2004:23) arvioi kunkin tehtävän ratkaisemiseen tarvittavan hallinnan määrittämällä, millä todennäköisyydellä testattava suorittaa tehtävän oikein. Näiden arvioiden kes-

kiarvosta tulee standardi, joka määrittää minimipisterajan eli taitotason katkaisukohtaan (Lynch 2003:99). Useimmat menettelyt edellyttävät empiiristä dataa ja useissa käytetään IRT -malleja. Pääongelmana IRT -menetelmien soveltamisessa taitotasojen asettamiseen on se, että koska mallit liittyvät todennäköisyyksiin, ne edellyttävät lisäharkintaa ratkaistaessa vaadittavaa osion hallinnan tasoa. Useimmissa IRT -malleissa osion vaikeustaso määritellään kohdaksi, jossa tämän tasoisella testattavalla on 50 %:n mahdollisuus saada ko. osio oikein. Monien testauksen asiantuntijoiden mielestä tämä on kuitenkin liian alhainen taso (Kaftandjieva 2004:17).

Taitotasojen asettamisessa on kuusi vaihetta:

1. Menetelmän valinta.
2. Arvioijien valinta.
3. Arvioijien koulutus.
4. Arviointiprosessi.
5. Pisterajojen määrittäminen.
6. Dokumentointi ja validointi (Kaftandjieva 2004: 28-30).

Menetelmän tai menetelmien valinta riippuu mm. tehtävätyypistä, osioiden lukumäärästä, otoskoosta, empiirisen datan käyttömahdollisuudesta, tutkinnon / kokeen tärkeydestä ja käytettävissä olevista resursseista. Arvioijien tulisi ihannetapauksessa olla aine-eksperttejä, tuntea testattavien yleistaso ja kyetä erottamaan erilaisia taitotasoja, osata arvioida osioiden vaikeustaso, tuntea testattavien opetussuunnitelma, ymmärtää taitotasojen asettamisen seuraukset ja edustaa kaikkia asianosaisia. Heidän koulutukseensa tulisi myös kiinnittää huomiota niin, että jokainen heistä tulkitsee kriteerit mahdollisimman samalla tavalla (Kaftandjieva 2004:28). Taitotasojen asettajalta edellytetään menetelmiin perehtymisen lisäksi laajaa ohjelmistojen tuntemusta ja niiden käyttötaitoa. Olennainen osa taitotasojen asettamisprosessia on arvioijien ja taitotasojen asettajan saumaton yhteistyö sekä arvioinnin ja asettamistoimenpiteiden huolellinen dokumentointi, mikä samalla tukee taitotasojen asettamispäätöksiä ja koko prosessin validointia.

Usein taitotasojen asettamista seuraa toinen agregointi, jolloin tarkoituksena on yhdistää eri tehtävien tai taitojen tulokset yhdeksi kokonaissuorituksen tasoa ilmaisevaksi tulokseksi. Tällaisia koulutuksen arvioinnissa käytettäviä taitotasojen asettamisstrategioita ovat mm. kompensoiva (*compensatory*), yhdistävä (*conjunctive*) ja yhdistetty (*mixed*) strategia. Ensimmäisessä näistä jonkin tehtävän suorittamisessa saavutettu hyvä tulos korvaa jonkin heikon tuloksen kuitenkin niin, että lopputulos on järkevässä suhteessa mitattavan konstruktin asettamiin vaatimuksiin. Yhdistävä strategia edellyttää jonkin ennalta asetetun minimipistemäärän saavuttamista jokaisessa tehtävässä. Jos tehtävät tai testattavat taidot eivät kuitenkaan ole kokonaisuuden kannalta yhtä tärkeitä, on hyvä ratkaisu yhdistää edellä esitetyt strategiat ja lähteä siitä, että tiettyihin tehtäviin ja taitoihin asetetaan minimirajat, mutta samalla tietyissä tapauksissa hyvä tulos kompensoi heikon. Yhtä ainoaa ylitse muiden olevaa taitotasojen asettamisstrategiaa ei ole (Kaftandjieva 2004:3).

Eri menetelmiä käyttäen päädytään yleensä jossakin määrin eri tuloksiin, mikä on aiheuttanut runsaasti hämmennystä ja epäluuloa niitä kohtaan. Jokaisella menetelmällä on etunsa ja puutteensa. Koska parasta menetelmää ei (vielä) ole olemassa, suositellaan käytettäväksi useaa, kahta tai jopa kolmea tapaa, jotka eroavat toisistaan erityisesti suhteessa arviointitehtävään, joka voi olla testi- tai oppilaskeskeinen⁸. Koska oppilaskeskeiset menetelmät (Kaftandjjeva 2004:12) edellyttävät testattavien tason hyvää (ainakin oletettua) tuntemista, laajoissa kansallisissa arvioinneissa niiden soveltaminen ei onnistune. Toisaalta se, että IRT -mallien käyttö vaatii suuren oppilas- ja osio-otoksen, vaikeuttaa menetelmän käyttöä sellaisissa tutkinnoissa / kokeissa, joissa osallistujamäärä on pieni (Kaftandjjeva 2004:17). Kaftandjjeva toteaa (1999:98), että osioiden määrän tulisi olla 20-60 ja testattavien määrän 200-1000. Keskustelua käydään myös siitä, tulisiko hallinnan taso määrittää ennen vai jälkeen taitotasojen asettamista (Kaftandjjeva 2004:22).

Taitotasojen asettamisen keskeisiä tekijöitä ovat palaute arvioijien taitotasotulkinnoista, empiirinen data osioiden vaikeustasosta ja pisteiden jakaumasta sekä ryhmäkeskustelu (Kaftandjjeva 2004:15). Kaftandjjevan artikkelissaan (2004) analysoimista empiiristä dataa edellyttävistä menetelmistä miltei kolmasosassa käytetään IRT -malleja (Kaftandjjeva 2004:16). Niiden käyttämisellä on monia etuja mm. riippumattomuus otoksesta ja kokeen suorittajista sekä se, että jokaisen taidon mittauksen keskivirhe tiedetään etukäteen asteikon jokaisella portaalla (Kaftandjjeva 1999:97, 2004:17). Kun nykyään on lisäksi käytettävissä runsaasti käyttäjäystävällisiä tämän kaltaisten analyysien tekemiseen tarkoitettuja ohjelmistoja, kiinnostus IRT -mallien käyttöä kohtaan on lisääntymässä.

Tulosten linkittäminen eri taitotasolle ja taitotasojen asettaminen on erityisesti Euroopan unionin alueella tapahtuvan kielitaidon arvioinnin vertailun mahdollistamiseksi ja tutkintojen / kokeiden läpinäkyvyyden takaamiseksi koettu viime aikoina tärkeäksi. Siihen on kuitenkin jo pitkään kohdistunut voimakastakin kritiikkiä johtuen sille tyypillisestä harkinnanvaraisuudesta, joka johtuu taitotasojen asettajan suuresta osuudesta menetelmän ja strategioiden valinnassa, mikä puolestaan edellyttää soveltajaltaan perehtyneisyyttä sekä eri taitotasojen asettamismenetelmiin että tietotekniikkaohjelmien tuntemusta. Epäilyksiä kohdistuu myös taitotasojen arvioijien laadukkuuteen. Suurimmat ongelmat, joihin taitotasojen asettamisen asiantuntijat hakevat ratkaisumalleja, ovat

- eräät arviointiprosessin yksityiskohdat ja tekijät, jotka vaikuttavat taitotasoihin,
- pisterajojen määrittämismenetelmät ja niiden vaikutus pisterajoihin,
- taitotasojen asettamisen ja suoritustandardien⁹ validointi,
- eri taitotasojen asettamismenettelyjen edut ja puutteet sekä kuhunkin arviointitalanteeseen soveltuvimman menetelmän valinta.

Se, että ei ole käytettävissä riittävästi evidenssiä mallin sopivuudesta aineistoon, on osoittautunut monien kielen testauksessa käytettyjen IRT -mallien

⁸ Useimmat taitotasojen asettamismenetelmät ovat testikeskeisiä (= niissä arvioidaan tehtävien vaikeustasoa). (Kaftandjjeva 2004:14).

⁹ Suoritusstandardi = selkeä määrittelmä siitä, mitä oppilaan tulee osata tehdä osoittaakseen taitonsa tietyllä sisältöalueella, taitotasokuvaus.

puutteeksi. Evidenssi aineiston sopivuudesta malliin, jolla ei tarkoiteta pelkääntään tilastollisten menetelmien käytön tuloksia, on erityisen tärkeitä silloin, kun IRT -malleja sovelletaan taitotasojen asettamiseen. Taitotasorajoja ei voi pitää luotettavina, elleivät niiden perustelut ole sellaisia.

7.4 Klassinen ja uusi testiteoria (Item Response Theory = IRT)

Oppimistuloksia voidaan analysoida sekä klassisen että uuden testiteorian (Item Response Theory) menetelmiä käyttäen. Edellinen sisältää osion vaikeustason ja erottelukyvyn laskemisen sekä monivalintatehtävissä distraktorianalyysin tekemisen. Tällöin vaikeustasolla tarkoitetaan sitä osuutta testattavista, jotka vastaavat osioon oikein. Se ilmaistaan klassisessa testiteoriassa ratkaisuprosentteina (P). Eräs tae luotettavasta kokeesta on suuri pistemäärien hajonta, joka tutkimusten mukaan saadaan, jos kysymykset ovat keskivaikeita, jolloin noin 50 % oppilaista vastaa oikein jokaiseen osioon (Ebel 1979:89). Monivalintatehtävissä on lisäksi otettava huomioon arvaamisen mahdollisuus (Ahlström: 170-172). Ihanteellisen oikein/väärin osion ratkaisuprosentti olisi 75 ja nelivaihtoehdoisen monivalintaosion 62,5 (Ebel 1965:89). Erottelukykyindeksi *D* ilmaistaan testattavan kokonaispistemäärän ja tietyn osion pistemäärän välisenä korrelaatiokertoimena, ja se ilmaisee, kuinka usein kokeessa hyvin menestyneet ovat antaneet osioon oikean vastauksen. Se osoittaa, miten hyvin osio erottelee testin suuntaisesti eli mittaako osio samaa asiaa kuin koko koe. Se siis kuvaa suhdetta testattavan tiettyyn osioon ja koko kokeeseen antamien vastausten välillä eli ilmaisee, onko henkilö, joka on saanut korkean pistemäärän koko kokeesta, vastannut oikein myös kyseiseen osioon (Lynch 2003:76-77). Mitä voimakkaammin osio korreloi summapistemäärän kanssa, sitä toimivampi osio on. Osioiden erottelukyvyn laatua arvioidaan perinteisesti seuraavasti:

$\geq 0,40$	Erittäin hyvä
0,30 - 0,39	Hyvä, voidaan vielä parantaa
0,20 - 0,29	Hyväksyttävä, mutta kaipaa tarkastelua ja parantamista
$\leq 0,19$	Heikko osio, usein syytä poistaa.

(Popham 1999:227; Ahlström 1971:173; Ebel 1979:267).

Jos kyseessä on normiviitteinen koe, jonka päätarkoitus on sijoittaa testattavat paremmuusjärjestykseen, ihanteellinen osio on sellainen, joka erottelee heikot ja hyvät tulokset tehokkaasti (Lynch 2003:77). Kriteeriviitteisessä mittaamisessa voidaan käyttää samoja osioanalyysimittauksia kuin normiviitteisessäkin, mutta tulokset tulkitaan eri tavalla. Jos vaikeustasoindeksi on korkea, ei osiota tulkita automaattisesti huonoksi, vaan se on nyt merkki tavoiteaineksen hyvästä hallinnasta. "Huono" erotteluindeksi ei myöskään ole merkki huonosta osiosta, vaan ainoastaan siitä, että se ei kyennyt erottelemaan hyvin. Osion tehtävänä on siis vain kyetä erottelemaan ne, jotka eivät osanneet, niistä, jotka osasivat (Lynch 2003:81-82). Tosin hyvin helppo osio voi omata hyvän erottelukyvyn, jos vain kaikkein heikoimmat suorittajat eivät ole osanneet ratkaista tehtävää oikein.

Klassisen testiteorian puutteena on se, että sen avulla ei voi ilmaista testattavan taitoa absoluuttisesti, vaan pelkästään suhteessa siihen ryhmään, otokseen, jonka tuloksia analysoidaan, ja niihin osioihin, joita arviointiin on sisällytetty. Se määrittää ainoastaan mittausvälin tai tulosten/pisteiden välin, jolle testattavan tulos hyvin todennäköisesti lähes ilman mittausvirhettä eli *true score* asettuu (Lynch 2003:91). Uusi testiteoria pyrkii poistamaan nämä rajoitukset mittaustarkkuudessa tarjoamalla arvion testattavasta taidosta, joka on, toisin kuin klassisen testiteorian tuottama ratkaisuprosentti, riippumaton testattavasta ryhmästä ja osioista, joihin data perustuu (Lynch 2003; Eurooppalainen viitekehys 2003). Mikäli aineisto sopii malliin, eri osioihin vastanneiden testattavien pistemäärät ovat vertailukelpoisia, joten arvioita voi verrata toisiinsa, vaikka niihin ovat vastanneet eri vastaajat. Kun jotkut testattavat ovat taidoiltaan heikompia ja jotkut osiot taas helpompia tai vaikeampia, testattavien lopulliset pisteet heijastavat näitä suhteellisia eroja.

Taitotasojen asettamiseen, joka on tämän tutkimuksen päätavoite, käytettiin IRT -analyysiä. Theta -arvojen määrittämistä varten on laskettava reseptiivisen kielitaidon kokeen vaikeustaso ja erottelukyky.

IRT:ssä mitattavaa kykyä, taitoa tai asennetta pidetään mahdottomana havaita, ts. latenttina l. piilevänä piirteenä. IRT määrittää parametrit eli osion ominaisuuksista saatavan tiedon tyypit mallintaakseen matemaattisesti osion vastaukset. Esimerkkejä näistä parametreista ovat 3-parametrisessa mallissa erottelukyky, vaikeustaso ja arvaaminen (so. olettamus siitä, että heikosti suoriutuva testattava antaa osioon oikean vastauksen). Malli olettaa edelleen, että nämä ominaisuudet kuvaavat tarkasti jokaisen osion havaitun suorituksen ja sen taustalla olevan latentin piirteen eli tehtävän vaikeustason ja testattavan todellisen taidon välisen suhteen. Jos data suoriutuu testistä ja sopii IRT:n matemaattiseen malliin, testattavan taitojen estimaattien sanotaan olevan *sample-free*, eli ne eivät riipu siitä testattavasta ryhmästä, jonka datasta oli kysymys (Lynch 2003:91-92.). Osioden vaikeustason estimaattien sanotaan olevan *item-free*, mikä tarkoittaa että ne eivät riipu kokeessa käytetyistä osioista. Vaikeustaso kuvataan viivalla, ja henkilöä edustaa tällä viivalla oleva piste tai arvioinnista kyseen ollen, pistemäärä. Mittauksen tavoitteena on sijoittaa jokaisen henkilön mittaustulos mahdollisimman tarkasti tälle viivalle. Tässä mielessä osioon annettua vastausta (*item response*), pistemäärää, pidetään latenttina muuttujana (Verhelst 2004b:1).

Yhtä näissä laskutoimituksissa yleisesti käytettävää kaavaa kutsutaan logistiseksi funktioksi (*logistic function*). Jos sitä käytetään luonnehtimaan osioiden käyttäytymistä (*item response functions*), sanotaan käytettävän logistista mallia. Tällaisessa kaavassa latenttia muuttujaa merkitään tavallisesti kreikkalaisella kirjaimella theta (θ) (Baker 2001:6-7). IRT-mallien yleismääritelmä on, että latentin muuttujan ja vastaustodennäköisyyden välillä on tietty suhde eli että jos jollakulla on korkea θ - eli logit -arvo, hän suorittaa osion todennäköisesti oikein (Verhelst 2004b:15; Lynch 2003:92). Asteikko lähtee arvosta 0, ja se voi vaihdella positiiviseen ja negatiiviseen suuntaan (Lynch 2003:92). Jos populaation θ -arvot ovat jakautuneet normaalisti, noin 99 % niistä asettuu välille $-2,5, +$

2,5, jolloin miinus-etumerkillä varustetut arvot merkitsevät helppoja ja plus-etumerkillä varustetut vaikeita osioita, parempaa tietoa – tai ankarampia arvioijia (Lynch 2003:92). Välille -0,42 ja +0,42 asetettujen henkilöiden prosentuaalinen osuus on 32, noin kolmasosa populaatiosta (Verhelst 2004b:20).

IRT -malleja on suuri määrä, ja ne eroavat toisistaan mm. siinä, kuinka monia mittausprosessin ominaisuuksia voidaan ilmaista matemaattisesti suhteessa kokeesta suoriutumiseen. Tunnetuimpia, ja samalla vahvin mutta ei helposti toteutettava, on yksiparametrinen Raschin malli, jossa ainoa käytettävä parametri on osion vaikeusaste (Lynch 2003:93). Käsillä olevassa tutkimuksessa IRT-mallina käytettiin OPLM -analyysia, joka on yksiparametrinen, mutta käytäytyy kaksiparametrisesti eli ottaa huomioon sekä osion vaikeusasteen että erottelukyvyn, jonka analyysin tekijä määrittelee itse kokeilemalla erilaisia vaihtoehtoja ja valitsemalla sen, joka tuottaa parhaan tuloksen mallin toteutumisen kannalta. θ -arvo (b), jonka keskiarvo on 0, vaihtelee välillä -3 - +3. Erottelukyky (a) ilmaistaan numeroin 1-15, jolloin osio on erottelukyvyltään sitä parempi, mitä korkeampi sen arvo on. Yli 5-6 arvoja saavutetaan harvoin (Sauli Takala, suullinen tiedonanto 25.8.2005).

7.5 Taitotasojen asettamisen käytännön toimenpiteet

Taitotasojen asettamismenettelyksi oli valittu ns. testikeskeinen jatkumometodi (*test-centered continuum method*), joka on aiemmin mainitun Angoff -menetelmän sovellus (jota Kaftandjjeva ja Takala ovat soveltaneet menestyksellisesti mm. ylioppilaskokeen englannin kokeen tulosten taitotasolinkittämisessä. Sauli Takala, suullinen tiedonanto 21.3.2005). Sen edellytyksenä on tieto empiirisistä koetuloksista ja arvioijien taitotasoarvioista. Puheen ja tekstin ymmärtämisen sekä kieliopillisen kompetenssin kohdalla tämä tarkoitti tehtävien ja osioiden taitotason arviointia käyttäen soveltuvia asteikkoja. Puhumisen ja kirjoittamisen kokeen osalta arvioitiin tuotoksia, jotka sijoitettiin sovittujen kriteerien määrittämille taitotasoille.

Luvussa 7.5.1 kuvaillaan tutkimusaineiston puheen ja tekstin ymmärtämisen sekä kieliopillisen kompetenssin koeosoiden taitotasoarviointi. Luku 7.5.2 käsittelee produktiivisen kielitaidon kokeiden tuotosten taitotasoarviointeja. Luku 7.5.3 kuvaa, miten näitä tietoja käytettiin hyväksi taitotasojen asettamisessa eli taitotasojen katkaisukohtien määrittämisessä.

7.5.1 Puheen ja tekstin ymmärtämisen sekä kielellisen kompetenssin kokeiden taitotasojen arviointi

Useat tutkijat (mm. Alderson 2000) ovat todenneet, että tekstin vaikeustason progressiota ei ole kyetty tyydyttävästi kuvaamaan. Tähän näytti tarjoutuvan mahdollisuus, kun Hollannin opetusministeriö (*Dutch Ministry of Education, Culture and Science*) rahoitti projektin “the Dutch CEFR Construct Project” (Grid

2005; myöhemmin Grid), jonka toimeksiantona oli kehittää Eurooppalaiseen viitekehykseen mahdollisimman suuressa määrin perustuva työkalu kuvaamaan kolmella (A, B, C) kaksiosaisella CEF -taitotasolla osioiden, tehtävien ja koko kokeen taustalla olevia englannin kielen tekstin ja puheen ymmärtämisen konstruktteja (Grid 2005:1). Tämä työkalu (myöhemmin Grid -taulukot) pyrkii tarjoamaan mahdollisuuden konstruoida uusia ja analysoida jo käytettyjä koe-tehtäviä suhteessa CEF- taitotaseihin. Se mahdollistaisi systemaattisen vastauksen hakemisen J. Charles Aldersonin tunnettuun kysymykseen ”How do I know that my Level B1 is your Level B1?” (CEF-käsikirja 2003: IX).

Projektiryhmä lisäsi CEFiin dimensioita, joiden katsottiin olevan välttämättömiä testispesifikaatioiden osia (Grid 2005:14). Grid -taulukot jakaantuvat kolmeen osaan: tekstin kuvailuun, osioiden kuvailuun sekä koko tehtävän vaikeustason arviointiin. Tekstin kuvailun neljä dimensioita on poimittu CEFistä: tekstilähde (CEF 2001:48-9), kielen käyttöalue CEF 2001:45), aihe (CEF 2001:52) ja taitotasoasteikot (CEF 2001: 65-72). Puheen ymmärtämisen tekstilähteissä erotetaan omiksi lajeikseen autenttinen ja laadittu/muokattu puhe. Muut tekstin kuvailun dimensiot esiintyvät myös CEFissä, mutta paikoitellen toisin muotoilutina. Grid -taulukoissa ne ovat:

- Diskurssityyppi
- Sisällön luonne
- Pituus/Kesto,
- Kielioppi,
- Tekstin nopeus,
- Aksentti,
- Artikulaation selkeys, ja
- Puhujien lukumäärä.

Diskurssityypin määrittely perustuu DIALANGin¹⁰ arviointispesifikaatioihin. Sisällön luonne kuvaa tekstin portaittain vaikeutuvaa abstraktiotasoa. Aihealueet on poimittu Euroopan Neuvoston kynnystaitoluettelon (Threshold Level 1990) aiheiden joukosta. Pituus on tekstin ymmärtämisen kokeen sanamäärä ja kesto puheen ymmärtämisen kokeen yhden kuuntelukerran pituus. Neljä viimeistä dimensioita koskevat luonnollisesti vain puheen ymmärtämistä (Grid 2005:10)

Osiokohtaisessa tekstin tarkastelussa määritetään

- Operaatiot
- Tehtävätyyppi
- Osion arvioitu taitotaso.

Operaation määrittämisessä Grid -projektiryhmä päätyi kehittämäänsä kolmi-osaiseen kuvaukseen

1. Tunnistaa ja esittää,
2. Tehdä johtopäätöksiä
3. Arvioida.

¹⁰ The DIALANG Assessment Framework: www.jyu.fi/DIALANG

Määrytykset perustuvat kuitenkin CEFin *Can do* -väittämiin ja tukeutuvat puheen ja tekstin ymmärtämisen teorioihin (Grid 2005:19-23).

Tässä tutkimuksessa puheen ja tekstin ymmärtämisen kokeiden sisältö ja vaikeustaso kuvattiin Grid -taulukoita käyttäen. Niiden avulla oli tarkoitus tehdä olettamuksia reseptiivisen kielitaidon kokeiden vaikeustasosta pohjaksi tulosten linkittämiseksi eri taitotasolle. Vaikeustaso määriteltiin arvioimalla CEFin taitotasoasteikoilla sen tekstikatkelman vaikeustaso, johon kysymys tai väittäjä kohdistui sekä kunkin osion vaikeustaso erikseen. Arvioinnissa käytettiin taulukon 43 asteikkoja.

TAULUKKO 43 Puheen ymmärtämisen kokeen arvioinnissa käytetyt taitotasoasteikot.

Asteikon lähde	Sivu	Asteikko
CEF 2001	66	Overall listening comprehension
	66	Understanding conversation between native speakers
	67	Listening as a member of a live audience
	67	Listening to announcements and instructions

Arvioitavan tekstin tyyppin perusteella ratkaistiin, mikä tai mitkä taulukot soveltuivat parhaiten yleistaulukon rinnalla käytettäväksi. Näin siis sovellettiin Eurooppalaisen viitekehysten suositusta useiden taitotasoasteikkojen käyttämisestä eri tehtävien arviointiin (Eurooppalainen viitekehys 2003:264). Puheen ymmärtämisen kokeen sisällön arviointitulokset esitetään liitteessä 4, jota tulkitaan liitettä 5 avaimena käyttäen.

Tekstin ymmärtämisen kokeen vaikeustaso määritettiin arvioimalla koko tekstin ja kunkin osion vaikeustaso. Tässä ei siis arvioitu tekstikatkelmien vaikeustasoa. Arviointiin käytettiin taulukon 44 asteikkoja.

TAULUKKO 44 Tekstin ymmärtämisen kokeen arvioinnissa käytetyt taitotasoasteikot.

Asteikon lähde	Sivu	Asteikko
CEF 2001	69	Overall reading comprehension
	70	Reading for orientation
	70	Reading for information and argument
	71	Reading instructions

Liitteestä 6 ilmenevät tekstin ymmärtämisen kokeen vaikeustason määrittelyjen tulokset, joita tulkitaan liitteenä 7 olevaa avainta käyttäen. Molemmat ymmärtämiskokeet arvioitiin myös OPS 2004 -asteikkoja käyttäen.

Tämän tutkimuksen tekijän ja arvioinnin ja taitotasojen asettamisen asian tuntija arvioivat yhdessä ymmärtämiskokeiden sisällön ja vaikeustason konsensusarviointina. Tekstin ja puheen ymmärtämisen kokeiden arviointi CEF -asteikkojen mukaan suoritettiin yhdessä samanaikaisesti mutta aluksi itsenäisesti. OPS 2004:än nojautuvat samoin kuin kieliopillisen kompetenssin taitotasoarviointit suoritettiin eri aikaan ja eri paikassa. Tämän tutkimuksen tekijä kokosi

arviot. Yksimielisyys saavutettiin valtaosassa tapauksia ja rajatapauksissakin lyhyen keskustelun jälkeen, kun käsityksiä kuvainten sisällöstä oli täsmennetty. Tämän vuoksi arvioijien välistä korrelaatiota ei ollut tarpeen laskea. Laajamittaisissa kokeissa arvioijia on yleensä useampia, ja asteikkojen katkaisukohtien määrittämisen pohjaksiin suositellaan noin kymmentä arvioijaa (Kaftandjieva 2004:23). Tässä tutkimuksessa päädyttiin kahden arvioijan konsensusarvioon erityisesti sen vuoksi, että uskottiin kahden yhteisen laajan CEF -kokemustaan omaavan arvioijan tulkitsevan reseptiivisen kokeen vaikeustasoja kriteereitä riittävän luotettavasti. Ongelmana on saada kokoon useita arvioitsijoita, jotka ovat todella kouliintuneet arviointien tekemiseen. Useiden mutta kokemattomien arvioijien käyttö ei ole hyvä ratkaisu ongelmaan. Tällaisten kokeiden vaikeustason arviointihan on vasta alkuvaiheessa. Jatkossa ongelma pätevien arvioijien löytämisestä lieneekin vähäisempi.

Kieliopillisen kompetenssin kokeiden taitotasojen määrittämisessä käytettiin pääasiassa YKIn validoituja kielitiedon tasokuvauksia. Lisäksi sovellettiin CEFin asteikkoja General Linguistic Range (CEF 2001:110) ja Grammatical Accuracy (CEF 2001:114). Kriteereiksi oli valittu laajuus ja tarkkuus siksi, että ne kuvasivat arvioijien mielestä olennaisimmin tämän kompetenssin sisältöä. Suomen vuoden 1994 opetussuunnitelman perusteissa kieliopillinen kompetenssi on sisällytetty reseptiivisen ja produktiivisen kielitaidon tasokuvauksiin. OPS 2004 -taitotasoa arvioitaessa poimittiin erilleen kielelliseen kompetenssiin viittaavat tavoitelauseet. Kaikki edellä mainitut kieliopillisen kompetenssin arviointiasteikot ja -perustelut on koottu liitteeseen 8.

7.5.2 Puhumisen ja kirjoittamisen kokeiden taitotason arviointi

Vuoden 1999 englannin kielen kansallisen arvioinnin tulokset päätettiin alkujaan linkittää A2001 -taitotasoasteikkoon. Puhumisen kokeen arviointia varten poimittiin jokaisesta koulusta vähintään yhden parin suoritus. Arvioiduksi tuli näin 143 paria. Kaikkia tuloksia ei kuitenkaan katsottu voitavan käyttää. Osittain se johtui äänitteiden kehnosta laadusta ja osittain siitä, että opettajat osallistuivat kokeen kulkuun siinä määrin, ettei oppilastuotosta saatu arvioitua. Kirjoittamisen kokeen arviointia varten poimittiin noin joka kymmenes kirjoitelma. Arvioiduksi tuli näin 539 kirjoitelmaa. Arvioinnin suoritti kaksi peruskoulun opetussuunnitelmaan 1994 ja taitotasoasteikkoihin hyvin perehtynyttä tutkijaa.

Kun Opetushallituksella ei näyttänyt olevan resursseja linkittää muiden vieraiden kielten oppimistuloksia CEF- ja/tai OPS 2004 -taitotasoasteikkoihin, päätettiin yrittää tämän tutkimuksen puitteissa linkittää myös A2001 -asteikot (liitteet 9 ja 10) CEF- ja OPS 2004 -taitotasoasteikkoihin. OPS 2004:n perustana on CEF, joten oli oletettavissa, että näiden kesken linkittäminen voisi onnistua hyvinkin. A2001 -asteikko sen sijaan oli tarkoitettu kattamaan vain perusopetuksen oppimäärä, vaikka se ei periaatteiltaan poikennut suuresti CEF- ja OPS 2004 -asteikoista. Sen ei välttämättä siis odotettu olevan yhteismitallinen muiden asteikkojen kanssa. Oli myös mielenkiintoista nähdä, miten linkittäminen

onnistuisi ainakin jossakin määrin eri pohjalle rakentuvaan asteikkoon. Otospopulaatiohan oli selvästi riittävä johtopäätösten tekemiseen.

Tämän tutkimuksen päätavoitteena oli asettaa vuoden 1999 oppimistulokset CEF - ja OPS 2004 -taitotasoasteikoille. Puhumisen kokeen tulosten linkittämistä varten tämän tutkimuksen tekijä sekä taitotasojen arviointiin monipuolisesti perehtynyt tutkija arvioivat yhdessä CEF- ja OPS 2004 -taitotasokuvauksia käyttäen 26 parin eli 52 oppilaan suoritukset. Myös kolmas asiantuntija arvioi nämä tuotokset. Oppilasparit valittiin aikaisemmin arvioidun suoritustason, kuntaryhmän ja koulun maantieteellisen sijainnin perusteella pyrkien samalla sukupuolten väliseen edustavuuteen. Tämän otoksen perusteella tuotettiin interpoloimalla koko otokselle puhumisen kokeen taitotasoarviot samalla tavalla kuin ylioppilastutkimuksen englannin kokeen taitotasojen asettamisprojektissa (Sauli Takala, suullinen tiedonanto 21.3.2005). CEF -taitotasoasteikkojen lisäksi käytettiin CEF -käsikirjan¹¹ asteikkoja sen periaatteen mukaan, että taulukot täydentävät toisiaan kunkin tehtävän tavoitteiden suuntaisesti. Taulukossa 45 esitetään käytetyt taitotasoasteikot.

TAULUKKO 45 Puhumisen kokeen arvioinnissa käytetyt taitotasoasteikot.¹²

Asteikon lähde	Sivu	Asteikko
CEF 2001	74	Overall Spoken Interaction
	117	Phonological Control
	124	Turn taking
CEF-käsikirja, taulukko 4.4: Relevant Qualitative Factors for Spoken interaction		Linguistic Range
		Linguistic Accuracy
		Linguistic Interaction

Arvioiden perustana käytettiin yleistä suullisen vuorovaikutuksen taitotasoasteikkoa, jota CEF -käsikirjan laajuus- ja tarkkuuskriteerit täsmensivät. Fonologia-asteikkoa käytettiin lähinnä lukemistehtävän arvioinnissa. Vaativimmassa keskustelutehtävässä sovellettiin lähinnä yleistä suullisen vuorovaikutuksen ja puheenvuoron vaihtamisasteikkoa. Suorituksia arvioitaessa OPS 2004 -tasokuvauksia käytettiin sellaisenaan.

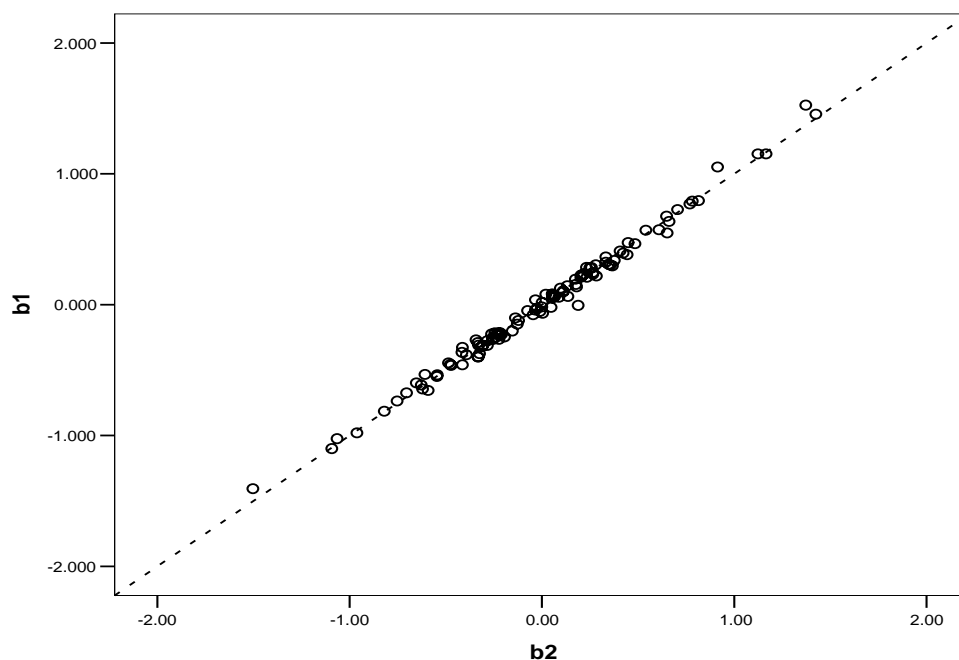
Puhetuotoksia arvioineet asiantuntijat arvioivat CEF- ja OPH -taitotasokuvauksia käyttäen 50 oppilaan kirjoitelmat. Tämän määrän oli katsottu riittävän yleistyksen koko populaation osalta (ks. edellä). Kirjoitelmat valittiin poimimalla aikaisemmasta 539 kirjoitelman otoksesta noin joka seitsemäs kirjoitelma. Suorituksia arvioitaessa sovittiin OPS 2004 tasokuvausten lisäksi käytettävän varsinaisia CEF-asteikkoja tarkempaa CEF-käsikirjan taulukkoa 5.5 "Written Assessment Criteria Grid" (liite 12). Sen arviointikriteerit ovat: yleinen kirjoittamisen taito (*overall*), laajuus (*range*), koherenssi (*coherence*) ja tarkkuus (*accuracy*). Kirjoitelmatyyppien luonnehdinta esitetään luvussa 8.

¹¹ Relating Language Examinations to the Common European Framework of Reference for Languages; Learning, Teaching, Assessment 2003.

¹² Lukijan kokonaiskuvan saamisen helpottamiseksi edellä mainitut asteikot on yhdistetty liitteeksi 11.

7.5.3 Aineistojen tilastollinen käsittely

Jotta koetulokset voitiin analysoida IRT -menetelmällä, mikä oli välttämätöntä θ -arvojen laskemiseksi, oli löydettävä matemaattinen malli, johon koetulokset sopisivat. Aineiston sopivuutta malliin testattiin OPLM -tilastotyökaluun kuuluvalla $R1c^*$ -analyysillä, jolloin saatiin tulokseksi p -arvo 0,0. Se osoittaa, että aineisto erosi merkitsevästi IRT -mallioletuksesta, koska p -arvon tulisi olla suurempi kuin 0. Tulos johtuu osittain oppilasotoksen erittäin suuresta koosta, jolloin ei ole tarkoituksenmukaista verrata jokaisen osion sopivuutta jokaiseen testattavaan, ja osittain vain yhden ankkuritehtävän käytöstä, mikä sellaisenaan ei ollut optimaalinen ratkaisu. Koska koe ei sopinut valittuun malliin, oli valittava toinen menetelmä. Kun aineiston laadun tarkastelua jatkettiin mallin hakemisen epäonnistuttua (todennäköisesti - ja odotetustikin - aineiston massiivisuuden vuoksi), osioparametrien vakaus arvioitiin jakamalla otos satunnaisesti kahteen osaan ja kalibroimalla ne erikseen. Kahden otoksen väliseksi korrelaatiokertoimeksi saatiin 0,996 (ks. Verhelst 2004b:6). Osioiden vaikeustason estimointi laskettiin käyttämällä Pearsonin korrelaatiokerrointa, joka ilmoittaa kahden muuttujan välisen lineaarisen yhteyden voimakkuuden. Arvo, joka on lähellä +1:tä, ilmaisee voimakkaasti positiivista yhteyttä eli toisen muuttujan kasvaessa myös toisen muuttujan arvot kasvavat (Nummenmaa 2004:268). Osaotosten välillä osoittautui olevan hyvin voimakas yhteys ($r = .99$), mikä osoittaa paitsi osien välisen myös osioiden välisen vertailukelpoisuuden, eli yksikään osio ei toisessa otoksessa ollut selvästi vaikeampi tai helpompi kuin toisessa, vaan osiot toimivat samalla tavalla molemmissa ryhmissä, mikä on vankka osoitus osioiden stabiilista mittauskyvystä (Verhelst 2004b:15). Yhteys voidaan kuvata myös sirontakuvion (kuvio 27) avulla, jossa jokainen piste kuvaa yhtä havaintoa eli osiota. Osion sijainti x-akselilla kuvaa toisen osaotoksen mittaustulosta (arvoa) ja vastaavasti sijainti y-akselilla toisen osaotoksen mittaustulosta. Mitä tiiviimmin pisteet sijoittuvat yhdelle viivalle, sitä voimakkaampi on muuttujien välinen yhteys (Nummenmaa 2004:265).



KUVIO 27 Kahden osaotoksen muuttujien välinen yhteys.

Sirontakuviassa 27 kaikki havainnot sijaitsevat lineaarisesti eli identiteettilinjalta. Kaikkien kuuden tehtävävihkon osioille suoritettiin myös pääkomponenttianalyysi vallitsevan faktorin löytämiseksi ja yksidimensionaalisuuden¹³ varmistamiseksi. Pääkomponenttianalyysi osoitti, että tehtävävihkot olivat käytännössä yksidimensionaalisia tilastolliselta kannalta, vaikka ne testasivat kolmea eri kielitaidon osa-aluetta. Jokaiselle testattavalle laskettiin IRT-arvoina hänen kokonaistuloksensa sekä erikseen tekstin ja puheen ymmärtämisen ja kielitiedon kokeiden tulokset. Oppilaskohtaisia tuloksia tässä tutkimuksessa ei käsitellä.

Faktorianalyysillä pyritään löytämään sellaisia muuttujien välisiä yhteyksiä kuvaavia faktoreita, joiden vaikutuksen poistaminen poistaa muuttujien väliset yhteydet. Faktoreiden ekstrakointi tarkoittaa matemaattista menettelyä, jolla määritetään malliin tulevat yhteydet eli lataukset mallissa oleviin muuttujiin (Nummenmaa 2004:344). Tässä tutkimuksessa käytettiin pääkomponenttimenetelmää, joka analysoi muuttujien varianssia. Siinä pyritään siihen, että ensimmäinen faktori selittää mahdollisimman suuren osan kokonaisvariانسista eli suoritusten vaihtelusta. Tämä voidaan tehdä tutkimalla faktorien ominaisarvoja. Taulukko 46 osoittaa kunkin tehtäväsarjan vaihtelun tärkeimpien selittäjien lukumäärän.

¹³ Kuitenkin: "There is a fair amount of controversy over the question of the assumptions of the IRT models, in particular the assumption of unidimensionality." (Lynch 2003:92)

TAULUKKO 46 Reseptiivisen kielitaidon kokeen eri tehtäväsarjojen pääkomponenttianalyysin tulos.

Komponentti	Tehtäväsarja 1			Tehtäväsarja 2			Tehtäväsarja 3		
	Yht.	% varianssista	Kumulatiivinen varianssi %	Yht.	% varianssista	Kumulatiivinen varianssi %	Yht.	% varianssista	Kumulatiivinen varianssi %
1	12,016	19,073	19,073	13,378	21,234	21,234	14,944	21,349	21,349
2	1,947	3,090	22,163	2,129	3,379	24,613	2,398	3,426	24,775
3	1,665	2,643	24,806	1,414	2,244	26,857	1,478	2,111	28,886
4	1,439	2,284	27,090	1,298	2,061	28,918	1,369	1,955	28,842
5	1,309	2,078	29,169	1,246	1,977	30,895	1,292	1,846	30,687
6	1,275	2,023	31,192	1,193	1,893	32,788	1,235	1,764	32,452
7	1,214	1,927	33,119	1,152	1,828	34,617	1,157	1,653	34,105
8	1,142	1,813	34,931	1,113	1,767	36,384	1,122	1,602	35,708
9	1,136	1,803	36,734	1,097	1,741	38,124	1,107	1,581	37,289
10	1,100	1,746	38,480	1,062	1,686	39,810	1,084	1,549	38,837
11	1,090	1,730	40,211	1,011	1,605	41,415	1,082	1,546	40,383
12	1,071	1,700	41,911	1,002	1,591	43,006	1,065	1,521	41,904
13	1,033	1,639	43,550				1,039	1,484	43,388
14	1,027	1,629	45,179				1,025	1,465	44,853
15	1,014	1,609	46,789						

Taulukko 46 osoittaa, että kunkin tehtäväsarjan ensimmäinen faktori selittää lähes 20 % ja ensimmäiset 12-15 faktoria lähes puolet yhteisvaihtelusta. Näin ollen ensimmäinen faktori on selvästi dominoiva. Pääkomponenttianalyysi vahvistaa havainnon, jonka mukaan myös kaikki kolme tehtäväsarjaa olivat tilastollisesti käytännöllisesti katsoen yksidimensionaalisia ja niiden tuloksia voidaan käsitellä summana.

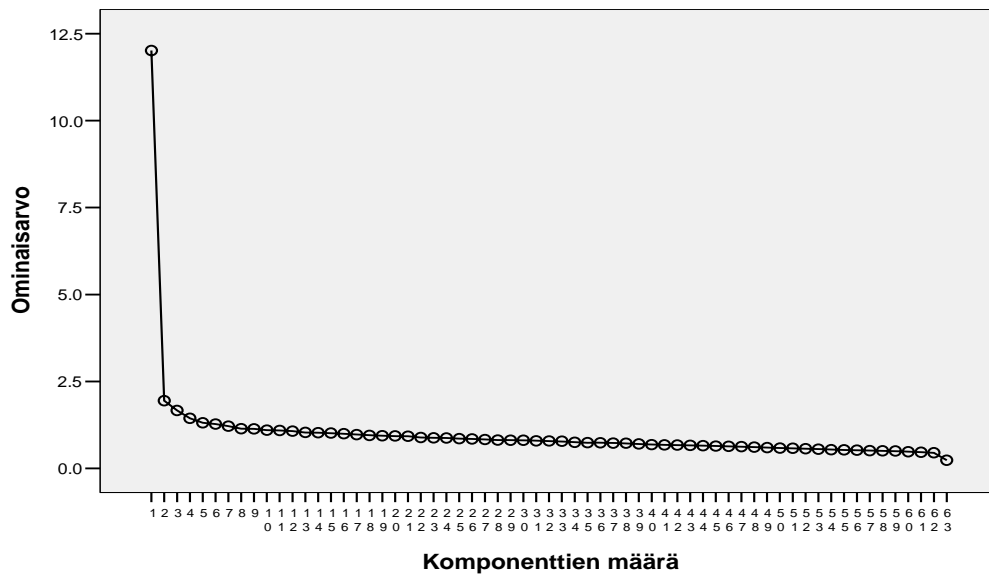
Mallilla tuotetun korrelaatiomatriisin ja alkuperäisen korrelaatiomatriisin vastaavuus tutkittiin Kaiser-Meyer-Olkinin ja Bartlettin testeillä (taulukko 47).

TAULUKKO 47 Faktorimallin sopivuuden tarkastelu χ -neliötestin avulla.

KMO Otoksen riittävyyden testi		Tehtäväsarja 1	Tehtäväsarja 2	Tehtäväsarja 3
		0,954	0,967	0,969
Bartlettin sfäärisyystesti	χ^2	22555,190	22644,695	26337,897
	vapausasteet	1953	1953	2415
	p-arvo	0,000	0,000	0,000

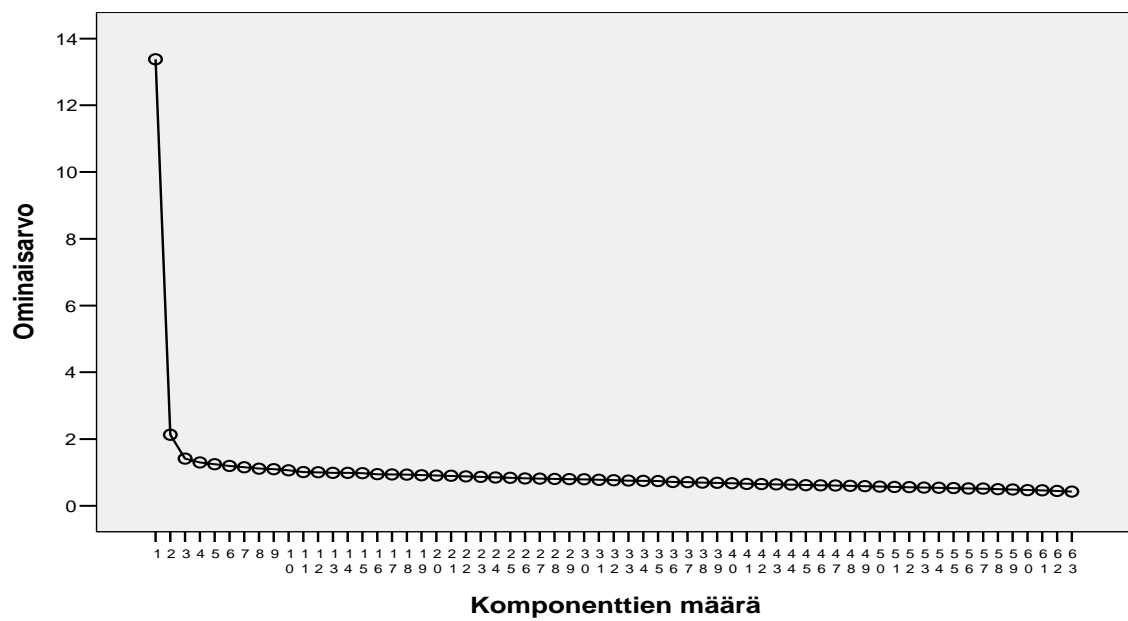
Mallin avulla tuotettu korrelaatiomatriisi ei siis poikkea tilastollisesti merkittävästi havaittujen muuttujien korrelaatiomatriisista. Edellä esitetty matriisin soveltuvuus malliin esitetään graafisesti kuvioissa 28, 29 ja 30, jotka ovat faktorin ominaisarvojen kuvaajia.

Scree-kuvio



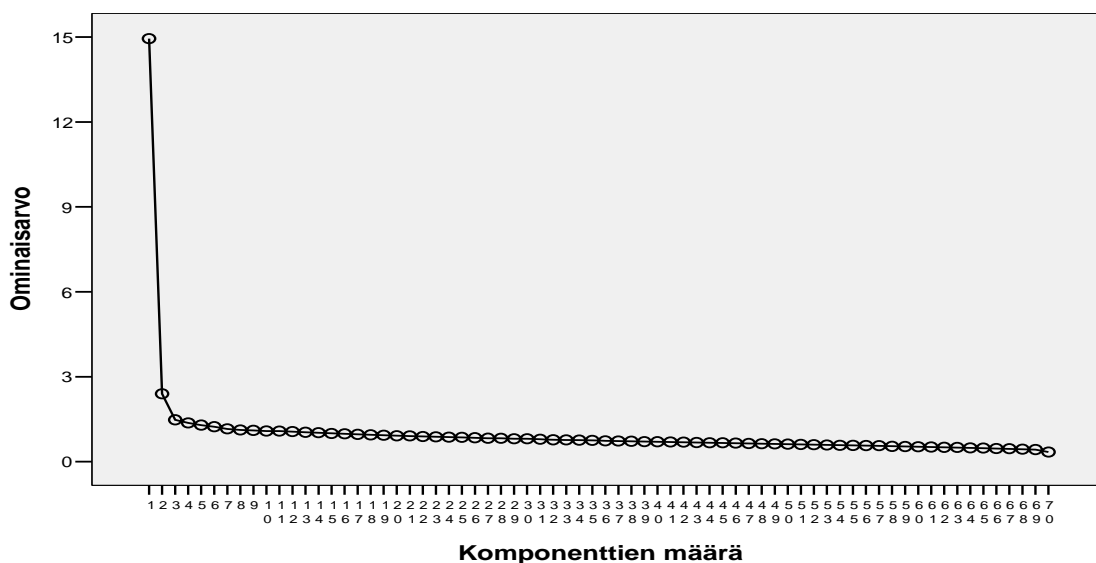
KUVIO 28 Tehtäväsarjan 1 faktoreiden ominaisarvot.

Scree-kuvio



KUVIO 29 Tehtäväsarjan 2 faktoreiden ominaisarvot.

Scree-kuvio



KUVIO 30 Tehtäväsarjan 3 faktoreiden ominaisarvot.

Kuvioissa faktorit on kuvattu x-akselilla ja faktoreiden ominaisarvot y-akselilla. Kukin kuvio osoittaa, että jokaisessa kuviossa on yksi selvästi muita selitysvoimaisempi faktori eikä muiden lisääminen tuo sanottavasti lisää selitysvoimaa. Täten tehtäväsarjat mittaavat tilastollisesti tarkasteltuina valtaosin samaa asia eli ovat käytännössä yksidimensionaalisia, mitä IRT -analyysin perusmalli (Raschin malli), joka on esitelty luvussa 7, ja myös OPLM -ohjelma edellyttävät.

Kun koe oli voitu todeta IRT -menetelmällä kelvolliseksi analysoitavaksi, ryhdyttiin valmistelemaan linkittämistä. Asiantuntijoiden suorittaman taitotasoarvioinnin jälkeen kullekin kirjain-numeroyhdistelmällä ilmaistulle reseptiivisen kielitaidon osa-alueen osiolle annettiin numeerinen arvo. Tilastollisissa analyyseissa kieliopillisen kompetenssin koeosioita käsiteltiin yhdessä puheen ja tekstin ymmärtämisen koeosioiden kanssa, jotta osioiden lukumäärä ja siten myös reliabelius saataisiin riittävän korkeaksi. Pistemäärien reliabeliuden selvittämiseksi laskettiin sekä koko kokeelle että puheen ja tekstin ymmärtämisen ja kielitiedon kokeille alfa-kerroin.

Reseptiivisen kielitaidon kokeen kaikki osiot arvioitiin dikotomisesti (0/1), jotta tulosten luotettavuuden varmistettaisiin (useampiportainen asteikko ei kokemuksen usein toimi loogisesti; Sauli Takala, suullinen tiedonanto 21.3.2005). IRT -analyysiin sisällytettiin aluksi kaikki osiot, mutta kalibroinnin jälkeen kolme osiota poistettiin heikosti toimivina. Tulokset analysoitiin OPLM -ohjelmalla. Osiokohtaista tarkastelua varten laskettiin sekä klassisen että uuden (IRT) testiteorian mukaiset erottelu- (D, a) ja vaikeustasoindeksit (P, b) koko reseptiivisen kokeen, kunkin tehtävävihkon ja kielitaidon eri osa-alueiden kaikkien osioiden osalta. Näiden analyysien tulokset esitetään luvussa 9, jossa vastataan ensimmäiseen, luonteeltaan metodologiseen, tutkimuskysymykseen osioiden vaikeustasosta ja erottelukyvystä koetulosten pohjana.

Ymmärtämisen ja kieliopillisen kompetenssin kokeiden taitotasojen asettamisessa käytettiin ns. testikeskeistä jatkumometodia (*test-centered continuum method*), jota voidaan – kuten edellä todettiin – pitää ns. klassisen Angoff-menetelmän (osaa/ei osaa) sovelluksena ja jatkokehittelmänä. Kokeiden laadun varmistamisen jälkeen osioiden saamat arvot yhdistetään taito taidolta, jolloin saadaan selville kullekin taitotasolle kuuluvien osioiden määrä. Taitotasojen katkaisukohtat määritellään osioiden kumulatiivisen frekvenssijakauman pohjalta (vrt. Takala 2005a:7). Empiiristen tulosten perusteella osiot ryhmiteltiin CEF- ja OPS -taitotasolle. Tämän jälkeen empiiriset tulokset yhdistettiin asian tuntija-arvioihin ja koodattiin uudelleen. Koodaus suoritettiin harkinnanvaraisesti, koska sitä ei voi tehdä minkään matemaattisen algoritmin perusteella. Taitotasojen asettamisprosessi kuvataan tulosluvussa 9, koska se liittyy kiinteästi tutkimuskysymyksiin vastaamiseen. Produktiivisen kielitaidon kokeet arvioitiin käyttämällä taitotasoasteikkoja. Kaikkien oppilaiden arviot yhdistettiin kriteereittäin, minkä jälkeen laskettiin niiden frekvenssit. Koska muuttujat (arvot) olivat järjestysasteikollisia, kunkin kriteerin keskimääräinen tulos ilmaistiin mediaanina keskiarvon sijasta. Koska OPS 2004 -taitotasoasteikko oli laadittu tihentämällä taitotasoa, sen ja CEF -asteikon vastaavuuden määrittäminen tapahtui harkinnanvaraisesti. Em. asteikkojen linkittäminen A2001 -asteikkoihin tapahtui klassisen testiteorian ristiintaulukointimenetelmällä. Tätä havainnollistettiin korrespondenssianalyysillä, joka osoittaa konkreettisesti eri arvojen yhteyden. Mitä lähempänä toisiaan arvot sijaitsevat kuviolla, sitä läheisempi niiden välinen yhteys on.

7.6 Luotettavan linkittämisprosessin edellytykset

Euroopan neuvoston kielipolitiikan jaos (*European Division for Language Policy*) haluaa varmistaa eri maissa käytettävien taitotasojen linkittämistoimenpiteiden laadun. Tämän tueksi se on tuottanut käsikirjan ”Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment”, jonka tarkoituksena on opastaa käyttäjiä

- kuvaamaan tutkintojen sisältöä, toimeenpanoa ja analyysimenetelmiä,
- linkittämään raportoituja tuloksia CEFiin ja
- raportoimaan käytettyjä menetelmiä.

CEF -käsikirjassa linkittäminen nähdään nelivaiheisena prosessina, jonka avulla perustellaan tutkinnon/kokeen ja CEF-taitotasojen vastaavuus (kuvio 31). Perehdyttämävaiheessa keskeistä on osoittaa, että validoijat tuntevat CEF -asteikot. On myös osoitettava, kuinka hyvin tutkinnon/kokeen sisältö vastaa CEFin kielitaidon tasokuvauksia. Standardointi sisältää sekä kielellisten suoritusten että koeosioiden arvioinnin CEF -asteikoilla ja empiirinen validointi arviointiaineistojen kokoamisen ja analysoinnin, joka osoittaa sekä kokeen että taitotasojen asettamisen luotettavuuden (Figueras ym. 2005:268-269). Linkittämisen tarkoituksena on osoittaa, että vastaavuuden saavuttamiseen tiettyyn CEF -taitotasoon vaadi-

taan tietty minimipistemäärä, joka on standardi. Standardien eli taitotasorajojen asettaminen on linkittämisprosessin keskeinen osa eri asteikkojen tasojen saattamisessa yhteismitallisiksi. Empiirisessä validoinnissa dataa analysoimalla validoidaan ja tarpeen vaatiessa muokataan pisterajat (Figueras ym. 2005: 269-70).

Teoreettinen selitys			
Tutkinnon sisällön spesifiointi	Arvioiden standardointi		Empiirinen validointi dataa analysoimalla
CEFiin perehtyminen	Perehdyttämiskoulutus		Datan kokoaminen
Sisäinen validius: - tutkinnon keskeisen sisällön, - testin laadintaprosessin, - pisteityksen ja tulosten, ja - kokeen analysoinnin ja sen jälkikäteen tutkimisen kuvailu ja analysointi	Arviointityöhön perehtyminen CEFin taitotasoasteikkojen mukaan standardoituja ja esimerkkejä käyttäen	Perehtyminen koeosioiden vaikeuden arviointiin CEFin taitotasoasteikoin standardoitujen osioiden avulla.	Sisäinen validointi: kokeen psykometrisen laadukkuuden varmistaminen
Ulkoinen validius: - kuvailla yleisellä tasolla tutkinnon liittyminen CEF -asteikkoon, - kuvailla kommunikatiiviset aktiviteetit, joita on testattu käyttäen CEF -asteikkoja, - CEF -asteikolla testattujen kommunikatiivisen kielitaidon aspektien kuvailu	Maakohtaisten suoritusten arvioiminen maamerkeiksi CEF -asteikoilla	Maakohtaisten kokeiden osioiden arviointi CEF -asteikoilla.	Ulkoinen validointi: vahvistaa kokeen liittyminen CEFiin ulkopuolisten mitausten avulla.
	Tiedon levittäminen ja käyttöönotto		
Väitetään, että koe/ tutkinto on liitetty viitekehukseen	Linkitysväitettä tuetaan arviointien standardisointivaiheen tulosten perusteella.		Linkitysvaiheelle esitetään empiiristä evidenssiä.

KUVIO 31 Linkittämisen kokonaiskuvaus. (Lähde: Figueras ym. 2005:267.)

Kuten kuvio 31 osoittaa, linkittämisprosessin sisäisen validiuden yhtenä edellytyksenä on testin laadintaprosessin kuvailu ja analysointi. Koska englannin kielen kansallinen arviointi oli laaja projekti, sen laadinta ja toimeenpano kuvataan erikseen luvussa 8.

7.7 Yhteenveto

Englannin kielen kansallisen arvioinnin toimeenpaneminen vuonna 1999 johtui paitsi uusien koululakien astumisesta voimaan, todennäköisesti myös opetussuunnitelman perusteiden uusimistarpeesta. Käsillä olevaan tutkimukseen taas johti tilastollisten työkalujen kehittyminen palvelemaan monipuolisesti oppimistulosten arviointitarpeita sekä mahdollisuus tukeutua niiden käytön hallitsevaan asiantuntijaan.

Tutkimuksen tavoitteeksi tuli arviointiprosessin kuvaamisen lisäksi

- selvittää kokeiden soveltuvuus IRT -analyysiin,
- reseptiivisen kielitaidon koetulosten ilmaiseminen θ -arvoina sekä CEF- ja OPS 2004 -taitotasojen katkaisukohtien määrittäminen niiden avulla,

- produktiivisen kielitaidon taitotasoina arvioitujen koetulosten ristiintaulukointi näiden taitotasojen katkaisukohtien määrittämiseksi CEF-, OPS 2004- ja A2001 -asteikoilla ja
- koetulosten ilmaiseminen asteikoilla.

Tutkimusaineistona oli

- 3 tekstin ymmärtämisen ja kirjoittamisen tehtävävihkoa,
- 3 puheen ymmärtämisen ja kieliopillisen kompetenssin tehtävävihkoa,
- 6 puhumisen tehtävävihkoa,
- perusopetuksen 9. vuosiluokan oppilaiden reseptiivisen kielitaidon kokeiden tulokset sekä
- asiantuntija-arviot em. tehtävien vaikeustasosta ja oppilaiden tuotoksista eri taitotasoasteikoilla.

Reseptiivisen kielitaidon kokeiden vaikeustason arviointiin käytettiin ns. Grid - taulukoita, joiden avulla voidaan määrittää monipuolisesti tekstien ja tehtävien vaikeustaso. CEF -vaikeustasojen arviointiin käytettiin kuitenkin varsinaisten viitekehysasteikkojen lisäksi tarkennettuja CEF -käsikirjan asteikkoja.

Taitotasojen asettamisessa käytettiin ns. testikeskeistä jatkumomenetelmää, jonka käytön edellytyksenä on tieto empiirisistä koetuloksista ja arvioijien taitotasoarvioista. Jotta menetelmää voidaan käyttää, koetulokset on ilmaistava theta -arvoina, joita tuotetaan IRT -analyysillä.

IRT -analyysien käyttö taitotasojen asettamiseen edellyttää aineiston soveltuvuutta valittuun matemaattiseen malliin. OPLM -analyysillä ei tällaista sopivuutta kuitenkaan löydetty. Siksi osioiden stabiili mittauskyky varmistettiin jakamalla otos sattumanvaraisesti kahtia, jolloin osaotosten välillä todettiin voimakas yhteys. Pääkomponenttimenetelmällä varmistettiin tehtävävihkojen tilastollinen yksidimensionaalisuus.

Euroopan neuvoston kielipolitiikan jaos on laatinut ohjeet siitä, mitä toimenpiteitä luotettavaan taitotasojen asettamisprosessiin liittyy. Ne on kuvattu viimeisinä tässä luvussa pohjana seuraavassa luvussa esitettävälle vuosina 1998-9 toteutetulle arviointiprosessille.

8 PERUSKOULUN 9. VUOSILUOKAN ENGLANNIN OPPIMISTULOSTEN KANSALLINEN ARVIOINTI

Tässä luvussa esitettävää arviointiprosessin suunnittelu ja toteutus perustuvat luvussa 2 esitettyihin yleisiin ja luvussa 4 kielitaidon koetta koskeviin arviointiperiaatteisiin. Toimeenpanon osalta tukeuduttiin luvussa 5 esitettyihin menettelyihin. Kokeiden laadinnan taustalla olivat luvussa 6 esitetyt näkemyksen kielikokeiden tehtävien laadinnassa huomioon otettavista asioista. Kuvaukseen on sisällytetty sekä aiottu että toteutunut prosessi, jotka ymmärrettävästi saattavat eri syistä poiketa toisistaan. Jotta prosessin kuvauksesta olisi hyötyä seuraavan prosessin käynnistyessä, mahdolliset poikkeamat on paikallaan tuoda esille, mihin Luomakin (2001:148) viittaa. Arvioinnin kielitaitokäsitys pohjautui opetussuunnitelman perusteiden A-englannin ainekohtaisiin tavoitteisiin, jotka taas perustuvat yleisesti hyväksytyihin kommunikatiivista lähestymistapaa korostaviin kielitaitokäsityksiin (luku 2; taustaa myös luvussa 3).

Valtaosa niistä teorioista ja käsityksistä, joita luvuissa 2 ja 4-6 esitetään, olivat tuttuja joko opettajan ja oppimateriaalin tuottajan uralta tai tulivat tutuiksi työssä oppimisen myötä. Vuonna 1999 järjestetyn arvioinnin haastavin osa oli päätösten tekeminen siitä, mitä tehtäviä kokeeseen sisällytetään (vrt. Luoma 2001:3). Kokeiden laatijat tuottivat runsaasti käyttökelpoista materiaalia, mutta valinnan tekeminen tästä runsaudesta jäi projektipäällikön (tämän tutkimuksen tekijän) tehtäväksi. Tulos ilmenee lopullisista tehtävistä. Tähän lukuun kirjatuissa huomioissa toteutuu kuitenkin vain osa niistä periaatteista, joihin kiinnitettiin huomiota valintaa tehtäessä. Tehtyjä valintoja perustellaan käsillä olevassa tutkimuksessa usein 2000-luvun lähteillä, vaikka varsinainen arviointi suoritettiin vuosina 1998-9. Tämä johtuu siitä, että on haluttu osoittaa, että kielitaidon määrittämisen teorian ja kokeiden laadinnan periaatteet ovat pitkän prosessin tulosta. 1900-luvun kokemukset ja käsitykset ovat käyttökelpoisia vielä 2000-luvullakin.

8.1 Alkuperäisen arvioinnin tausta ja tavoite

Vuoden 1999 englannin oppimistulosten kansallinen arviointi oli osa Opetushallituksen opetusministeriön kanssa tehdyn tulossopimuksen perusteella toimeenpanemaa oppimistulosten arviointijärjestelmää. Sen tavoitteena oli selvittää, miten hyvin koulutus vastaa sille asetettuja tavoitteita, sekä tukea oppilaitoksia ja opettajia opetuksen jatkuvassa kehittämisessä. Arviointijärjestelmä perustui silloin uusiin koululakeihin ja sen toteuttamisessa noudatettiin Opetushallituksessa kehitettyä ohjeistoa (luku 2).

Englannin arvioinnin tavoitteeksi tuli arvioida Peruskoulun opetussuunnitelman perusteiden 1994 määrittämien tavoitteiden saavuttamista ja antaa kuva siitä, mitä oppilas osaa tai ei osaa kielitaidollaan tehdä (luvut 4 ja 6). Tavoitteeseen pyrittiin taulukossa 48 esitetyillä toimenpiteillä.

TAULUKKO 48 Englannin kansallisen arvioinnin tavoitteet ja toimenpiteet.

Kysymys	Toimenpiteet
1. Millainen 9. luokan oppilaiden englannin kielen taso oli vuonna 1999?	1.1 Kielitaidon määrittelemine Peruskoulun opetussuunnitelman perusteiden 1994 ja Perusopetuksen päättöarvioinnin hyvän arvosanan (8) kriteerien ¹ perusteella 1.2 Kokeiden laadinta 1.3 Kyselykaavakkeiden laadinta 1.4 Tulosten kokoaminen ja analysointi
2. Toteutuiko englannin kielen oppimistuloksissa koulutuksellinen tasa-arvo?	Tulosten analysointi 2.1 koko maata koskevina. 2.2 läänikohtaisesti. 2.3 EU-tukialueittain. 2.4 kieliryhmittäin. 2.5 sukupuolen mukaan. 2.6 kuntaryhmittäin. 2.7 koulun koon mukaan.
3. Mitkä tekijät olivat yhteydessä oppimistuloksiin?	3.1 Oppilaskyselyn 3.2 Opettajakyselyn 3.3 Rehtorikyselyn laatiminen.

Arvioinnin kohteena oli edustava (suurempi kuin 5 %) otos lukuvuonna 1998-1999 perusopetuksen piirissä olleista englantia A-kielenä opiskelleista suomen- ja ruotsinkielisistä oppilaista Ahvenanmaan koulut mukaan lukien.

Ainekohtaisessa oppimistulosten arvioinnissa keskitytään opetussuunnitelman perusteissa tavoitteeksi asetettujen valmiuksien saavuttamisen arviointiin (OPH 1998a:6). Koska perusteet sisältävät ainekohtaisten tavoitteiden lisäksi myös muita tavoitteita, kuten esimerkiksi kiinnostuksen herättämisen oppiainetta kohtaan (luku 2; OPH 1994: 68), arviointiin sisällytetään yleensä myös oppilas-, opettaja- ja rehtorikysely. Rehtorikyselyllä pyritään selvittämään

¹ Arviointia suunniteltaessa oli käytettävissä luonnos kriteereistä.

esimerkiksi koulun toiminnan taloudellisia edellytyksiä, opettajakyselyillä opetussuunnitelman perusteiden, kunnan / koulun opetussuunnitelman ja oppimateriaalin yhteensopivuutta sekä opetustilojen ja -välineiden käyttömahdollisuuksia. Oppilaskyselyillä selvitetään esimerkiksi oppilaiden suhtautumista arvioitavaan oppiaineeseen sekä heidän käsitystään omasta kielitaidostaan. Työssä pyrittiin eettiseen arviointiin ja siinä otettaisiin huomioon arviointitoimeksiannon puitteissa ohjelma-arvioinnin standardit (luku 2).

8.2 Arvioinnin suunnittelu

Arvioinnin valmistelu kattoi noin puoli vuotta ja se koostui ajallisesti seuraavista päävaiheista:

- maaliskuussa 1998 asetettiin ohjaus- eli asiantuntijaryhmä,
- elokuussa 1998 nimitettiin projektipäällikkö,
- elokuussa 1998 nimettiin kokeenlaatijaryhmät,
- lokakuussa 1998 saatiin ensimmäiset koeversiot muokattaviksi,
- helmikuun 1999 lopussa saatiin valmiiksi lopulliset koeversiot, joten
- maaliskuussa 1999 voitiin toimeenpanna arviointi (luku 5).

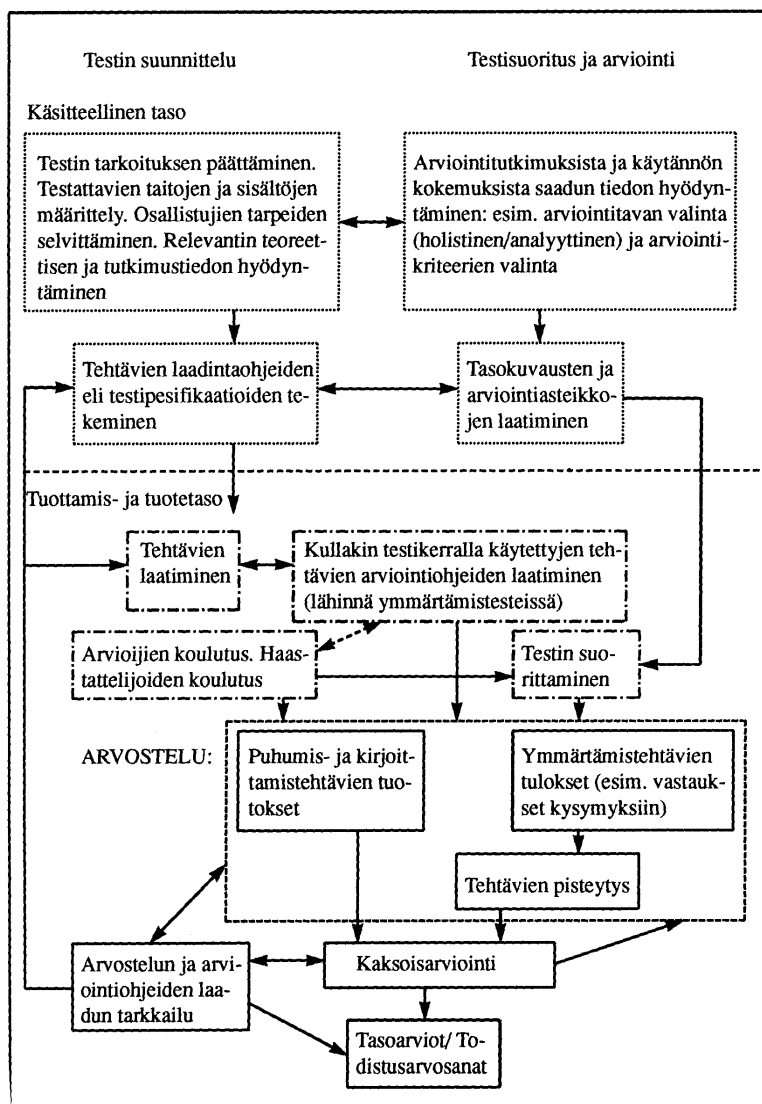
Koko arviointiprojektin toteutumisen aikataulu esitetään liitteessä 13.

Projektin tueksi nimitettiin ohjausryhmä. Sen tehtävänä oli seurata kokeiden kehittämistä huolehtien niiden pedagogisesta, opetussuunnitelmallisesta ja arviointiteoreettisesta tasosta. Ryhmään kutsuttiin lehtori Liisa Havola Muuramen lukiosta, kasvatustiet. tri Hanna Jaakkola Helsingin II Normaalikoulusta, tutkija Kaija Kärkkäinen ja tutkimusprofessori Sauli Takala Jyväskylän yliopiston Soveltavan kielentutkimuksen keskukselta (SOLKI), opetusneuvos Kalevi Pohjala Opetushallituksesta, asiamies Riitta Sarras Opetusalan Ammattijärjestö r.y.:stä, puheenjohtaja Kristiina Sirén Suomen Englanninopettajat r.y.:stä sekä puheenjohtaja Terttu Valojärvi Suomen kieltenopettajien liitosta (SUKOL). Tri Jaakkola pyysi kuitenkin eroa ryhmästä jo syksyllä 1998 runsaiden työtehtäviensä vuoksi ja siksi, että saksan kieli ei tuolloin ollut arvioitavana. Arvioinnin toimeenpanijaksi valittiin tämän tutkimuksen tekijä, fil. maist., erikoissuunnittelija Eeva Tuokko, jonka kanssa yhteistyössä toimi projektisuunnittelija, kasvatustiet. kand. Aulikki Etelälahti. Aineiston tilastollisesta käsittelystä vastasi erikoistutkija, kasvatustiet. tri Jari Metsämuuronen. Oppilas-, opettaja- ja rehtorikyselyt laadittiin yhteistyössä eri oppiaineiden erikoissuunnittelijoiden ja Arviointivirkamiesten opetusneuvosten kanssa.

Kokeiden laadinta käynnistettiin kutsumalla koolle kokeenlaatijat. Kaikilla kahdeksalla ryhmän jäsenellä oli usean vuoden kokemus peruskoulun opettajana toimimisesta. Ryhmän jäsenillä oli lisäksi kokemusta opettajankouluttajina, oppimateriaalin laatijoina ja lukion opettajina. Projektipäälliköllä oli kokemusta toimimisesta peruskoulun ja lukion englannin opettajana, kunnan ohjaavana opettajana ja lääninkouluttajana, suullisten kokeiden laatijana ja toimeenpanijana.

na, ylioppilastutkintolautakunnan englannin kielen sensorina sekä niin painetun kuin elektronisen oppimateriaalin laatijana.

Arviointi toteutettiin pääpiirteissään kuvion 32 mukaisena kuitenkin sillä tavoin pidemmälle vietyä, että asianosaiset (koulut) saivat tuloksista pikapalutteen ja myöhemmin arviointiraportin. Kuvio noudattelee ohjelma- arvioinnin ja kielikokeen luvuissa 2 ja 5 esitettyjä periaatteita.



KUVIO 32 Kielitaidon arvioinnin vaiheet. (Lähde: Huhta 1997:21)

Vastuuhenkilöt

Opetushallituksessa arvioinnin suunnittelusta, kokeiden teettämisestä ja käytännön järjestelyistä arvioinnin toimeenpanossa samoin kuin raportoinnista vastaisi projektipäällikkö. Oppimistulosten tilastollisesta käsittelystä ja analyseistä vastaisivat projektisihteri ja erikoistutkija.

Tiedottaminen

Arvioinnista päätettiin informoida tiedotusvälineitä syyskuun alussa ja kouluja otannan valmistuttua marras-joulukuussa. Rehtorit ja opettajat tulisivat saamaan tarkennettua tietoa seuraavan vuoden alussa. Opetushallituksen johtokuntaa informoitaisiin tulosten valmistuttua.

Kokeen toimeenpanijan ohjeet

Koska kokeen toimeenpano perustui lakiin, koulujen rehtorit vastaisivat toimeenpanosta ja koulujen englanninopettajat kokeiden korjaamisesta ja pisteittämisestä. Opettajan ohjeisiin sisällytettäisiin koko arvioinnin aikataulu, arviointiohjeet ja puheen ymmärtämisen kokeen käsikirjoitus. Ohjeisiin kirjattaisiin myös opastusta suullisen kokeen järjestämisestä, arviointikaavakkeet sekä valvojan ohjeet. Koetulokset oli tarkoitus analysoida JOTOS -ohjelmalla, mikä edellytti kirjaamista optiselle lomakkeelle.

Aikataulu

Kirjallinen koe suunniteltiin pidettäväksi 31.3.1999 kahden peräkkäisen oppitunnin (à 45 min.) aikana. Näin tekstin ja puheen ymmärtämisen, kieliopillisten rakenteiden (kieliopillinen kompetenssi) ja kirjoittamisen kokeen suorittamiseen jäisi 20-25 minuuttia kuhunkin. Oppituntien väliin jäävällä välitunnilla oppilaiden oli tarkoitus täyttää vastauslomakkeen oppilaskyselyn tiedot (vrt. luku 3). Suullisen kokeen ajankohdaksi määräytyi 30.3.1999. Koe toteutettaisiin parityönä siten, että kullakin parilla olisi käytettävänä 15 minuuttia tehtäviin tutustumiseen ja roolien omaksumiseen, ja varsinainen koe kestäisi maksimissaan 15 minuuttia. Oppilaat tulisivat kokeeseen muilta oppitunneilta siten, että edellisen parin ollessa kokeessa seuraava pari olisi valmistautumassa omaansa.

Taustakyselyt

Koetulokseen mahdollisesti vaikuttavien taustatekijöiden selvittämiseksi projektipäälliköt laativat yhteistyössä oppilas-, opettaja- ja rehtorikyselyt, joista ensimmäinen sisältyi oppilaan optiseen vastauslomakkeeseen. Kyselyt olivat verrattain niukkoja, koska tulokset haluttiin toimittaa kohteisiin nopealla aikataululla, jolloin monipuolisiin analyysihin ei olisi jäänyt aikaa etenkin, kun äidinkielen kansallisen arvioinnin tulokset oli analysoitava samaan aikaan.

Äänitteet, painatus ja kielentarkastus

Tarvittavien äänitteiden (C-kasetit) tuottamisesta sekä tekstien kielellisestä tarkistuksesta pyydettiin tarjouksia. Äänityskertoja arvioitiin tarvittavan vähintään kaksi. Tehtävävihkojen esikokeiluversiot toimitettaisiin Editan toimittamina monisteina, kun taas lopulliset tehtävävihkot painatettaisiin Yliopistopainossa.

Esikokeilu

Tehtävien valinnan pohjaksi päätettiin järjestää esikokeilu.

Tulosten raportointi

Koulut tulisivat saamaan pikapalautteen toukokuun 1999 loppuun mennessä. Arviointiraportin koulut saivat saman vuoden loppuun mennessä.

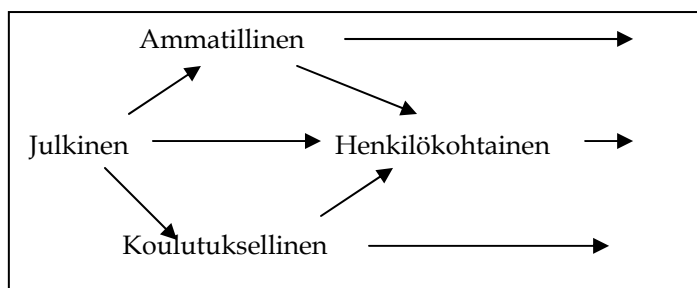
Arvioinnin kielitaitokäsitys

Arviointiin sovellettava kielitaitokäsitys oli jo määritelty Opetussuunnitelman perusteissa (luku 2; vrt. McNamara 2000:25). Sen pohjana oli kielen tutkijoiden toistaiseksi yleisesti hyväksymä näkemys, jonka mukaan kielitaidon voidaan katsoa jakaantuvan neljään osa-alueeseen (luku 2; esim. CEF 2001, Lynch 2001), ja se heijastaa voimakkaasti niiden tutkijoiden ja teorioiden käsityksiä, jotka painottavat suorituksen merkitystä kielitaidon määrittämisessä ja sen arvioinnissa (luku 5; Bachman ja Palmer 1996, Alderson 2000, CEF 2001). Sekä reseptiivisen että produktiivisen kielitaidon tason arviointiin kuuluu olennaisena osana kieliopillisten rakenteiden tarkkuus (luvut 5 ja 6; Canale ja Swain 1980; Jaakkola 1997; Köhlmyr 2003; Purpura 2004) ja sanaston laajuus (Takala 1984; Read 2000:74). Vuoden 1994 opetussuunnitelman perusteiden tavoitteiden jakaantuminen kielitaidon osa-alueille on koottu liitteeseen 14.

Arvioinnin tuli kattaa koko perusopetuksen oppimäärä, mutta sen tarkoitus ei ollut tulosten käyttäminen oppilasarvostelun pohjana eikä oppilaiden edistymisen arviointiin eli heidän asettamisensa järjestykseen kielitaidon perusteella (luku 2). Siksi pyrittiin kriteeri-, ei normiviitteiseen arviointiin (luvut 2 ja 3), jolloin voidaan saada tarkka kuva oppilaan suoritustasosta tietyllä käytäytymis- ja sisältöalueella (Takala 1980:2).

Kielen käyttöalueet

Jokainen kielenkäyttötapahtuma on kiinteässä yhteydessä tiettyyn tilanteeseen ja edellyttää erilaisia prosessointitarkoituksia. Kielenkäyttötilanteet ryhmitellään usein seuraaviksi kielenkäyttöalueiksi, jotka toimivat keskenään vuorovaikutuksessa kuvion 33 osoittamalla tavalla, kuten myös Eurooppalaisessa viitekehysessä (2003:76) todetaan.



KUVIO 33 Kielenkäyttöalueet CEFin mukaan (Lähde: Modern Languages 1995.)

Näihin alueisiin liittyvistä kielenkäyttötilanteista on esitetty lukuisia luetteloita (luku 4; POPS I 1970, VIKKE 1976, Threshold Level 1990, Kouluhallitus 1985, Eurooppalainen viitekehys 2003; Grid 2005). Perusopetuksen 9. luokan oppilaan

elämänpiiriin voidaan katsoa kuuluvan jo suurimman osan esimerkiksi CEFIn sivuilla 48-49 esitetyistä tilannekategorioista.

Kokeisiin sisältyvien aihepiirien tuli olla 9.-luokkalaisten poikien ja tyttöjen kannalta kiinnostavia (luku 6). Niiden tuli myös käsitellä eri kielenkäyttöalueita, jotta eri tavoin harrastuneet oppilaat löytäisivät itseään kiinnostavia aiheita.

Konstruktien laatimisen periaatteita

Arvioinnin lähtökohtana oli, että ymmärtäminen koostuu sekä kielitaidosta että strategioiden hallinnasta (luku 5), vaikka vieraan kielen hallinnassa ymmärrettävästi kielellisellä kompetenssilla on suurempi merkitys (Buck 2001:105) Arvioiden konstruktien kehittämisessä sovellettiin näkemystä ”helpommasta vaikeampaan” tai ”suppeammasta laajempaan” (Buck 2001:93). Reseptiivisen kieli- taidon kokeiden laadinnan pohjana päätettiin käyttää pohjana Weirin (1993) jakoa operaatioihin (luku 4) tekstien prosessoinnissa (taulukko 49).

TAULUKKO 49 Ymmärtämisen operaatiot. (Lähde: Weir 1993:70.)

Ymmärtämisen operaatiot	
Tekstin ymmärtäminen	Puheen ymmärtäminen
1. Tekstin silmäily pääasian ymmärtämiseksi. (Skimming)	1. Pääasian ymmärtäminen.
2. Selaaminen tietyn asian selville saamiseksi. (Scanning)	2. Puhujan tarkoituksen ymmärtäminen.
3. Tarkan merkityksen selvittäminen.	3. Yksityiskohdan ymmärtäminen.
4. Tekstin eri osien välisten yhteyksien ymmärtäminen.	4. Päättely.
5. Implisiittisesti ilmaistun merkityksen ymmärtäminen. (Päättely)	5. Sanelu.

Tehtävävihkojen kokoaminen

Koska arvioitava alue oli laaja, lähes seitsemän vuoden opiskelujakso, kattavan arvioinnin yksittäisten konstruktien lukumäärä oli korkea (luku 6). Näin laajamittaisen kokeen suorittaminen veisi liikaa koulujen opetusaikaa. Silti oli tärkeää, että mittaväline kattaisi mahdollisimman laajasti hallittavaksi oletetun alueen. Tähän voidaan päästä kahdella tavalla. Joko mittauksia toimeenpannaan niin usein, että ajan mittaan suuri osa hallittavasta aineksesta tulee arvioiduksi tai oppilasotos on niin suuri ja tehtäviä niin runsaasti, että yhdellä mittauskeralla saadaan riittävästi tietoa. (luku 4). Koska oli oletettavissa, että perusopetuksen oppilaiden englannin taitoa ei tulisi kovin pian arviomaan uudelleen, mitattava alue oli laaja ja mittaukseen käytettävissä oleva aika rajallinen, mutta otantapopulaatio suuri, päädyttiin usean tehtäväversion käyttöön. Koska suunniteltu otos tuottaisi joka tapauksessa suuren määrän oppilaita suorittamaan koetta, jokaisen oppilaan ei olisi tarpeen tehdä samoja tehtäviä luotettavan kansallisen tiedon tuottamiseksi (luku 3).

Yksi kunkin kirjallisen kokeen kolmen tehtävävihkon ymmärtämisen tehtävistä tulisi olemaan ankkuritehtävä, yhteinen kaikille oppilaille (luku 5). Muut tehtävät olisivat erilaisia. Näin rakennetehtävät olisi tarvittaessa mahdol-

lista linkittää toisiinsa monimatriisitekniikalla, joka tarjoaa mahdollisuuden kattaa laaja sisältöalue ja on taloudellinen, koska sen avulla voidaan hankkia mahdollisimman suuri tietomäärä mahdollisimman vähäisellä koeajan käytöllä (Takala 1980:32-33; luvut 2-3). Koeohjeet annettaisiin oppilaan äidinkielellä, mikä tässä tapauksessa tarkoitti suomea, ruotsia ja otannan tuloksesta riippuen saamea.

Ymmärtämisen kokeiden tekstityypit

Tässä yhteydessä tekstillä tarkoitetaan sekä painettua että kuultua tekstiä, mikä ei kuitenkaan tarkoita sitä, että puhuttua kieltä pidettäisiin samanlaisena kuin kirjoitettua kieltä (luku 4). Kuitenkin molempien ymmärtämisessä sekä tekstin että vastaanottajan ominaisuudet vaikuttavat ymmärtämisen konstruktin määrittelyyn ja siihen, mitä ja miten testataan (luvut 4-6). Jotta mikään näistä ominaisuuksista (esim. vastaanottajien taustatiedot, äidinkielen taito, tekstin vaikeustaso), ei vaikuttaisi kohtuuttomasti testattavan suoritukseen, kokeeseen pyrittiin sisällyttämään mahdollisimman erilaisia aihepiirejä sekä teksti- ja tehtävätyyppejä (luvut 3 ja 6; Alderson 2000:81). Tekstien valinnassa ei pyrittäisi autenttisiin näytteisiin, mutta sen sijaa korostettiin ”realistisia tekstejä” (Buck 2001:115), kommunikatiivisuutta ja todellisen elämän käyttötilanteita (luku 6) Tekstien valinnassa luotettiin ryhmän asiantuntemukseen ja siihen, että runsaasta aineistosta saataisiin esikokeilun jälkeen monipuolinen otos lopulliseen kokeeseen.

Ymmärtämisen kokeiden tehtävätyypit

Usein oletetaan, että jos vieraan kielen koeosiot annetaan äidinkielellä, testattavat suoriutuvat kokeesta paremmin kuin jos ne annettu kohdekielellä. Kuitenkin vaikeasta tekstistä voidaan tehdä helppoja koeosioita ja päinvastoin (Alderson 2000) niiden kielestä riippumatta.

Tässä arvioinnissa tavoitteena oli käyttää sekä monivalinta-, oikein/väärin- ja avotehtäviä kuitenkin niin, että jälkimäiset annettaisiin ja niihin vastattaisiin äidinkielellä. Tällöin testattavan ei tarvitse yhdistää ymmärtämistä ja tuottamista eikä vastauksen ymmärrettävyys myöskään riipu testattavan vieraan kielen taidosta, jolloin kokeen arvioijan on helpompi tulkita vastauksen oikeellisuus. Monivalinta- ja oikein/-väärin tehtäviä voisi olla sekä kohde- että äidinkielellä. Arviointiin sisällytettäisiin myös sanelu, koska sen on todettu olevan hyvä puheen ymmärtämisen mittari (luku 6; Buck 2001:74). Tavoitteena oli, että tehtävät olisivat käytännöllisiä ja taloudellisia eli yksiselitteisesti ja mahdollisimman helposti tarkistettavissa ja pisteitettävissä (luku 5; Bachman & Palmer 1996:35).

Puhumisen koe

Kommunikatiivista suullista koetta ei ole useinkaan sisällytetty pakollisena osana Suomessa toimeenpantuihin laajamittaisiin vieraan kielen arviointeihin. Koska oli oletettavissa, että koetyyppi ei olisi tuttu kaikille osallistujille, tehtävät tuli laatia sellaisiksi, joita peruskoulun oppimateriaalit olivat perinteisesti tarjonneet.

Koesalaisuuden säilymiseksi (ks. Puhumisen koe, s. 130) tavoitteeksi asetettiin laatia useita tehtävävihkoja. Tehtävistä yksi olisi kaikissa sama. Tehtävien laadinnan pohjana käytettäisiin soveltuvin osin Threshold Level -kynnystaitoluettelon (1990) ja CEFin luonnoksen (Modern Languages 1996) funktionaalisten taitojen luetteloita. Aihepiirit pyrittäisiin valitsemaan oppilaan välittömään elämänpäiriin liittyvistä tilanteista.

Kirjoittamisen koe

Kirjoittamisen kokeessa testattaville päätettiin tarjota useita aiheita (luku 6). Ne oli tarkoitus valita esikokeilun jälkeen siinä suosituimmiksi osoittautuneista. Kirjoittamisen kokeessa jokaiseen tehtävävihkoon sijoitettaisiin useita aiheita, joista yksi olisi kaikille yhteinen. Otsikot olisivat sekä ohjattuja että ohjaamattomia (= pelkkä otsikko). Useiden aiheiden tarjoamisen ongelmana on kuitenkin konstruktin määrittämisen vaikeus, eli koska samaan pistemäärään voi päästä kirjoittamalla eri aiheista, jää epäselväksi, mitä koe mittaa (luku 6; Weigle 2002:153). Arvioinnissa pidettiin kuitenkin tärkeämpänä saada tietoa siitä, minkä tasoisia tuotoksia eri tekstilajeja edellyttävät aiheet generoisivat (luku 5).

Kieliopillisen kompetenssin koe

Kieliopillista kompetenssia päätettiin testata monivalinnalla, ohjatulla aukko-täydennyksellä ja käännöksellä. Tehtävien tuli täyttää luvussa 6 käsitellyt edellytykset tehtävätyyppien monipuolisuudesta ja kieliopillisten rakenteiden sijoittumisesta todellisen elämän kielenkäyttökonteksteihin. Kokeen käytännöllisyyden takaamiseksi (luku 5) päätettiin kuitenkin käyttää runsaasti monivalintaa, joka takaisi myös konstruktin edustavuuden eikä uhkaisi validiutta (luku 3).

8.3 Varsinaisen arvioinnin valmistelu

Otanta

Kansallisessa oppimistulosten arviointijärjestelmässä perusopetuksen arvioinnin otantakooksi suositellaan 3 000-4 000 oppilasta eli noin 5 % ikäluokasta (OPH 1998a:12). Resurssien salliessa otantakokoa pyritään kuitenkin kasvattamaan, jotta mahdollisimman monet koulut tulisivat arvioinnin piiriin. Koulutuksellisen tasa-arvon toteutumiseksi otannassa on otettava huomioon alueelliset ja taloudelliset olosuhteet samoin kuin kuntaryhmä (OPH 1998a:9). Englannin arvioinnin otannassa otettiin huomioon alueellinen edustavuus (läänien asukasluku), taloudellinen edustavuus (EU-tukialueet), kuntaryhmä (kaupunki, taajama, maaseutu) ja koulun koko (vrt. Törmäkangas 2004:78). Esimerkiksi ulkomailla perusopetusta antavat koulut jäivät otannan ulkopuolelle.

Otannassa sovellettiin monivaiheista ositettua otantaa (*multi-staged stratified sampling*). Se suoritettiin kahdessa vaiheessa. Oppimistulosten arviointiraportissa otantaa kuvataan seuraavasti: "Ensimmäisessä osituksessa suomenkielisten koulujen määrä suhteutettiin lääneittäin tasaisesti asukasmäärään näh-

den. Toisen osituksen yhteydessä oppilaitokset suhteutettiin kussakin läänissä edustettuina oleviin EU:n tukialueisiin (Suomessa esiintyvistä alueista 0-tunnuksella merkitty ei saa lainkaan EU-tukea, 2-tukea saavat taantuvat teollisuusalueet, tukialue 5 on maaseudun kehittämisaluetta ja alue 6 harvaan asutua seutua.) Kolmannessa osituksessa oli mukana asutusrakenne: otettiin huomioon koulujen lukumäärä kaupunki-, taajama- ja maaseutualueilla edellä mainittujen tukialueiden osalta (Tuokko 2000:22).

Näiden toimenpiteiden perusteella perusopetuksen 9. luokalla opetusta antavien oppilaitosten joukosta valikoitui koko maan kouluja edustava 124 suomenkielisen koulun otos ja 20:n, tai oikeammin 19:n (kaksi ahvenanmaalaisista yläastetta toimi yhdessä) ruotsinkielisen koulun otos (Tuokko 2000:22). Toisessa vaiheessa oppilaat valittiin satunnaisesti koulun sisältä. Rehtorien johdolla koottujen 9. luokkien oppilaista saatujen aakkosellisten luetteloiden pohjalta kirjallisia kokeita suorittamaan valittiin oppilaat ikäluokan koon perusteella taulukon 50 mukaisesti.

TAULUKKO 50 Kirjallisiin kokeisiin osallistuneiden oppilaiden otostaminen.

Oppilaiden lukumäärä	Otantaperiaate
> 31	Kaikki osallistuvat
31-66	Joka kolmas oppilas jäi pois
67-100	Joka 2. oppilas osallistui
< 101	Joka 3. oppilas osallistui

Otokseen tuli 8 % maan suomenkielisistä ja 17 % ruotsinkielisistä 9. vuosiluokan oppilaista, yhteensä 5027 suomen- ja 614 ruotsinkielistä oppilasta (Tuokko 2000:25). Puhumisen kokeeseen valittiin otosoppilaista joka neljäs, mutta kuitenkin vähintään neljä paria. Siihen valikoitui 1303 sellaista oppilasta, joilta saatiin mittaustulokset. Osallistujien lukumäärä katsottiin riittäväksi johtopäätösten tekemiseksi koko valtakunnan tasolla. Otosvirheen kokoa ei ennakoitu tarkemmin.

Kokeenlaatijat

Kokeenlaatijat työskentelivät pareittain siten, että kukin pari laati kokeet yhteen kielitaidon osa-alueeseen. Kirjoittamisen kokeen aiheista vastannut ryhmä laati kuitenkin myös kieliopillisten rakenteiden kokeet. Kaikkien kokeenlaatijoiden oli myös tarkoitus kommentoida toistensa työtä. Koska kokeenlaatijat olivat kokeneita opettajia, he tunsivat käytössä olleet oppimateriaalit ja tarvittaessa tutustuivat niihin, jotta mikään materiaali ei painottuisi liikaa. Ryhmän teoreettista tietämystä tuettiin jakamalla kielitaidon arviointiin liittyvää aineistoa.

Testispesifikaatioiden tarkentaminen

Tehtävien laatijat toimittivat ehdotuksensa Opetushallitukseen lokakuussa. Heidän oli toivottu varustavan tehtäväehdotuksensa testispesifikaatioilla, jotka noudattaisivat väljästi Aldersonin ym. (1995) mallia. Niistä tuli ilmetä testattavan kielitaidon osa-alueen lisäksi soveltuvin osin aihepiiri, tekstilaji, funktio,

viestintätehtävä, tehtävätyyppi, tehtävän ja ohjeiden kieli, osioiden ennakoitu vaikeustaso ja tietynlaista ymmärtämistä edellyttävä operaatio, pisteitys, ja avotehtävissä hyväksyttäväksi ehdotetut vastausvaihtoehdot (luku 5)

Tekstin ymmärtämisen kokeet olivat varustetut tehtäväkohtaisilla spesifikaatioilla, joten niiden konstruktaja tarkennettiin, mutta muihin osakokeisiin oli testispesifikaatiot laadittava. Esimerkki testispesifikaatioista on liitteenä 15. Lisäksi määriteltiin kieliopillisen kompetenssin tehtävien sisällöt sekä kirjoittamisen kokeen tekstilajit. Ohjausryhmän tarkastelun jälkeen suullinen koe koottiin uudelleen, sillä saadut koe-ehdotukset perustuivat osaksi kuvien tulkinnalle, ja koska kuvista oli käytettävissä vain kirjallinen selostus, ei voitu olla varmoja lopputuloksen yksiselitteisyydestä.

Äänitykset

Puhumisen kokeen äänitykset olivat tehtävävihkojen kokoamisen pahin pulonkaula. Koska äänityksiin ei ollut aikaisemmin ollut mahdollista varautua, Opetushallitus oli tuottajan valinnan jälkeen täysin riippuvainen tämän aikataulusta.

Ensimmäistä äänityskertaa varten oli saaduista tehtävähdotuksista valittava pikaisella aikataululla noin seitsemän esikokeiluun toimitettavan vihkon aineisto, joista kukin sisälsi kolme puheen ymmärtämisen tehtävää. Tuottaja toimitti sen äänitettäväksi Englantiin, jossa lukijoiksi oli saatavissa eri ikäisiä, britti- ja amerikanenglannin variantit hallitsevia ammattitaitoisia näyttelijöitä. Vasta äänitteiden valmistuttua voitiin todeta kunkin tehtävän kesto. Koska kaikki tehtävät oli suunniteltu kuunneltavaksi kahteen kertaan, ensin tauottamatta ja sitten 1-2 osion jälkeen tauotettuina, tekstejä jouduttiin yleensä lyhentämään ja vastaavasti osioita poistamaan ja/tai muokkaamaan.

Esikokeiluversioiksi työstettiin 18 tehtävää. Jokaiseen laadittiin suoritusohje, suunniteltiin tarvittaessa uusi tauotus, arvioitiin ja merkittiin taukojen paikat äänityskäsikirjoituksiin, jotka sisälsivät sekä tauottamattoman että tauotetun käsikirjoituksen. Nämä toimitettiin tuottajalle, joka vastasi äänitteiden digitoimisesta ja uudelleen koostamisesta.

Kirjallisen kokeen esikokeilu

Esikokeiluvihko A

Tehtävävihkoihin, joita tässä kutsutaan A-vihkoiksi, sisällytettiin tekstin ymmärtämisen ja kirjoittamisen kokeet. Esikokeiluun koottiin viisi A-vihkoa, joissa kussakin oli kolme tekstin ymmärtämisen tehtävää. Näiden valintaperusteina käytettiin konstruktien sisältöä, projektipäällikön arviota osioiden vaikeustasosta, tekstityyppien ja aihepiirien vaihtelua sekä tehtävätyyppien käytännöllisyyttä. Erityisesti aihepiirien monipuolisuuden takaamiseksi joitakin tekstejä ja jouduttiin lyhentämään. Tekstien sanamäärää ei laskettu (vrt. Grid -taulukot, luku 7). Tehtävät varustettiin äidinkielisillä suoritusohjeilla ja opettajille tarkoitetuilla pisteitysohjeilla. Opettajan ohjeisiin laadittiin myös ehdotukset hyväksyttäväksi vastauksiksi, joista myös pyydettiin kommentteja.

Kirjoittamisen koe oli kaksiosainen. Kunkin oppilaan piti kirjoittaa yhteisen aiheen lisäksi joko taulukon 51 A- tai B-ryhmän kirjoitelma(t).

TAULUKKO 51 Esikokeilun kirjoittamisen kokeen aihe-ehtotusten lukumäärät ja vaihtoehdot.

1. Yhteinen aihe, maksimi 50 sanaa					
Ohje	A	Pituus	Ohje	B	Pituus
2. Joko			2. tai		
Valittava kaksi aihetta	1. aihe 2. aihe 3. aihe	30-50 sanaa	Valittava yksi aihe	1. aihe 2. aihe 3. aihe 4. aihe 5. aihe	50-100 sanaa

A-vaihtoehdon valinnet saattoivat siis kirjoittaa 110-150 sanaa ja B-vaihtoehdon valinnet 100-150 sanaa.

Esikokeilua varten ei laadittu vastauskaavakkeita, vaan oppilaat vastasivat konseptiarkeille. Kokeen järjestäneet opettajat arvostelivat kokeet Opetushallituksesta toimitettujen ohjeiden mukaisesti ja toimittivat tulokset sekä koeaineiston Opetushallitukseen. Tehtävävihko A pystyttiin toimittamaan esikokeiluun sellaisena ajankohtana, jolloin syyslukukauden koeruuha ei ollut vielä alkanut². Tulokset tulivat siis ajallaan, ja osioanalyysin jälkeen saatettiin alkaa valmistella lopullista koetta.

Esikokeiluvihko B

Tehtävävihkoon B sisällytettiin puheen ymmärtämisen ja kieliopillisen kompetenssin tehtävät. Esikokeiluun koottiin kuusi vihkoa, joissa kussakin oli kolme puheen ymmärtämisen tehtävää ja 20-25 kieliopillista kompetenssia mittaavaa tehtävää. Viimemainitut valittiin arvioidun vaikeustason, oletetun (= opettajakokemukseen perustuvan) frekvenssin sekä tehtävätyypin perusteella. Puheen ymmärtämisen kokeen suoritusohjeisiin kiinnitettiin erityistä huomiota (luku 6). Niiden tuli olla mahdollisimman lyhyitä, mutta myös ehdottoman selkeitä ja riittäviä. Samankaltaisissa tehtävissä tuli olla samansanaiset ohjeet.

B-vihkon esikokeilu pääsi äänitteiden valmistumisen myöhäisestä ajankohdasta johtuen alkamaan vasta tammikuussa 1999, koska opettajat eivät muiden velvoitteidensa vuoksi ehtineet järjestää ulkopuolisia kokeita joulukuussa. Esikokeilua varten ei laadittu vastauskaavakkeita, vaan oppilaat vastasivat konseptiarkeille. Kokeen järjestäneet opettajat arvostelivat kokeen Opetushallituksessa laadittujen ohjeiden mukaisesti ja toimittivat tulokset sekä koeaineiston Opetushallitukseen.

Kun puheen ymmärtämisen esikokeilutulokset oli analysoitu, koostettiin lopulliset kolmen tehtävävihkon yhdeksän tehtävää. Prosessi oli käytännöllisesti katsoen sama kuin edellä kuvattu. Projektipäällikkö kävi lukemassa nauhalle

² Syksyllä 1998 useat perusopetusta antavat koulut eivät olleet siirtyneet jakso- tai kurssimuotoiseen järjestelmään.

esikokeilutehtävien suoritusohjeet. Lopullisten äänitteiden suoritusohjeita lukivat myös Opetushallituksen suomen- ja ruotsinkieliset äidinkielen erikoissuunnittelijat. Saamenkielisiä ohjeita ei kokeeseen opettajan suostumuksella laadittu eikä äänitetty. Lopullisten äänitteiden kestot olivat 21'18", 21' ja 19'45".

Puhumisen kokeen esikokeilu

Suullisen kokeen laatiminen ja esikokeilu jäivät myöhäisimmiksi siksi, että tehtäviä oli laadittava lisää, jotta niitä olisi ollut riittävästi koosalaisuuden säilyttämiseksi. Esikokeilu voitiin näin ollen järjestää vasta tammikuun lopulla. Käytännön koejärjestelyjen mallina käytettiin Opetushallituksen tuottamien vapaaehtoisten lukion suullisten kokeiden ohjeita.

Esikokeilun suorittajat

Esikokeilu suoritettiin pääsääntöisesti kokeenlaatijoiden kouluissa. Kokeenlaatijat eivät kokeilleet itse laatimiaan kokeita. Tässä vaiheessa eräät edellisten kokeiden esikokeilussa mukana olleet opettajat eivät enää voineet osallistua kokeiluun, mutta heidän tilalleen onnistuttiin saamaan muutaman muun, otokseen kuulumattoman, koulun opettajia.

Kokeenlaatijoiden käyttäminen esikokeilijoina johtui siitä, että Opetushallituksessa ei ollut periaatepäätöstä esikokeilukoulujen ja -opettajien valitsemistavasta eikä myöskään varattuna taloudellisia resursseja tähän tarkoitukseen. Aikataulu ei sallinut ratkaisujen etsimistä näihin kysymyksiin, ja kun kokeenlaatijat suostuivat menettelyyn, päätettiin toimia näin. Myös yksi ohjausryhmän jäsen toimi esikokeilijana, ja projektipäällikkö kävi pitämässä puheen ymmärtämisen kokeen kolmelle oppilasryhmälle.

Kukin opettaja piti kokeen keskimäärin kahdelle opetusryhmälleen. Kun opetusryhmän keskikoko rehtorikyselyn tulosten mukaan oli 17,5 oppilasta (Tuokko 2000:87), esikokeiluun osallistuneita oppilaita oli minimissään noin 300. Tätä tutkimusta tehtäessä aineisto ei enää ollut Opetushallituksessa käytettävissä.

Esikokeilupalautteen aiheuttamat toimenpiteet

Osioiden esitestausten tulokset käsiteltiin Opetushallituksessa laskemalla osion korrelaatio mittarin summaan. Tällä perusteella valittiin toimivimmilta vaikuttaneet osiot tekstin ja puheen ymmärtämisen sekä kieliopillisten rakenteiden osakokeisiin. Koska koe oli laadittu kriteerikokeeksi, ei vältetty helppojen tai vaikeiden osioiden sisällyttämistä kokeeseen (luvut 2 ja 5). Osioiden oli tarkoitus erotella testattavalla kielenkäyttöalueella ne, jotka osaavat ja ne, jotka eivät osaa. Tätä tutkimusta tehtäessä osioanalyysien tulokset eivät enää olleet Opetushallituksessa käytettävissä.

Esikokeilupalaute osoitti, että kirjoittamistehtäviä oli ollut liian monta erityisesti siksi, että monien eri aihe-ehdotusten lukeminen vei liiaksi aikaa. Tällä perusteella tehtävävihkoihin päätettiin sijoittaa kolme aiheetta. Puhumisen kokeen dialogi- ja lukemistehtävät olivat suhteellisen toimivia sellaisenaan. Kesustelutehtäviä lyhennettiin ja niiden ohjeita selkeytettiin.

8.4 Kokeet

Opetushallitus ei antanut lupaa koko koeaineiston julkistamiseen, minkä vuoksi seuraavassa esitellään tehtävien sisältöjä erilaisista näkökulmista.

Tekstin ymmärtämisen ja kirjoittamisen tehtävät oli sijoitettu samoihin tehtävävihkoihin, joita oli kolme erilaista versiota. Samoin oli yhdistetty puheen ymmärtämisen ja kieliopillisten rakenteiden kokeet. Sanaston testaus sijoitettiin tekstin ymmärtämisen kokeen yhteyteen. Sanojen merkitystä testattiin kontekstissa eli testattavalla oli mahdollisuus tietämisen ohella tai sijasta käyttää sanapäätelytaitojaan (luku 6).

Jokaisessa vihkossa oli yksi jokaista kielitaidon osa-alueita testaava ankkuritehtävä, joka näkyy seuraavissa ymmärtämisen tehtävätyypejä kuvaavissa taulukoissa rastina vastaavan kokeen esittelyn kohdalla. Kieliopillista kompetenssia testaavista tehtävistä on koottu erillinen taulukko. Kirjoittamisen ja puhumisen kokeen tehtävätyypit esitellään tarkemmin asianomaisissa kohdissa. Ankkuritehtäväksi pyrittiin löytämään jokin aikaisemmissa arvioinneissa esiintynyt teksti, jonka arviointituloksia olisi voinut verrata tässä kokeessa saatuihin. Ehdolla oli mm. vuoden 1971 IEA:n englannin arvioinnissa käytettyjä tekstejä. Kokeenlaatijaryhmä piti niitä kuitenkin sisällöltään vanhahtavina testattavien mielenkiinnon herättämisen kannalta. Tehtäviä yritettiin saada myös Ruotsista, mutta Skolverketiin nimitettiin arvioinnista vastaava henkilö meidän arviointimme kannalta siinä määrin myöhään, että emme voineet jäädä odottamaan. Puheen ymmärtämisen kokeeseen saatiin sen sijaan sisällytetyksi yksi aikaisemmin käytetty tehtävä.

Kvantitatiivisissa analyyseissä lähdetään yleensä siitä, että oikein/väärin- ja monivalintatehtävät arvostellaan dikotomisesti (0/1). Nämä tehtävätyypit eivät kuitenkaan välttämättä ole testattavalle helpompia kuin ns. avokysymykset, joihin testattava vastaa muutamien sanoin lukemansa tai kuulemansa perusteella. Kun kokeesta lasketaan kokonaistulos, jossa avokysymyksestä on annettu esimerkiksi kolme pistettä, oikein/väärin- ja monivalintatehtävän painoarvo jää dikotomisesti pisteitettäessä liian vähäiseksi. Alkuperäisessä, vuonna 1999 suoritettussa arvioinnissa oikein/väärin- ja monivalintaosioista on annettu 0-2 p/osio, jotta sitten esimerkiksi helpoksi tehtäväksi arvioidun yhden sanan käännöksestä tai avovastauksen osasta saattoi antaa suhteellisesti oikeudenmukaisemman yhden pisteen (ks. kuitenkin luku 11).

Tekstin ymmärtämisen koe

Tekstin ymmärtämisen kokeessa oppilaat merkitsivät monivalinta- ja oikein/väärin-tehtävien vastaukset annettujen ohjeiden mukaan optisille lomakkeille. Avovastaukset kirjoitettiin konseptiarkille, jolta opettajat siirsivät pisteiden merkitsemisen jälkeen tulokset optisille lomakkeille.

Taulukko 52 osoittaa tekstin ymmärtämisen kokeen eri tehtäväsarjoissa käytetyt tehtävätyypit, kunkin tehtävän osioiden lukumäärän sekä tehtävätyypin prosentuaalisen osuuden.

TAULUKKO 52 Tekstin ymmärtämisen koeversioiden sisältö.

Tehtävätyyppi	Osioiden lukumäärä	Tehtäväsarja			Tehtävätyyppien %-osuus
		1	2	3	
Monivalinta - 3 äidinkielistä vaihtoehtoa	5	x	x	x	14
- 4 äidinkielistä vaihtoehtoa	4	x			11
- 3 kohdekielistä vaihtoehtoa	3		x		8
	4			x	11
Avokysymykset äidinkielellä	6	x			17
O/V -väittämät	3			x	8
Sanojen kääntämisen äidinkieleen	4			x	11
Sanojen/Sanontojen kääntäminen äidinkieleen	4		x		11
Pistemäärä	36 ³	32	30	28	100 %

Tarkasteltaessa tehtävätyyppejä taulukon 52 pohjalta voidaan todeta, että niistä helpoksi oletettuja O/V -väittämiä oli 8 % ja äidinkieliisiä monivalintatehtäviä, joissa oli kolme vaihtoehtoa 14 % eli yhteensä 22 %. Keskivaikeiksi oletettuja äidinkieliisiä neljän vaihtoehdon äidinkieliisiä monivalintatehtäviä oli 11 %, kohdekieliisiä kolmen vaihtoehdon tehtäviä 19 % ja äidinkieliisiä avokysymyksiä 17 %, eli yhteensä 47 %. Vaikeiksi oletettuja osioita, joiden katsottiin edellyttävän ongelmanratkaisutaitoja, olivat tekstiyhteydessä esiintyvien sanojen (11 %) ja sanontojen (11 %) kääntäminen äidinkieleen.

Aldersonin ehdotus (2000:257) kommunikatiivisen tekstin ymmärtämisen kokeen laadinnalle on valita ensin sellaisia tekstejä, joita testattavat haluaisivat lukea (luku 6). Tähän haasteeseen on vastattu sisällyttämällä kokeeseen mahdollisimman monia erilaisia aihepiirejä (taulukko 53), joiden sisällöt on poimittu kynnystaitoluettelosta (luku 4). Tekstityypin valinnan perustana oli Weirin (1993) jaottelu tekstin viestintätehtäviin taulukon 53 mukaisesti. Viestintätehtävien jaottelu voidaan rinnastaa Linnakylän (Linnakylä 1995:14) esittämään Werlichiltä peräisin olevaan tekstien luokitteluun seuraavasti:

- | | |
|----------------------------------|--------------------------------|
| 1. Lukijan toiminnan ohjaaminen | = instruktiivinen |
| 2. Lukijalle kertominen | = narratiivinen |
| 3. Tiedon välittäminen lukijalle | = ekspositorinen eli selittävä |

Muita tekstityyppejä ei valittuihin kokeisiin sisällynyt. Perinteisten suorasanaisten tekstien ohella tehtävinä oli esimerkiksi aikataulun ja taulukon tulkintaa.

³ Jo ennen tätä tutkimusta varten suoritettua kalibrointia tekstin ymmärtämisen datasta jätettiin pois yksi osio, jonka alkuperäisen (polytomisen) pistemäärän muuttaminen dikotomiseksi osoittautui hankalaksi.

TAULUKKO 53 Tekstin ymmärtämisen kokeen aihepiirit ja viestintätehtävät.

Aihepiiri	Tehtäväsarja 1		Tehtäväsarja 2			Tehtäväsarja 3			
	Tehtävä		Tehtävä			Tehtävä			
	1	2	3	1	2	3	1	2	3
Vapaa-aika ja harrastukset	x			x			x		
Ihminen ja luonto		x							
Liikenne			x						
Kierrätys					x				
Pukeutuminen						x			
Muut maat ja kansat/ruoka								x	
Terveys									X
Viestintätehtävä									
(Lukijan) toiminnan ohjaaminen	x	.		x			x		
(Lukijalle) kertominen		x				x		x	
Tiedon välittäminen			x		x				X

Seuraava kysymys on, miksi testattava lukisi tekstejä. Oletetusta vastauksesta riippuen tehtäviksi valittiin kysymys/-tehtävätyyppi, johon vastaaminen edellyttää jonkin taulukon 54 tekstin ymmärtämisen operaation soveltamista. Taulukosta ilmenee myös kutakin operaatiota edellyttävien tehtävien lukumäärä

TAULUKKO 54 Tekstin ymmärtämisen operaatiot (Weir 1993:73)

Operaatio	Osioiden lukumäärä
1. Tekstin silmäily pääasian ymmärtämiseksi.	10
2. Paikallistaminen tietyn asian selville saamiseksi.	11
3. Tarkan merkityksen ymmärtäminen	11
4. Tekstin eri osien välisten yhteyksien ymmärtäminen	7
5. Implisiittisesti ilmaistun merkityksen ymmärtäminen pääättelemällä.	-

Puheen ymmärtämisen koe

Puheen ymmärtämisen kokeessa oppilaat merkitsivät monivalinta- ja oikein/väärin-tehtävien vastaukset annettujen ohjeiden mukaan optisille lomakkeille. Avovastaukset kirjoitettiin konseptiarkille, jolta opettajat siirsivät pisteiden merkitsemisen jälkeen tulokset optisille lomakkeille.

Kunkin kolmen tehtäväsarjan ankkuritehtävänä oli yhdeksän erillistä äänitettyä keskustelutilannetta, joiden ymmärtämistä testattiin monivalintatehtävin, joissa oli neljä kohdekielistä vaihtoehtoa. Sitä oli käytetty aikaisemmissa tutkimuksissa jo kahdesti, joten se antoi taustaa ja vertailupohjaa kansallisen arvioinnin tuloksille. Uniikkitehtävinä käytettiin monivalinnan lisäksi oikein/väärin-väittämiä, äidinkieliä avokysymyksiä sekä sanelua (luku 6). Ankkuritehtävä testasi myös kaikkia tärkeimpiä Weirin (1993) jaotteluun sisältyviä viestintätaitoja.

Taulukko 55 osoittaa puheen ymmärtämisen kokeen eri tehtäväsarjoissa käytetyt tehtävätyypit, kunkin tehtävän osioiden lukumäärän ja kunkin tehtävätyypin prosenttiosuuden.

TAULUKKO 55 Puheen ymmärtämisen koeversioiden sisältö.

Tehtävätyyppi	Osioden lukumäärä	Tehtäväsarja			Tehtävätyyppien %-osuus
		1	2	3	
O/V -väittämät äidinkielellä	2	x			7
Monivalinta					30
- 3 äidinkielistä vaihtoehtoa	3			x	10
	3		x		10
- 4 englanninkielistä vaihtoehtoa	9	x	x	x	29
Avokysymykset äidinkielellä	4	x			13
	2		x		7
Sanelu	8			x	26
Pistemäärä	31	29		34	100 %

Helpoiksi arvioituja tehtäviä olivat äidinkieliset O/V -väittämät (7 %) sekä monivalintatehtävät, joissa oli kolme äidinkielistä vaihtoehtoa (20 %), eli 27 % tehtävistä. Keskivaikeiksi oli arvioitu monivalinnat, joissa oli neljä englanninkielistä vaihtoehtoa (29%) ja äidinkieliset avokysymykset (20 %) eli 49 % tehtävistä. Vaikeaksi arvioitiin sanelu, 26 %.

Arvioinnin viitekehys edellytti tehtävien aihepiirien olevan monipuolisia, ikäkaudelle sopivia ja molempia sukupuolia kiinnostavia. Taulukko 56 osoittaa puheen ymmärtämisen kokeen aihepiirit. Jaottelussa on käytetty pääasiassa Threshold Level -kynnystaitoluettelon (1990) luokittelua, kun taas jaottelu viestintätehtäviin perustuu Weirin (1993) tulkintaan.

TAULUKKO 56 Puheen ymmärtämisen kokeen aihepiirit ja viestintätehtävät.

Aihepiiri	Tehtäväsarja 1			Tehtäväsarja 2			Tehtäväsarja 3		
	Teht. 1	Teht. 2	Teht. 3	Teht. 1	Teht. 2	Teht. 3	Teht. 1	Teht. 2	Teht. 3
Jokapäiväinen elämä	x			x			x		
Urheilu		x			x				
Muut maat ja kansat			x						
Nuoriso						x		x	
Ruoka									x
Viestintätehtävä									
Sosiaalinen vuorovaikutus	x			x			x		
Tiedon välittäminen (kuulijalle)		x	x					x	
(Kuulijalle) kertominen					x				
(Kuulijan) toiminnan ohjaaminen						x			x

Arviointi pyrittiin kohdentamaan niihin piirteisiin, jotka ovat tyypillisiä kuuntelemiselle ja joita ei siis voitaisi luontevasti testata kommunikatiivisesti esimerkiksi tekstin ymmärtämisen kokeella (luku 6). Tehtävät pyrittiin englannin arvioinnissa kokoamaan niin, että niissä mitattaisiin kykyä

- prosessoida realistisia näytteitä / katkelmia kielestä automaattisesti ja reaaliajassa,
- ymmärtää kaikki se kielellinen informaatio, joka yksiselitteisesti sisältyy tekstiin, ja
- tehdä ne johtopäätökset, jotka yksiselitteisesti ilmenevät näytteen/katkelman sisälöstä (luvut 4-6; Buck 2001:250).

Ankkuritehtävissä⁴ osioiden laatiminen lähti kuitenkin toisesta lähestymistavasta, kohdekielen käyttötilanteista identifioimalla kussakin tilanteessa tyypillisimmät kommunikatiiviset tehtävät ja arvioimalla sen jälkeen tärkeimmät taidot (luku 5; Buck 2001:251).

Tehtävien tulo

- liittyä kiinteästi kuultuun katkelmaan, joten niitä ei voi suorittaa kuuntelematta sitä (ovat siis *passage dependent*),
- edellyttää ymmärtämistä, so. tehtävä ja kuultu eivät saa olla samansanaisia, ja
- testata niitä tietoja tai taitoja, jotka oli määrätty konstruktissa tai jotka oli katsottu keskeisiksi testaamisen kohteiksi sovitussa kielenkäyttötilanteissa (luku 6).

Taulukossa 57 esitetään operaatiot, joita puheen ymmärtämisen kokeessa edellytettiin sekä niiden osioiden lukumäärä, jotka testasivat kunkin em. operaation hallintaa.

TAULUKKO 57 Puheen ymmärtämisen operaatiot. (Lähde: Weir 1993:97.)

Operaatio	Osioiden lukumäärä
1. Pääasian ymmärtäminen	2
2. Puhujan tarkoituksen ymmärtäminen	7
3. Yksityiskohtainen ymmärtäminen	9
4. Päättely	4
5. Sanelu ⁵	8

Operaatiot voidaan myös ryhmitellä sen mukaan, kuinka suorasti tai epäsuorasti puhujan tarkoitus ilmenee (*direct meaning comprehension* ja *indirect meaning comprehension*). Näistä edelliseen kuuluisivat taulukon 57 ryhmät 1-3 ja jälkimmäiseen ryhmä 4 (luku 4). Weir (1993:98) erottaa vielä eri ryhmäksi sanelun, koska siihen liittyy kuullun ymmärtämisen lisäksi tuottamista eli kirjoittaminen.

Arvioinnissa käytetty sanelu oli ns. osittais- eli aukkosanelu (*partial dictation*). Tässä tehtävässä täydennettiin yksittäisiä sanoja tai lyhyitä, muutaman sanan mittaisia katkelmia kuultuun ruuanlaitto-ohjeeseen. Pisteitys, lukuun ottamatta kahta kohtaa, oli dikotominen. Väärin kirjoitetusta sanasta rangaistiin, jos

⁴ Tehtävä oli saatu käyttöön Kasvatustieteiden tutkimuslaitokselta, jonka arvioinneissa sitä oli käytetty vuosina 1994 ja 1995.

⁵ Sanelussakin on luonnollisesti kysymys yksityiskohtaisesta ymmärtämisestä. Kohdassa 5 testaus kuitenkin kohdistui yhteen sanaan ja tuotoksen arvioinnissa otettiin huomioon myös oikeinkirjoitus

sitä oli mahdoton ymmärtää tai jos tuloksena oli täysin muuta merkitsevä sana. Periaatteessa täydennykseksi olisi hyväksytty myös synonyymi tai toisin sanoin kuvattu ohjeen osa, mutta tämän tyyppisessä tekstissä korvaavia sanoja ja sanontoja ei yleensä löydy, niitä ei ohjeissa rohkaistu etsimään eivätkä oppilaat niitä myöskään tässä kokeessa pyrkineet käyttämään.

Kuuntelu jaetaan usein kaksivaiheiseksi prosessiksi, joista ensimmäinen on kielellisen perusinformaation erottaminen ja toinen informaation hyödyntäminen kommunikatiiviseen tarkoitukseen (Buck 2001:51). Buck (2001) katsoo, että sanelu operationaalistaa kuuntelun vain edellisellä tasolla ja testaa sen lisäksi myös lyhytmuistia ja kirjoittamista (emt. s. 78). Lyhyiden jaksojen tunnistaminen testaa tuskin muuta kuin sanojen tunnistamista (emt. s. 77). Mm. Finocchiaro ja Sacks (1983), Weir (19923) ja Hughes (1989/2004) ovat kuitenkin pitäneet sanelua hyvänä tapana testata puheen ymmärtämistä.

Kieliopillinen kompetenssi

Kussakin tehtäväsarjassa oli 30 monivalinta- ja 4 aukkotäydennystehtävää sekä 6 käännösvirkettä. Monivalintatehtävistä 20 ensimmäistä oli kaikille tehtäväsarjoille yhteisiä ja loput joko esiintyivät kahdessa tehtäväsarjassa tai olivat uniikkitehtäviä (taulukko 58). Aukko- ja käännöstehtävät olivat uniikkitehtäviä. Käännöstehtävät lisättiin kokeeseen jälkikäteen tehtävätyyppivalikoiman monipuolistamiseksi. Kuten aikaisemmin perusteltiin, monivalintatehtäviä käytettiin kokeen käytännöllisyyden lisäämiseksi. Ne nopeuttavan testattavan vastaamista, testaavat hyvin laadittuina tarkoitettua konstruktia, ovat objektiivisia ja nopeasti korjattavia.

Monivalintaosiot ja aukkotäydennykset arvosteltiin dikotomisesti. Käännöslauseet oli tarkoitus pisteittää antamalla piste vain oikeasta kieliopillisesta rakenteesta vähentämättä pisteitä mahdollisista muista virheistä. Esimerkiksi virkkeen "Oletko nähnyt häntä?" käännöksestä olisi voitu antaa yksi piste, vaikka vain objektimuoto *him/her* olisi ollut oikein edellyttäen, että tarkoitus olisi ollut testata vain sitä.

TAULUKKO 58 Kieliopillisen kompetenssin kokeen ankkuri- ja uniikkiosioden lukumäärät.

Tehtäväsarja 1	Tehtäväsarja 2	Tehtäväsarja 3
1-20	1-20	1-20
21	21	
22	22	
23	23	
24	24	
25	25	
	26	21
	27	22
	28	23
	29	24
26		25
27		26
28		27
29		28
30		
	30	29
		30

Useimmat monivalintatehtävien osiot testasivat verbiopin ja pronomien hallintaa. Substantiivien testauksen yhteydessä saatettiin tarkkailla esimerkiksi säännöllisen ja epäsäännöllisen monikon sekä genetiivin hallintaa. Artikkelitehtävissä testattavan piti esimerkiksi huomata valita vokaalilla alkavan sanan eteen epämääräinen artikkeli *an*, ja preesens-väittämässä liittää verbin perusmuotoon yksikön kolmannen persoonan ortografisesti oikea päätte.

Puhumisen koe

Koe suoritettiin parityönä ja äänitettiin kouluissa niihin toimitetuille kaseteille. Parityön käyttö perustui ajatukseen ajankäytön minimoinnista sekä kokemukseen, joita lukion suullisen kielitaidon kokeiluissa 1990-luvun alussa oli saatu lukion suullisen kielitaidon opetuskokeilusta (OPH 1993).

Itse kokeen kesto oli 15 minuuttia ja sitä ennen pareilla oli 15 minuuttia kokeeseen tutustumista varten. Oppilaat tulivat kokeeseen eri aineiden oppitunneilta. Koesalaisuuden turvaamiseksi pyrittiin siihen, että peräkkäisille pareille ei tulisi samoja tehtäviä, mikä edellytti useiden tehtävien ja tehtäväsarjojen laatimista. Puhumisen kokeessa oli käytössä yhteensä kuusi tehtäväsarjaa, joissa osa tehtävistä oli samoja, kuten taulukko 59 osoittaa. A-tehtävät (10) olivat roolitehtäviä, B-tehtävät (6) lukemistehtäviä ja C-tehtävät (4) keskustelutehtäviä. Kummallakin parilla oli oma tehtävävihkonsa, jossa olivat heidän osuutensa äidinkielliset ohjeet. Parit päättivät itse, kumman roolihenkilön osan he esittäisivät.

A-tehtävät olivat melko tarkasti strukturoituja (luku 6). Keskustelustrategioiden käytöstä (siirtyminen aiheesta toiseen, toiston pyytäminen jne.) ei erikseen muistutettu koeohjeissa, mutta asia sisältyi testattaville erikseen annettavaksi tarkoitettuun informaatioon. Kussakin C-tehtävässä oli virikkeenä kolme väittämää, joita testattavien tuli käsitellä mahdollisimman kattavasti. Näistä kahden tehtävän väittämät liittyivät luettuihin tekstikatkelmiin, jolloin oppilaita

ohjattiin myös hyödyntämään lukemaansa. Yhdessä tehtävässä oli ohje käsitellä vain yhtä annetuista väittämistä. Osassa keskusteluja edellytettiin ratkaisun esittämistä annettuun ongelmaan. Keskusteluaiheet olivat

- kahden maan nuorison vertailu,
- myymälävarkauksien estäminen,
- nuorisorikollisuuden vähentäminen, ja
- nuoriso ja vapaa-aika.

Keskustelutehtävien voidaan katsoa olleen osin puolistrukturoituja, osin avoimia (luku 6).

Opettajat arvioivat suullisen kokeen niukasti määritellyllä kuusiportaisella asteikolla 0-5 käyttäen kolmea kriteeriä⁶ à 5 p. Kouluille oli toimitettu kaavakkeita, joille oppilaiden suoritukset merkittiin arvioinnin yhteydessä ja joilta ne siirrettiin optisille lomakkeille Opetushallitukseen toimittamista varten. Siellä osa tuotoksista pisteitettiin uudelleen. Puhumisen kokeen toimeenpanosta sisällytettiin opettajan ohjeisiin informaatiota sekä koetta vastaanottaville opettajille että siihen osallistuville oppilaille. Tehtävät jakautuivat eri tehtävävihkoihin taulukossa 59 esitetyllä tavalla.

TAULUKKO 59 Puhumisen kokeen tehtävien jakaantuminen eri tehtävävihkoihin.
(Lähde: Tuokko 2000:26.)

Tehtävä	Tehtäväsarja					
	1	2	3	4	5	6
1	A1					
2	A2	A3	A2	A3	A2	A2
3	A3	A4	A5	A4	A3	A5
4	A6	A7	A8	A9	A10	A8
5	B1+B2	B3+B4	B1+B2	B5+B6	B1+B2	B3+B4
6	C1	C2	C3	C4	C1	C2

Puhumisen kokeen tehtävissä testattiin kielitaidon funktioiden hallintaa seuraavissa jokapäiväisen elämän kielenkäyttötilanteissa:

- puhelimeen vastaaminen, puhelinnumeron ilmoittaminen, puhelun päättäminen,
- päivämäärän ja kellonajan ilmoittaminen,
- ulkonäön ja vaatetuksen kuvaus (numerot, värit, vaatekappaleet),
- suunnan ilmaiseminen,
- pahoittelu,
- avun tarjoaminen,
- tuntemattoman henkilön puhuttelu,
- tapaaminen ja esittäytyminen,
- ruuan ja juoman tilaaminen,
- sään kommentoiminen,
- perheestä ja harrastuksista kertominen,
- vierailuohjelman tulkitseminen,
- liikkeessä asiointi,
- keskustelu tullivirkailijan kanssa,
- juhlapäivistä kertominen.

⁶ sujuvuus, rakenteet ja sanasto, ääntäminen

Funktioiden lähteenä on käytetty Threshold Level -kynnystaitoluettelo (1990) ja Suomen Peruskoulun opetussuunnitelman perusteet 1985 oppimääräkuvauksia. Ne tai niissä käsitellyt aiheet sisältyvät myös perusopetuksessa käytettyihin oppimateriaaleihin.

Kirjoittamisen koe

Aiheet valittiin esikokeilussa suosituimmiksi osoittautuneiden joukosta. Testattavat saivat kirjoittaa yhdestä annetuista kolmesta aiheesta, joista ensimmäinen oli yhteinen kaikille tehtäväsarjoille. Taulukosta 60 käyvät ilmi kirjoittamisen laji, aiheen kuvaus ja ohjannan määrä.

TAULUKKO 60 Kirjoittamisen kokeen aiheiden kuvaus.

Tekstilaji	Aiheen kuvaus	Ohjannan määrä
Narratiivinen	- Mieluisin päivä/vuodenaika/kuukausi	Ei ohjantaa
	- Kotipaikkakunnan esittely	Ei ohjantaa
	- Mieleen painuneen päivän tapahtuma(t)	Ei ohjantaa
Deskriptiivinen	- Kirje ystävälle matkalta	Englanninkielinen kysymyskiekko
	- Musiikin (tai muun harrastuksen) merkitys	Ei ohjantaa
Argumentoiva	- Vastaaminen nuorten palstan kirjoitukseen	Englanninkielinen nuorten palstan kirjoitus koulukiusaamisesta

Oppilaat kirjoittivat kirjoitelmat konseptiarkeille. Opettajat arvioivat ne suhteellisen väljin ohjein kuusiportaisella asteikolla käyttäen kolmea kriteeriä⁷, joista kukin oli viiden pisteen arvoinen, ja siirsivät tulokset optisille lomakkeille. Niille tehtiin myös merkintä, jos tuotos oli alle 25 sanaa tai yli 125 sanaa. Opetushallituksessa osa tuotoksista pisteitettiin uudelleen.

8.5 Arvioinnin toimeenpano

Tiedotusvälineitä informoitiin arvioinnista syyskuussa 1998. Opetushallitus tiedotti otoskouluille otantaan osumisesta 3.12.1998 ja arvioinnin edellyttämistä järjestelyistä 25.2.1999 lähetetyllä kirjeellä. Koulujen rehtoreita pyydettiin myös vahvistamaan koulun osallistuminen arviointiin. Koeaineisto ohjeineen toimitettiin kouluille siten, että se Postin kanssa tehdyn sopimuksen mukaan oli perillä viikkoa ennen koepäivää.

Tässä vaiheessa neljä koulua ilmoitti olevansa estyneitä osallistumaan arviointiin. Koska arviointiin osallistuminen oli lakisäätöistä, siitä kieltäytymiseen tuli olla vahvat perusteet. Yhdessä koulussa ei opiskeltu englantia A-kielenä, yksi oli yhdistetty toiseen kouluun, yhdessä koulun voimavarat hupenivat ponnisteluihin saada koululle todistuksenanto-oikeus ja yhdessä oppilaat suorittivat kyseisenä ajanjaksona muun aineen opintoja toisen paikkakunnan kou-

⁷ sisältö, rakenteet, sanasto

lussa. Varakoulujen valinta ei kuitenkaan aiheuttanut muutoksia otokseen, koska ne voitiin poimia samoilta alueilta käyttäen samoja kriteereitä kuin millä pois jääneetkin oli otostettu.

Arvioinnin toimeenpanon tueksi laadittiin Opettajan ohjeet -vihko, jonka sisällysluettelo on liitteenä 16. Liitteenä 17 ovat suulliseen kokeeseen osallistuville oppilaille etukäteen annetut ohjeet kokeen sisällöstä, siinä toimimisesta ja käytännön järjestelyistä sekä liitteenä 18 opettajan muistilista ja vihjeet suullisen kokeen järjestämiseksi. Liitteenä 19 ovat valvojan ohjeet, joita tuli noudattaa koelosuhteiden yhdenvertaisuuden takaamiseksi eri kouluissa.

Koulut olivat saaneet ohjeen valita suulliseen kokeeseen osallistuvat oppilaat siten, että ensimmäinen osallistuva oppilas arvottiin. Jos varsinaiseen otokseen oli osunut esimerkiksi 30 oppilasta ja arpakuutio osoitti numeroa neljä, aakkosissa neljäs oppilas oli ensimmäinen osallistuja, seuraava kahdeksas oppilas jne. Kokeeseen tuli kuitenkin osallistua vähintään neljän oppilasparin, joten tarvittaessa poimittiin joka kolmas tai joka toinen oppilas, jotta lukumäärä saatiin täyteen. Koulun oli määrä alleviivata Opetushallituksen toimitettavaan oppilasluetteloon suulliseen kokeeseen osallistuvat oppilaat. Otokseen osuneiden oppilaiden joukosta opettajat valitsivat oppilasparit. Opetushallitus suositti, että taidoiltaan suunnilleen samantasoiset oppilaat toimisivat yhdessä, mutta päätäntävalta tästä jäi opettajille.

Kouluihin toimitettavat suullisen kokeen tehtävävihkot pakattiin siihen järjestykseen, jossa ne oli tarkoitus jakaa oppilaille. Aakkosjärjestyksessä ensimmäisen koulun ensimmäisen oppilasparin tehtävävihko arvottiin. Oletetaan, että arpakuutio osoitti numeroa 3 ja että koulussa oli arvioitu neljän oppilasparin osallistuvan kokeeseen. Koulun lähetykseen tuli silloin päällimmäiseksi tehtäväsarjan 3 oppilasvihkot A ja B ja alimmaiseksi tehtäväsarjan 6 vihkot. Seuraavalle koululle, jossa oli ehkä myös neljä oppilasparia osallistujina, toimitettiin tehtäväsarjoja 1, 2, 3 ja 4 jne. Puhumisen kokeen arviointia varten toimitettiin vastaustaavakkeita, joilta opettajan tuli siirtää tulokset optisille lomakkeille.

Kirjallisen kokeen kolmesta tehtäväsarjasta kullekin koululle toimitettiin kahta tehtäväsarjaa, jotta pientenkin koulujen oppilaat tulisivat saaneeksi riittävän määrän tehtäviä vastattavakseen koulun tason johtopäätösten tekemisen mahdollistamiseksi. Tehtävävihkojen toimittaminen kouluille aloitettiin tehtäväsarjasta 1. Jos otoskoulujen luettelon ensimmäisessä koulussa oli esimerkiksi 47 otosoppilasta, kouluille toimitettujen tehtäväsarjan 1 vihkojen lukumäärä oli 23 ja tehtäväsarjan 2 puolestaan 24. Aakkosjärjestyksessä seuraavaan otoskouluun, jossa oli esimerkiksi 54 otosoppilasta, toimitettiin 27 tehtäväsarjan 3 vihkoa ja 27 tehtäväsarjan 1 vihkoa.

Kokeiden kouluille toimittamista varten niistä oli laadittu lista, josta ilmeni koulun nimen lisäksi koulun koodi ja postiosoite, koko kokeeseen osallistuvien oppilaiden sekä arvioitu suulliseen kokeeseen osallistuvien oppilaiden määrä sekä arvonnin perusteella saatu tieto siitä, mikä tehtäväsarja oli kunkin koulun ensimmäinen.

Kokeiden toimittaminen kouluille oli työ, jonka parissa uurasti muutaman päivän käytännöllisesti katsoen koko Arviointivirkamiesten väki. Kokeet sijo-

tettiin siihen järjestykseen, jossa ne oli tarkoitus jakaa oppilaille. Laatikkoihin pakattiin tehtävävihkojen ja optisten lomakkeiden lisäksi kuuntelukasetit, tarpeellinen määrä C-kasetteja suullisen kokeen tuotosten äänittämistä varten, opettajan ohjeet, sablonit monivalintatehtävien tarkistamista varten ja omavastauskuoret, joissa koulut palauttaisivat aineiston Opetushallitukseen. Laatikot odottivat käytävällä kuljetuskärryissä muutaman päivän ennen kuin Postin edustajat tulivat ne noutamaan. Lähetykset toimitettiin pikapaketteina.

Puhumisen koe järjestettiin ennen kirjallista koetta 30.3.1999. Työtapa oli parityö, ja oppilaille annettiin 15 minuuttia aikaa valmistautua kokeeseen, minkä jälkeen he osallistuivat 15 minuuttia kestävään kokeeseen. Koulu sai itse valita ajankohdan, jolloin suulliset kokeet aloitettiin, koska kokeen toimeenpaneminen vaati runsaasti järjestelyjä: valmistautumis- ja kokeen suoritus-tilan, nauhurin/videokameran sijoittamisen koetilaan sekä kahden (suositus) opettajan vapauttaminen luokkatyöskentelystä koetta toimeenpanemaan. Koulut saivat myös tarpeelliseksi arvioidun määrän C-kasetteja oppilastuotosten äänittämistä varten. Vaihtoehtoisesti tuotokset saattoi myös videoida, mutta tätä mahdollisuutta ei käytetty. Videokasetteja ei myöskään tarjottu kouluille. Kirjallinen koe järjestettiin seuraavana päivänä, 31.3.1999. Kokeeseen tuli käyttää 45+15+45 minuuttia, jolloin itse kokeen suorittamiseen jäi kaksi 45 minuutin jaksoa ja niiden väliin 15 minuuttia, joka oli tarkoitus käyttää ohjeiden antamiseen, henkilökohtaisten tietojen merkitsemiseen optisille lomakkeille ja taustakyselyyn vastaamiseen.

Jotta tulosten tilastollinen käsittely nopeutuisi ja helpottuisi, lopullisina vastauskaavakkeina käytettiin optista lomaketta, jota luettiin JOTOS -ohjelmalla. Jokaista reseptiivisen kokeen tehtävävihkoa varten oli Opetushallituksessa suunniteltu oma lomake kutakin koekokonaisuutta varten. Oppilaat mustasivat itse oikeat vaihtoehdot monivalintatehtäviin ja oikein/väärin-väittämiin sekä taustakyselyyn ja asenneväittämiin. Opettajat mustasivat lomakkeille avokysymyksistä antamansa pisteet heille toimitettujen ohjeiden mukaisesti. Oppilaille ei varattu tilaisuutta harjoitella optisen lomakkeen käyttöä.

Huolimatta siitä, että kokeiden korjaaminen osui osittain pääsiäislomalle, suurin osa tuloksista toimitettiin määräaikaan mennessä, kahden viikon kuluessa, Opetushallitukseen, joten ensimmäiset tulokset saatettiin toimittaa kouluille ennen lukukauden loppua. Otsokoulujen rehtoreille 19.5.1999 päivätyn kirjeen liitteenä oli pikapalautte arvioinnin tuloksista tehtäväsarjoittain koulukohtaisesti ja koko maan osalta seuraavasti:

- taustamuuttujat,
- oppilaiden asennekyselyn tulokset,
- pistekeskisarvot ja
- pistemäärien jakaumat tekstin ymmärtämisen, kuullun ymmärtämisen ja kielipillisen kompetenssin osakokeista.

Kokeiden tulokset analysoitiin OPLM-ohjelmalla (luku 8) laskemalla klassisen testiteorian mukainen osioiden vaikeustaso ja erottelukyky sekä oppilaiden ratkaisuprosentit. Regressioanalyysillä pyrittiin selvittämään, löytyisikö oppilastiedoista tekijää tai tekijöitä, jo(i)lla olisi yhteyttä oppimistuloksiin. Pikapalaut-

teet perustuivat opettajien ja oppilaiden optiselle lomakkeelle merkitsemiin pistemääriin ja tulosten käsittelyyn optisella lukijalla ja tilasto-ohjelmilla. Oppilaan saamat tulokset suullisesta kokeesta ilmoitettiin sen kirjallisen kokeen yhteydessä, johon oppilas osallistui. Niissä kouluissa, joissa oppilaat tekivät kokeessa A eri tehtäväsarjat kuin kokeessa B, taustamuuttajat ja asennekyselyjen tulokset ilmoitettiin molempien kokeiden tulosten yhteydessä. Tiedot tuloksista toimitettiin ainoastaan kouluihin. Opetushallituksessa suoritettujen analyysien perusteella kouluja eikä oppilaita voi identifioida.

Vuoden 2000 alkupuolella projektipäällikkö esitteli alustavasti raportin pääjohtaja Jukka Sarjalalle, minkä jälkeen se esiteltiin opetusministeri Maija Raskille. Opetusministeriö informoi tuloksista tiedotusvälineitä, ja Opetushallituksessa järjestettiin niille tiedotustilaisuus.

8.6 Yhteenveto

Englannin kansallisen arvioinnin toimeenpano voidaan jakaa neljään osaan. Ensimmäinen niistä liittyy arvioinnin taustoittamiseen. Siihen liittyi kokeenlaatijoiden nimeäminen ja heidän työnjaostaan ja kokeiden sisällöstä ja niiden toimittamisaikataulusta sopiminen. Tässä yhteydessä pidettiin useita palavereita ohjausryhmän kanssa mietittäessä mahdollisimman hyviin tuloksiin pääsemistä melko vähäisin ajallisin resurssein. Itse kokeiden laadinnan aikana Opetushallituksessa selvitettiin mm. äänitysmahdollisuuksia, painatuspaikkoja ja niiden aikatauluja sekä valmisteltiin optisen lomakkeen koostamista. Tällöin laadittiin myös kyselykaavakkeet oppilaille, opettajille ja rehtoreille.

Toisessa vaiheessa, jolloin kokeenlaatijat olivat toimittaneet koe-ehdotuksensa Opetushallitukseen, täsmennettiin melko laveasti hahmotellut koekonstruktit, täydennettiin toimitettuja koespesifikaatioita, valittiin tehtävät ja niihin liittyvät osiot ja sijoitettiin ne tehtävävihkoihin esikokeilua varten. Samalla laadittiin myös suoritus- ja pisteitysohjeet sekä äänityskäsikirjoitukset.

Kolmas vaihe koostui esikokeilupalautteen analysoinnista ja kokeiden muokkaamisesta. Osioanalyysin perusteella valittiin lopullisiin tehtävävihkoihin paitsi keskivaikeita myös kriteeriviitteisen arvioinnin mukaisia ”helppoja” tehtäviä vaikeiksikaan osoittautuneita unohtamatta. Useita tehtäviä jouduttiin lyhentämään, ja osioita tämän vuoksi muokkaamaan. Kirjoittamisen kokeen tehtäviä vähennettiin ja rakennekokeen tehtävätyyppivalikoimaa monipuolistettiin. Erityisesti puheen ymmärtämisen kokeen valmistelu vei runsaasti aikaa, koska muokkaamisen jälkeen käytännöllisesti katsoen kaikki aineisto oli kirjoitettava uudelleen äänitystä varten.

Puhumisen koe saatiin esikokeiluun vasta tässä vaiheessa. Palaute siitä tuli kuitenkin nopeasti, ja koska tehtävät näyttivät toimivan, kokeen muokkaaminen oli suhteellisen helppoa. Neljännessä eli viimeisessä vaiheessa keskityttiin raportin työstämiseen. Vaikka työ oli haastava, raportti on jo kirjoitettu eikä sen kommentointi enää kuulu tämän tutkimuksen piiriin.

Seuraavassa siirrytään kuvaamaan tutkimuskysymyksiin tuotettuja vastauksia. Kuten edellä sanotusta on käynyt ilmi, tässä tutkimuksessa on monella tavalla edetty alkuperäistä tulosraporttia pidemmälle.

9 TULOKSET

Tässä luvussa esitetään perusopetuksen englannin kielen kansallisen arvioinnin koetulokset vastauksena esitettyihin tutkimuskysymyksiin. Tuloksia tarkastellaan eri näkökulmista. Yksi tarkastelunäkökulma koskee jakoa reseptiiviseen ja produktiiviseen kielitaitoon. Keskeinen syy tämän näkökulman mukaan ottamiseen oli se, että yhdistämällä varsinaisen reseptiivisen kielitaidon kokeen ja kieliopillisen kompetenssin kokeen osiot voitiin varmistaa mittauksen riittävä reliaaбелиustaso. Kuten jäljempänä käy ilmi, sekä puheen että tekstin ymmärtämisen osakokeiden reliaaбелиus ei ollut niin korkea kuin olisi ollut toivottavaa, mikä johtuu pääasiassa vähäisestä osiomäärästä.

Reseptiivisen kielitaidon koeosioiden tarkoitukseensa soveltuvuuden arvioimiseksi esitellään ensin kaikkien koeosioiden vaikeustaso ja erottelukyky sekä klassisen että uuden testiteorian analyysien mukaisina. Sen jälkeen vastaavat tiedot esitetään tehtävöivihkoittain ja kielitaidon osa-alueittain. Seuraavaksi kuvataan käytetty taitotasojen asettamismenettely ja esitetään reseptiivisen kielitaidon kokeiden tulokset sekä niiden sijoittaminen valitun taitotasojen asettamismenettelyn avulla CEF- ja OPS 204-taitotasolle osoittaen niiden vastaavuus ja oppilaiden kielitaidon taso näillä taitotasoilta arvioituna.

Lopuksi käsitellään produktiivisen kielitaidon kokeiden taitotasojen asettamismenettely, koetulokset CEF-, OPS 2004- ja A2001-taitotasoasteikoilla sekä näiden linkittäminen osoittamaan, missä määrin ne ovat vertailukelpoisia keskenään ja mikä oppilaiden taito oli näillä taitotasolla arvioituna.

9.1 Reseptiivisen kielitaidon kokeet

Ensimmäisessä tutkimuskysymyksessä kysyttiin: Millaisia ovat reseptiivisen kielitaidon (puheen ja tekstin ymmärtämisen ja kieliopillisen kompetenssin) kokeiden osiot klassisen ja uuden testiteorian menetelmillä analysoituina? Seuraavassa vastataan tähän kysymykseen.

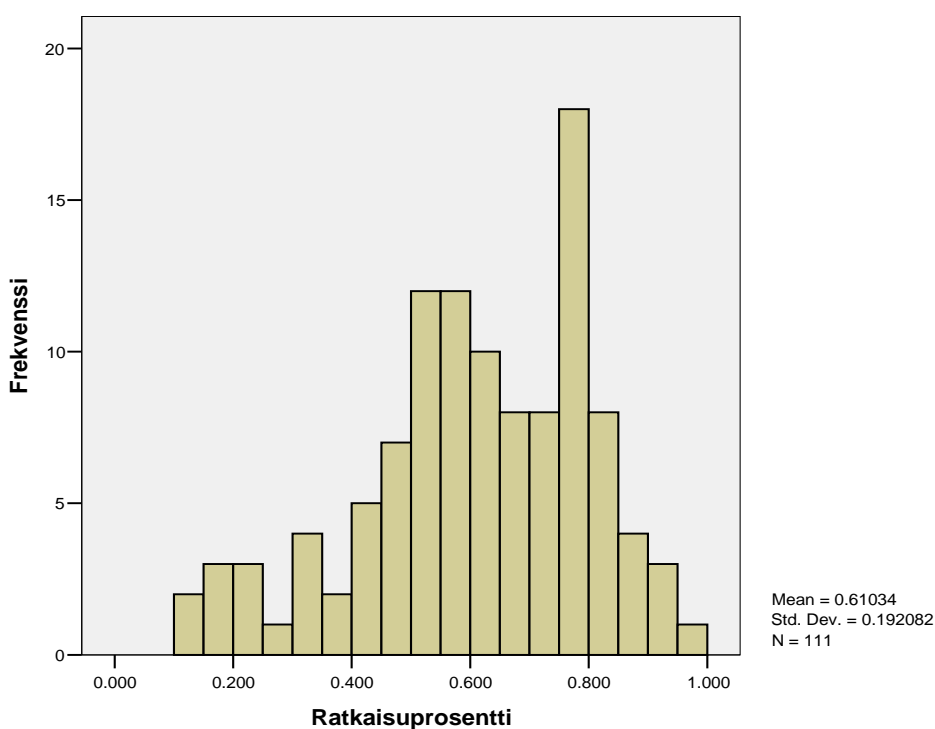
9.1.1 Koko kokeen vaikeusaste ja erottelukyky

Käsillä olevassa tutkimuksessa klassisen testiteorian analyysien mukainen reseptiivisen kielitaidon kokeen osioiden keskimääräinen vaikeustaso eli ratkaisuprosentti oli 61 ja uuden testiteorian mukainen θ -arvo odotetusti 0,0 (taulukko 61).

TAULUKKO 61 Reseptiivisen kielitaidon kokeen osioiden vaikeustaso ja erottelukyky.

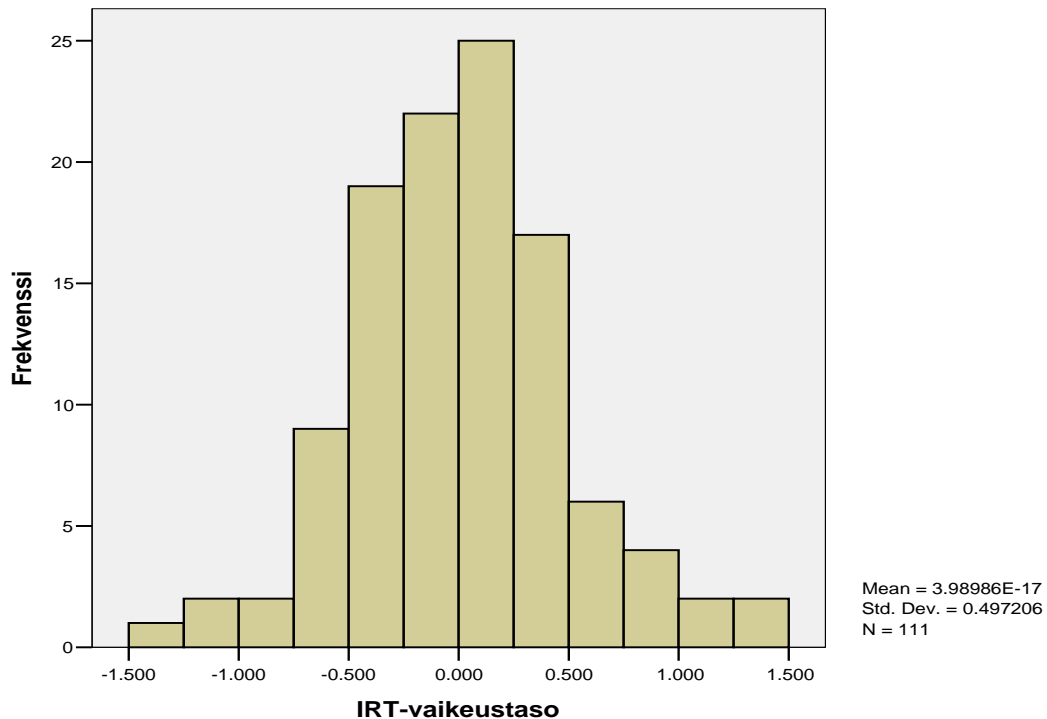
	Osioiden lukumäärä	Minimi	Maksimi	Keskiarvo	Keskihajonta
Vaikeustaso (P, %)	111	11,5	95,5	61,0	19,21
Eroittelukyky (D)	111	0,24	0,8	0,5	0,10
Vaikeustaso (θ , b)	111	-1,45	1,5	0,0	0,50
Eroittelukyky (a)	111	1	5	2,7	1,04

Kuvion 34 x-akselilla olevat arvot osoittavat osioiden vaikeustason jakauman klassisen testiteorian analyysien mukaisena. Y-akseli osoittaa havaintojen lukumäärän. Suurin osa vastauksista sijoittuu välille 40 - 80 %.



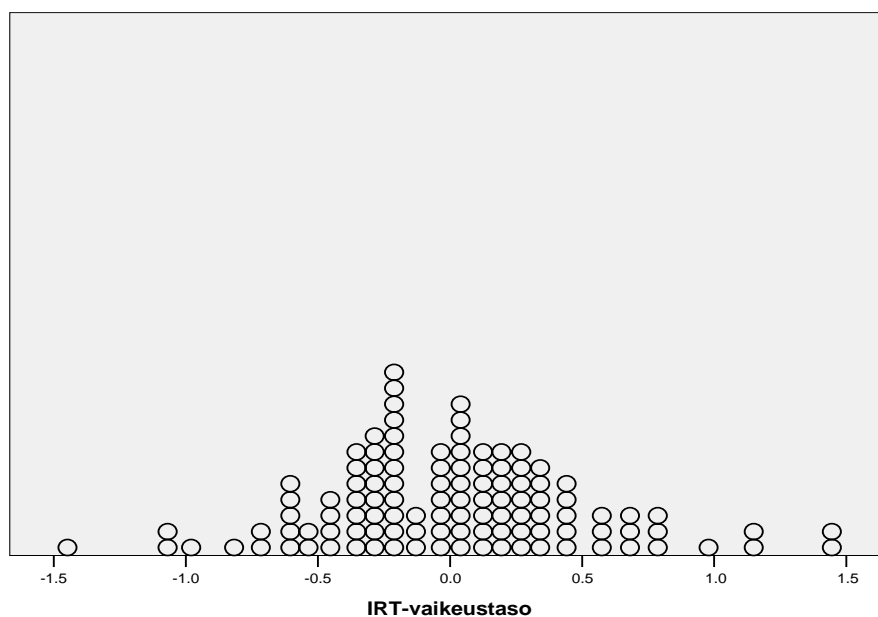
KUVIO 34 Reseptiivisen kielitaidon kokeiden ratkaisuprosenttien (P) jakauma.

Kuvio 35 osoittaa osioiden vaikeustason uuden testiteorian analyysien mukaisena. Suuri osa osioista sijoittuu melko tasaisesti välille -0,5 - + 0,5.



KUVIO 35 Reseptiivisen kielitaidon koeosioiden θ -arvojen jakauma (b).

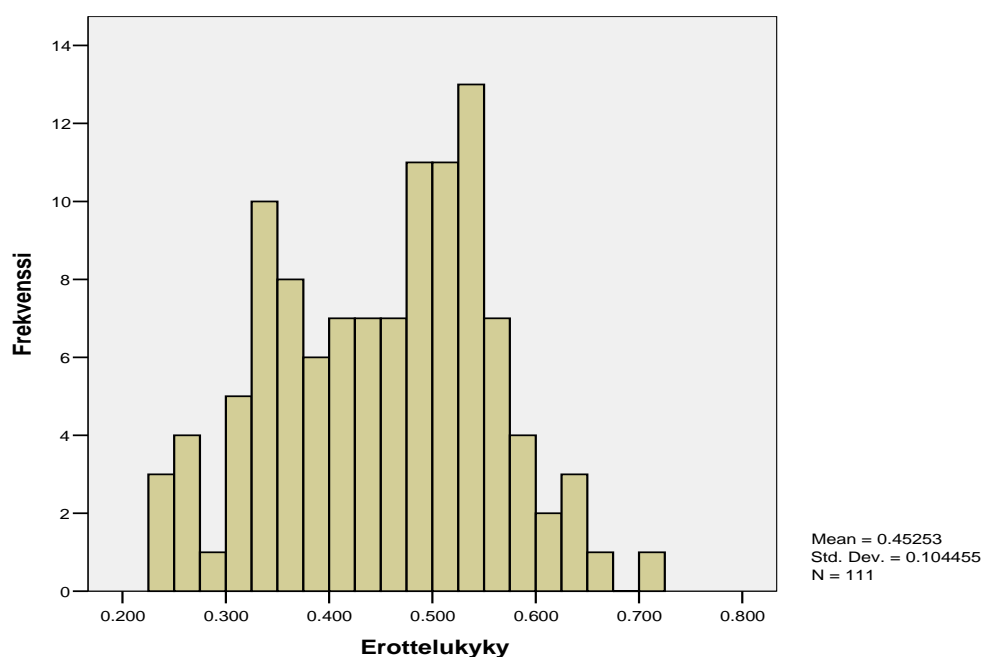
Kuvio 36 tarkoittaa IRT -analyysin tuottamaa tulkintaa reseptiivisen kielitaidon koeosioiden vaikeustasosta. x -akselilla plus-merkkiset arvot merkitsevät vaikeita ja miinus -merkkiset helppoja osioita. Vaikeus lisääntyy oikealle ja vähenee vasemmalle. Kukin ympyrä vastaa yhtä osiota, jonka vaikeustaso on IRT -analyysin tulos oppilaiden keskimääräisestä kyvystä ratkaista kyseinen osio. Se osoittaa myös osion vaikeustason. Pistekeskittymän -0,5 ja 0,5 ulkopuolelle jää sama määrä helppoiksi että vaikeiksi osoittautuneita osioita. Myöhempi kielitaidon osa-alueittain tehty tulosten tarkastelu osoittaa, miten vaikeustaso jakaantuu niiden välillä.



KUVIO 36 Reseptiivisen kielitaidon osiokohtainen θ -arvojen jakauma.

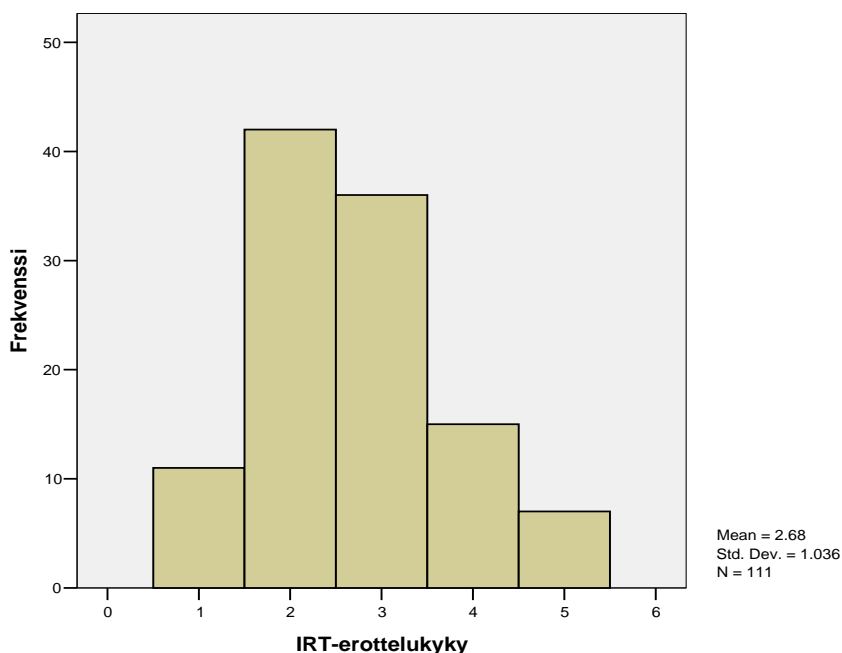
Kuvioita verrattaessa nähdään, että IRT -analyysi tuottaa vaikeustasojakauman, joka on selvästi lähempänä normaalijakaumaa kuin klassiseen testiteoriaan perustuva jakauma.

Kokeen keskimääräisen erottelukykyindeksi on klassisen testiteorian mukaan 0,45 ja uuden testiteorian mukaan 2,7, mikä osoittaa, että osiot erottelivat heikot oppilaat hyvistä keskimäärin hyvin. Kuvio 37 osoittaa osioiden erottelukyvyn klassisen testiteorian analyysien mukaan. Vaaka-akseli osoittaa erottelukykyindeksin ja pystyrivi osioiden lukumäärän. Valtaosa osioista sijoittuu erottelukyvyltään hyvin tai erinomaisesti erottelemaan ryhmään.



KUVIO 37 Reseptiivisen kielitaidon koeosioiden keskimääräinen erottelukykyindeksi en (D) jakauma.

Kuvio 38 osoittaa osioiden erottelukyvyn uuden testiteorian mukaan. Vaaka-akselilla näkyy erottelukyky ja pystyakselilla osioiden lukumäärä. Valtaosa osioista sijoittuu erottelukyvyltään arvojen 2-3 välille, mikä on normaali tulos OPLM -ohjelmistoa käytettäessä (luku 7). Myös erotteluindeksin kohdalla IRT -analyysin tulos on lähempänä normaalijakaumaa.



KUVIO 38 Reseptiivisen kielitaidon koeosioiden erottelukykyindeksien (a) jakauma.

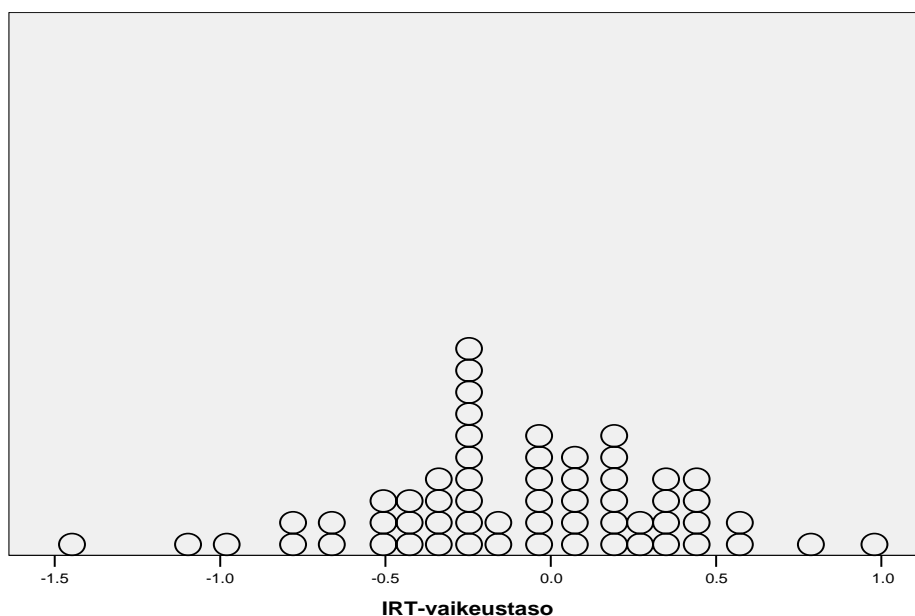
9.1.2 Eri tehtäväsarjojen osioiden vaikeustason ja erottelukyvyn vertailua

Koska jokaiseen kouluun toimitettiin kahta tehtäväsarjaa kolmesta, oli tärkeätä, että tehtäväsarjat olivat vaikeusasteeltaan ja erottelukyvyltään mahdollisimman samantasoisia. Taulukko 62 osoittaa pyrkimyksessä onnistutun melko hyvin.

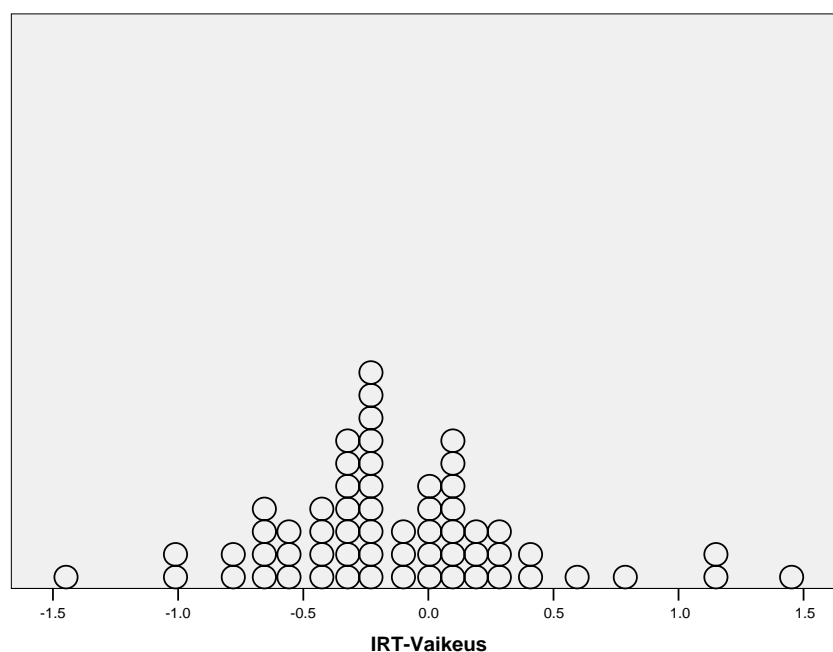
TAULUKKO 62 Reseptiivisen kielitaidon kokeen tehtäväsarjojen osioiden vaikeustason ja erottelukyvyn vertailu.

	Tehtäväsarja 1					Tehtäväsarja 2					Tehtäväsarja 3				
	Ratk %	Erott.-kyky	Theta	IRT-erott.-kyky	N.	Ratk %	Erott.-kyky	Theta	IRT-erott.-kyky	N.	Ratk %	Erott.-kyky	Theta	IRT-erott.-kyky	N.
N	60	60	60	60	60	61	61	61	61	61	68	68	68	68	68
Min.	23	0,24	-1,12	1		12	0,25	-1,45	1		12	0,26	-1,45	1	
Maks.	96	0,63	0,98	5		96	0,67	-1,45	5		96	0,71	1,44	5	
Ka	65	0,44	0,10	2,68		67	0,45	-0,13	2,72		64	0,45	-0,08	2,75	
Haj.	16,91	0,11	0,45	1,16		18,38	0,10	0,51	1,02		18,90	0,10	0,47	1,03	

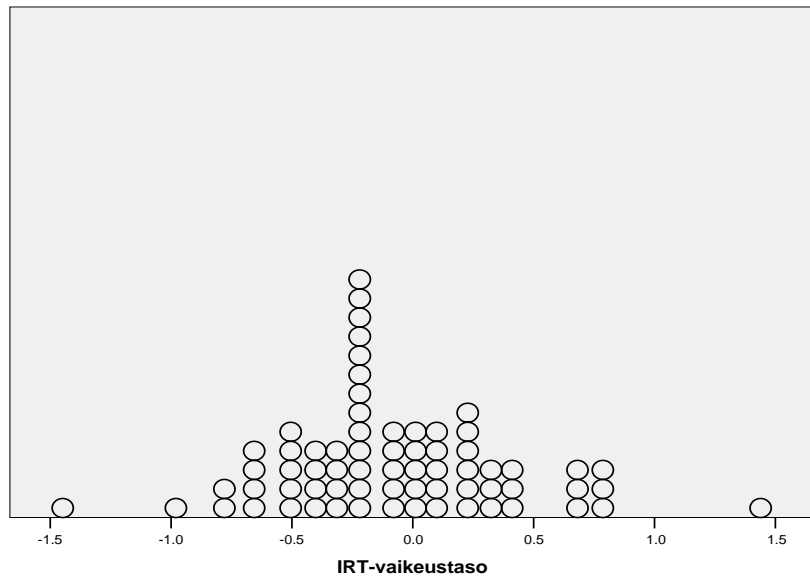
Kuviot 39, 40 ja 41 havainnollistavat eri tehtäväsarjoihin sisältyvien osioiden IRT -analyysin mukaista vaikeusastetta. Kaikissa tehtäväsarjoissa jakauma on hyvä. θ -arvojen $-0,5$, $+0,5$ välillä on ensimmäisen tehtäväsarjan osioista 82 %, toisen 72 % ja kolmannen 78 %.



KUVIO 39 Tehtäväsarjan 1 reseptiivisen kielitaidon koeosioden IRT -vaikeustasojakauma.

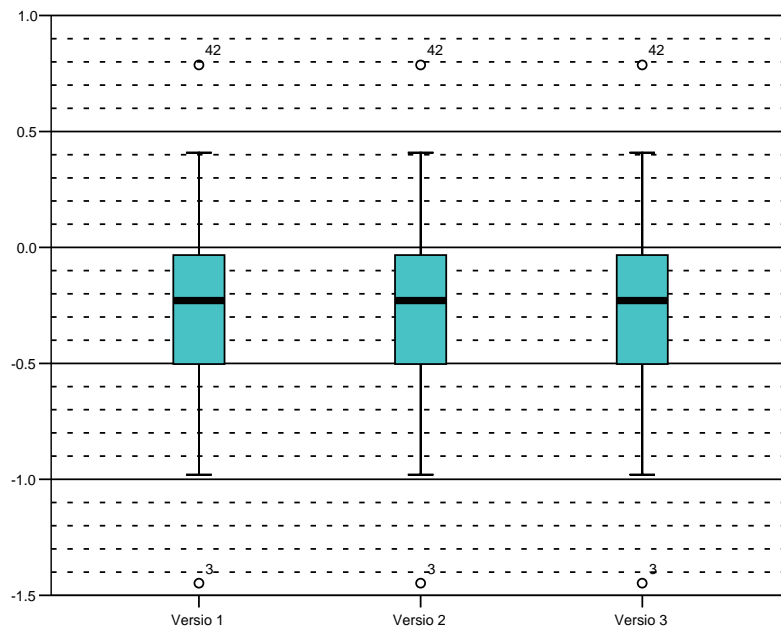


KUVIO 40 Tehtäväsarjan 2 reseptiivisen kielitaidon koeosioden IRT-vaikeustasojakauma.



KUVIO 41 Tehtäväsarjan 3 reseptiivisen kielitaidon koeosioiden IRT -vaikeustasojakauma.

Laatikko-jana -kuvaus¹ (kuvio 42) havainnollistaa edelleen sitä, että kaikki kolme tehtäväsarjaa ovat vaikeusasteeltaan hyvin samanlaiset. Kuviossa esitetään jakauman sijainti ja hajonta. Kuvaajassa laatikko sisältää kvartiilivälin eli 50 % havainnoista, ja sen keskellä oleva viiva kuvaa mediaania. Poikkeavat havainnot on merkitty kuvioon ympyröillä, ja janojen päiden väliin mahtuvat kaikki muut paitsi voimakkaimmin poikkeavat havainnot.



KUVIO 42 Reseptiivisen kielitaidon kokeiden tehtäväsarjojen vaikeustason vertailu.

¹ Kuvauksesta käytetään myös nimitystä laatikko - viiksi -kuvaus, engl. box plot.

Liitteestä 20 käyvät ilmi eri tehtäväsarjojen osiokohtaiset ratkaisuprosentit.

Taulukossa 63 esitetään reseptiivisen kokeen osakokeiden keskiarvot, -hajonnat ja alfa-reliabeliuskertoimet. Se osoittaa, että koko kokeen samoin kuin kieliopillisen kompetenssin kokeen reliabelius on varsin korkea, kun taas puheen ja tekstin ymmärtämisen ovat kohtalaisia. Syynä jälkimmäisten alhaisempaan reliabeliuteen on pääasiassa näiden kokeiden osioiden vähäinen määrä.

TAULUKKO 63 Eri tehtäväsarjojen ja niiden osakokeiden keskiarvo, -hajonta ja reliabelius.

Tehtävä-sarja	Oppilaat N	Koko koe			Puheen ymmärtäminen			Tekstin ymmärtäminen			Kieliopillinen kompetenssi		
		Ka	kh	α	ka	kh	α	ka	kh	α	ka	kh	α
1	1699	39,1	11,44	,925	7,5	2,79	,715	9,2	2,85	,714	22,4	7,05	,896
2	1585	40,6	12,07	,936	8,1	2,87	,758	10,7	3,19	,793	21,7	7,0	,894
3	1570	43,2	13,76	,943	9,3	4,36	,845	11,0	3,69	,822	22,9	6,9	,890

9.1.3 Kielitaidon eri osa-alueiden koeosioiden vaikeustason ja erottelukyvyn vertailua

Osiokohtaisten tarkastelujen yhteydessä heikot osiot tavallisesti poistetaan ja koetta tarkastellaan kelpollisten eli kalibroittujen osioiden pohjalta. Tässä arvioinnissa jouduttiin heikon erottelukyvyn vuoksi poistamaan vain kolme osiota, kaksi puheen ymmärtämisen ja yksi kieliopillisen kompetenssin osa-alueelta, mikä osoittaa että sekä tehtävien laadinta että esitestaustulosten tuloksena suoritettu osioiden valinta oli ollut laadukasta. Liitteessä 21 esitetään kaikkien kokeen osioiden IRT -erottelukykyindeksit, kunkin osion suorittaneiden oppilaiden lukumäärä, osion ratkaisuprosentti sekä kuinka moni vastasi kuhunkin tehtävään väärin (0) tai oikein (1).

Taulukossa 64 esitetään tiivistetyssä muodossa reseptiivisen kielitaidon osakokeiden osioiden lukumäärät erottelukykyindeksiryhmittäin klassisen ja uuden testiteorian mukaisten analyysien perusteella. Se osoittaa, että valtaosa osioista oli erottelukyvyltään hyviä tai erittäin hyviä. Parhaiten näyttävät erottelleen kieliopillista kompetenssia testaavat tehtävät, joista 60 % sijoittui erottelukyvyltään uuden testiteorian mukaan tässä arvioinnissa saavutettuihin kolmeen ylimpään luokkaan.

TAULUKKO 64 Reseptiivisen kielitaidon koeosioiden lukumäärät erottelukykyindeksi-ryhmittäin.

Klassinen testiteoria				Uusi testiteoria			
Erottelukyky- indeksiryhmät (D)	Osioiden lukumäärä			Erottelukyky- indeksiryhmät (a)	Osioiden lukumäärä		
	TY*	PY*	K*		TY*	PY*	K*
- 0,40	24	18	32	5	1	2	4
0,30 – 0,39	7	9	13	4	2	4	9
0,20 – 0,29	4	2	2	3	13	8	15
0,19 -				2	15	11	16
				1	4	4	3
Yhteensä	35	29	47		35	29	47

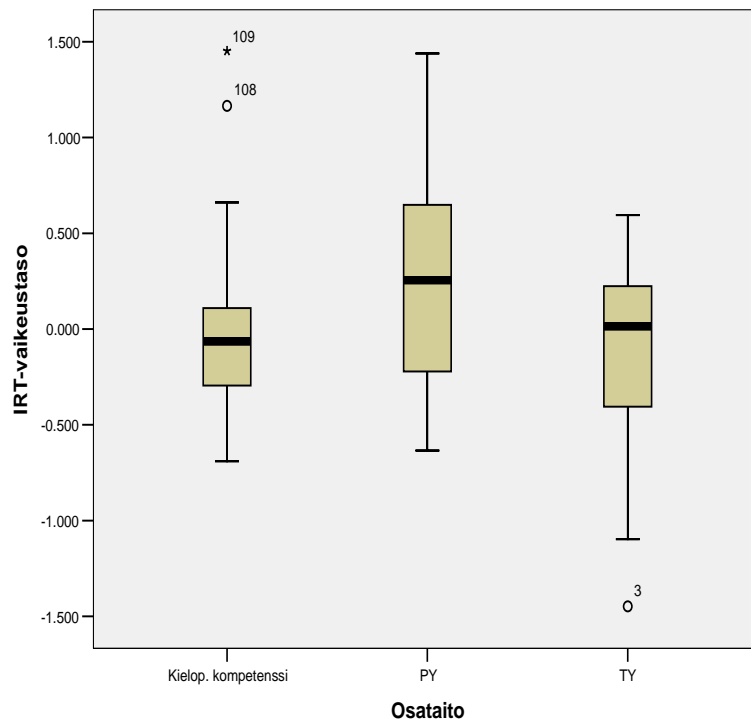
* TY = Tekstin ymmärtäminen PY = Puheen ymmärtäminen K = Kieliopillinen kompetenssi

TAULUKKO 65 Reseptiivisen kielitaidon kokeen ja sen osa-alueiden koeosioiden vaikeustaso ja erottelukyky

Kielitaidon osa-alue	Tunnusluvut	Klassinen testiteoria		Uusi testiteoria	
		Ratkaisuprosenttien keskiarvo (Osioiden vaikeustaso, P)	Erotteluindeksi (D)	Osioiden vaikeustaso (b)	Erotteluindeksi (a)
Kieliopillinen kompetenssi	Keskiarvo	64	0,46	-0,04	2,89
	Keskihajonta	17,75	0,10	0,43	1,07
	Minimi	0,12	0,25	-0,70	1
	Maksimi	90	0,64	1,45	5
	Osioiden N	47	47	47	47
Puheen ymmärtämi- nen	Keskiarvo	51	0,44	0,24	2,62
	Keskihajonta	22,03	0,10	0,52	1,12
	Minimi	12	0,24	-0,63	1
	Maksimi	94	0,59	1,44	5
	Osioiden N	29	29	29	29
Tekstin ymmärtämi- nen	Keskiarvo	65	0,45	-0,15	2,46
	Keskihajonta	15,70	0,12	0,50	0,89
	Minimi	33	0,24	-1,45	1
	Maksimi	96	0,71	0,60	5
	Osioiden N	35	35	35	35
Koko reseptiivinen kielitaito	Keskiarvo	61	0,45	0,00	2,68
	Keskihajonta	19,21	0,10	0,50	1,04
	Minimi	12	0,24	-1,45	1
	Maksimi	96	0,71	1,45	5
	Osioiden N	111	111	111	111

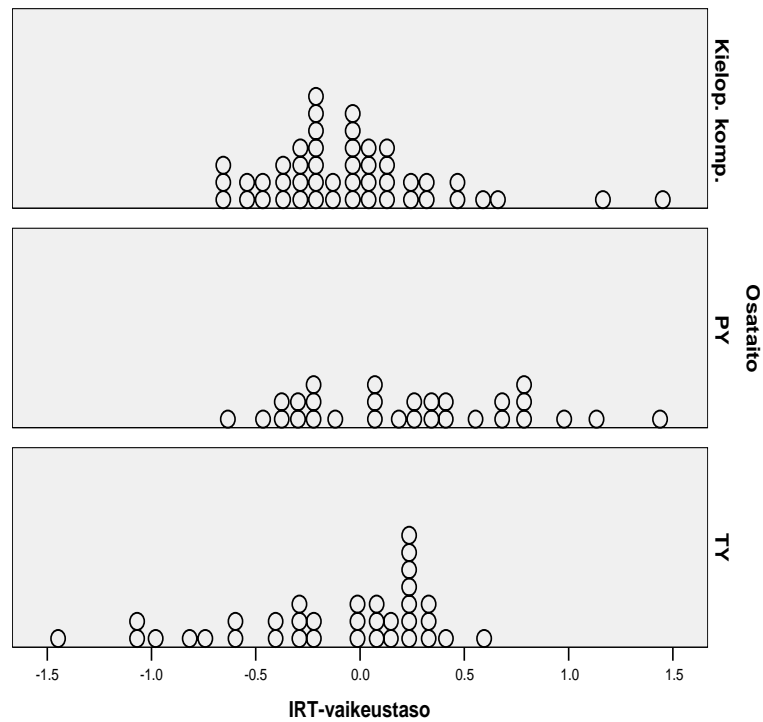
Verrattaessa kielitaidon eri osa-alueita mittaavien kokeiden vaikeustasoja (taulukko 65) voidaan todeta, että puheen ymmärtämisen koe on ollut reseptiivisiä taitoja mittaavista kokeista vaikein (keskimääräinen θ -arvo 0,24) ja tekstin ymmärtämisen helpoin (θ -arvo -0,15). Kieliopillisen kompetenssin kokeen keskimääräinen θ -arvo oli -0,04. Luvussa 7 selostetun θ -arvon määritelmän perusteella on siis osoitettu, että puheen ymmärtämisen kokeessa saavutettu tekstin ymmärtämisen ja kielitiedon kokeita heikompi tulos johtuu siis vaikeammasta kokeesta eikä esimerkiksi siitä, että testattavien taito olisi tällä osa-alueella heikompi.

Asiaa havainnollistaa laatikko-jana-kuvio 43, jossa on esitetty jakauman sijainti ja hajonta. Kuvio osoittaa, että eri osa-alueiden kokeiden välillä on jonkin verran eroa. Kun IRT -analyysissä plus-merkkiset arvot osoittavat vaikeita tehtäviä, kuvioista nähdään, että puheen ymmärtämisen kokeen (PY) mediaani jää tälle puolelle. Sen hajonta on myös suurin ja vaikeisiin osioihin viittaava jana pisin, kun taas tekstin ymmärtämisen kokeen laatikko on jakaumaltaan miltei päinvastainen.



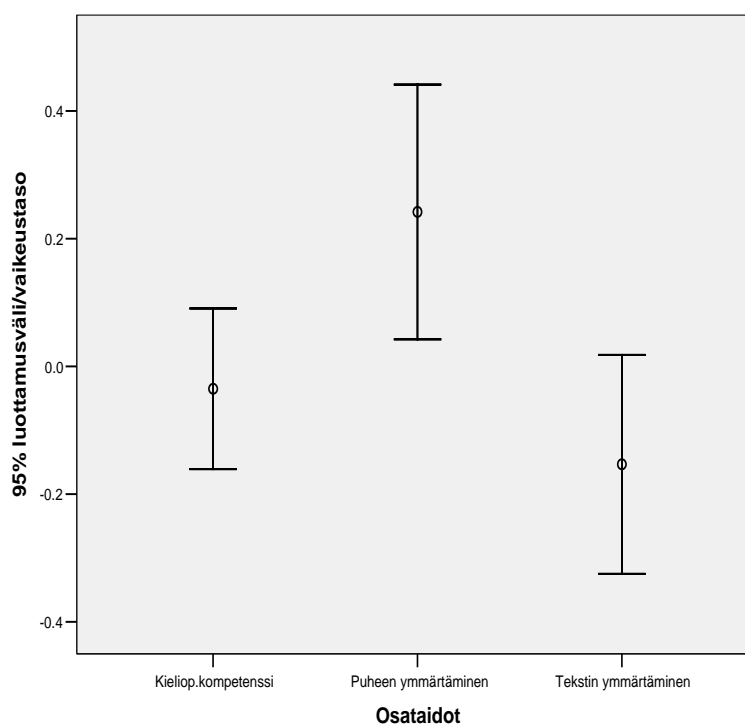
KUVIO 43 Reseptiivisen kielitaidon osa-alueiden kokeiden vaikeustason vertailu.

Kuvio 44 osoittaa, kuinka monta havaintoa kullekin θ -arvolle saatiin kielitaidon eri osa-alueiden kokeissa. Siitä myös näkyy, että puheen ymmärtämisen (PY) kokeessa on ollut suhteellisesti runsaammin vaikeita osioita (plus -merkkiset) kuin muissa osakokeissa.



KUVIO 44 Tekstin ja puheen ymmärtämisen sekä kielipillisen kompetenssin koeosioiden θ - arvot.

Luottamusväli ilmaisee, millä välillä perusjoukon todellinen tunnusluvun arvo on tietyllä todennäköisyydellä. Kuvion 45 tuloksen voidaan siis katsoa koskevan 95 %:n todennäköisyydellä kaikkia vuoden 1999 yhdeksäsluokkalaisia. Kuvioista näkyy, että puheen ja tekstin ymmärtämisen janat eivät osu kohdakkain (edellisen keskimääräinen θ -arvo noin +0.2, jälkimmäisen noin -0.2), joten näiden kokeiden vaikeustasot eroavat tilastollisesti merkitsevästi toisistaan. Ero on siis todellinen eikä johdu sattumasta.



KUVIO 45 Kieliopillisen kompetenssin , puheen ja tekstin ymmärtämisen kokeiden keskiarvot ja niiden 95 % luottamusväli.

Yhteenvedona ensimmäiseen tutkimuskysymykseen tuotetuista vastauksista voidaan esittää seuraavaa: 1) Reseptiivistä kielitaitoa mittavien osioiden laadun kannalta tärkeän ominaisuuden, erottelukyvyn, suhteen osiot osoittautuivat hyviksi, koska vain kolme vähän yli sadasta osiosta jouduttiin jättämään jatkoanalyysien ulkopuolelle, 2) aineisto osoittautui tilastolliselta kannalta (pääkomponenttianalyysissa) käytännössä yksidimensionaaliseksi ja muutenkin kelvolliseksi IRT -analyysien tekemiselle, 3) kuusi eri tehtäväsarjaa osoittautuivat riittävässä määrin vertailukelpoisiksi ja 4) kielitaidon eri osa-alueiden kokeista puheen ymmärtämisen koe osoittautui muita osakokeita vaativammaksi. Ensimmäiseen tutkimuskysymykseen saadut vastaukset loivat hyvän pohjan jatkoanalyysille.

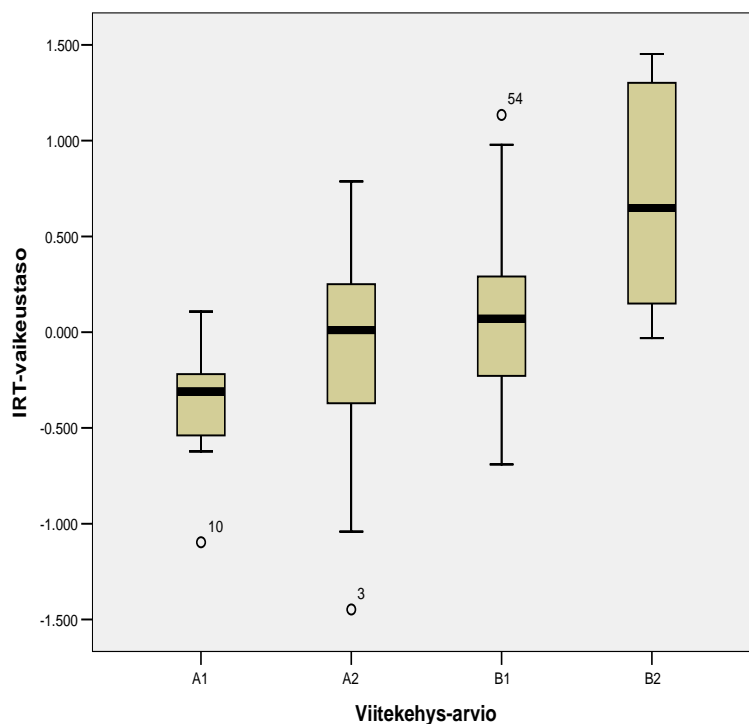
9.2 Reseptiivisen kielitaidon kokeiden taitotasojen asettaminen

Toisessa tutkimuskysymyksessä kysyttiin: Miten CEF- ja OPS 2004 -taitotasoasteikot vastaavat toisiaan reseptiivistä kielitaitoa arvioitaessa? Seuraavassa vastataan tähän kysymykseen.

9.2.1 Asiantuntijoiden suorittamien taitotasoarviointien tulokset CEF -asteikolle sijoitettuina

Tämän tutkimuksen tekijä ja taitotasojen arvioinnin asiantuntija arvioivat reseptiivisen kielitaidon kokeiden vaikeustason. CEF -taitotasoarvion korrelaatio reseptiivisen kielitaidon kokeiden empiiriseen tulokseen oli 0,348**, joka on verrattain alhainen, mutta vastaa niitä korrelaatioita, joita vastaavissa tutkimuksissa on saatu. Koetulosten sijoittamiseksi CEF -taitotasoasteikolle laskettiin asiantuntijoiden arvioiden ja empiiristen tulosten keskiarvo kullekin osiolle ja sijoitettiin ne vaikeusjärjestyksessä tasoille, käytännössä tasoille A1 - B2. Tulos ei tällaisenaan ollut tyydyttävä, koska tasojen A2 ja B1 välille ei saatu riittävää eroa, vaikka keskimääräinen osioiden vaikeusaste lisääntyikin toivotulla tavalla alemmalta tasolta ylemmälle siirryttäessä.

Yleensä yliarvioidaan monivalintatehtävien ja aliarvioidaan avotehtävien vaikeustaso (Sauli Takala, suullinen tiedonanto, 2.6.2005). Syynä tähän lienee se, että kokemusta ja vastaavasti myös palautetta tästä työstä puuttuu. Taitotasoarvioiden keskimääräinen IRT -vaikeustaso oli kuitenkin nouseva, kuten laatikko-jana -kuvio 46 osoittaa.



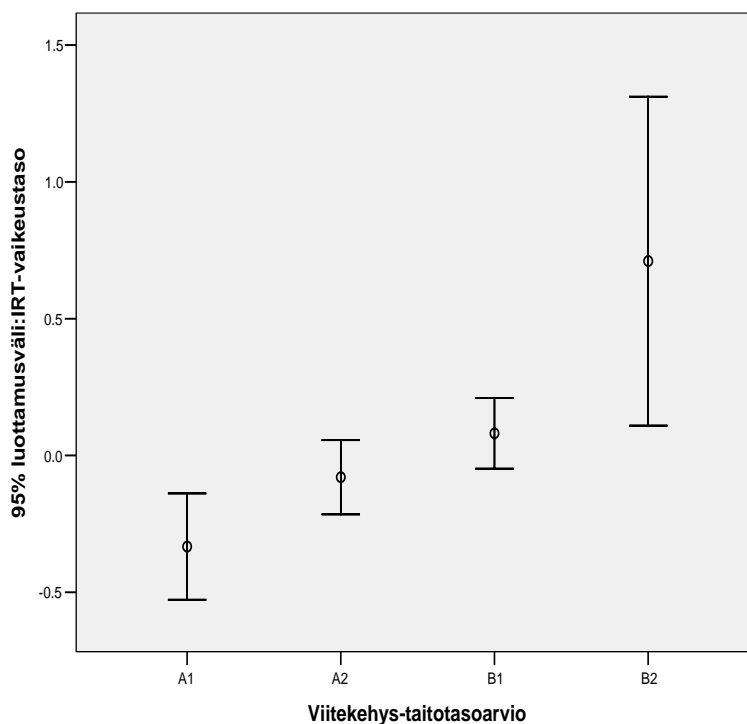
KUVIO 46 Asiantuntijoiden CEF -taitotasoasteikolla suorittama reseptiivisen kielitaidon koeosoiden vaikeustason arviointi.

Taulukko 66 osoittaa, että asiantuntijat olivat arvioineet suurimman osan osioista edustavan CEF -taitotasoja A2 ja B1.

TAULUKKO 66 Asiantuntijoiden arvioima koeosoiden jakaantuminen CEF -taitotasasteikoilla.

CEF - taitotaso	Osioiden lukumäärä	Minimi	Maksimi	Keskiarvo	Keskihajonta	Keskiarvon keskivirhe
A1	13	-0,1	0,1	-0,3	0,32	0,1
A2	50	-0,5	0,8	0,1	0,48	0,1
B1	41	-0,7	1,1	0,9	0,41	0,1
B2	7	0,0	1,1	0,7	0,6	0,2
Yhteensä	111	-1,4	1,5	0,0	0,5	0,0

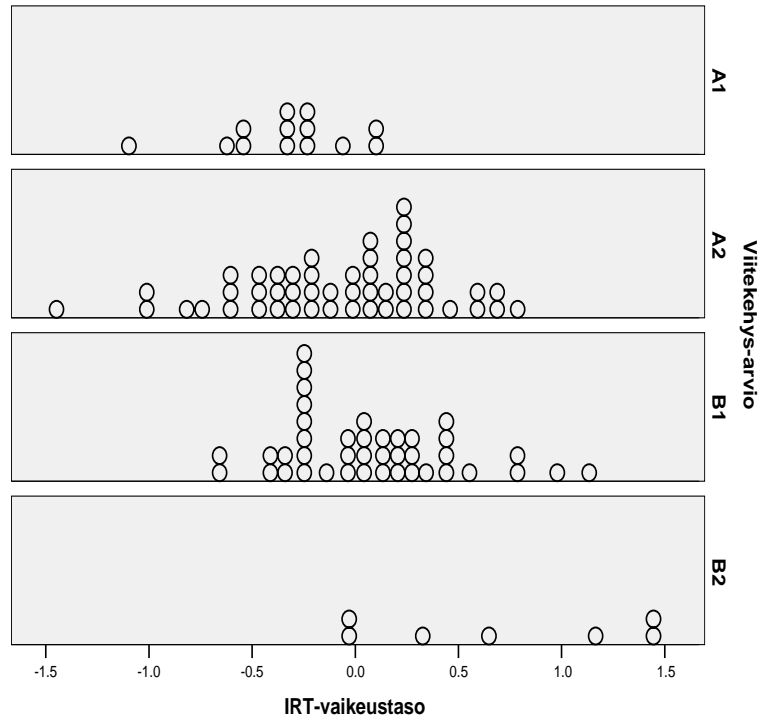
Myös kuvio 47 osoittaa, että kun keskimääräiselle viitekehysarviointitulokselle on laskettu 95 %:n luottamusväli, taitotasoa osoittavat janat ovat osittain samoilla kohdilla eivätkä siis erottele taitotasoa. Näin on erityisesti taitotasojen A2 ja B1 kohdalla. B2-tason arvioiden suuri luottamusväli johtuu todennäköisesti tälle tasolle arvioitujen osioiden vähäisestä määrästä ja tähän luonnostaan liittyvästä arvioinnin reliaaabeliuden puutteesta.



KUVIO 47 Reseptiivisen kielitaidon kokeen osioiden CEF- taitotasoarvioiden keskiarvot ja 95 % luottamusväli.

Yksittäisten osioiden sijoittumista eri taitotasolle asiantuntijoiden arvion perusteella havainnollistaa kuvio 48. Arviointi on ollut sillä tavoin oikean suuntainen, että käytännöllisesti katsoen kaikki B2-taitotasolle arvioidut osiot ovat saaneet IRT -vaikeustasoarvokseen korkeamman arvon kuin 0,0, ja A1-tasoisiksi arvioidut osiot jäävät tämän keskimääräisen tason alapuolelle. B1-tasoisiksi arvioiduista osioista suurin osa sijoittuu IRT -vaikeustasovälille $-0,5 - + 0,5$, mutta näin sijoittuvat myös A2-taitotasolle arvioidut osiot, jopa niin, että hieman suu-

rempi osuus niistä edustaa vaikeita tehtäviä. Tulos havainnollistaa myös vaikeustason arvioinnin ongelmallisuutta sekä useaan otteeseen mainittua tarvetta koulutukseen, jota tutkimusten mukaan varsin kokoneetkin arvioijat vielä kaipaavat.



KUVIO 48 Reseptiivisen kielitaidon kokeiden yksittäisten koeosioiden sijoittuminen CEF-taitotasolle asiantuntijoiden arvioinnin perusteella.

Tarkistettu taitotasojen määrittäminen CEF -asteikolla

Kullekin taitotasolle laskettiin asiantuntijoiden arvioiden ja empiiristen koetulosten mukaan niille asettuneiden osioiden keskimääräinen IRT -vaikeustaso ja sijoitettiin nämä tulokset taitotasolle A1 – B2. Taitotasojen välyksiksi muodostuivat taulukon 67 θ -arvot.

TAULUKKO 67 Reseptiivisen kielitaidon kokeiden CEF -taitotasoasteikon tarkistetut välykset (θ).

CEF-taitotaso	θ -arvojen välyys
A1	<-0,30
A2	-0,30 - 0
B1	0 - +0,30
B2 -	> +0,30

Kun tällä perusteella laskettiin raakapisteiden kumulatiivinen summa kullakin tasolla ja muunnettiin raakapisteet θ -asteikkoon, saatiin seuraavat reseptiivisen kielitaidon kokeen CEF -taitotasojen katkaisukohtat:

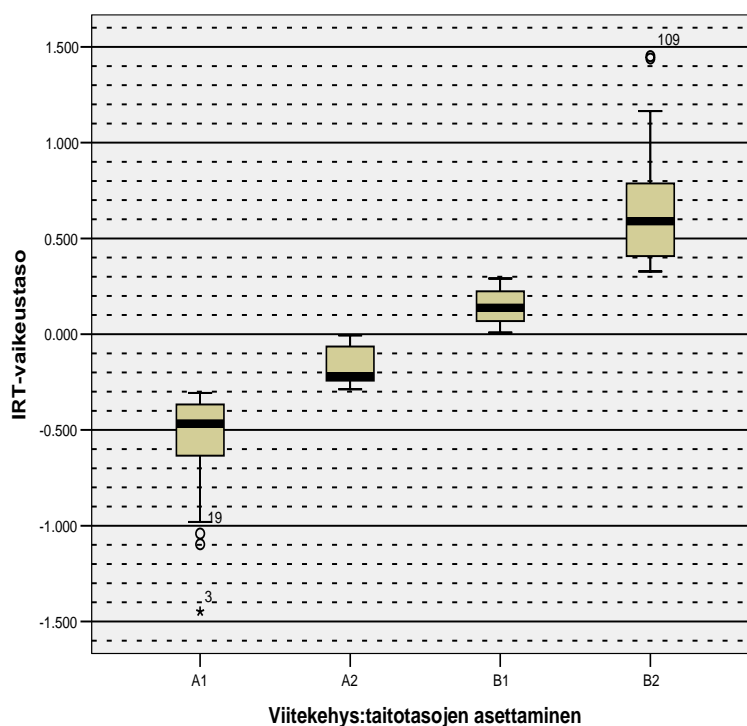
< =A1	-0,513	A2	-0,012	B1	+0,562	= >B2
-------	--------	----	--------	----	--------	-------

Tämän tarkistetun taitotasojen määrittämisen jälkeen osioiden sijoittuminen CEF -taitotasoille muodostui taulukon 68 mukaiseksi.

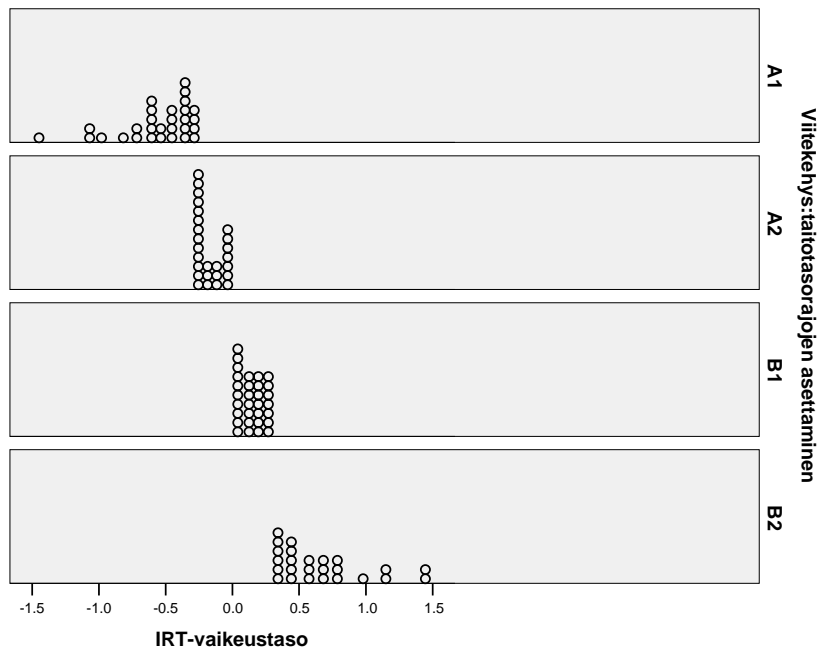
TAULUKKO 68 Reseptiivisen kielitaidon koeosioiden tarkistettu CEF -taitotaso (θ).

CEF -taitotaso	Osioiden lukumäärä	Minimi	Maksimi	Keskiarvo	Keskihajonta	Keskiarvon keskivirhe
A1	29	-1,4	-0,3	-0,6	0,28	0,1
A2	26	-0,3	-0,0	-0,2	0,10	0,0
B1	31	0,0	0,3	0,1	0,09	0,0
B2	25	0,3	1,5	0,7	0,34	0,1
Yhteensä	111	1,4	1,5	0,0	0,50	0,0

Sekä laatikko-jana-kuvio 49 että kuvio 50 osoittavat havainnollisesti sen, että osiot ryhmittyvät taitotasojen asettamisen tuloksena melko kauniisti omille taitotasoilleen, kaikki osiot sijaitsevat θ -arvojen -1,5 ja + 1,5 välillä ja että suurin osa osioista on IRT -vaikeustasoltaan välillä -0,5 ja +0,5.

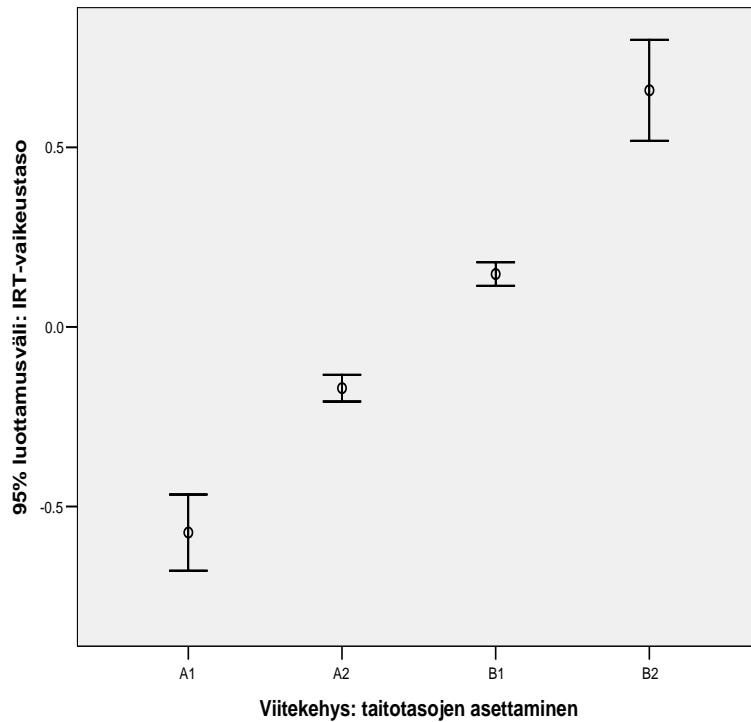


KUVIO 49 CEF -taitotasoille asettuneiden osioiden tarkistettujen θ -arvojen jakaumat.



KUVIO 50 Reseptiivisen kielitaidon kokeen yksittäisten koeosioiden tarkistettu jakautuminen CEF -taitotasoille.

Kun keskimääräiselle tulokselle on laskettu 95 %:n luottamusväli, voidaan todeta, että kaikki janat ovat eri tasoilla, joten vaikeusasteiden erot ovat todellisia. Lisäksi vaikeusaste (taitotasoarvio) nousee tasaisesti helpommista osioista vaikeampiin siirryttäessä (kuvio 51). Tulokset ovat tutkimuksen laatua tukevia.



KUVIO 51 Tarkistetun CEF -taitotasojen määrittämisen jälkeen eri taitotasoille asettuneiden reseptiivisen kielitaidon koeosioiden θ -keskiarvot ja 95 % luottamusväli.

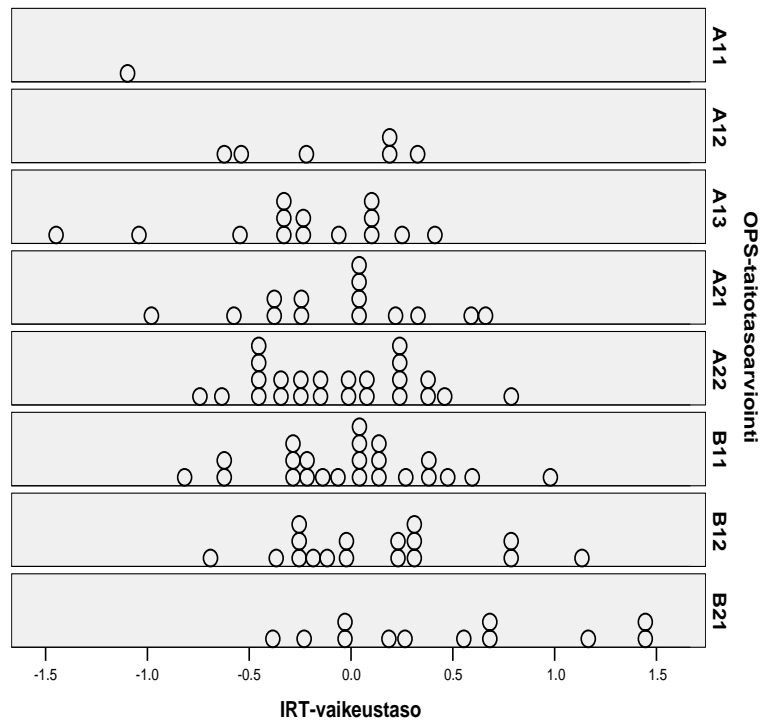
9.2.2 Asiantuntijoiden suorittamien taitotasarviointien tulokset OPS 2004 -asteikolle sijoitettuina

Reseptiivisen kielitaidon koeosoiden linkittäminen OPS 2004:n taitotasoihin tehtiin samalla menetelmällä kuin linkittäminen CEF -taitotasoihinkin. Asiantuntijoiden arvioiden korrelaatio osioiden empiiriseen vaikeustasoon oli alempi kuin CEF -linkittämisessä, 0,294, mikä on kuitenkin tilastollisesti merkitsevä korrelaatio. Tiheämpiportainen asteikko ei siis tukenut arvioinnin tarkkuutta. Taulukko 69 osoittaa OPS 2004:n asiantuntijoiden eri taitotasoihin sijoittamien osioiden lukumäärän, θ -arvojen väläyksen, niiden keskiarvon, keskihajonnan ja keskiarvon keskivirheen.

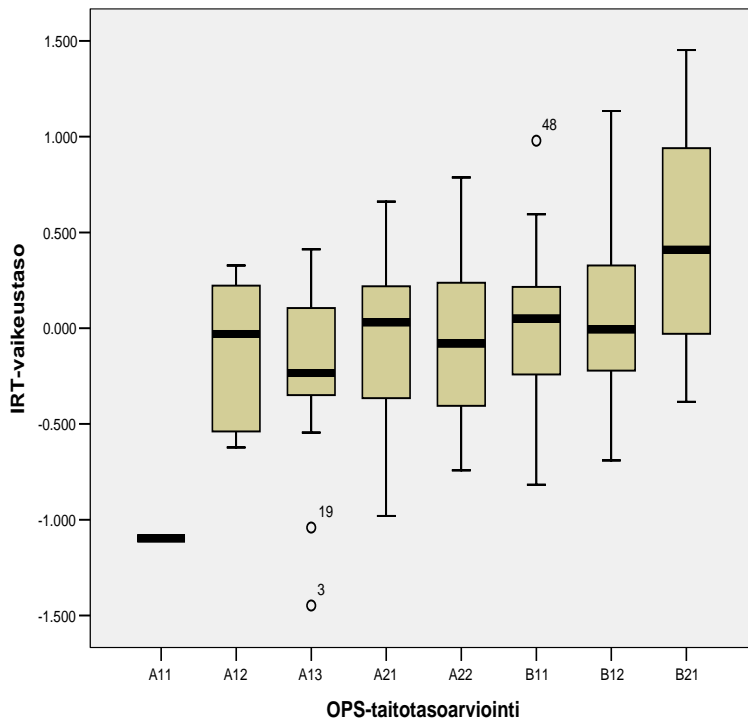
TAULUKKO 69 Asiantuntijoiden OPS 2004 -taitotasoasteikon mukaan arvioimat reseptiivisen kielitaidon koeosoiden jakaumat.

OPS 2004 - taitotasot	Osioiden lukumäärä	Minimi	Maksimi	Keskiarvo	Keskihajonta	Keskiarvon keskivirhe
A1.1	1	-1,1	-1,1	-1,1	.	.
A1.2	6	-0,6	0,3	-0,1	0,41	0,2
A1.3	14	-1,4	0,4	-0,3	0,50	0,1
A2.1	14	-1,0	0,7	-0,1	0,45	0,1
A2.2	24	-0,7	0,8	-0,1	0,39	0,8
B1.1	23	-0,8	1,0	0,0	0,42	0,1
B1.2	17	-0,7	1,1	0,1	0,47	0,1
B2.1	12	-0,4	1,5	0,5	0,62	0,2
Yhteensä	111	-1,4	1,5	0,0	0,50	0,5

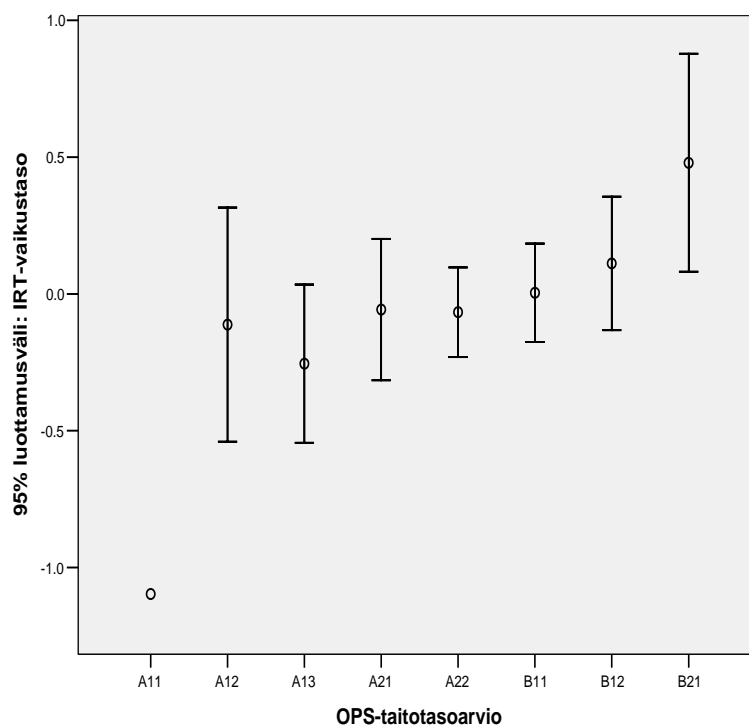
Kuten kuvioista 52 ja 53 havaitaan, tulos ei ole tyydyttävä, koska arvioitu vaikeustaso ei kohoa johdonmukaisesti empiirisen θ -arvion mukaisesti. Kuvio 54 osoittaa keskimääräiselle tulokselle lasketun 95 %:n luottamusvälin, mistä näkyy, että jakaumat menevät päällekkäin. Taitotasojen uudelleen asettaminen jouduttiinkin tekemään harkinnanvaraisesti, ja näin ollen tulokset ovat vain suuntaa antavia.



KUVIO 52 Asiantuntijoiden OPS 2004-taitotasoarvioiden mukainen reseptiivisen kokeen yksittäisten osioiden sijoittuminen taitotasoille.



KUVIO 53 Asiantuntijoiden OPS 2004-taitotasoarvioiden mukainen reseptiivisen kokeen osioiden sijoittuminen eri taitotasoille.



KUVIO 54 Asiantuntijoiden arvioiden mukaiset OPS 2004 -taitotasojen θ -keskiarvot ja niiden 95 % luottamusväli.

Tarkistettu taitotasojen määrittäminen OPS 2004 -asteikolla

Taitotasojen uudelleen asettamisen periaatteena oli, että CEF- ja OPS 2004-taitotasot olisivat vertailukelpoisia ja että kunkin tason osioiden lukumäärä olisi lähellä normaalijakaumaa. Jokaiselle tasolle pyrittiin saamaan riittävästi osioita, mikä takaisi mahdollisimman luotettavan piste- tai tasoarvioinnin. Osioiden keskimääräisen vaikeuden eri tasoilla toivottiin olevan mahdollisimman samanlaisen kuin asiantuntijoiden arvioissa, mikä ei kuitenkaan ollut usein mahdollista. Kullekin taitotasolle laskettiin taaskin asiantuntijoiden arvioiden ja empiiristen koetulosten mukaan sille asettuneiden osioiden keskimääräinen IRT -vaikeustaso ja osiot sijoitettiin uudelleen OPS 2004-taitotasolle. Näin saatiin taitotasojen välyksiksi taulukon 70 mukaiset arvot.

TAULUKKO 70 Reseptiiviseen kielitaidon kokeiden OPS 2004 -taitotasojen tarkistettut välykset (θ).

OPS 2004 - taitotaso	θ -arvojen välyys
A1.1	≤ -1
A1.2	-1 - -0,7
A1.3	-0,7 - -0,4
A2.1	-0,4 - -0,3
A2.2	-0,3 - 0
B1.1	0 - 0,2
B1.2	0,2 - 0,5
B2.1	0,5 - 1,5
B2.2	$> 1,5$

Kun tällä perusteella laskettiin raakapisteiden kumulatiivinen summa kullakin tasolla ja muunnettiin raakapisteet θ -asteikollisiksi, saatiin seuraavat reseptiivisen kielitaidon taitotasojen katkaisukohdat:

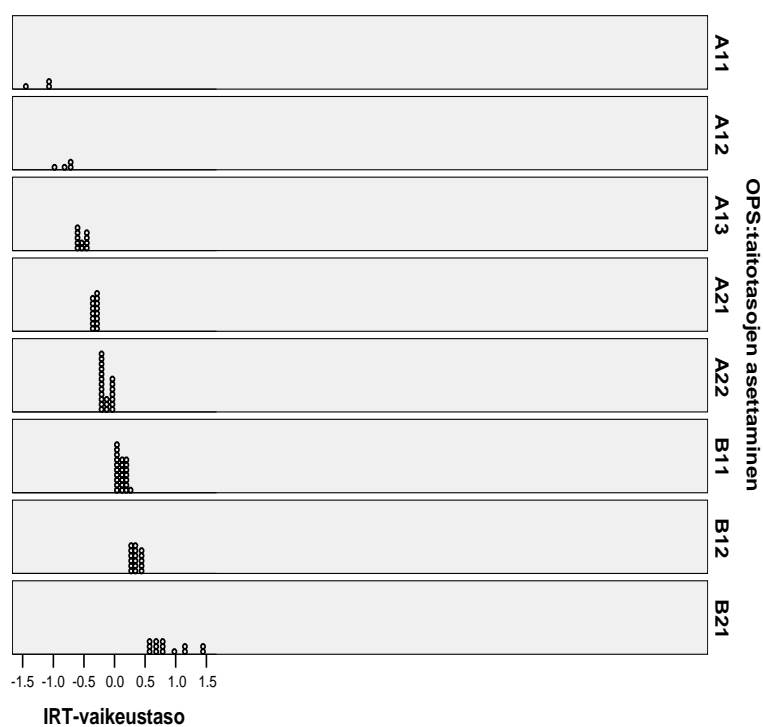
A1.1		A1.2		A1.3		A2.1		A2.2		B1.1		B1.2		B2.1
	-1,8		-1,4		-0,8		-0,4		-0,0		+0,4		+0,9	

Tarkistetun taitotasojen määrittämisen jälkeen osioiden sijoittuminen OPS 2004 -taitotasoille muodostui taulukon 71 mukaiseksi.

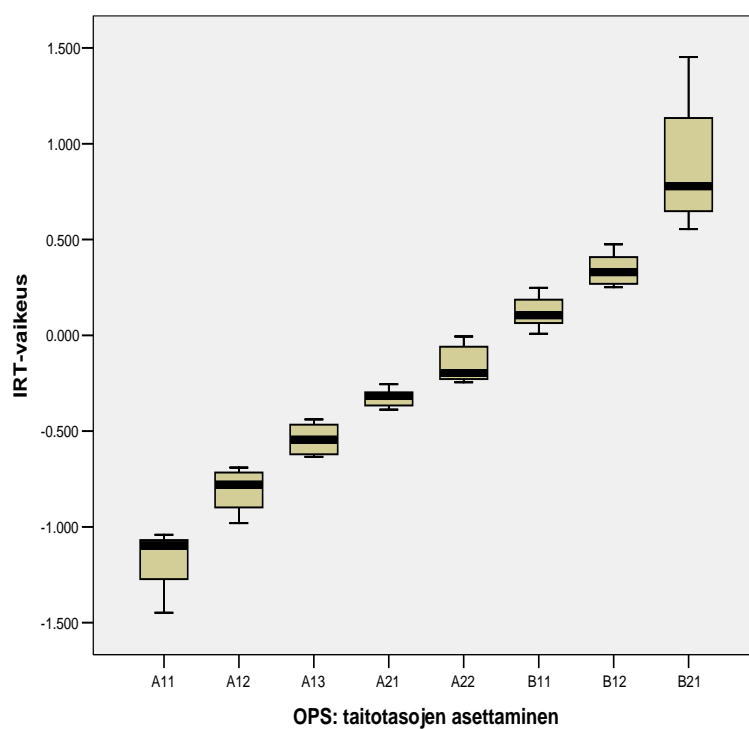
TAULUKKO 71 Reseptiivisen kielitaidon koeosioiden tarkistettu OPS 2004 -taitotaso-asteikko (θ).

OPS 2004-asteikko	Osioiden lukumäärä	Minimi	Maksimi	Keskiarvo	Keskihajonta	Keskiarvon keskivirhe
A1.1	3	-1,5	-1,0	-1,2	0,22	0,13
A1.2	4	-1,0	-0,7	-0,8	0,13	0,1
A1.3	11	-0,6	-0,4	-0,5	0,08	0,0
A2.1	15	-0,4	-0,3	-0,3	0,04	0,0
A2.2	22	-0,2	-0,0	-0,2	0,09	0,0
B1.1	25	0,0	0,2	0,1	0,07	0,1
B1.2	17	0,3	0,5	0,3	0,07	0,1
B2.1	14	0,6	1,5	0,9	0,31	0,1
Yhteensä	111	-1,4	1,5	0,0	0,50	0,0

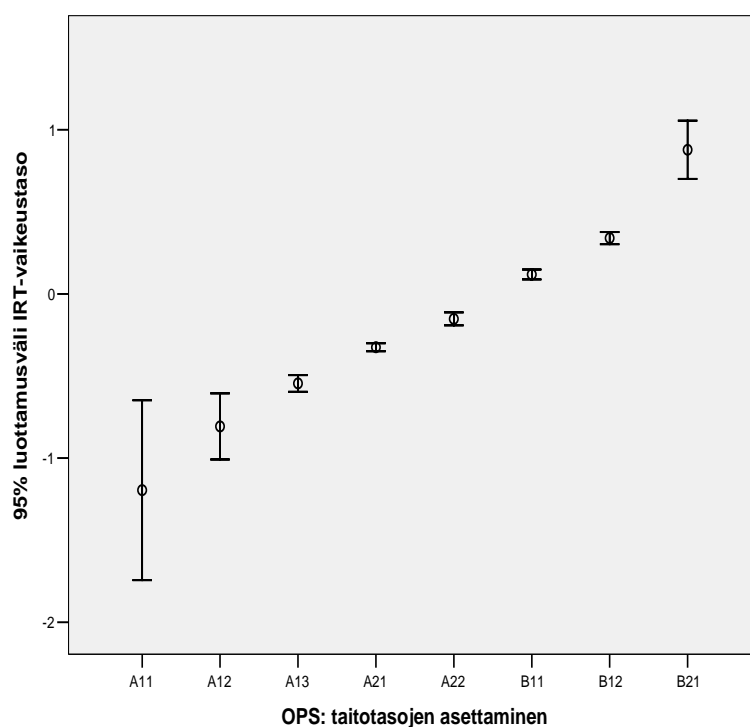
Edellä kuvattujen menettelyjen tulosta havainnollistavat kuviot 55, 56 ja 57, joissa osioiden vaikeustaso lisääntyy taitotasolta toiselle siirryttäessä. Tuloksen tulkinnassa on edelleen korostettava sitä, että osioita on vaikea sijoittaa luotettavasti OPSin eri taitotasoille.



KUVIO 55 Reseptiivisen kielitaidon kokeen yksittäisten osioiden tarkistettu jakaantuminen OPS 2004 -taitotasoille.



KUVIO 56 Eri OPS 2004 -taitotasoille asettuneiden osioiden tarkistetut θ -arvojen jakaumat.



KUVIO 57 Tarkistetun OPS 2004-taitotasojen määrittämisen jälkeen eri taitotasolle asetuneiden reseptiivisen kielitaidon koeosioiden θ -keskiarvot ja niiden 95 % luottamusväli.

On syytä edelleen korostaa, että OPS -pohjainen taitotasokuvaus on tuntuvasti heikommalla pohjalla kuin CEF -pohjainen arviointi, ja se on vain alustava arvio. Hienojakoisen OPS -asteikon käyttö edellyttää CEFiäkin enemmän koulutusta sekä osioiden laatijoille että taitotason arvioijille.

Reseptiivisen kielitaidon kokeiden taitotasovastaavuus muodostui edellä esitettyjen toimenpiteiden seurauksena taulukon 72 mukaiseksi.

TAULUKKO 72 CEF- ja OPS 2004 -taitotasosteikkojen vastaavuus reseptiivisen kielitaidon kokeita arvioitaessa.

OPS 2004		CEF	
Taitotaso	θ -arvot	Taitotaso	θ -arvot
A1.1	-1,8	A1	-0,5
A1.2	-1,4		
A1.3	-0,8		
A2.1	-0,4	A2	-0,0
A2.2	-0,0		
B1.1	+0,4	B1	+0,6
B1.2	+0,9		

Taulukkoon 72 on lihavoitu ne θ -arvot, jotka ovat lähinnä vastaavia verrattaessa CEFin harvempaa ja OPS 2004:n tiheämpää asteikkoa toisiinsa. Tulos osoittaa, että asteikot näyttävät varsin vertailukelpoisilta: A2-tasoa vastaava θ -arvo on molemmilla taitotasosteikoilla -0,0. Tasot A2.1 (-0,4) ja A1 (0,5) ovat hyvin lähellä toisiaan, mikä oli havaittavissa myös vaikeustasoja arvioitaessa. CEFin B1 -tason katkaisukohta taas asettuu jotakuinkin OPS 2004 -asteikon tasojen

B1.1 (+0,4) ja B1.2 (+0,9) väliin. Tietyin varauksin voidaan kuitenkin verrata CEF:in ja OPS-2004:n välisiä tuloksia.

Asiantuntijoiden OPS 2004-asteikkoon perustuvat tulkinnat oppilaiden kielitaidosta korreloivat empiirisiin tuloksiin heikommin kuin CEF -asteikkoa käytettäessä. Tässä tutkimuksessa ongelmallisimmaksi osoittautui tasojen A1.1. ja A1.2 erottaminen toisistaan. Sama koskee osin myös tasoa A1.3. Pääsyyinä tähän lienee se, että näille tasoille sijoittuu vain suhteellisen vähäinen määrä oppilaita, koska perusopetuksen päättövaiheen odotustason keskiarvo on korkeampi, B1.1. Lisäksi arvioijat eivät olleet kovin harjaantuneita ala-asteen oppilaiden arviointiin, mitä alimpien tasojen määrittäminen edellyttäisi. Ebelin (1969) mukaan tulokset ovat sitä luotettavampia mitä useampia taitotasojä käytetään. Tässä tutkimuksessa vaikeustasojen lukumäärien keskinäinen epäsuhta vaikeutti taitotasojen erottamista. Syy voi tietysti olla, että taitotasokuvauksissa ei ilmaista riittävän selvästi, mikä erottaa tasot toisistaan. Toinen mahdollisuus on, että taitotasokuvausten ilmauksia ei tulkita kuvauksen laatijan tarkoittamalla tavalla, mikä paranee ainakin osittain koulutuksella.

Yhteenvetona voidaan toisesta tutkimuskysymyksestä todeta, että 1) asiantuntijoiden arviot reseptiivisen kielitaidon kokeen vaikeustasosta korreloivat heikosti empiiristen tulosten kanssa, 2) CEF -asteikolla taitotasot A2 ja B1 osoittautuivat vaativimmiksi määrittää, 3) tarkistettu taitotason määrittäminen CEF -asteikolla onnistui jokseenkin luotettavasti, kun taas OPS 2004 -asteikon taitotasorajat ovat vain suuntaa antavia, ja että 4) CEF- ja OPS 2004 -taitotasoasteikot saatiin jokseenkin yhteismitallisiksi.

9.3 Reseptiivisen kielitaidon kokeiden tulokset taitotasoittain

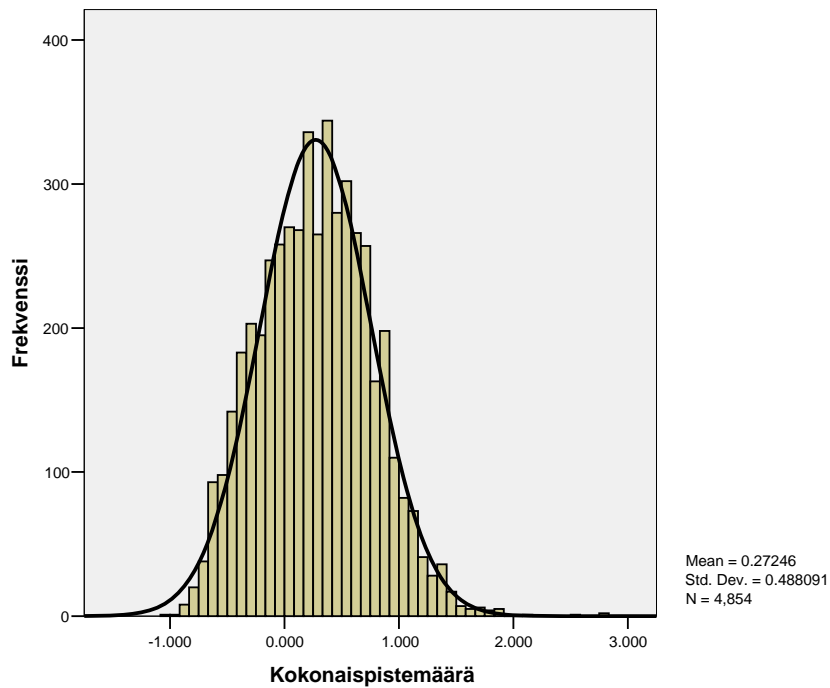
Kolmannessa tutkimuskysymyksessä kysyttiin: Mikä perusopetuksen päättövaiheen oppilaiden englannin kielen reseptiivinen kielitaito on CEF- ja OPS 2004 -taitotasoasteikoilla ilmaistuna? Seuraavassa vastataan tähän kysymykseen.

Taulukossa 73 esitetään IRT -analyysiin perustuvat koko reseptiivisen kielitaidon kokeen tulokset samoin kuin siihen sisältyvien kielitaidon osalueiden kokeiden tulokset.

TAULUKKO 73 Koko reseptiivisen kielitaidon kokeen ja sen osakokeiden tulokset (θ).

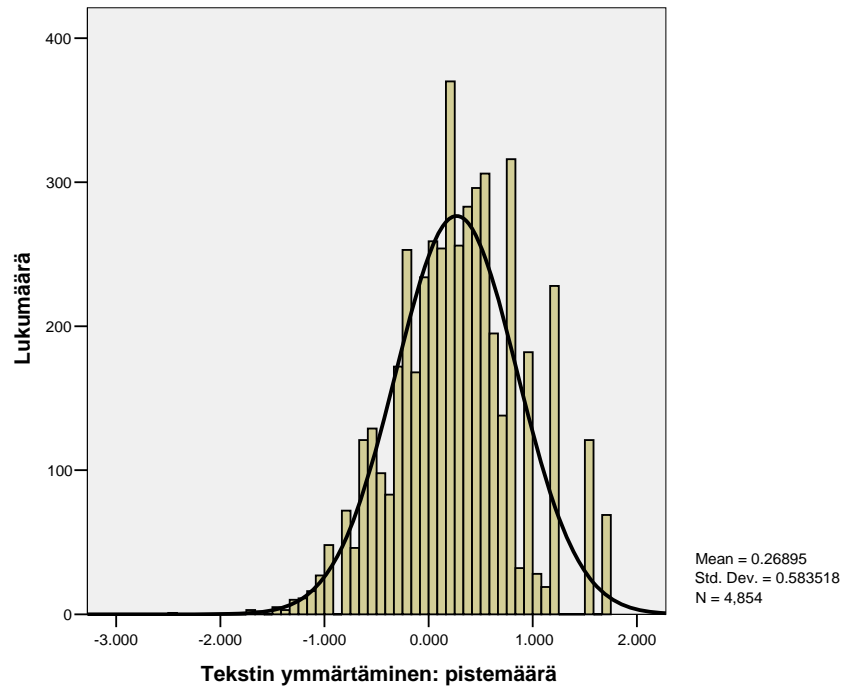
Koko koe	N	Minimi	Maksimi	Keskiarvo	Keskihajonta
	4854	-1,0	2,8	0,27	0,49
Tekstin ymmärtäminen	4584	-2,4	1,7	0,27	0,58
Puheen ymmärtäminen	4854	-1,5	2,1	0,28	0,56
Kieliopillinen kompetenssi	4854	-2,2	2,4	0,28	0,52
N	4854				

Koko kokeen keskimääräinen θ -arvo oli 0,27. Kuvio 58 osoittaa, kuinka tasaisesti θ -arvot ovat jakautuneet arvon 0 molemmille puolille.



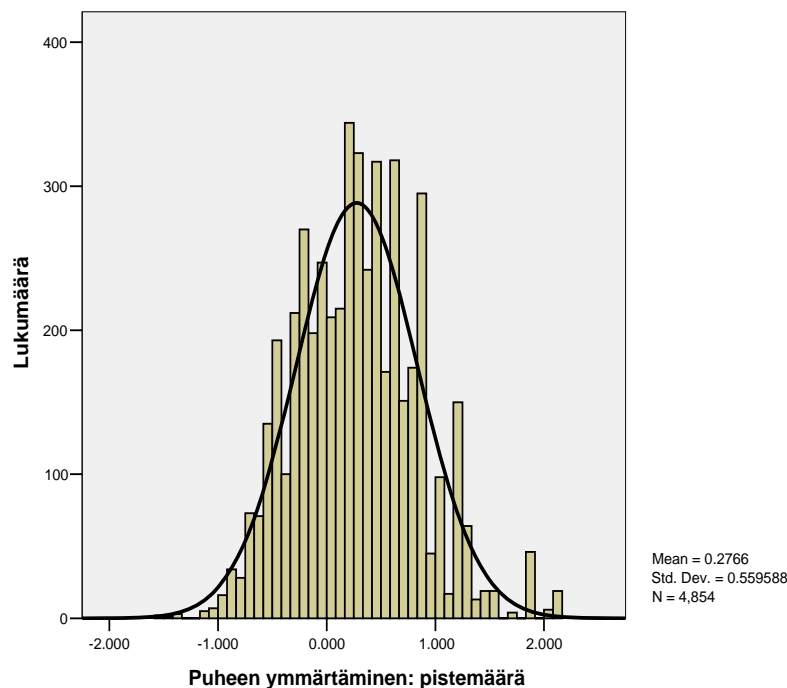
KUVIO 58 Koko reseptiivisen kielitaidon kokeen tulos (θ).

Kuten koko reseptiivisen kielitaidon kokeessa, myös tekstin ymmärtämisen kokeessa valtaosa θ -arvoista oli välillä -1,0 - +2 (kuvio 59), mutta koko kokeen kuvio on huipukkaampi.



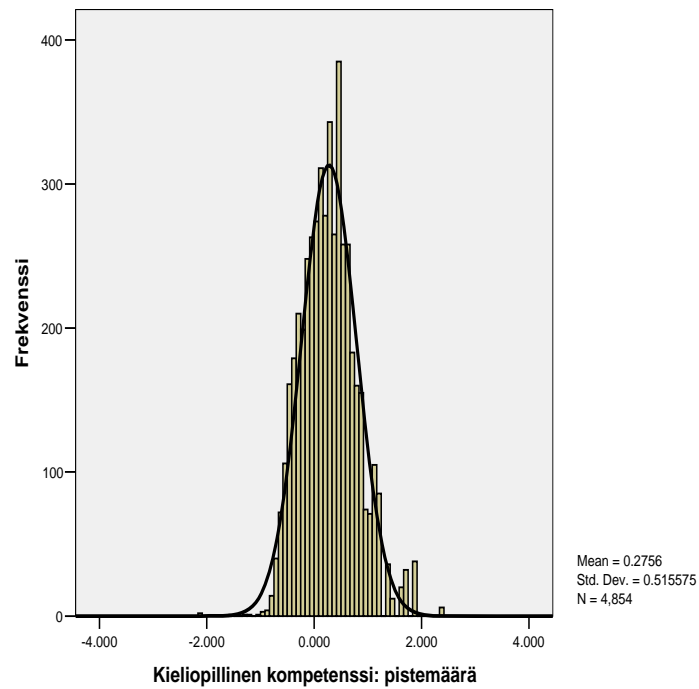
KUVIO 59 Tekstin ymmärtämisen kokeen tulos (θ).

Puheen ymmärtämisen kokeen IRT -keskiarvo oli samoin 0,27 ja kuvio 60 muodoltaan hyvin samankaltainen kuin tekstin ymmärtämisen tuloksia kuvaava kuvio. Tekstin ymmärtämisen kokeen tuloksiin verrattuna heikkoja arvoja on vähemmän ja hyviä arvoja enemmän.



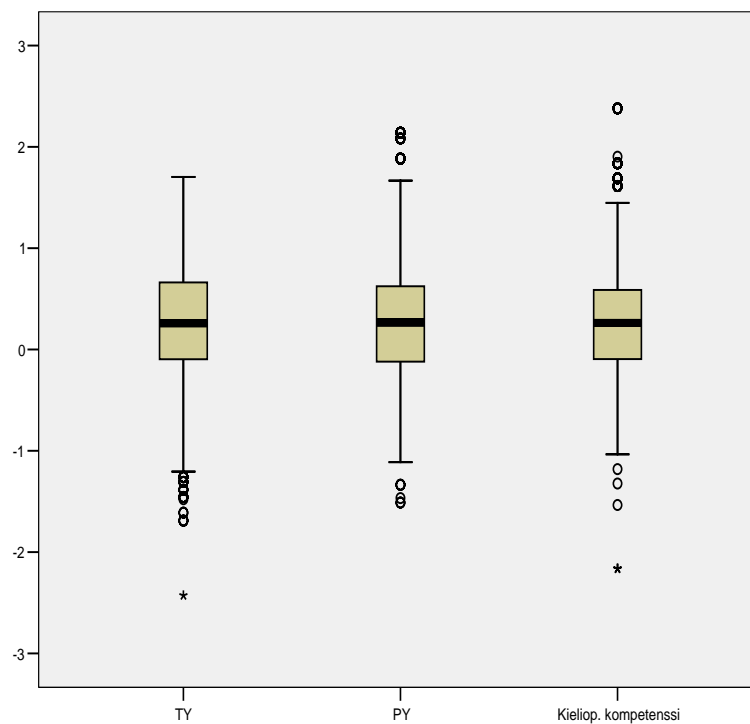
KUVIO 60 Puheen ymmärtämisen kokeen tulos (θ).

Kieliopillisen kompetenssin kokeessa oppilaat suoriutuivat samantasoisesti kuin muissakin reseptiivistä kielitaitoa mittaavissa kokeissa eli sen θ -keskiarvo oli 0,28 (kuvio 61). Valtaosa θ -arvoista sijoittuu välille -0,50 - +0,1. Kuvio noudattelee normaalijakaumaa, kuitenkin siten, että keskiarvon ympärille on kerääntynyt runsaasti suorituksia



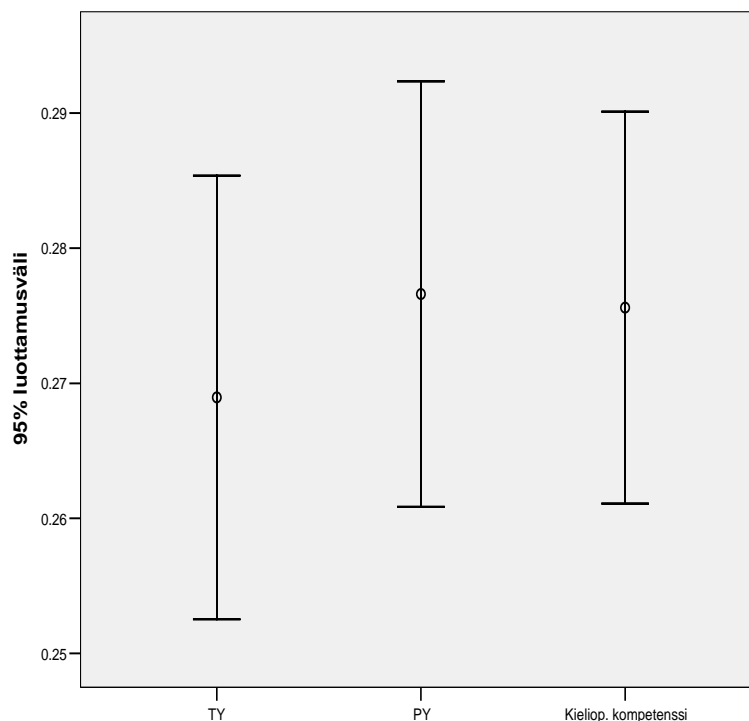
KUVIO 61 Kieliopillisen kompetenssin kokeen tulos (θ).

Laatikko-jana -kuvio 62 osoittaa havainnollisesti, kuinka samantasoisia reseptiivisen kokeen osakokeet olivat toisiinsa verrattuina.



KUVIO 62 Reseptiivisen kielitaidon osakokeiden tulosten (θ) vertailu.

Kun keskimääräiselle θ -tulokselle on laskettu 95 %:n luottamusväli, havaitaan, että tässäkin suhteessa osakokeet ovat toisiinsa verrattuina samanlaisia, koska niiden kuviot ovat miltei rinnakkaisia (kuvio 63).



KUVIO 63 Reseptiivisen kielitaidon osakokeiden keskiarvot (θ) ja niiden 95 % luottamusväli.

TAULUKKO 74 Koko reseptiivisen kielitaidon kokeen ja sen osakokeiden korrelaatiot ja niiden merkitsevyyden testaus (2-suuntainen, N = 4854).

Kokonais-pistemäärä	Kokonais-pistemäärä	TY*-pistemäärä	PY*-pistemäärä	K*-pistemäärä
	1	0,832**	0,870**	0,941**
		0,000	0,000	0,000
TY*-pistemäärä	0,832**	1	0,679**	0,702**
		0,000	0,000	0,000
PY*-pistemäärä	0,870**	0,679**	1	0,735**
		0,000	0,000	0,000
K*-pistemäärä	0,941**	0,702**	0,735**	1
		0,000	0,000	0,000

** Korrelaatio on merkitsevä tasolla 0,01

*TY = tekstin ymmärtämisen koe, PY = puheen ymmärtämisen koe, K = kieliopillisen kompetenssin koe

Taulukon 74 tulokset osoittavat, että tekstin ymmärtämisen, puheen ymmärtämisen ja kieliopillisen kompetenssin kokeet korreloivat välillä .68 - .74, mikä on kohtalainen ja jopa korkeahko korrelaatio ja korrelaatiot ovat luonnollisesti myös tilastollisesti merkitsevät. Korrelaatiot merkitsevät, että kullakin kokeella voidaan "ennustaa" (korottamalla korrelaatio toiseen potenssiin = determinaatikerroin) puolet (50%) muiden kokeiden tuloksen vaihtelusta.

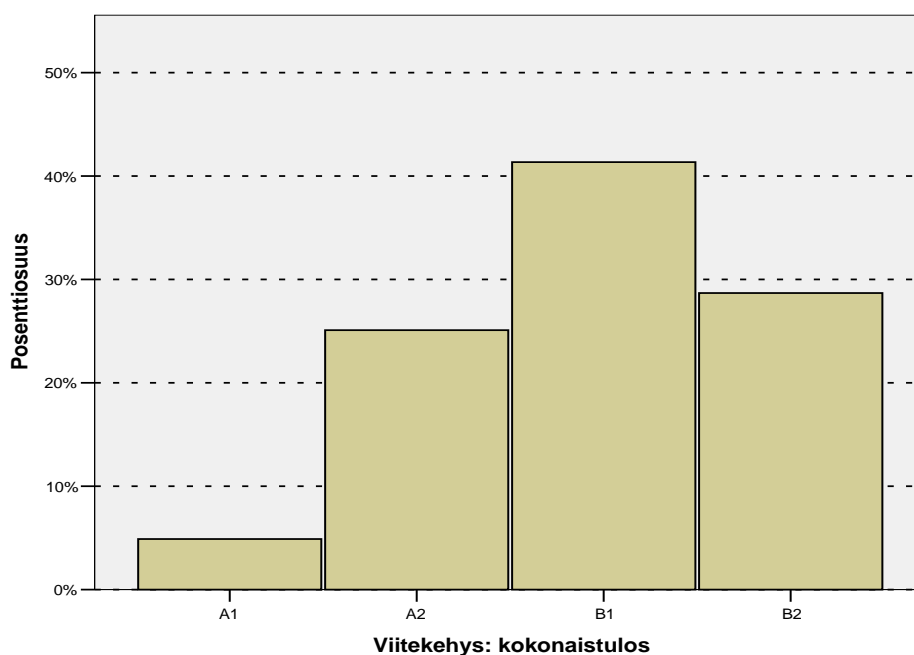
Kun taitotasojen katkaisukohtat on määritetty, tulokset voidaan ilmaista taitotasoina. Ne ovat järjestysasteikollisia, joten keskilukuna voidaan käyttää mediaania. Taulukko 75 osoittaa, että CEF -asteikolla mediaani on B1 ja OPS 2004-asteikolla B1.1. Samat arvot ovat myös moodeja eli taitotasolla useimmin esiintyviä arvoja. Taulukko osoittaa myös, että heikoimpien oppilaiden taso CEF -asteikolla on A1 ja OPS 2004 -asteikolla A1.1 – A1.3. Vastaavat arvot parhailla oppilailla ovat B2 ja B1.2. Alakvartiilin (heikoimmin menestyneen 25 %:n ryhmä) mediaanit ovat A2 ja A2.2, kun taas parhaat yltävät keskimäärin em. tasoille B2 ja B1.2.

TAULUKKO 75 Reseptiivisen kielitaidon kokeen CEF- ja OPS 2004 -taitotasoasteikkojen mediaanit, moodit ja kvartiilit.

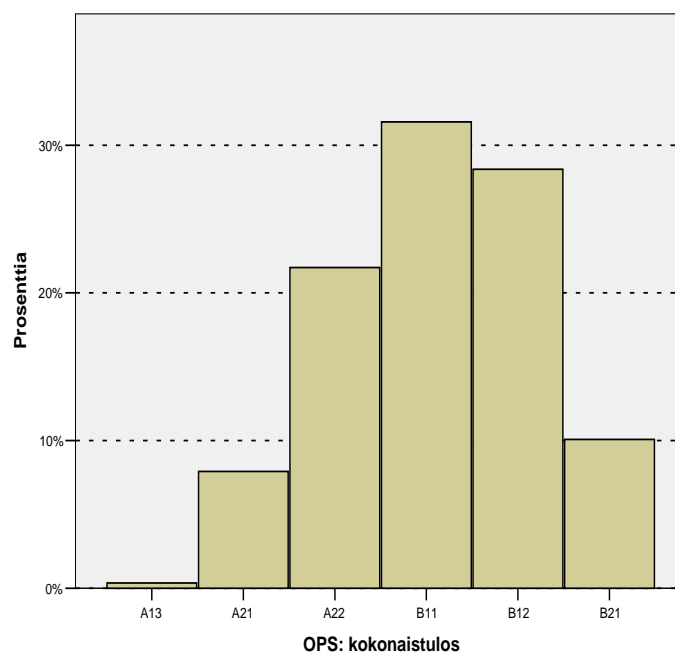
Koe	Oppilaiden lukumäärä		Mediaani	Moodi	Minimi	Maksimi	Kvartiilit		
	Validi	Kato					25	50	75
CEF, koko koe	4854	787	B1	B1	A1	B2	A2	B1	B2
CEF, tekstin ymmärtäminen	4854	787	B1	B1	A1	B2	A2	B1	B2
CEF, puheen ymmärtäminen	4854	787	B1	B1	A1	B2	A2	B1	B2
CEF, kieliopilliset rakenteet	4854	787	B1	B1	A1	B2	A2	B1	B2
OPS 2004, koko koe	4854	787	B1.1	B1.1	A1.3	B2	A2.2	B1.1	B1.2
OPS 2004, tekstin ymmärtäminen	B1.2	787	B1.1	B1.1	A1.1	B2	A2.2	B1.1	B1.2
OPS 2004, puheen ymmärtäminen	4854	787	B1.1	B1.1	A1.2	B2	A2.2	B1.1	B1.2
OPS 2004, Kieliopilliset rakenteet	4854	787	B1.1	B1.1	A1.1	B2	A2.2	B1.1	B1.2

9.3.1 Koko koe

Kuten kuvio 64 osoittaa, suuri osa oppilaista (noin 40 %) sijoittui CEF -tasolle B1. Tasolle A1 sijoittui vajaa 5 % ja tasolle B2 miltei 30 % oppilaista.



KUVIO 64 Koko reseptiivisen kielitaidon kokeen tulos CEF -taitotasosteikolle sijoitettuna.



KUVIO 65 Koko reseptiivisen kielitaidon kokeen tulos OPS 2004 -taitotasosteikolle sijoitettuna.

Kuvio 65 osoittaa reseptiivisen kokeen tuloksen OPS 2004-taitotasojen mukaisena. Noin 60 %, oppilaista sijoittui tasoille B1.1 ja B1.2. Noin 10 % oppilaista saavutti tason B2.1. Käytännöllisesti katsoen kaikki oppilaat olivat kuitenkin vähintään tasolla A2. Kuten edellä on korostettu, tulosta on taitotasoluokittelun pulmista johtuen pidettävä vain alustavana arviona.

Koko kokeen θ -arvo oli 0,27, joka edellä esitetyllä CEF-taitotasolla merkitsee korkeata B1-tasoista taitoa. OPS 2004- taitotasoasteikkoa sovellettaessa tulos on B1.1.

Saavutetun reseptiivisen kielitaidon tason havainnollistamiseksi esitetään taulukossa 76 kielitaidon yleiskuvaus CEF -asteikon B1-tasolla. Sen vieressä on vertailun vuoksi OPS 2004:n kuvaus siitä, mitä oppilaan tulisi osata 9. vuosiluokan päättyessä, minkä on tarkoitus vastata odotettua keskimääräistä tavoitetasoa B1.1. Suomen opetussuunnitelman perusteisiin 2004 ei ole sisällytetty taitotasoittaista kielitaidon yleiskuvausta.

TAULUKKO 76 Kielitaidon yleiskuvaus CEF -asteikon B1-tasolla ja OPS 2004:n kuvaus perusopetuksen päättövaiheen kielitaidosta.

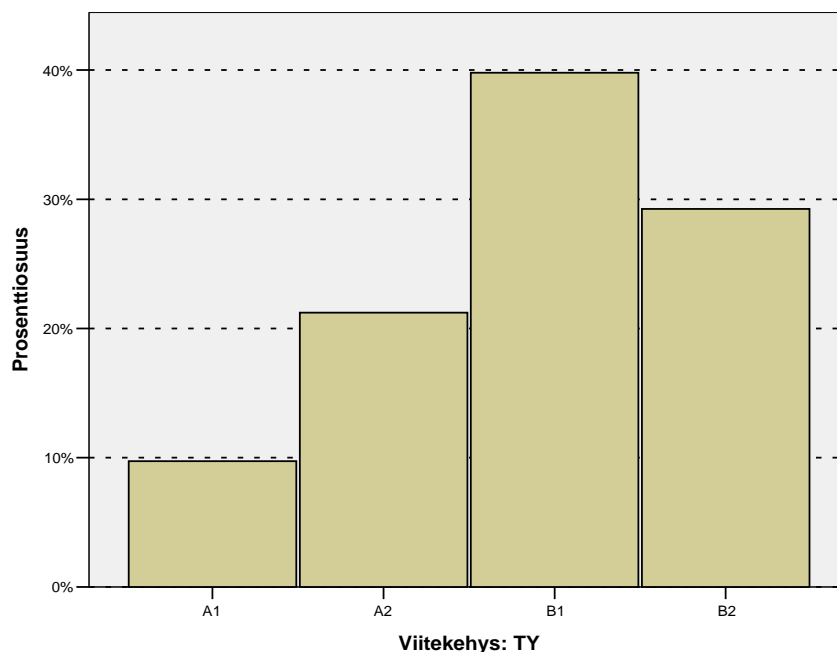
CEF B1 -taso	OPS 2004 Yleiskuvaus
Ymmärtää pääkohdat selkeistä yleiskielisistä viesteistä, joita esiintyy usein esimerkiksi työssä, koulussa ja vapaa-aikana. Selviytyy useimmista tilanteista matkustaessaan kohdekielisillä alueilla. Pystyy tuottamaan yksinkertaista, johdonmukaista tekstiä tutuista tai itseään kiinnostavista aiheista. Pystyy kuvaamaan kokemuksia ja tapahtumia, unelmia, toiveita ja tavoitteita. Pystyy perustelemaan ja selittämään lyhyesti mielipiteitä ja suunnitelmia.	Oppilas oppii <ul style="list-style-type: none"> - ymmärtämään pääajatuksia ja keskeisiä yksityiskohtia myös laajempaa yleistietoa sisältävästä kullusta tai luetusta, selvästi jäsennetystä tekstistä - selviytymään myös hiukan vaativammista epävirallisista keskustelutilanteista sekä kertomaan suullisesti tai kirjallisesti arkisista asioista, jotka sisältävät myös jonkin verran yksityiskohtia tiedostamaan joitakin keskeisiä eroja englannin kielen eri varianteista.

(Eurooppalainen viitekehys 2003:48)

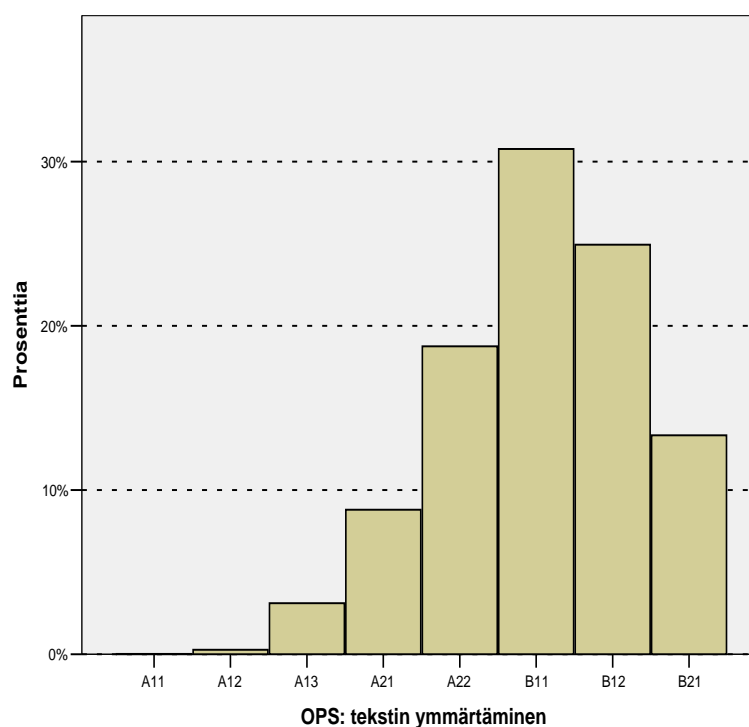
(OPS 2004:141)

9.3.2 Tekstin ymmärtämisen koe

Kuvio 66 osoittaa, että suuri osa oppilaista (n. 40 %) oli CEF -tasolla B1. Tasolle A1 sijoittui lähes 10 % ja tasolle B2 lähes 30 % oppilaista.



KUVIO 66 Tekstin ymmärtämisen kokeen tulos CEF -taitotasoasteikolle sijoitettuna.



KUVIO 67 Tekstin ymmärtämisen kokeen tulos OPS 2004 -taitotasosteikolle sijoitettuna.

Kuvio 67 osoittaa tekstin ymmärtämisen kokeen tuloksen OPS 2004-taitotasolla. Suurin osa oppilaista oli tasoilla B1.1 (31 %) ja B1.2 (25 %). A1-tasolle jäi noin 3,5 % oppilaista. Edellä mainittu varauma tulosten alustavuudesta on syytä pitää mielessä.

Myös tekstin ymmärtämisen kokeen θ -keskiarvo oli 0,27, joka vastaa CEF -taitotasoa B1 ja OPS 2004-tasoa B1.1.

Taulukossa 77 kuvataan oppilaiden keskimääräistä tekstin ymmärtämisen taitoa saavutetuilla CEF -asteikon B1-tasolla ja OPS 2004 -asteikon saavutetulla tavoitetasolla B1.1.

TAULUKKO 77 Oppilaiden tekstin ymmärtämisen taidon kuvaus CEF -asteikon B1-tasolla ja OPS 2004 -asteikon tasolla B1.1.

CEF B1	OPS 2004 B1.1
Pystyy lukemaan yksinkertaisia asiatekstejä aiheista, jotka liittyvät omaan alaan ja omiin mielenkiinnon alueisiin, ja ymmärtää niitä tyydyttävästi.	Pystyy lukemaan monenlaisia, muutaman sivun pituisia tekstejä (taulukot, kalenterit, kurssiohjelmat, keittokirjat) tutuista aiheista ja seuraamaan tekstin pääajatuksia, avainsanoja ja tärkeitä yksityiskohtia myös valmistautumatta. Pystyy seuraamaan tuttua aihetta käsittelevän parisivuisen tekstin pääajatuksia, avainsanoja ja tärkeitä yksityiskohtia. Arkikokemuksesta poikkeavien aiheiden ja tekstin yksityiskohtien ymmärtäminen voi olla puutteellista.

(Eurooppalainen viitekehys 2003:106). (OPH 2004. Liite 2)

Esimerkki 1 edustaa tekstin ymmärtämisen osiota, joka osoittautui CEF -asteikolla B.1 -tasoiseksi ja OPS 2004 -asteikolla B1.1 -tasoiseksi. Teksti on katkelma pitemmästä tekstistä, jonka ymmärtämistä testattiin neljällä monivalintaväittämällä, joista tämä oli toiseksi viimeinen.

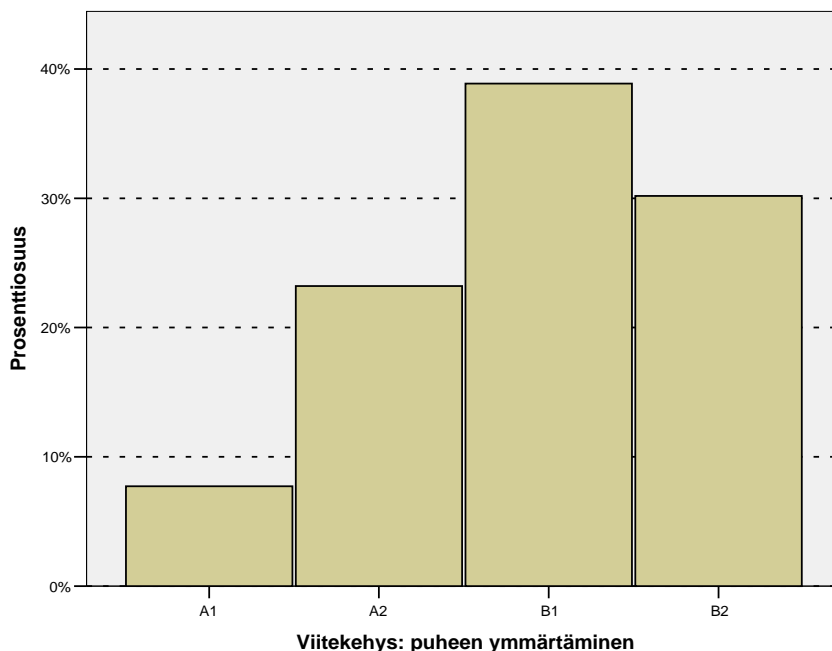
ESIMERKKI 1 CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava tekstin ymmärtämisen tehtävä.

Have you ever stopped to really think about what you and your family eat every day, and why? What about other people's food? In the movie *Indiana Jones and the Temple of Doom*, there are two scenes in which the two lead characters are offered meals from different cultures. One meal, meant to break the ice, consisted of insects. The second meal was a great banquet with such "delicacies" as live snakes, eyeball soup and chilled monkey brains for dessert. Some people eat things like rattlesnakes, bush rats, dog meat, horsemeat, bats, animal heart, liver, eyes and bugs of all sorts. Does it sound good?.

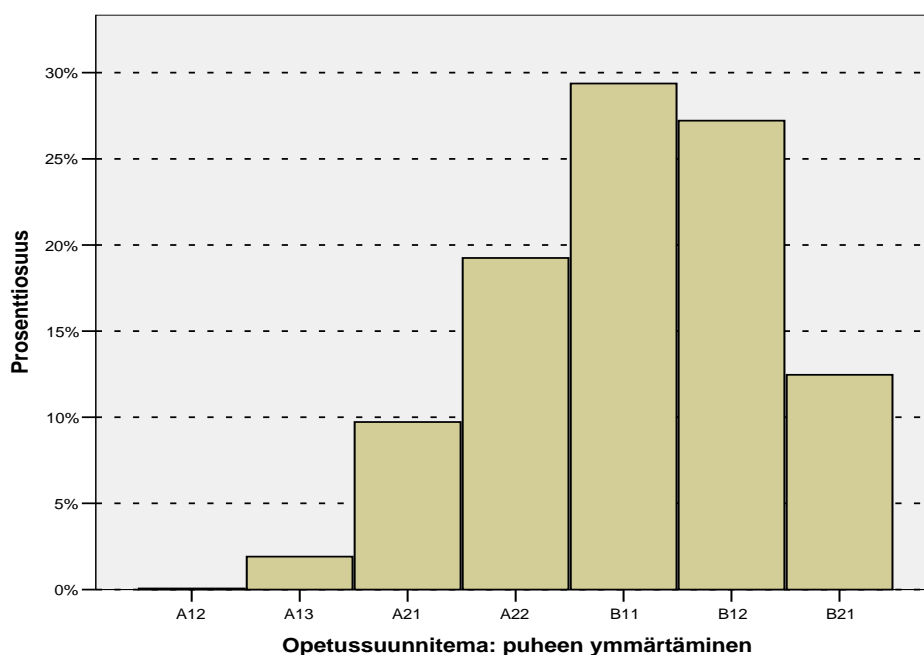
- A Food varies a lot in different countries.
- B Indiana Jones is an expert on many kinds of foods.
- C In some cultures meat is served as dessert.

9.3.3 Puheen ymmärtämisen koe

Kuvio 68 osoittaa, että suuri osa oppilaista oli CEF -taitotasolla B1 (lähes 40 %). Tasolle A1 sijoittui noin 8 % ja tasolle B2 noin 30 % oppilaista.



KUVIO 68 Puheen ymmärtämisen kokeen tulos CEF --taitotasoasteikolle sijoitettuna.



KUVIO 69 Puheen ymmärtämisen kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna.

Vastaavat luvut OPS 2004-taitotasoilla ovat, kuten kuvio 69 osoittaa: A1 2 %, A2 29 %, B1 57 % ja B2 13 %. Suurin osa oppilaista oli tasolla B1.1. OPS 2004:n tuloksiin on kuitenkin suhtauduttava aikaisemmin mainituin varauksin.

Myös puheen ymmärtämisessä saavutettu keskimääräinen taso vastaa CEF -taitotaso B1 ja OPS 2004-tasoa B1.1.

Taulukossa 78 kuvataan oppilaiden keskimääräistä puheen ymmärtämisen taitoa saavutetuilla CEF -asteikon B1-tasolla ja OPS 2004-asteikon tavoitetasolla B1.1 ja saavutetulla B1.1 tasolla.

TAULUKKO 78 Puheen ymmärtämisen taidon kuvaus CEF -asteikon B1-tasolla ja OPS 2004 -asteikon tasolla B1.1.

CEF B1	OPS 2004 B1.1
Ymmärtää pääajatuksat ja keskeiset yksityiskohdat selkeällä yleiskielellä esitetystä puheesta, joka käsittelee koulussa, työssä tai vapaa-aikana säännöllisesti toistuvia teemoja. Puhe voi olla myös lyhyt kertomus. Ymmärtää selväpiirteistä asiantietoa, joka liittyy tuttuihin seikkoihin tai työhön liittyviin aiheisiin. Ymmärtää puheesta sekä pääkohdat että tärkeimmät yksityiskohdat. Ymmärtäminen edellyttää kuitenkin selkeää puhetta ja tuttua puhetyyliä.	Ymmärtää pääajatuksat ja keskeisiä yksityiskohtia puheesta, joka käsittelee koulussa, työssä tai vapaa-aikana säännöllisesti toistuvia teemoja mukaan lukien lyhyt kerronta. Tavoittaa radiouutisten, elokuvien, tv-ohjelmien ja selkeiden puhelinviestien pääkohdat. Pystyy seuraamaan yhteiseen kokemukseen tai yleis-tietoon perustuvaa puhetta. Ymmärtää tavallista sanastoa ja rajallisen joukon idiomeja. Pitemmän viestin ymmärtäminen edellyttää normaalia hitaampaa ja selkeämpää yleiskielistä puhetta. Toistoa tarvitaan silloin tällöin.

(Eurooppalainen viitekehys 2003:102)

(OPH 2004. Liite 2:5-6)

Esimerkki 2 edustaa puheen ymmärtämisen tehtävää, joka osoittautui CEF -asteikolla B1 -tasoiseksi ja OPS 2004 -asteikolla B1.1 -tasoiseksi. Teksti on katkelma pitemmästä tekstistä, jonka ymmärtämistä testattiin kolmella monivalintaväittämällä, joista tämä oli viimeinen. Teksti kuultiin kahdessa osassa, ensin kokonaan ja sitten osina.

ESIMERKKI 2 CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava puheen ymmärtämisen tehtävä.

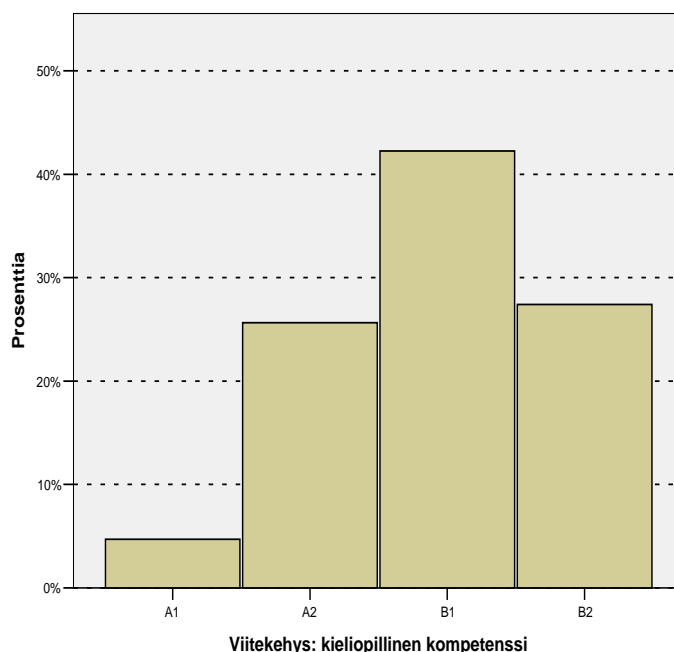
You have a visitor coming to your house, a friend you haven't seen for a long time. You just met her in town and asked whether she had time to come to your house. She agreed to try and find time, and now she's just arrived.

What do you say?

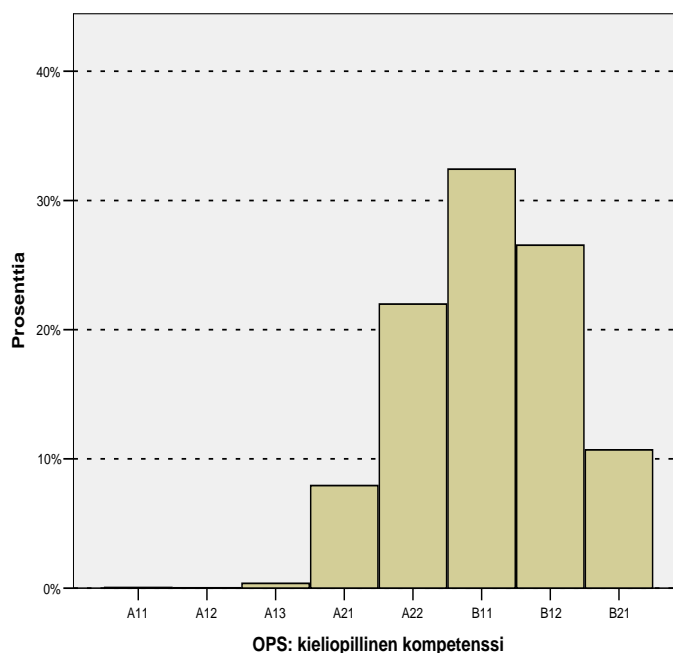
- A I'm glad you could do it.
- B I'm glad you could make it.
- C I'm glad you could stay.
- D I'm glad you could agree.

9.3.4 Kieliopillisen kompetenssin koe

Kuvio 70 osoittaa, että suurin osa oppilaista (n. 42%) oli CEF -taitotasolla B1. Noin neljäsosa oppilaista sijoittui tasolle A2 ja suunnilleen sama määrä tasolle B2.



KUVIO 70 Kieliopillisen kompetenssin kokeen tulos CEF- taitotasoasteikolle sijoitettuna.



KUVIO 71 Kieliopillisen kompetenssin kokeen tulos OPS 2004 -taitotasoasteikolle sijoitettuna.

Kuvion 71 mukaan suurin osa oppilaista (32 %) sijoittui kieliopillisen kompetenssin² kokeessa OPS 2004-taitotasolle B1.1. A1-tasolle tässä kokeessa sijoittuneiden oppilaiden määrä oli häviävän pieni. Sen sijaan tasoille A2.1 (8 %) ja A2.2 (22 %) sijoittui yhteensä noin 1/3 oppilaista ja hieman suurempi määrä tasoille B1.2 (27 %) ja B2.1 (11 %). Myös tässä tulos on nähtävä alustavana. Näin ollen myös kieliopillisen kompetenssin saavutettu keskimääräinen taso vastaa CEF -taitotasoa B1 ja OPS 2004-tasoa B1.1.

Esimerkki 3 edustaa kieliopillisen kompetenssin kokeen osiota, joka osoittautui CEF -asteikolla B1-tasoiseksi ja OPS 2004 -asteikolla B1.1 -tasoiseksi.

ESIMERKKI 3 CEF -tasoa B1 ja OPS 2004 -tasoa B1.1 edustava kieliopillisen kompetenssin tehtävä.

- What will happen to computers on January 1st 2000?
- A good question. And not ___ easy one to answer.

- a) -
- b) a
- c) an
- d) the

² Ks. liite. 8, jossa selitetään kieliopillisen kompetenssin OPS 2004 -taitotasoasteikon koostumus.

TAULUKKO 79 Tulosten jakaumat kokeissa (%) CEF- ja OPS 2004 -asteikoilla.

Taitotaso		Yleistaso		Puheen ymmärtäminen		Tekstin ymmärtäminen		Kielitieto	
CEF	OPS 2004	CEF	OPS 2004	CEF	OPS 2004	CEF	OPS 2004	CEF	OPS 2004
Alle A1	alle A1.1	-	-	-	-	-	-	-	-
A1	A1.1	5	-	8	-	10	-	4,7	-
	A1.2		-		-		-		
	A1.3		-		2		3		-
A2	A2.1	25	8	23	10	21	9	25,6	8
	A2.2		22		19		19		22
B1	B1.1	41	32	39	29	40	31	42,3	34
	B1.2		28		27		25		27
B2	B2.1	29	10	30	13	29	13	27,4	11
	B2.2		-		-		-		-

Taulukossa 79 esitetään reseptiivisen kielitaidon koko kokeen ja osakokeiden eri taitotasolle sijoittuneiden oppilaiden prosentuaaliset osuudet. Suomen opetussuunnitelman perusteiden tavoitetaso on lihavoitu. Reseptiivisen kielitaidon kokeen osioanalyysit tehtäväsarjoittain esitetään liitteessä 20 ja oppilasvastausten osiokohtaiset jakaumat liitteessä 21.

Yhteenvedona kolmanteen tutkimuskysymykseen saaduista vastauksista todetaan, että koko reseptiivisessä kokeessa ja sen kaikissa osakokeissa oppilaiden taito oli korkealla CEF -taitotasolla B1 ja OPS 2004-taitotasolla B1.1, mikä on vuoden 2004 perusopetuksen opetussuunnitelman perusteiden keskimääräinen tavoitetaso.

9.4 Produktiivisen kielitaidon kokeet

Neljännessä tutkimuskysymyksessä kysyttiin: Miten CEF-, OPS 2004- ja A2001-taitotasoasteikot vastaavat toisiaan produktiivista kielitaitoa arvioitaessa? Seuraavassa vastataan kysymykseen.

Sekä CEFissä että OPS 2004:ssä taitotasot on ilmaistu kirjainnumeroyhdistelmin. CEFissä käytetään 6-portaista asteikkoa, joista kolme (A2, B1 ja B2) on joissakin tapauksissa jaettu kahteen osaan. Seuraavassa esityksessä niitä käytetään viittaamaan "alemman" tason suoritukseen, kun taas ylempää tasoa tarkoitettaessa niiden jälkeen on lisätty +-merkki. Suullisen kielitaidon arvioimiseksi tarjotaan 25 asteikkoa, joista on mahdollista valita käytettäväksi se tai ne, jotka parhaiten soveltuvat koetehtävän aiheeseen ja kielenkäyttötilanteeseen. CEF -käsikirja tarjoaa lisäksi tarkennettuja asteikkoja, joita tässäkin tutkimuksessa on käytetty. Kirjoittamisen arviointiin on käytettävissä tulkinnoista riippuen ainakin 11 asteikkoa. Näiden lisäksi on mahdollista arvioida tuotosta kieliopillisen tarkkuuden ja sanaston osalta. OPS 2004:ssä produktiivisen kielitaidon tasokuvaukset esitetään 10-portaisena asteikkona kielitaidon osaluottain. A2001-taitotasot on ilmaistu numeroin 0-5.

Produktiivisen kielitaidon kokeiden tuotoksista arvioitiin eri taitotasoa-
steikkoja käyttäen otokset, ei siis koko sen populaation tuotoksia, joka osallistui
englannin arviointiin vuonna 1999. Näiden perusteella oli mahdollista estimoi-
da kaikkien produktiivisen kielitaidon kokeeseen osallistuneiden taitotasoa.
Asteikon A2001 mukaan oli jo aikaisemmin arvioitu 286 (143 paria) puhumisen
kokeen ja 539 kirjoittamisen kokeen tuotosta eli noin 22 % puhumisen kokee-
seen ja runsaat 10 % kirjalliseen kokeeseen osallistuneiden oppilaiden tuotoksis-
ta. Lisäksi arvioitiin CEF-, OPS 2004- ja A2001 -asteikkojen mukaan 52 (26 paria)
suullisen kokeen ja 50 kirjoittamisen kokeen tuotosta. Tämän määrän arvioitiin
aikaisempien vastaavien linkittämishankkeiden perusteella riittävän taito-
tasojen ja tulosten vastaavuuden määrittämisen pohjaksi. Liitteessä 22 esitetään
tähän otokseen perustuen produktiivisen kielitaidon kokeiden CEF-, OPS 2004-
ja A2001 -taitotasolle sijoittuneiden oppilaiden frekvenssijakaumat ja prosen-
tuaaliset osuudet kriteereittäin. Liitteessä 23 esitetään vastaavat tiedot laajem-
masta otoksesta. Seuraavassa selostetaan suppeampaan otokseen pohjautuvaa
taitotasojen asettamismenettelyä.

9.4.1 Produktiivisen kielitaidon kokeiden taitotasojen asettaminen

Taitotasojen asettamiseksi laskettiin asiantuntijoiden arvioiden ja empiiristen
tulosten yhdistämisen jälkeen CEF-, OPS 2004- ja A2001 -taitotasoa-
steikkojen kriteerien korrelaatiot. Seuraavassa ei esitetä kaikkien käytettyjen kriteerien
keskinäisiä korrelaatioita, koska produktiivisen kielitaidon taitotasoa-
steikkojen katkaisukohtien asettamisessa ei käytetty vastaavanlaista uutta menetelmää
kuin reseptiivisten kokeiden taitotasojen asettamisessa. Lopullisissa tuloksissa
kaikki kriteerit on kuitenkin otettu huomioon. Menetelmän selostamisesimer-
keissä on käytetty CEF -taitotasoa-
steikon yleistaso -kriteeriä (*overall*) ja A2001 -
taitotasoa-
steikon sujuvuus -kriteeriä.

Koetulokset linkitettiin taitotasoa-
steikkoihin ristiintaulukoimalla eri taito-
tasoa-
steikkoja käyttäen saadut frekvenssit ja arviointikriteerit. Tulosta havain-
nollistetaan korrespondenssianalyysin avulla saatavilla kuvioilla. Korrespon-
denssianalyysillä voidaan havainnollistaa yksinkertaisia kaksi- tai useam-
pisuuntaisia taulukoita, joihin sisältyy jokin rivien ja sarakkeiden välinen vas-
taavuus (Phillips 1995). Korrespondenssianalyysiä käytetään datan, esimerkiksi
juuri ristiintaulukointien tulosten, visuaalisessa esittämisessä (Greenacre 1993).
Analyysin tuloksena saadaan havainnollinen kaksi- tai useampiulotteinen ku-
vio, joka perustuu χ^2 :n neliöön. Tulokseksi saatu kuvio tulkitaan niin, että läh-
tekkäin olevat ryhmät ovat jakautumaltaan eniten samankaltaisia ja etäällä toi-
sistaan olevat vastaavasti erilaisia.

Puhumisen koe

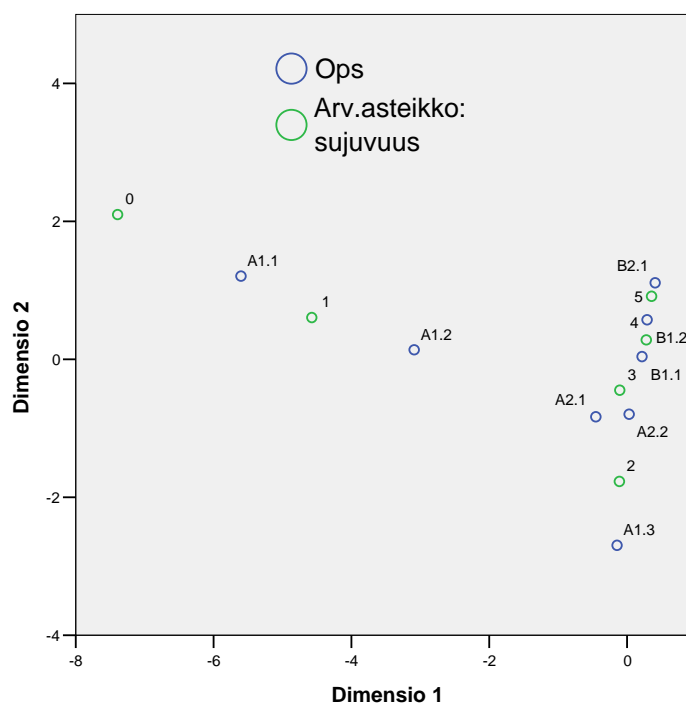
Esimerkkinä puhumisen kokeen linkittämisprosessista esitetään A2001- (sujuvuus) ja OPS 2004 -taitotasotaulukoiden³ linkittämismenettely. Ensinnäkin A2001-taitotasoasteikon sujuvuus -piirteen frekvenssit ja prosentuaaliset osuudet ristiintaulukoitiin OPS 2004-taitotasoasteikon kanssa (taulukko 80).

TAULUKKO 80 Puhumisen kokeen A2001- ja OPS 2004 -taitotasojen frekvenssien ristiintaulukointi (n=52).

		A2001-taitotasoasteikko; sujuvuus							Summa
		0	1	2	3	4	5		
OPS 2004	0	Lkm	1	1	0	0	0	0	2
		% summasta	0,4	0,4	0,0	0,0	0,0	0,0	0,8
	A1.1	Lkm	1	3	0	1	0	0	5
		% summasta	0,4	1,2	0,0	0,4	0,0	0,0	1,9
	A1.3	Lkm	0	0	5	1	0	0	6
		% summasta	0,0	0,0	1,9	0,4	0,0	0,0	2,3
	A2.1	Lkm	0	2	6	16	3	2	29
		% summasta	0,0	0,8	2,3	6,2	1,2	0,8	11,2
	A2.2	Lkm	0	0	9	20	12	2	43
		% summasta	0,0	0,0	3,5	7,7	4,6	0,8	16,5
	B1.1	Lkm	0	0	5	20	47	7	79
		% summasta	0,0	0,0	1,9	7,7	18,1	2,7	30,4
	B1.2	Lkm	0	0	0	16	33	24	73
		% summasta	0,0	0,0	0,0	6,2	12,7	9,2	28,1
	B2.1	Lkm	0	0	0	1	7	13	21
		% summasta	0,0	0,0	0,0	0,4	22,7	5,0	8,1
	Yht.	Lkm	2	7	25	76	102	48	260
		% summasta	0,8	2,7	9,6	29,2	39,2	18,5	100,0

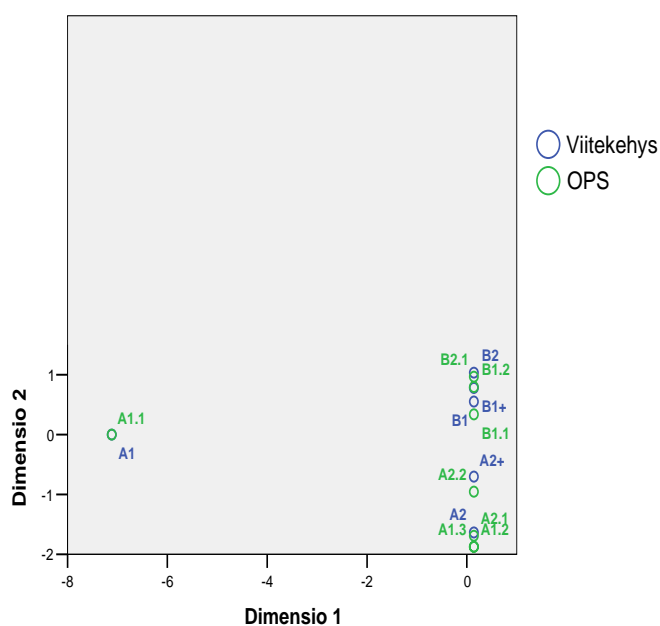
Spearmanin järjestyskorrelaatiokerroin näiden kahden taitotasoasteikon lukujen välillä oli 0,61, mikä on melko korkea yhteys ja luonnollisesti tilastollisesti erittäin merkitsevä. Ristiintaulukoinnin pohjalta suoritettu korrespondenssianalyysi tuotti kuvion 72, josta nähdään OPS 2004- ja A2001 (sujuvuus) -taitotasoisten vastaavuus eli että esimerkiksi puhumisen A2001 -taso 1 vastaa jossakin määrin OPS 2004 -tasoa A1.2, mutta että näiden etäisyydet toisistaan kuviolla ovat selvästi kauempana toisistaan kuin mitä tasot 4-5 ovat B-tasoista.

³ A2001 -taitotasoasteikkoa käytettäessä arviointi suoritetaan neljää kriteeriä käyttäen (luku 6). OPS 2004 antaa vain tuottamistaitojen yleiskuvauksen taitotasoittain.



KUVIO 72 Puhumisen kokeen A2001- ja OPS 2004-taitotasojen korrespondenssianalyysi. (A2001: 0-5, OPS 2004: A1.1 – B2.1.)

Ristiintaulukointi suoritettiin siis myös muiden A2001-asteikon kriteerien ja OPS 2004-asteikon välillä. Niiden korrelaatiot olivat kuitenkin heikkoja, joten näiden kahden asteikon vastaavuus jää epävarmaksi. Spearmanin järjestyskorrelaatiokerroin 0,57 osoitti, että CEF (yleistaso) -kriteerin korrelaatio A2001 -asteikon sujuvuus -kriteeriin oli suhteellisen heikko.



KUVIO 73 Puhumisen kokeen OPS 2004- ja CEF -taitotasojen korrespondenssianalyysi.

Taulukossa 81 esitetään edellä esitettyjen analyysien pohjalta laadittu taitotasoasteikkojen muuntotaulukko, joka kuitenkin on luonteeltaan alustava varsinkin siksi, että taitotasoasteikon A2001 tulokset korreloivat heikosti kahden muuhun asteikkoon. Toinen syy on, että yleensäkin on melko mahdotonta verrata 6- ja 10 -tasoisia asteikkoja toisiinsa. Tässä tapauksessa luotettavan arviointin tekeminen oli vaikeata myös pienen otoskoon vuoksi, koska tällöin kullekin tasolle saattaa tulla, ja tässä tapauksessa tulikin, hyvin harvoja tuloksia.

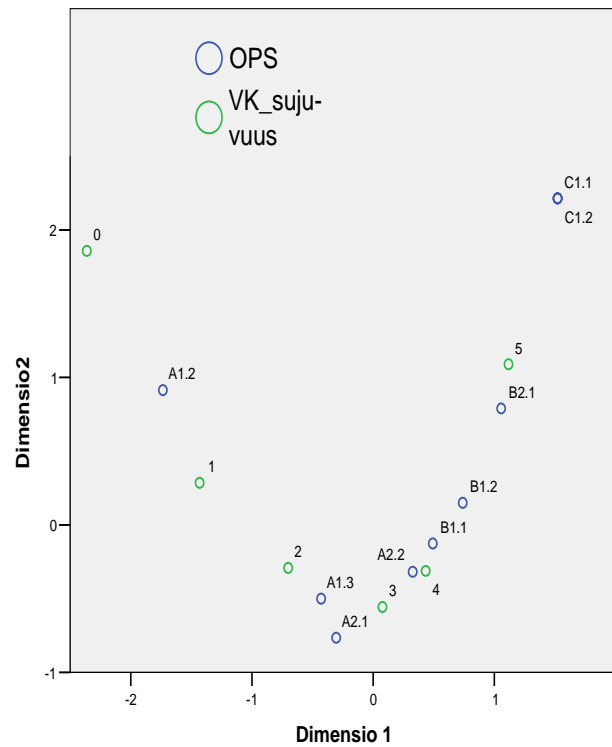
TAULUKKO 81 Puhumisen kokeen taitotasojen vastaavuus.

Taitotasot		
CEF	OPS 2004	A2001
A1	A1.1	1
	A1.2	
	A1.3	2
A2	A2.1	3
A2+	A2.2	
B1	B1.1	4
B1+	B1.2	
B2	B2.1	5
C1	B2.2	
C2	C1.1	

Kirjoittamisen koe

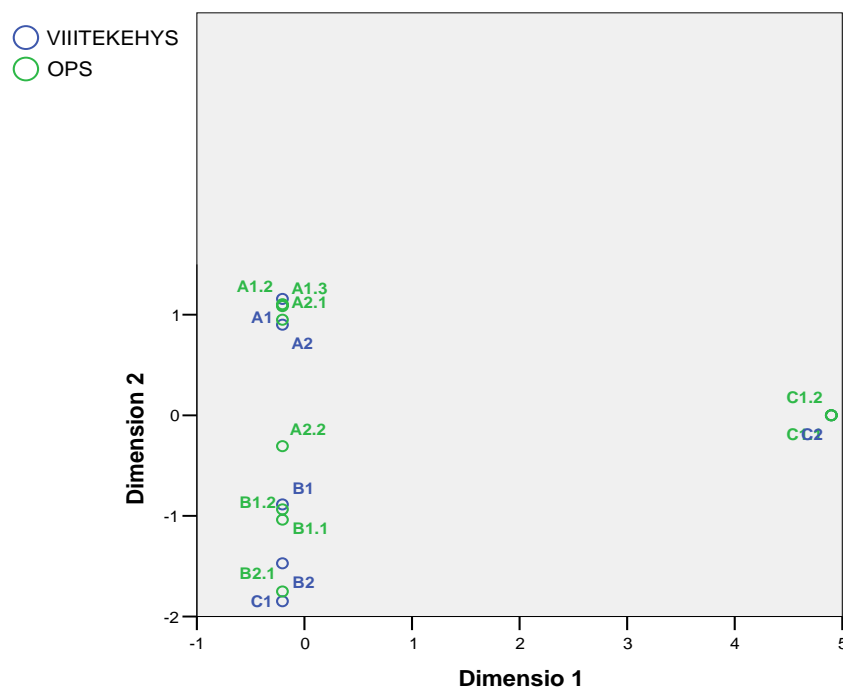
Taitotasojen asettaminen suoritettiin samalla tavalla kuin puhumisen kokeessa. A2001 -taitotasoasteikon sujuvuus -kriteerin katsottiin kattavan laajimmin kirjoittamisenkin taidon, joten sitä on käytetty esimerkkinä korrespondenssianalyysikuviossa.

Spearmanin järjestyskorrelaatiokerroin A2001 (sujuvuus)- ja OPS 2004 -taitotaso-asteikon välillä oli 0,68, mikä on melko korkea yhteys ja tilastollisesti erittäin merkitsevä. Kirjoittamisen kokeen kriteerien korrelaatiot olivat kauttaaltaan korkeampia kuin puhumisen kokeen korrelaatiot, mikä johtuu varmasti jossakin määrin siitä, että kirjoittamisen kokeiden arvioinnista on jo esimerkiksi ylioppilastutkinnon osalta noin 30 vuoden kokemus. Tämä korrespondenssianalyysi tuotti kuvion 74, joka osoittaa kirjoittamisen kokeen OPS 2004- ja A2001 -taitotasojen (sujuvuus) vastaavuuden, eli että esimerkiksi tasot 1 ja A1.2 ovat lähinnä vastaavia, vaikkakin sijaitsevat melko etäällä toisistaan, kun taas 2 ja A1.3 vastaavat toisiaan, että A2.1 vastaavat toisiaan ja että OPS 2004:n kolme tasoa A2.2, B1.1 ja B1.2 vastaavat lähinnä toisiaan.



KUVIO 74 Kirjoittamisen kokeen A2001- (sujuvuus)- ja OPS 2004-taitotasojen korrespondenssianalyysi.

Kuvio 75 osoittaa CEF (yleistaso)- ja OPS 2004 -taitotasojen vastaavuuden korrespondenssianalyysillä kuvattuna.



KUVIO 75 Kirjoittamisen kokeen OPS 2004- ja CEF -taitotasojen (yleistaito) korrespondenssianalyysi.

Taulukossa 82 esitetään edellä esitettyjen analyysien pohjalta laadittu kirjoittamisen taitotasosteikkojen muuntotaulukko, jossa opetussuunnitelman perus-

teiden tavoitetaso on lihavoitu. Muunnos on kuitenkin epätarkka varsinkin siksi, että taitotasosteikon A2001 tulokset korreloivat melko heikosti kahteen muuhun asteikkoon. Toinen syy on, kuten aikaisemminkin jo on mainittu, toisistaan suuresti poikkeavien asteikkojen tasojen vastaavuuden määrittämisen vaikeus. Tässä tapauksessa luotettavan arvion tekeminen oli vaikeata myös pienen otoskoon vuoksi, koska tällöin kullekin tasolle saattaa tulla hyvin harvoja tuloksia.

TAULUKKO 82 Kirjoittamisen kokeen taitotasojen vastaavuus.

Taitotaso		
CEF	OPS 2004	A2001
A1	A1.1	0
	A1.2	
	A1.3	1
A2	A2.1	2
	A2.2	3
B1	B1.1	4
	B1.2	
B2	B2.1	5
C1	B2.2	
C2	C1.1	

Vastauksena neljänteen tutkimuskysymykseen todetaan, että sekä suullisessa että kirjallisessa kokeessa CEF- ja OPS 2004 -taitotasosteikkojen linkittäminen on melko luotettavalla pohjalla, kun taas A2001 -asteikon linkittäminen niihin jouduttiin tekemään pitkälti harkinnanvaraisesti.

9.5 Produktiivisen kielitaidon kokeiden tulokset taitotasoittain

Viidennessä tutkimuskysymyksessä kysyttiin: Mikä perusopetuksen päättövaiheen oppilaiden englannin kielen produktiivinen kielitaito on CEF-, OPS 2004- ja A2001 -taitotasosteikoilla ilmaistuna? Seuraavassa vastataan kysymykseen.

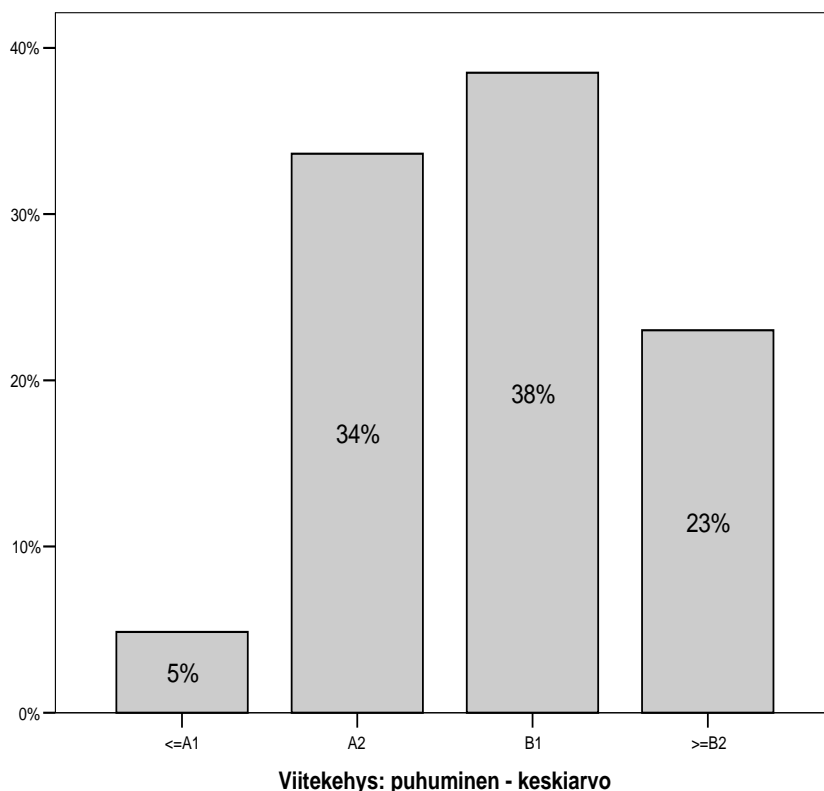
9.5.1 Puhumisen kokeen tulokset

Puhumisen kokeen suppeata aineistoa arvioitaessa käytettiin kolmea CEFin asteikkoa ja kolmea CEF-käsikirjan asteikkoa. Analysointivaiheessa CEFin Turn Taking- ja CEF -käsikirjan Turntaking -asteikkoa käsiteltiin yhtenä asteikkona, joten lopulta arvioinnissa käytettiin viittä kriteeriä: vuorovaikutus (*overall spoken interaction*), fonologia (*phonological control*), puheenvuoron vaihtaminen (*turntaking*) laajuus (*range*) ja tarkkuus/virheettömyys (*accuracy*). Kuten liittees-

tä 11 käy ilmi, kolme ensimmäistä ovat varsinaisia CEFin taulukoita, kun taas kaksi muuta ovat yhdistelmiä ja tarkennuksia useista niistä. Näiden kriteerien käytön katsottiin soveltuvan parhaiten sellaisen kokeen arviointiin, jossa tuotokset ovat pääasiassa vuorovaikutteisia, ja joissa odotetaan käytettävän tuttuihin tilanteisiin sopivia fraaseja. Ääntämistä seurattiin äänneiden osalta erityisesti ääneen lukemisen kokeessa, vaikka tiedetäänkin, että hyvätkin lukijat saattavat suoriutua alle oman tasonsa "sisäluvussa". Intonaatiota seurattiin lähinnä verbillä alkavissa kysymyslauseissa.

Suuremman aineiston arvioinnissa käytettiin "alkuperäistä" A2001-taitotasoasteikkoa (liite 9), jonka kriteereihin ei kuulunut, toisin kuin suppeamman aineiston arvioinnissa, fonologia. OPS 2004 -asteikkoon linkitetyt koetulokset ovat kuuden osion (tehtävän) pyöristettyjä keskiarvoja. Vastaavuuden määrittäminen CEF- ja OPS 2004 -asteikkoihin tapahtui eri arviointikriteerien tuottamisen arviointien yhdistämisen (agregoinnin) jälkeen. Ekstrapoloidut CEF-, OPS 2004- ja A2001 -keskiarvot (taulukko 85) perustuvat käytettyjen kriteerien pyöristettyihin keskiarvoihin.

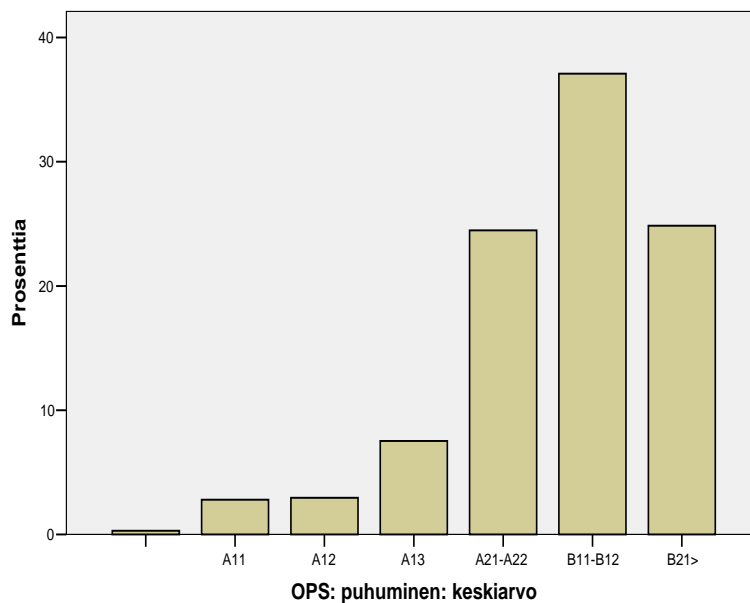
Kuviot 76, 77 ja 78 osoittavat suullisen kokeen tulosten keskimääräisen jakautumisen OPS 2004-, CEF (*overall*)- ja A2001 (sujuvuus) -taitotasoilla. Suurin osa oppilaista oli CEF -taitotasolla B1 (39 %), mutta miltei yhtä suuri osuus oppilaista jäi tasolle A2 (34 %). Saavutettu keskimääräinen puhumisen taso oli B1.



KUVIO 76 Puhumisen kokeen tulosten ekstrapoloitu jakauma CEF -taitotasoasteikolle sijoitettuna.

Saavutettua keskimääräistä puhumisen taitoa B1 kuvaillaan CEF -asteikolla seuraavasti: Pystyy esittämään suhteellisen sujuvasti ja lähinnä peräkkäisenä luettelona kuvauksen erilaisista tutuista, itseään kiinnostavista aiheista (Eurooppalainen viitekehys 2003:93).

OPS 2004-taitotason mukaan arvioituna suurin ryhmä oppilaita sijoittui B1.1 - B1.2 -tasolle (37 %) ja B1.2-tasolle (28 %). Hyväksyttävän suorituksen alapuolelle jäi edelleen vajaa 1 % osallistujista; B2.1- tai sitä korkeammalle tasolle ylsi noin 25 % oppilaista (kuvio 77). Toisin kuin reseptiivisen kokeen kohdalla, taitotasoarviota voi pitää varsin luotettavana.



KUVIO 77 Puhumisen kokeen tulosten ekstrapoloitu jakauma OPS 2004- taitotasasteikolle sijoitettuna.

Suullisen kielitaidon keskimääräinen taso OPS 2004 -asteikolla arvioituna oli B1.1. Taulukossa 83 esitetään sekä sen että opetussuunnitelman perusteiden 2004 odotettu taitotasokuvaus. A2.2.

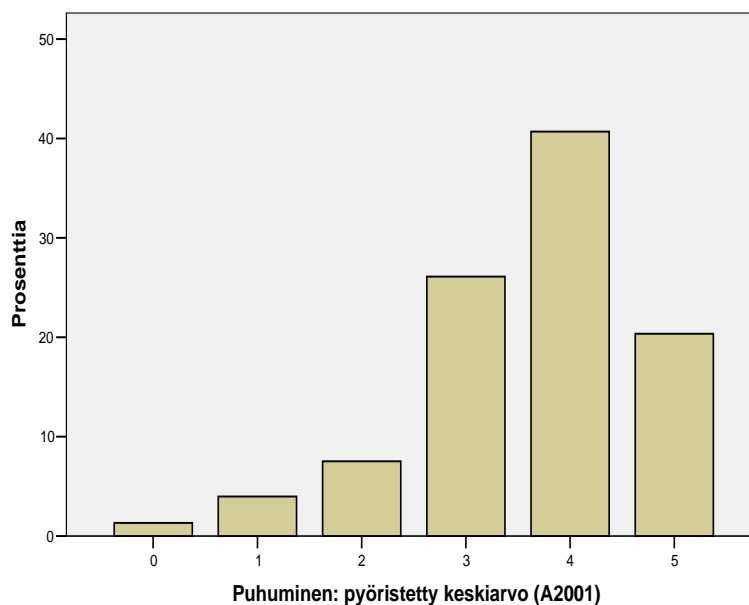
TAULUKKO 83 Saavutetun puhumisen keskimääräisen taitotason B1.1 ja tavoitetason A2.2 kuvaus OPS 2004 -asteikolla.

OPS 2004 A2.2	OPS 2004 B1.1
<p>Osaa esittää pienen, luettelomaisen kuvauksen lähipiiristään ja sen jokapäiväisistä puolia. Pystyy osallistumaan rutiinomaisiin keskusteluihin omista tai itselleen tärkeistä asioista. Voi tarvita apua keskustelussa ja vältellä joitakin aihepiirejä. Puhe on välillä sujuvaa, mutta erilaiset katkokset ovat hyvin ilmeisiä. Ääntäminen on ymmärrettävää, vaikka vieras korostus on ilmeistä ja ääntämisvirheitä esiin-</p>	<p>Osaa kertoa tutuista asioista myös joitakin yksityiskohtia. Selviytyy kielialueella tavallisimmista arkitilanteista ja epävirallisista keskusteluista. Osaa viestiä itselleen tärkeistä asioista myös hieman vaativammissa tilanteissa. Pitkäkestoinen esitys tai käsitteelliset aiheet tuottavat ilmeisiä vaikeuksia. Pitää yllä ymmärrettävää puhetta, vaikka pitkäkestoisissa puhejaksoissa esiintyy taukoja ja epärointiä.</p>

<p>tyy. Osaa kohtalaisen hyvin tavallisen, jokapäiväisen sanaston ja jonkin verran idiomaattisia ilmaisuja. Osaa useita yksinkertaisia ja myös joitakin vaativampia rakenteita. Laajemmassa vapaassa puheessa esiintyy paljon virheitä perusasioissa (esim. verbien aikamuodoissa) ja ne voivat joskus haitata ymmärrettävyyttä.</p>	<p>Ääntäminen on selvästi ymmärrettävää, vaikka vieras korostus on joskus ilmeistä ja ääntämisvirheitä esiintyy jonkin verran. Osaa käyttää melko laajaa jokapäiväistä sanastoa ja joitakin yleisiä fraaseja ja idiomeja. Käyttää useita erilaisia rakenteita. Laajemmassa vapaassa puheessa kielioppivirheet ovat tavallisia (esim. artikkeleita ja päätteitä puuttuu), mutta ne haittaavat harvoin ymmärrettävyyttä.</p>
--	--

(OPH 2004. Liite 2:4-5.)

A2001-taitotasoasteikon tasoja kuvataan neljällä kriteerillä: tehtävän suorittaminen ja sujuvuus, ilmaisun laajuus, virheettömyys ja vuorovaikutus. Analyysit perustuvat kaikkiin käytettyihin kriteereihin ja tulokset on ilmaistu niiden pyöristettyinä keskiarvoina. Kuvio 78 osoittaa oppilaiden tulosten sijoittumisen tälle asteikolle. Oppilaista 41 % sijoittui tasolle 4, 26 % tasolle 3 ja 20 % tasolle 5. Vajaa 1 % ei yltänyt hyväksyttävään suoritukseen.



KUVIO 78 Puhumisen kokeen tulosten ekstrapoloitu jakauma A2001 -taitotasoasteikolle sijoitettuna.

Puhumisen saavutettua keskimääräistä tasoa (4) A2001-taitotasoasteikolla kuvaillaan seuraavasti: Viesti välittyy tarkoitetulla tavalla. Oppilas suoriutuu tehtävästä suhteellisen vaivattomasti, vaikka toisinaan epäröiden (Tuokko 2002:141). Esimerkki 4 on kaikilla taitotasoasteikoilla keskimääräistä tasoa (B1.1, B1 ja 4) edustavaksi arvioitu transkriboitu puhumisen kokeen tuotos. Kauttaviivoin on osoitettu selvät ääntämisvirheet (ks. luku 8).

ESIMERKKI 4 Transkriboitu keskimääräistä puhumisen tasoa kaikilla taitotasoasteikoilla (CEF, OPS 2004 ja A2001) edustava tuotos.

1.
(hihittelevät)
B - Fifty-five, thirty-eight and twenty.
A - Hi! My name is Msrie (hihittelevät)
B - I've wait your call.
A - I'm coming to Oulu fourteen of April ten past ((lyhyt a)) six.
B - Ten past (lyhyt a). How I /resaiks/ you?
A - Mm. I have (hihittää; rykäisee) long blond here and and blue /p/ jeans /tz/ and blue /p/ eyes and brown jacket.
B - (äännähdys) Say something to what you look like.
A - (vinkaisee) I've just say it.
B I have blond hair and black jacket and /Velkam/ and bye-bye.
A (Hyvin reippaasti) I think we can recognise each other when we meet two weeks later. Bye-bye.
B - (Hämmentynyt naurahdus). Bye-bye.

-
2.
B - Hi. You are Mari.
A - Yes, I am. Hello (miltei tympeästi).
B - It's nice to meet you.
A - Same thing. You too.
B - The weather is not really good this time year. It's always so windy and cold. (Aidoilta kuulostavat äänneet)
A - Well, it's nearly (rykäisy) the same that in Greek. It was raining there when I left.
B - We have to hurry up. Bus is waiting.
A - Ok. I'm ready. I just take my bags.
B - Can I help you (tauko) to your bags. (ei nouseva intonaatio)
A - Thank you but they are not heavy. I can make it on my own.

-
3.
B - What kind is your family? A
- Well, I live with my mom and my little brother and my little sister. My little brother is twelve years old and my little sister is only two. And we have also dog. What about your family?
B - I live with my dad and mom and my big brother.. (jotakin, mistä ei saa selvää, naurahdus)
A - That's nice. Do you have any hobbies?
B - (naurahdus) I ... is painting and art. What do you ...?
A - I don't have any hobbies. Well ... sometimes I read and I go walking out ... side.
B - What do you do in your freetime?
A - I just hang hanging out with my friends. What about you?
B - Mm, I spend time with my friends.
A - Oh, that's nice (nauravat).

-
4.
A - Hello!
B - Hi! Can I have menu?
A - This is the menu.
B - Do you want something to drink?
A - Yes. I would like to have a Coca-Cola.
B - Thanks. Coca Cola. Do ... What do you want to eat?
A - I would like to have this Four Seasons.
B - Do you like something else?
A - I decide that by later. (tauko ja sitten kuiskaa: LOPPU!)
-

- B - Do you like food?
 A - Yes. Can I have check.
 B - This is check.
 A - Thank you (nouseva äänenpaino). Here's the money. You can keep change.
 B - Thanks and goodbye.

A - Bye-bye (kuulostaa kyllästyneeltä).

5. Lukemistehtävä

- B (Kohdekielen mukaiset äänteet.)
 Intonaatio puuttuu jotakuinkin täysin, mikä tekee luetun vaikeaksi ymmärtää.
 Vieraampien sanojen ääntämys ontuu.
 (Ujo esiintyjä)
 A Parempi intonaatio, minkä vuoksi luettu on ymmärrettävämpää.
 (Rohkea esiintyjä.)
-

6.

- A - I think that parents doesn't have to pay all the damage their children have done.
 (Virke koostuu annetusta kysymyksestä)
 B - I think ... (ei kuulu) It's not the parents they don't have to pay for their children
 ...
 A - Yes, it's the children's /suhu-s/ own fault /fo:t/ (tauko) if they have done some
 crime.
 B - I think that ...
 A - (vaihtaa reippaasti aihetta) Well, do you think that children should be punished in
 the same way as adults? (Virke koostuu annetusta kysymyksestä.)
 B - (tauko). I think that young people have to be some opportunities for children.
 A - Yes, I think to the too. They are as (ei kuulu) as adults, so they may have do
 /stupid/ things.
-

Koska kaikki arviointivariaabelit olivat järjestysasteikollisia, tulosten ilmaisemissa käytettiin keskiarvojen sijasta mediaania ja moodia.

TAULUKKO 84 Puhumisen kokeen mediaanit ja moodit pyöristettyinä keskiarvoina CEF-, OPS 2004- ja A2001 -taitotasoasteikkojen mukaan.

Taitotasoasteikko	n	Mediaani	Moodi
OPS 2004	226	B1.1	B1.1
CEF			
- yleistaso	226	B1	B1
- laajuus	226	B1	B1
- tarkkuus	226	A2	A2
- vuorovaikutus	226	B1	>= B2
A2001			
- sujuvuus	226	4	4
- laajuus	226	4	4
- tarkkuus	226	3	3
- vuorovaikutus	226	4	5

Vertailtaessa taulukon 84 osoittamia eri taitotasoasteikoin arvioituja suullisen kielitaidon piirteitä voidaan todeta, että CEFiä ja OPS 2004:ä käytettäessä tulokset vastaavat suurin piirtein toisiaan. Koska OPS 2004 kuvaa taitotasot holistisesti niihin sisältyviä piirteitä erittelemättä, eroja muihin asteikkoihin ei ole

mahdollista kommentoida. Arvioituista piirteistä kieliopillinen tarkkuus/virheettömyys tuottaa heikoimmat tulokset ja vuorovaikutus parhaat.

Taulukossa 85 esitetään eri taitotasoasteikkojen mukaisten arviontien vastaavuus ja taitotasoille asettuneiden suoritusten frekvenssit ja suhteelliset osuudet.

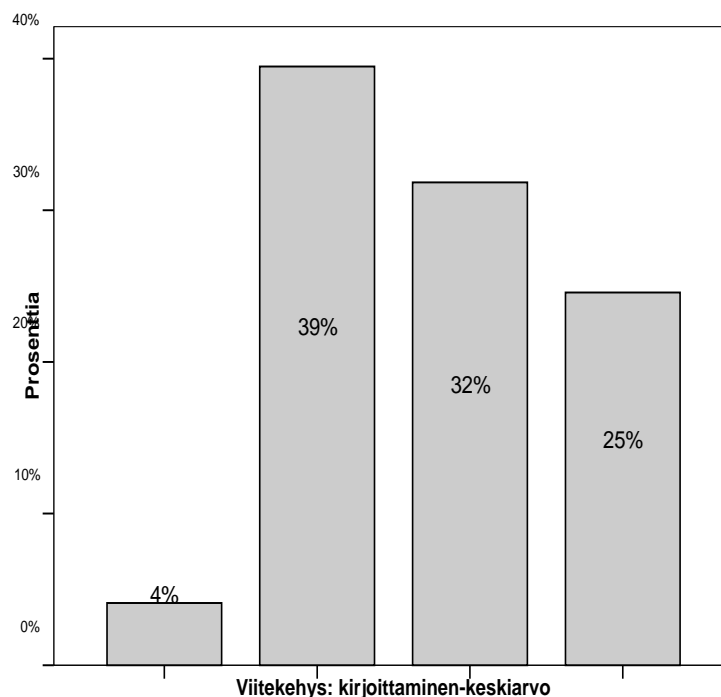
TAULUKKO 85 Puhumisen kokeen taitotasoasteikkojen vastaavuus sekä kunkin tason suoritusten ekstrapoloidut frekvenssijakaumat ja prosenttiosuudet.

Taitotasot			Frekvenssi			Validi prosentti		
CEF	OPS 2004	A2001	CEF	OPS 2004	A2001	CEF	OPS 2004	A2001
<=A1	0	0	11	1	3	5	-	1
	A1.1			6			3	
	A1.2	1		7	9		3	5
	A1.3	2		17	17		8	8
A2	A2.1 - A2.2	3	76	55	59	34	24	26
B1	B1.1 - B1.2	4	87	84	92	39	37	41
>= B2	>= B2.1	5	52	56	46	23	25	20
Yhteensä			226	226	226	100	100	100

CEF -asteikkoa käyttäen puhumisen kokeen mediaani oli B1, OPS 2004 -asteikolla B1.1 ja A2001 -asteikolla 4. Opetussuunnitelman perusteissa puhumisen tavoitetaso on A2.2. Tasoilla A2.1 - A2.2 oli 26 % oppilaista, sen alapuolelle jäi 13 % oppilaista ja sen ylitti 61 % oppilaista. Kuten taulukko 85 osoittaa, tavoite ylittyi.

9.5.2 Kirjoittamisen kokeen tulokset

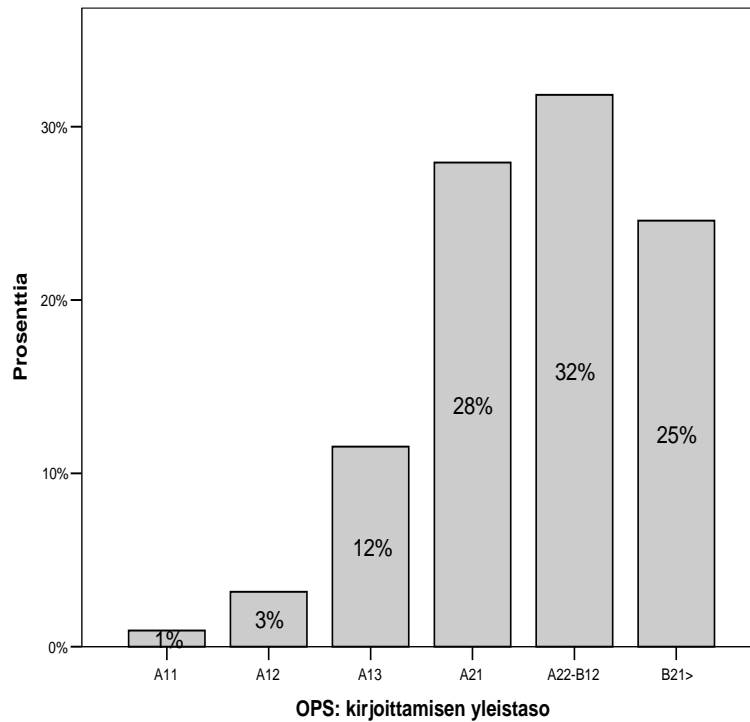
Kirjoittamisen kokeen tulosten arvioinnissa käytettiin CEF -käsikirjan taulukkoa Written Assessment Criteria Grid (liite 12) sekä OPS 2004 ja A2001-taitotasokuvauksia. Kuvioissa 79, 80 ja 81 esitetään kirjoittamisen kokeen tulokset CEF-, OPS 2004- ja A2001 -taitotasoasteikoilla. OPS -asteikolla arvioituna eniten oppilaita oli A2.2 - B1.2 -tasolla (32 %), CEF -asteikolla tasolla A2 (39 %) ja A2001 -asteikolla 4 (32 %).



KUVIO 79 Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma CEF -taitotaso-asteikolle sijoitettuna.

Oppilaiden saavuttamaa keskimääräistä kirjoittamisen tasoa B1 kuvaillaan CEF -asteikolla seuraavasti: Pystyy kirjoittamaan yksinkertaisia, yhtenäisiä tekstejä tavallisista, itseään kiinnostavista aiheista yhdistämällä lyhyempiä, irrallisia, yksinkertaisia ilmauksia yhtenäiseksi tuotokseksi (Eurooppalainen viitekehys 2003:96).

Kuviossa 80 OPS 2004 -taitotasojen A2.2 ja B1.2 tulokset esitetään yhtenä pylväänä samoin kuin tasoa B1.2 korkeammat tulokset. Tämä selittyy sillä, että taitotasojen määrittäminen oli OPS -tasolla epävarmempaa kuin CEF -tasolla.



KUVIO 80 Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma OPS 2004 -taitotaso-asteikolle sijoitettuna.

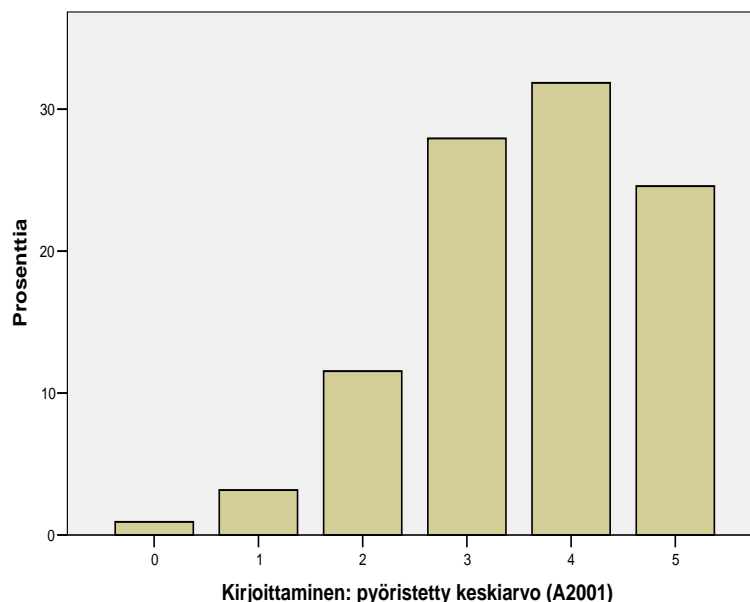
Taulukossa 86 esitetään OPS 2004 -asteikon taitotasojen A2.2 (tavoitetaso) ja B1.1 (saavutettu taso) kuvaukset.

TAULUKKO 86 Kirjoittamisen taitotasojen A2.2 ja B1.1 kuvaukset OPS 2004 -asteikolla.

OPS 2004 A2.2	OPS 2004 B1.1
<p>Selviytyy kirjoittamalla tavanomaisissa arkitilanteissa.</p> <p>Osaa kirjoittaa hyvin lyhyen, yksinkertaisen kuvauksen tapahtumista, menneistä toimista ja henkilökohtaisista kokemuksista tai elinympäristönsä arkipäiväisistä puolista (lyhyet kirjeet, muistilaput, hakemukset, puhelinviestit).</p> <p>Osaa arkisen perussanaston, rakenteet ja tavallisimmat sidoskeinot.</p> <p>Kirjoittaa yksinkertaiset sanat ja rakenteet oikein, mutta tekee virheitä harvinaisemmissa rakenteissa ja muodoissa ja tuottaa kömpelöitä ilmaisuja.</p>	<p>Pystyy kirjoittamaan ymmärrettävän, jonkin verran yksityiskohtaistakin arkitietoa välittävän tekstin tutuista, itseään kiinnostavista todellisista tai kuvitelluista aiheista.</p> <p>Osaa kirjoittaa selväpiirteisen sidosteisen tekstin liittämällä erilliset ilmaukset peräkkäin jaksoiksi (kirjeet, kuvaukset, tarinat, puhelinviestit). Pystyy välittämään tehokkaasti tuttua tietoa tavallisimmissa kirjallisen viestinnän muodoissa.</p> <p>Osaa useimpien tutuissa tilanteissa tarvittavien tekstien laadintaan riittävän sanaston ja rakenteet, vaikka teksteissä esiintyy interferenssiä ja ilmeisiä kiertoilmaisuja.</p> <p>Rutiinomainen kieliaines ja perusrakenteet ovat jo suhteellisen virheettömiä, mutta jotkut vaativammat rakenteet ja sanaliitot tuottavat ongelmia.</p>

(OPS 2004:237)

A2001-taitotasosteikolla arvioituna eniten oppilaita sijoittui tasolle 4 (32 %), joka oli myös mediaani. Seuraavaksi eniten oppilaita oli tasoilla 3 (28 %) ja 5 (25 %). Tasolle 1 jäi 17 oppilasta ja viideltä ei saatu tuotosta lainkaan (kuvio 81).



KUVIO 81 Kirjoittamisen kokeen tulosten ekstrapoloitu jakauma A2001 -taitotasolle (sujuvuus) sijoitettuna.

A2001 -asteikolla saavutettua keskimääräistä kirjoittamisen taitotasoa 4 kuvataan seuraavasti: Helppolukuista ja sujuvaa. Sopiva, mutta ei kovin monipuolinen sanasto. Melko vaihtelevia rakenteita. Selkeää, mutta melko tavanomaista. Joitakin virheitä ja epäaitoja ilmauksia. (Tuokko 2002:143).

Esimerkki 5 edustaa kaikilla taitotasasteikoilla keskimääräistä tasoa (A2.2, B1 ja 4) edustavaksi arvioidusta kirjoittamisen tuotoksesta.

ESIMERKKI 5 Kaikilla taitotasolla keskimääräiseksi (B1.1, B1 ja 4) arvioitu kirjoittamisen tuotos (Tuokko 2000:130).

What music means to me

Music is very important to me. I listen music all time, except in school ofcourse. But when I am at home I put my stereos on even if I didn't want to listen it. Sounds stupid huh? Yeah but it's the truth. I just like to do things when music's playing. I can't handle if there's too quiet. People got nervous then. You can learn much about person by knowing what kind of music he/she likes. I like old music like the Beatles, the Doors, Black Sabbath ... I hate that kind of new Heavy Metal stuff what everybody listen and you can find from every musicshop in town. There is no sense in it. Every song hears the same and every band hears the same. Maybe its just in my ear, maybe I can't hear it "right". They listen their music and I listen my ... and that's it..

Taulukko 87 osoittaa käytettyjen arviointiasteikkojen avulla arvioitujen kirjoittamisen kokeeseen osallistuneiden otosoppilaiden lukumäärän, mediaanit ja moodit.

TAULUKKO 87 Kirjoittamisen kokeen mediaanit ja moodit CEF-, OPS-2004- ja A2001-taitotasoasteikoilla.

Taitotasoasteikko	Oppilasvastausten lukumäärä		Mediaani	Moodi
	Validi	Kato		
OPS 2004	537	0	B1.1	B1.1
CEF, yleistaso	537	0	B1	B1
CEF, laajuus	537	0	B1	B1
CEF, koherenssi	537	0	B1	B1
CEF, tarkkuus	537	0	A2	B1
A2001, sujuvuus	532	5	4	4
A2001, laajuus	537	0	4	4
A2001, koherenssi	537	0	4	4
A2001, tarkkuus	535	2	3	3

Samoin kuin puhumisen kokeessa, kirjoittamisen kokeessa heikoimmat tulokset tulivat arvioitaessa tarkkuutta/virheettömyyttä. Aihepiireittäin tarkasteltuna korkeimmalle tasolle päästiin useimmiten kirjoitettaessa argumentointia edellyttävästä aiheesta

Taulukossa 88 esitetään kirjoittamisen CEF-, OPS 2004- ja A2001-taitotasoasteikkojen ekstrapoloitu vastaavuus ja kullekin taitotasolle saatujen tulosten frekvenssit ja prosenttiosuudet. OPS 2004:n tavoitetaso A.2.2 on lihavoitu.

TAULUKKO 88 Kirjoittamisen kokeen ekstrapoloitu frekvenssijakauma ja prosenttiosuudet CEF-, OPS 2004- ja A2001-taitotasoilla.

Taitotasojen vastaavuus			Frekvenssi			Validi prosentti		
CEF	OPS 2004	A2001	CEF	OPS 2004	A2001	CEF	OPS 2004	A2001
<=A1	A1.1	0	22	5	5	4	1	1
	A1.2			17			3	
	A1.3	1		62	17		12	3
A2	A2.1	2	212	150	62	40	28	12
		3			150			
B1	A2.2 - B1.2	4	171	171	171	32	32	32
>=B2	>= B2.1	5	132	132	132	25	25	25
Tuotoksia yhteensä			537	537	537	537	537	537

Opetussuunnitelman perusteiden kirjallisen tuottamisen keskimmäisen arvonsanan (8) tavoitetaso on A2.2. Mediaaniksi saatiin kuitenkin B1.1, joten tavoite ylittyi selvästi. Johtuen suppeasta otoksesta, jonka pohjalta ekstrapolointi suoritettiin, tulosta voidaan kuitenkin pitää vain suuntaa antavana.

Yhteenvedona viidenteen tutkimuskysymykseen saaduista vastauksista todetaan, että 1) puhumisen kokeen mediaaniksi tuli CEF -taitotasoasteikolla taso B1, OPS 2004 -asteikolla B1.1 ja A2001 -asteikolla taso 4, kun perusopetuksen opetussuunnitelman perusteiden tavoitetaso on A2.2, ja että 2) kirjoittamisen kokeessa mediaaniksi tuli CEF -taitotasoasteikolla B1, OPS 2004 -asteikolla B1.1 ja A2001 -asteikolla tasolla 4, kun perusopetuksen opetussuunnitelman perusteiden tavoitetaso on A2.2. Opetuksessa on siis ylletty tavoitetasoa korkeammalle.

9.6 Yhteenveto

Vastauksena ensimmäiseen tutkimuskysymykseen, joka koski käytettyjen tehtävien laatua, voitiin todeta, että tehtävät osoittautuivat hyvin niiden kannalta tärkeimmän ominaisuuden, erottelukyvyn suhteen. Aineisto osoittautui soveltuvan IRT -analyysien tekemiseen ja tehtäväsarjat riittävässä määrin vertailukelpoisiksi keskenään. Puheen ymmärtämisen koe osoittautui muita osakokeita vaativammaksi. Sijoitettaessa reseptiivisen kielitaidon (puheen ja tekstin ymmärtäminen sekä kieliopillinen kompetenssi) osioita CEF- ja OPS 2004 -taitotasoasteikoille ilmeni, että asiantuntijoiden arviot korreloivat heikosti empiiristen tulosten kanssa, ja että CEF -taitotasoasteikolla tasot A2 ja B1 olivat vaativia määrittää. Huolimatta siitä, että tarkistettu taitotasojen määrittäminen ei OPS 2004 -asteikolla onnistunut optimaalisesti, asteikot saatiin jokseenkin yhteismitallisiksi. Kun reseptiivisen kokeen tulokset sijoitettiin taitotasolle, voitiin todeta, että oppilaiden taitotaso CEF -asteikolla oli B1 ja OPS 2004 -asteikolla Suomen opetussuunnitelman perusteiden tavoitetasolla B1.1.

Produktiivisen kielitaidon osalta käytettäviä asteikkoja oli kolme: CEF, OPS 2004 ja A2001. Kahden ensiksi mainitun taitotasojen vastaavuus onnistuttiin määrittämään melko hyvin, kun taas A2001 -asteikon mukaiset taitotasoluokittelut jouduttiin tekemään jokseenkin harkinnanvaraisesti. Puhumisen kokeen mediaaniksi saatiin CEF -asteikolla B1, A2001 -asteikolla 4 ja OPS 2004 -asteikolla B1.1, mikä on tavoitetasoa yhtä tasoa korkeampi tulos. Vastaavat kirjoittamisen kokeen arvot ovat B1, 4, ja B1.1, mikä on myös noin yhden tason korkeampi kuin Suomen opetussuunnitelman perusteiden tavoitetaso A2.2. Koska ekstrapoloinnin pohjana oli suppea aineisto, produktiivisen kielitaidon koetuloksia on pidettävä vain suuntaa antavina.

Taulukoissa 89 ja 90 esitetään reseptiivisen kokeen frekvenssijakaumat ja prosenttiosuudet (oppilaat) CEF- ja OPS -taitotasoasteikoilla.

TAULUKKO 89 Reseptiivisen kokeen frekvenssijakauma ja prosenttiosuudet (oppilaat) CEF -taitotasoasteikolla arvioituina.

Taito- taso	Yleistaso (koko koet)		Puheen ymmärtäminen		Tekstin ymmärtäminen		Kieliopillinen kompetenssi	
	%	Fr	%	Fr	%	Fr	%	Fr
Alle A1	-	-	-	-	-	-	-	-
A1	5	237	8	375	10	472	5	228
A2	25	1218	23	1127	21	1030	26	1245
B1	41	2007	39	1887	40	1932	42	2051
B2	29	1392	30	1465	29	1420	27	1330

TAULUKKO 90 Reseptiivisen kokeen frekvenssijakauma ja prosenttiosuudet (oppilaat) OPS 2004 -taitotasoasteikolla arvioituina.

Taito- taso	Yleistaso (koko koe)		Puheen ymmärtäminen		Tekstin ymmärtäminen		Kieliopillinen kompetenssi	
	%	Fr	%	Fr	%	Fr	%	Fr
Alle A1-	-	-	-	-	-	-	-	-
A1.1	-	-	-	-	-	1	-	2
A1.2	-	-	-	3	-	13	-	1
A1.3	-	17	2	93	3	151	-	18
A2.1	8	384	10	472	9	427	8	385
A2.2	22	1054	19	934	19	910	22	1067
B1.1	32	1533	29	142631	1494	32	1574	
B1.2	28	1377	27	1321	25	1211	127	1288
B2.1	10	489	13	605	13	647	11	519

10 TUTKIMUKSEN LUOTETTAVUUDEN ARVIOINTI

Käsillä olevan tutkimuksen tavoite oli saada tietoa perusopetuksen yhdeksäsluokkalaisten englannin kielitaidon tasosta eli mille tasolle eri taitotasoasteikkoja käyttäen päästiin. Sen arvioimiseksi tarvittiin arviointimenetelmä, jonka tulosten avulla saatiin tietoa tästä tasosta. Arviointimenetelmän (esim. kokeen) käyttökelpoisuuden toteamisen kannalta olennaisia ominaisuuksia ovat sen reliaabelius ja validius (Bachman & Palmer 1996:23; Nunan 2001:119).

Mittaaminen arvioinnin muotona on arvioinnin laadun varmistamista. Siihen liittyy kvantifiointi, arvioinnin tulosten pisteittäminen ja näiden tulosten tutkiminen erilaisin matemaattisin ja tilastollisin mallein sen selvittämiseksi, sisältyvätkö arviointiin siltä edellytettävät ominaisuudet (McNamara 2000:56).

Tämän tutkimuksen toinen tavoite oli englannin kokeen uudelleen arvioinnin lisäksi löytää eri taitotasoasteikkojen vastaavuus. Seuraavassa kuvataan myös menetelmät, joilla tähän pyrittiin.

10.1 Reliaabelius

Reliaabelius määritellään yleensä mittauksen johdonmukaisuudeksi (Bachman & Palmer 1996:19), ja se osoittaa, missä määrin kokeen perusteella tehdyt erilaiset johtopäätökset, yleistykset, ovat perusteltuja (Linn & Gronlund 1995:81). Johdonmukaisuutta on kuitenkin tarkasteltava useasta eri näkökulmasta. Näistä näkökulmista perinteisesti keskeisiä ovat vakaus (*stability*), vertailtavuus (*alternate-form*) ja kokeen sisäinen johdonmukaisuus (*internal consistency*) (Popham 1999:24; Nitko 2001:77). Vakaus viittaa siihen, että mittaustulokset pysyvät eri mittauskerroilla samoina, koska kunkin kerran arviointitulokset ovat vain tietynä aikana saatu otos testattavien kaikista mahdollisista suorituksista. Tulosten luotettavuuden osoittamiseksi niiden on oltava ainakin tietyssä määrin johdonmukaiset eri suorituskerroilla (Linn & Gronlund 1995:81). Vertailtavuudella tarkoitetaan johdonmukaisuuden osoittamista kahden samantasoiseksi oletetun

kokeen (rinnakkaistestin) välillä. Sisäinen johdonmukaisuus kohdistuu kokeen osioiden toimivuuteen, jolloin osiot mittaavat samanlaista konstruktiota.

Reliaabeliutta arvioitaessa on otettava huomioon, että se viittaa mittarilla saatuihin tuloksiin eikä itse mittariin. Kokeen tai arvioinnin reliaabeliuden sijasta on siis puhuttava mieluummin koepisteiden tai arvioinnin tulosten reliaabeliudesta (Linn & Gronlund 1995:82; Nitko 2001:63). Reliaabelius viittaa myös aina tietyn tyyppiseen johdonmukaisuuteen eli tulokset voivat olla joltakin osin reliaabeleita ja joltakin toiselta taas eivät. Vaikka joskus asiasta käydään keskustelua (mm. Moss 1994:7), vallitseva käsitys on, että reliaabelius on myös välttämätön, vaikkakaan ei riittävä validiuden edellytys. Matala reliaabelius merkitsee samalla matalaa validiutta, vaikka korkea reliaabelius ei välttämättä takaa korkeaa validiutta.

10.1.1 Reseptiivisen kielitaidon kokeen reliaabelius

Taulukko 63 osoittaa eri tehtäväsarjojen reliaabeliuden sisäisen konsistenssin (Cronbachin α). Koko kokeen samoin kuin kielitiedon kokeen alfa -kertoimet ovat varsin korkeat. Puheen ja tekstin ymmärtämisen kokeiden kertoimet ovat selvästi matalammat, mikä johtunee ensi sijassa osioiden vähäisestä lukumäärästä. Reliaabelius on kuitenkin tutkimuksen kannalta riittävä.

10.1.2 Produktiivisen kielitaidon kokeen reliaabelius

Produktiivisen kielitaidon kokeet arvioitiin käyttäen kolmea eri taitotasoasteikkoa. Vuosina 2001 ja 2002 vieraiden kielten kansallisissa arvioinneissa käytetyllä A2001 -asteikolla arvioitiin 286 suullista ja 539 kirjoittamisen suoritusta. Nämä arvoinnit teki kaksi asiantuntijaa itsenäisesti ja heidän arviointiensa vastaavat korrelaatiot olivat korkeat kuten taulukosta 91 käy ilmi.

TAULUKKO 91 Puhumisen arvioinnin reliaabeliuskerroin (α) ja kirjoittamisen kokeen arvioijien välinen korrelaatio.

Arvioitava piirre	Arvioija				α	Arvioijien välinen korrelaatio
	Arvioija 1		Arvioija 2			
	Keskiarvo (maksimi 5)	Hajonta	Keskiarvo (maksimi 5)	Hajonta		
Puhuminen	3,8	1,2	3,9	1,2	.928	
Kirjoittaminen	3,7	1,2	3,8	1,2		.950
Luettavuus						.885
Ilmaisovarasto	3,6	1,1	3,6	1,0.		.889
Aiheen käsittely	3,7	1,3	3,5	1,2		.941
Kielelliset ansiot	3,2	1,2	3,2	1,2		

Arvioijien väliset korrelaatiot olivat siis korkeat. Puhumisen kokeessa yksimielisyys oli hyvinkin suuri, samoin kirjallisten suoritusten luettavuuden ja kielellisten ansioiden arvioinnissa. Näin ollen analyysien pohjana oleva arviointidata täytti laatuvaatimukset. Kahdella muulla, CEF- ja OPS 2004 -taitotasoasteikolla arvioitujen taitotasoasteikolla produktiivisen kielitaidon kokeiden taitotasoar-

viointien korrelaatioita ei laskettu siksi, että arviot olivat ns. konsensusarvioin- teja eli arvioijat arvioivat 52 suullista ja 50 kirjoittamisen tuotosta yhdessä siten, että ensiksi kumpikin arvioi suoritukset itsenäisesti ja sitten päädyttiin keskus- telun jälkeen konsensusarvioon, mitä esim. Dutch CEF Construct -projektissa (Grid) suositellaan. Molemmilla arvioijilla on pitkä ja monipuolinen kokemus arvioinnista. Toisella arvioijista on lisäksi pitkä ja laaja kokemus myös taito- tasoasteikkojen soveltamisesta ja arvioinnista yleensä. Työ suoritettiin mahdol- lisimman perusteellisesti ja huolellisesti. Tämän vuoksi tuottamisen taitotasojen asettamisen voidaan katsoa perustuneen riittävän luotettavaan aineistoon.

10.2 Validius

Validiudella tarkoitetaan sitä, että arviointituloksista tehdään oikeita tulkintoja ja että niitä käytetään asianmukaisella tavalla (Popham 1999:42; Nitko 2001:36). Edellinen liittyy kokeen pätevyYTEEN, sen mittauskykyyn, ja jälkimmäinen sii- hen, mihin toimenpiteisiin arviointitulosten perusteella ryhdytään. Arviointi- asiantuntijat katsovat, että validius on kokonaisuus, johon sisältyy useita vali- doinnin aspekteja (Nitko 2001:74). Popham (1999:46) pitää kolmijaossa erot- taen sisältö-, kriteeri- ja konstruktivalidiuden. Kriteerivalidius liittyy sellaisten taitotasokokeiden validointiin, joiden avulla ennustetaan kielitaidon tulevaa kehitystä (*proficiency tests*). Englannin arviointi oli saavutuskoee (*achievement test*), jossa arvioitiin opetussuunnitelman perusteiden määrittämää kielitaitoa, joten tässä tutkimuksessa keskitytään validoimaan tuloksia kokeen sisällön ja konstruktien osalta.

Erityisesti yleistyksiin pyrittäessä sisältövalidiuden toteamiseksi tulee vasta- ta seuraaviin kysymyksiin:

1. Miten hyvin osiot edustavat koko osioallasta, mitattavaa kielitaidon aluetta?
2. Miten hyvin osiot heijastavat nykynäkemyksiä arvioinnista? (Nitko 2001:46.)

McNamara (2000:52) korostaa koko kokeenlaadintaprosessin kuvausta sekä edellisten Nitkon esittämien kysymysten lisäksi tehtävätyyppien valinnan pe- rustelemista. Sekä Popham (1999:67) että Nitko (2001:36) ottavat lisäksi esille arvioinnin eettisen näkökulman edellyttäen testattavien tasapuolisen kohtelun osoittamista validoinnissa.

Käsitevalidointi (*construct validation*) on prosessi, jolla määritetään, missä määrin koesuoritus voidaan tulkita osoitukseksi konstruktin hallinnasta (mm. Luoma 2001). Konstruktit ovat yksilöllisiä ominaisuuksia, joiden oletetaan selit- tävän käyttäytymistä, esimerkiksi tekstin ymmärtämistä, ja ne määritetään koe- spesifikaatioissa. Kun arviointituloksia tulkitaan tietyn konstruktin mittarina, lähdetään siitä olettamuksesta, että on olemassa konstrukti, joka eroaa muista, ja että kokeen tulokset ovat juuri tämän konstruktin mittaustuloksia (Linn & Gronlund 1995:67-70).

Taitotasojen asettamiseen liittyy päätöksentekoprosessi, jossa koetulokset pyritään luokittelemaan tietylle määrälle peräkkäisiä portaittain vaikeutuvia suoritustasoja. Kun validius liittyy koetulosten tulkintaan ja koska taitotasojen asettamisessa on kysymys vasta pisterajojen määrittämisestä, katkaisukohtien validointi edellyttää ehdotettujen pisterajojen soveltavuuden ja uskottavuuden perustelemista (Kaftandjieva 2004:19). Taitotasojen asettaminen on prosessiluonteista ja perustuu arvopäätelmiin, joten sen uskottavuutta voidaan arvioida parhaiten kuvaamalla asettamisprosessi. Lisätukea validoinnille antaa se, että taitotasojen asettaja tuntee asteikot, hallitsee käyttämänsä menetelmät ja on puolueeton (Kaftandjieva 2004:20).

10.2.1 Kokeen sisältö- ja käsitevalidius

Kokeen järjestämisen tarkoituksena oli osoittaa, millainen 9. -luokan oppilaiden englannin kielen taso oli vuoden 1994 Perusopetuksen opetussuunnitelman perusteiden tavoittein arvioituna. Nämä perusteet edustavat kommunikatiivista kielitaitokäsitystä, mikä sisältää ajatuksen siitä, että taitoja mitataan produktiivisen kielitaidon alueella suorituserviöintinä ja reseptiivisen kielitaidon alueella tarjoamalla lukijalle ja kuulijalle autenttista, ikäkaudelle sopivaa, kiinnostavaa ja opiskeltuun ainekseen perustuvaa aiheiltaan monipuolista tekstiä (luvut 4 ja 6). Arviointitavaksi valittiin kriteerivitteinen arviointi, koska analysoitava kielitaidon alue oli periaatteessa rajattu ja tietyllä tarkkuustasolla määritelty (luku 5).

Kokeen laadinta noudatti Bachmanin ja Palmerin (1996) kokeen laadinnan mallia, jolle Yleisten kielitutkintojen arviointimalli perustuu. Arviointiprosessiin sisältyivät suunnittelu-, operationaalistamis- ja toimeenpanovaiheet. Kokeen laatijat olivat opetustyössä kokeneita ja ansioituneita opettajia, ja kokeen kehittäjä seurasi kieltenopetukseen ja arviointiin perehtynyt asiantuntijaryhmä. Kokeenlaatijat suorittivat myös kouluissaan esikokeilun, johon osallistui lisäksi muutama muukin koulu. Esikokeilu, ohjausryhmän asiantuntemus ja kokeenlaatijoiden kokemus kohdealueesta edesauttoivat sitä, että koetta voidaan pitää riittävän edustavana otoksena opiskellusta alueesta, vaikka tästä ei ole esitettävänä yksityiskohtaista evidenssiä.

Vuoden 1994 opetussuunnitelman perusteiden tueksi laadittu opas jakoi kielitaidon perinteisen tavan mukaan neljään osa-alueeseen. Tätä jakoa noudatettiin myös arvioinnissa, johon kuitenkin lisättiin myös kieliopillisen kompetenssin koe mm. siksi, että kirjoittamisen tuotosten ei katsottu antavan riittävästi tietoa tarkasta kieliopillisten rakenteiden hallinnasta. Tekstin ja puheen ymmärtämisen sekä kieliopillisen kompetenssin kokeista laadittiin testispesifikaatiot Aldersonin ym. (1995) ja Bachmanin ja Palmerin (1996) ohjeita soveltaen. Ymmärtämisen konstruktit perustuivat Weirin (1993) luokitteluihin operaatioihin. Kieliopillisen kompetenssin sisällöt koottiin kaikista käytössä olleista A-englannin oppikirjoista ja valittiin esikokeiluun oletetun frekvenssin ja vaikeustason perusteella. Produktiivista kielitaitoa mittaavat tehtävät valittiin samoin oppimateriaaleissa esiintyvien aiheiden joukosta.

Puheen ja tekstin ymmärtämisen sekä kielitiedon kokeiden tehtävätyypit olivat äidin- ja kohdekielistä monivalintaa, oikein/väärin-väittämiä, täydennys-

tä ja lyhyitä vastauksia edellyttäviä äidinkielisiä kysymyksiä. Monivalinnan sisällyttämistä perusteltiin kokeen korjaamisen nopeuttamisella, O/V-väittämiä riittävän helpon vaihtoehdon tarjoamiselle myös heikommille oppilaille, jotta saataisiin riittävästi näyttöä myös alempitaisoisesta ymmärtämisestä. Usean tehtävätyypin käyttämisen tarkoituksena oli myös välttää sitä, että ns. tehtävätyyppiefekti vinouttaisi tuloksia. Erilaiset tehtävätyypit johtivat kuitenkin jossakin määrin ongelmiin vastausten merkitsemisessä. Negatiivista palautetta tuli optisen lomakkeen täyttämistä, mihin oppilaat eivät olleet tottuneet ja jota ei etukäteen opastettu harjoittelemaan.

Eettiset ohjeet pyrittiin ottamaan huomioon mm. aihepiirien valinnassa. Vain otokseen osuneet koulut saivat tiedon osallistumisestaan. Ohjeissa myös korostettiin arvioinnin tarkoitusta, joka oli kartoittaa kielitaidon tasoa, ei olla oppilasarvioinnin peruste. Kaikki koulut saivat tiedot kokeen keskimääräisistä tuloksista, mutta koulu- ja oppilaskohtaiset tiedot toimitettiin vain asianomaisille kouluille, ei esimerkiksi kouluvirastoihin. Oletettua opettajien kokemattomuutta puhumisen kokeen järjestämisessä ja oppilaiden tottumattomuutta kokeeseen osallistumisessa yritettiin helpottaa toimittamalla näistä asioista ohjeita.

Kokeen laadinnassa noudatettiin siis kommunikatiivisen kielitaidon kokeen¹ periaatteita, joiden mukaan kieltä käytetään kommunikointiin tietyssä tilanteessa ja tiettyyn tarkoitukseen kiinnittäen huomiota siihen, että tilanteet ja tarkoitukset olisivat mahdollisuuksien mukaan sellaisia, jotka voivat esiintyä todellisessa elämässä. Sisällön analyysi tehtiin konstruktit testispesifikaatioissa määrittäen, joten kokeen rakenteen voidaan katsoa vastanneen riittävän hyvin sen taustalla olleita teorioita.

10.2.2 Taitotasojen asettamisen validius

Taitotasojen asettamisen tavoitteena oli sijoittaa englannin kansallisen arvioinnin reseptiivisen kielitaidon ja kielitiedon koetulokset CEF - ja OPS 2004 -taitotasolle ja produktiivisen kielitaidon tulosten sijoittaminen CEF-, OPS 2004 ja A2001-taitotasolle. Tätä varten oli arvioitava em. kokeiden vaikeustaso näillä asteikoilla. Tämän arvioinnin suorittivat samat asiantuntijat, jotka arvioivat produktiivisen kokeen otoksen taitotasot.

Tilastollisia analyyseja varten asiantuntijoiden eri taitotasoasteikoille sijoittamat arviot ymmärtämis- ja kielitiedon koetehtävien vaikeustasosta yhdistettiin empiiriseen tietoon tehtävien todellisesta vaikeustasosta, joka perustui alkuperäiseen dataan, mutta joka oli nyt osioiden osalta koodattu analyyysien toimivuuden vuoksi dikotomiseksi niiltä osin, jossa oli alun perin käytetty polytomista asteikkoa. Tämän jälkeen siirryttiin taitotasojen asettamisen menettelyyn, mikä, johtuen yleisesti muissakin yhteyksissä todetusta suhteellisen heikosta korrelaatiosta asiantuntija-arvioiden ja todellisten koetulosten välillä, vaati useamman yrityksen, koska taitotasojen asettaminen on suuressa määrin harkin-

¹ Buck (2001:85) toteaa, että kommunikatiivinen koe on vaikeasti määriteltävissä joutuessa siitä, että eri ihmiset käyttävät kieltä kovin eri tavoin.

nanvarainen eikä vain matemaattinen toimenpide. Tämän jälkeen oli mahdollisuus todeta, miten oppimistulokset sijoittuivat eri asteikkojen taitotasolle ja miten nämä vastasivat toisiaan.

Kun asiantuntijoiden arviot ja empiiriset koetulokset yhdistettiin, todettiin, että näiden yhdistelmien korrelaatio CEF -asteikkoon oli 0,348 ja OPS 2004 -asteikkoon 0,294. Vaikka korrelaatiot ovat alhaisia, ne ovat kuitenkin samaa tasoa tai korkeampia kuin useissa muissa vastaavissa arvioinneissa. Osioiden sijoittaminen CEF -asteikolle onnistui paremmin, sillä taitotasoarviot heijastavat johdonmukaisesti vaikeustason nousua, kun taas OPS 2004 -asteikolla tällaista täysin johdonmukaista nousua ei ollut havaittavissa. Tämä johtuneen ensisijassa siitä, että taitotasojen määrä on suurempi OPS skaalassa.

Tämän jälkeen laskettiin osioiden vaikeustasojen keskiarvo tasoittain ja järjestettiin osiot molemmille taitotasoasteikolle vaikeustasoittain. Kun sitten laskettiin tasoittain raakapisteiden kumulatiivinen summa ja muutettiin raakapisteteet theta -asteikollisiksi, voitiin määrittää taitotasorajat. Oppilaiden tulokset sijoitettiin sitten näille tasoille, jolloin saatiin selville oppilaiden kokonaiskielitaidon ja osa-alueiden taso sekä CEF - että OPS 2004 -asteikolla. Koska asiantuntija-arvioiden korrelaatio empiiriseen suoritustasoon oli alhainen, OPS 2004 -taitotasorajoja voidaan pitää vain suuntaa antavina. Sen sijaan CEF -arviointia voidaan pitää varsin luotettavana.

Reseptiivisen kielitaidon kokeen taitotasojen asettamisen validius

Termi ´taitotasojen asettaminen` viittaa päätöksentekoprosessiin, jossa pyritään luokittelemaan tutkinto- tai koetulokset rajalliselle määrälle peräkkäisiä portaittain vaikeutuvia suoritustasoja. Suuri määrä koepisteitä siis ikään kuin puristetaan joukolle taitotasoa. Taitotasojen asettaminenkin on validoitava. Kysymykseen tulevat prosessiin ja yleistettävyyteen liittyvät todisteet. Kaftandjieva (2004:20) katsoo, että prosessi voidaan validoida selostamalla taulukossa 92 luetellut menettelyt.

TAULUKKO 92 Taitotasojen validointi (Kaftandjieva 2004)

Taitotasojen asettamisprosessi
Taitotasojen asettamisen tarkoituksen ja vastaavien konstruktien määrittely.
Käytettyjen suoritusstandardien määrittely.
Käytettyjen taitotasojen asettamisen menetelmien kuvaus ja perustelut niiden käyttämiselle.
Arvioijien valinta
Arvioijien koulutus
Arvioijapalaute koskien taitotasojen asettamisen ja arviointityön sisäistämistä sekä sitä, miten tyytyväisiä ollaan tasoa sekä prosessia että lopullisia pisterajojen katkaisukohtia kohtaan.
Tietojen keruun kuvaus.
Taitotasojen määrittämisen kuvaus.
Mahdollisesti käytettyjen taitotasorajojen muokkaus-/täsmennysmenettelyjen kuvaus.

Koska taitotasojen asettaminen on suuressa määrin harkintaa vaativaa toimintaa, tulosten johdonmukaisuus ja toistettavuus eivät takaa ehdotettujen katkaisukohtien riidattomuutta, vaikkakin johdonmukaisuuden puute saattaa vaaran-

taa vakavasti katkaisukohtien uskottavuuden. Mitä useampia lähteitä käytetään yleistettävyyden takaamiseksi, sitä vahvempia todisteet ovat ja tukevat näin ehdotettujen tulkintojen oikeellisuutta. Kaftandjieva (2004:21-22) luettelee seuraavia lähteitä yleistettävyyden takaamiseksi:

- katkaisukohtien arvioinnin tarkkuus, johon vaikuttaa myös itse kokeen reliabiliteetti,
- arvioijien välinen johdonmukaisuus,
- arvioijien oman työn johdonmukaisuus,
- päätöksenteon johdonmukaisuus ja täsmällisyys sekä
- tiedot arvioitavien aikaisemmasta menestyksestä.

Nämä toimet, viimeistä lukuun ottamatta, pyrittiin suorittamaan mahdollisimman huolellisesti. Katkaisukohtien mittavirhe voidaan estimoida empiirisesti ja verrata sitä koko kokeen mittavirheeseen, mutta sitä ei tässä vaiheessa suoritettu. Jatkossa myös siihen on syytä ryhtyä.

Taitotasojen asettamisen ulkoinen validointi edellyttäisi tyypillisesti sitä, että arvioitujen oppilaiden tuotokset jostakin muusta kokeesta, joka olisi jo linkitetty CEF:iin, olisivat tiedossa. Tämä ei ole ollut mahdollista, koska vastaavia kansallisia arviointeja ei ole OPS 2004:n voimassaoloaikana järjestetty, ja niin muodoin tämä tutkimus on ensimmäinen laatuaan Suomessa. Linkityksen uskottavuus on siis arvioitava kaiken sen tiedon perusteella, mikä odotetusta kokeiden tasosta on ollut käytettävissä.

Reseptiivisten kokeiden linkittäminen suoritettiin CEF -käsikirjassa suositellun prosessin mukaisesti. Osioiden vaikeustason määrittämiseen käytettiin apuna CEF:iin liittyvän Gridin antamaa konkreettista ja yksinkertaisempaa mallia.

Edellä on raportoitu menettelytavat, ja niiden voi arvioida onnistuneen yleisesti ottaen riittävän hyvin. Näin ollen reseptiivisen kokeen taitotasolinkitystä voi pitää riittävän validina.

Produktiivisen kielitaidon kokeiden taitotasojen asettamisen validius

Koska tutkimusaineistoon kuuluvat puhumisen ja kirjoittamisen suoritukset arvioitiin käyttämällä suoraan CEF -, OPS 2004 - ja A2001 -taitotasoasteikkoja, taitotasojen asettamista ei tarvinnut suorittaa samalla tavalla epäsuorasti kuin reseptiivisten taitojen kohdalla. Erityisesti CEF - A2001 -korrelaatio (taulukko 93) oli melko matala, mikä johtunee siitä, että CEF -asteikon kuvaukset kattavat kielitaidon alkeista hyvin kieltä hallitsevan, äidinkielenään kieltä puhuvan henkilön kielitaitoon, kun taas A2001 pyrki kattamaan vain perusopetuksen oppilaiden oletetun kielitaidon. On melko todennäköistä, että koska maassamme on jo pitkä traditio kirjoittamisen arvioinnissa, sen tulokset muodostuvat helpommin yhtäpitäviksi kuin puhumisen tulokset. Toisaalta OPS 2004-taitotasot on laadittu pitkälti CEF -taitotasojen mukaisiksi.

TAULUKKO 93 Eri taitotasoasteikkojen väliset korrelaatiot asiantuntijoiden suorittamassa puhumisen ja kirjoittamisen arvioinnissa.

Taitotasoasteikot	Puhuminen	Kirjoittaminen
OPS2004 - A 2001, sujuvuus	0,607	0,680
CEF - A2001, sujuvuus	0,565	0,701

Asteikot tuli linkittää toisiinsa. Tässä käytettiin korrespondenssianalyysia, joka tuottaa havainnollisen kuvion osoittamaan, mitkä taitotasot lähinnä vastaavat toisiaan. Vastaavuus on vähintään tyydyttävä. Suurempi otos oppilaiden tuotoksia olisi kuitenkin johtanut tässä menettelyssä ainakin jossakin määrin luotettavampaan tulokseen.

Kuten edellä todettiin, taitotasojen asettamisprosessi noudatti suurelta osin niitä periaatteita, jotka on kirjattu CEF -käsikirjaan ohjeeksi Eurooppalaisen viitekehyksen taitotasojen linkittämiseksi muihin taitotasoasteikkoihin. Oppilasmäärien jakauma oli oikean suuntainen niin, että keskiarvo- tai mediaanitasolle asetui suurin määrä tuloksia, joiden määrä laski asteikon ala- ja yläpään suuntaan. CEF -taitotasoarviointeja voi pitää varsin luotettavina reseptiivisen kielitaidon arvioinnin osalta ja OPS 2004:n arviointeja vartenotettavina ja suuntaa antavina.

10.3 Tutkimuksen ongelmia

Luotettava taitotasojen asettaminen pohjautuu huolelliseen arviointiprosessiin ja laadukkaisiin koeosioihin. Huolellinen arviointiprosessi taas vaatii perustakseen arvioinnin eri tavoitteiden ja käsitteiden selventämisen. Tämän vuoksi tutkimukseen on sisällytetty huomattava määrä edellä mainittujen aiheiden teoreettista käsittelyä, joka saattaa lukijasta tuntua liialliseltakin. Ne ovat kuitenkin aspekteja, joita arviointiprosessin aikana on jouduttu runsaasti pohtimaan ennen kuin arviointi toimeenpantiin valitsemalla näistä teorioista olosuhteisiin ja resursseihin parhaiten soveltuva yhdistelmä.

Kielitaidon teorioita ja taksonomioita pyrittiin hyödyntämään mahdollisimman monipuolisesti, vaikka koeosioiden rajallisen määrän vuoksi vain pieni osa kyettiin operationaalistamaan. Toteutunutta koetta on niin muodoin mahdollisuus verrata teoriapohjaan ja kytkeä koeaineiston kuvaukset sekä pohjana olleen opetussuunnitelman perusteisiin että teoriaosan esitykseen. Siihen ei ole paneuduttu tässä tutkimuksessa siksi, että sen pääkohde oli kuitenkin taitotasojen asettamisessa ja IRT -analyysin avulla saaduissa koetuloksissa.

Taitotasojen asettaminen edellyttää tavallisesti usean asiantuntijan käyttöä osioiden vaikeustasoa arvioitaessa. Tämän tutkimuksen puutteeksi voidaan katsoa asiantuntijoiden vähäinen määrä, ja siitä johtuen, näin voitaisiin olettaa, asiantuntijoiden reseptiivisen kielitaidon kokeiden koeosioiden vaikeusarvioiden heikot korrelaatiot empiirisiin tuloksiin, suhteessa CEF- asteikkoon 0,348 ja OPS 2004 -asteikkoon 0,294. Koska tällaista arviointia ei Suomessa vielä ole juuri tehty

johtuen taitotasoarviointien soveltamisen uutuudesta, kokeneita arvioijia ei olisi ollut käytettävissä. Muissakin arviointitutkimuksissa, joissa on ollut käytettävissä paljon useampia arvioitsijoita, vastaavat korrelaatiot ovat olleet yhtä matalia tai jopa matalampia. Linkitettäessä B-ruotsin ylioppilaskoetuloksia CEF -asteikkoon 14 arvioijaa arvioi koeosiot, ja heidän korrelaationsa oli 0,239. He suorittivat työn pääosin itsenäisesti, kun taas tämän tutkimuksen yhteydessä käytettiin konsensusarviointia, mitä mm. CEFin Grid -projekti pitää suositeltavana menetelmänä. Konsensusprosessin ”laadun” toivottiin korvaavan määrän.

IRT -analyysien edellyttämää yksidimensionaalisuutta ei kyetty todentamaan tavallisesti näissä yhteyksissä käytetyillä analyyseilla, koska suuren kokonsa vuoksi aineisto ei sopinut mihinkään matemaattiseen malliin. Yksidimensionaalisuus voitiin kuitenkin todeta pääkomponenttianalyysillä. Produktiivisen kielitaidon kokeissa otoksen pienuus puolestaan johti epätarkkuuteen OPS 2004 -taitotasorajojen asettamisessa, koska kullekin 10:sta tasosta sijoittui noin 50 oppilaan otoksesta liian vähän oppilaita luotettavien johtopäätösten tekemiseksi. A2001 -taitotasoasteikkoa käytettäessä taas matala korrelaatio teki tämän asteikon taitotasorajat epätarkoiksi. A2001 -asteikko poikkesi muista siinä, että se oli alkujaan tarkoitettu vain perusopetuksen oppilaiden arviointiin, kun taas molemmat muut taitotasoasteikot kattavat kielitaidon kuvaukset alkeistasolta mestarin tasolle. Otoksen määrän tulee siis olla sitä suurempi mitä useammalle taitotasolle tuloksia tulee sijoittaa. Ilmeisesti taitotasoasteikkojenkin tulee olla suunniteltuja samantasoisille testattaville taitotasojen asettamisen ja tulosten linkittämisen mahdollistamiseksi.

Kuten luvussa 7 todettiin, eri taitotasojen asettamismenetelmillä saatetaan saada erilaisia tuloksia. Tässä tutkimuksessa oli käytetty vain yhtä menetelmää. Prosessi on kuitenkin siinä määrin vaativa, että resurssit eivät riittäneet monipuolisempaan käsittelyyn. Tämä tutkimus osoittaa kuitenkin sen, miten menetelmä toimii, ja koska sitä on käytetty useissa muissakin arvioinneissa, se on todettu käyttökelpoiseksi.

10.4 Yhteenveto

Reseptiivisen kielitaidon kokeiden koko kokeen ja kieliopillisen kompetenssin kokeen eri tehtäväsarjojen reliabeliuskertoimet (α) olivat varsin korkeat. Puheen ja tekstin ymmärtämisen kokeiden kertoimet olivat selvästi matalammat, mikä johtunee ensi sijassa osioiden vähäisestä lukumäärästä. Reliabelius on kuitenkin riittävä.

Produktiivisen kielitaidon kokeiden suuremman otoksen A2001 -asteikolla arvioi kaksi asiantuntijaa itsenäisesti, ja heidän keskinäiset korrelaationsa olivat varsin korkeat. CEF- ja OPS 2004 -taitotasoasteikkoja käytettäessä arvioijat, joista vain toinen oli sama kuin edellisessä arvioinnissa, arviot olivat ns. konsensusarviointeja, jolloin korrelaatiota ei ollut tarpeen laskea. Yksimielisyys saavutettiin yleensä varsin helposti.

Tämän tutkimuksen käsitevalidiutta eli sitä, että koetuloksista on tehty oikeita johtopäätöksiä (koesuoritus voidaan tulkita osoitukseksi konstruktin hallinnasta), on pyritty takaamaan esittelemällä tutkimuspohjaisia näkemyksiä arviointiprosessista ja kielitaidon käsitteestä sekä vastaavasti kuvaamalla arviointiprosessi ja kokeiden sisältö. Taitotasojen asettamisen validointi suoritettiin sille asetettujen vaatimusten mukaisesti.

Tutkimuksen ongelmiksi kirjattiin reseptiivisen kielitaidon kokeen otospopulaation suuri koko, joka vaikeutti jossakin määrin IRT -analyysien tekemistä sekä produktiivisen kielitaidon kokeen otospopulaation pienuus, mikä johti OPS 2004 -taitotasojen asettamisen epätarkkuuteen. Puheen ja tekstin ymmärtämisen kokeiden reliabelius olisi myös edellyttänyt useampia osioita. A2001 -taitotasoasteikon linkittämistulos on epätarkka, mikä johtunee siitä, että sen kohderyhmä poikkesi muista. Taitotasojen asettamisessa edellytetään yleensä vähintään kymmentä arvioijaa samoin kuin useamman kuin yhden arviointimenetelmän käyttämistä. Tässä tapauksessa vähempi määrä pääsi yleisesti ottaen melko hyvään keskinäiseen korrelaatioon. Resurssit eivät puolestaan riittäneet useamman taitotasojen asettamismenetelmän soveltamiseen. Prosessin voidaan kuitenkin katsoa kokonaisuudessaan onnistuneen varsin hyvin.

Tässä on kommentoitu selvimmin ilmenneitä tämän tutkimuksen puutteita. Johtopäätökset -luvussa otetaan kantaa myös joihinkin varsinaisen arvioinnin toimeenpanemisessa nousseihin kysymyksiin ja esitetään eräitä lisätutkimusaiheita.

11 JOHTOPÄÄTÖKSET

Tämän tutkimuksen laadinnan kuluessa ja tulkittaessa sen analyyssejä sekä jo suoritettaessa alkuperäistä arviointia vuosina 1998-9 nousi esille muutamia sekä yksityiskohdaisia että laajempaankin pohdintaan aihetta antavia ajatuksia. Tässä luvussa tehdään ehdotuksia arviointiprosessin ja kansallisten arviointien kehittämistä sekä eräistä tämän tutkimuksen yhteydessä nousseista tutkimusaiheista, ja perustellaan IRT -analyysien käyttöä kriteeriviitteisen arvioinnin tulosten esittämisessä ja linkittämisessä taitotasoasteikkoihin.

11.1 Tutkimuksessa todettua

Kokeet oli pyritty laatimaan huolellisesti seuraten Aldersonin (2000) ohjeita testispesifikaatioiden sisällöstä. Kuten luvussa 5 todettiin, eräiden kohtien huomiointiin ottaminen jäi kuitenkin puutteelliseksi. Produktiivisen kielitaidon tuotosten arviointiin ei ollut käytettävissä kriteerejä, minkä vuoksi arviointi kouluissa oli suoritettava niiden itsensä määrittämän tasoon nojautuen. Toisaalta taitotasoasteikkojen ja kriteerien soveltaminen oli vielä 1999 siinä määrin uutta, että tulosten luotettavuus ei ehkä olisi parantunut kovinkaan paljon. Koulujen arvioinnin tueksi ei myöskään voitu tarjota maamerkkituotoksia tai -tehtäviä, eikä myöskään tukea optisen vastauslomakkeen käytön harjoittelua. Tällaista tukiaineistoa tulisi olla käytettävissä jokaisessa arvioinnissa, ja se tulisi toimittaa kouluihin yhdellä kertaa ja riittävän aikaisin ennen koetta.

Jos reseptiivisen kielitaidon koetehtävinä käytetään ns. avokysymyksiä, niiden vaikeustason hierarkiaan (ks. luku 6.1.1), yksiselitteisyyteen ja pisteitykseen tulisi kiinnittää vakavaa huomiota. Polytomisen pisteityksen (1-3 p.) käyttö ei välttämättä johda luotettaviin tuloksiin. Siksi tulisi harkita pitäytymistä dikotomisessa pisteityksessä (0/1 p.) ja sisällyttää kokeisiin mahdollisimman paljon strukturoituja tai puolistrukturoituja tehtäviä.

Jos oppimistuloksista halutaan kattavampi kuva, käytettävien tehtäväsarjojen (versioiden) määrää tulisi lisätä kolmesta vähintään viiteen, mikä oli tä-

mänkin tutkimuksen pohjana olevan arvioinnin alkuperäinen tavoite. Vaihtoehtoisesti tulisi pohtia nykyistä laajempien kokeiden rotatointia oppilaiden kesken tai arviointisyklin nopeuttamista. Näin voitaisiin lisätä puheen ja tekstin ymmärtämisen koeosioden määrää, millä laajennettaisiin konstruktiedustusta ja parannettaisiin näiden osa-alueiden kokeiden reliaabeliutta. Näin koeaika ei lisäännä vaikka osiomääriä voitaisiin nostaa. Kielitaidon kaikkia osa-alueita tulisi edelleenkin arvioida.

Tekstin ja puheen ymmärtämisen tehtävävihkoissa oli yhteinen ankkuri-tehtävä seuraavaan tapaan (a = yhteinen tehtävä):

Tehtävävihko 1

a	b	c
-----	-----	-----

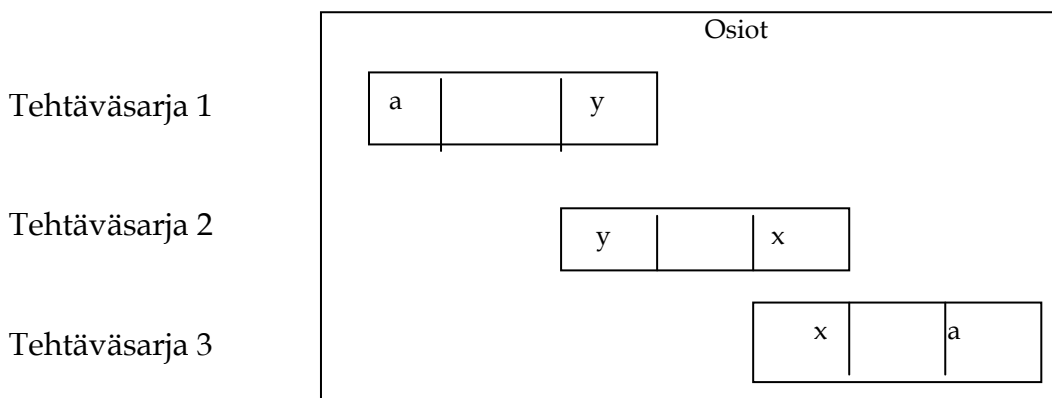
Tehtävävihko 2

a	d	e
-----	-----	-----

Tehtävävihko 3

a	f	g
-----	-----	-----

Yhden yhteisen tehtävän sijasta parempi asetelma olisi ollut linkittää vihkot toisiinsa esimerkiksi niin, että vihkoissa 1 ja 2 olisi ollut yksi yhteinen tehtävä, vihkoissa 2 ja 3 toinen ja vihkoissa 3 ja 1 kolmas (Kaftandjieva 1999:97; Verhelst 2004b:6).



Tässä tutkimuksessa kävi selvästi ilmi, miten vaikeata reseptiivisen kielitaidon tekstien ja koeosioden luotettava arviointi on. Se johtuu kokemuksen ja vastaavasti palautteen puutteesta. Ongelma on tunnistettu kansainvälisestikin, mitä osoittaa mm. Gridien laadintaprojekti. Niissä tutkimuksissa, joissa ymmärtämistehtävien vaikeustaso on arvioitu, arvioijia on ollut yleensä suositusten mukaisesti runsaat kymmenen. Käsillä olevassa tutkimuksessa arvioinnin suoritti konsensusarviointina kaksi asiantuntijaa, kuten Gridien laatijat ehdottavat. Näiden arvioiden suhde koeosioden empiiriseen vaikeustasoon oli jopa hieman korkeampi kuin joissakin vastaavissa muissa arvioinneissa. Arviointi perustuu siihen, että arvioijat tulkitsevat kriteereitä samalla tavalla. Tähän päästään vain kokemuksen kautta. Sen taustalla on kuitenkin oltava yhteistyötä eri arvioijien kanssa, mihin päästään vain koulutuksen avulla. Kouluttajat ovat ihannetapauksessa henkilöitä, joiden korrelaatiot tiettyihin empiirisiin koetuloksiin ovat mahdollisimman korkeita. Koulutuksen suuntaaminen on tosin

nykyisellään vaikeata, koska ei ole olemassa kriteereitä sille, kuka voi toimia arvioijana ja kokeiden tai oppimateriaalien laatijana. Koulutus tulisi suunnata ensi sijassa heihin.

Perustaksi reseptiivisen kielitaidon koetehtävien ja -osoiden vaikeustason arvioinnille tarvitaan ymmärtämisprosessien tutkimusta. Niitä ei tunneta riittävästi, vaikka sekä kielitieteen että neurologian asiantuntijat työskentelevät aktiivisesti päästäkseen lähemmäs ratkaisua. On tiettyä näyttöä siitä, että avotehtävien vaikeustaso usein aliarvioidaan ja monivalintatehtävien taso yliarvioidaan. Samoin on todettu, että pyrittäessä laatimaan C -tason tekstin ymmärtämisen tehtäviä, ne osoittautuvat usein empiirisessä tarkastelussa helpommiksi. Mm. näissä tapauksissa ymmärtämisprosessin tunteminen tukisi kokeen laadintaa. Myös tuottamisprosesseja tulisi tutkia paitsi tuotosten vertailtavuuden mahdollistamiseksi myös tuottamistaidoissa esiintyvien vaikeuksien vähentämiseksi. Näin arviointia voitaisiin tulosten ja tuotosten arvioinnin lisäksi suunnata prosessien arviointiin, erityisesti muutamilla pedagogisesti kriittisillä luokka-asteilla. Niitä ovat mm. 4. ja/tai 6. luokka (vrt. Ruotsi; luku 3), jolloin oppimisvaikeuksiinkin kyettäisiin puuttumaan riittävän aikaisin. Näin saataisiin tietoa myös siitä, millaisia kielellisiä piirteitä ja erityisiä ongelmia eritasoisilla oppijoilla esiintyy. Sekä Suomessa että muualla on jo alkamassa tutkimuksia tästä (ks. luku 11.4). Niistä toivotaan saatavan kielikohtaista, tarkempaa tietoa mm. oppimateriaalien ja diagnostisen arvioinnin tarpeisiin. Käsillä olevassa tutkimuksessa osoittautui myös vaikeaksi arvioida Suomen opetussuunnitelman perusteiden 2004 taitotasoa A1.1-A1.3. Nämä kuvaavat lähinnä alakoulun oppilaiden taitoja. Tutkimuksella saataisiin tähänkin taitotasoarviointiin lisävalaistusta. Entisen ala-asteen oppimistulosten kansallista arviointia suunniteltiinkin vuonna 2001, mutta tuolloin asia raukesi. Nyt se olisi paikallaan ottaa uudelleen esille.

Noin kolmasosa oppilaista oli vuonna 1999 käsillä olevan tutkimuksen mukaan vuoden 2004 opetussuunnitelman perusteiden tavoitteen mukaisella reseptiivisen kielitaidon tasolla B1.1. Lähes 40 % oppilaista ylitti tämän tason. On positiivista, että prosentuaaliset osuudet vähenevät alemmille ja nousevat ylemmille taitotasolle siirryttäessä. Lukion päättövaiheen tavoitetaso on uusien opetussuunnitelman perusteiden mukaan B2.1 eli taso, jolla vuonna 1999 oli runsaat 10 % peruskoulun päättävistä oppilaista. Jos kuitenkin perusopetuksen taso on näin korkea, loppuuko asteikko parhaiden lukiolaisten osalta? Toisaalta voisi myös kysyä, miksi ei edes osa tasoilla B1.2.- B2.1 olleista n. 40 %: sta oppilaita yltänyt taitotasolle B2.2. Yksi selitys voi olla, että osaavimmille oppilaille ei ollut tarjolla heidän kykyjään vastaavia tehtäviä. Kokeiden valinnassa oli pyritty noudattamaan opetussuunnitelman perusteiden määritelmää helpohkon tekstin ymmärtämisestä (luku 9) ja valittu Grideillä ja ennakoarvioinninkin mukaan korkeintaan B1 -tasoisia tehtäviä. Koska tulokset ovat melko hyviä, olisi syytä panostaa vaativampienkin reseptiivisen kielitaidon tehtävien sisällyttämistä kokeisiin.

Produktiivisen kielitaidon kokeiden tulokset ylittivät opetussuunnitelman perusteiden tavoitetason sekä puhumisessa että kirjoittamisessa keskimäärin

yhdellä tasolla. Otos, johon suurempi otos ekstrapoloitiin, oli tosin suppea, joten tuloksia voidaan pitää vain suuntaa antavina. Ne antavat kuitenkin aiheen kysyä, pitäisikö tasokuvauksien vaatimustasoa nostaa välillä B1 - B2 ja B1.1 - B2.1. Kirjallisen tuottamisen odotuksia selvästi korkeampi tulos panee miettimään, onko se kenties takaistusvaikutusta ylioppilaskokeesta, jossa kirjoitelman tuottaman pistemäärän osuus on noin 1/3 kokonaispistemäärästä, mutta jossa puhumista ei testata lainkaan. Puhumisenkin taso oli tosin tavoitetasoa korkeampi, mutta se voi johtua myös helpoista tehtävistä. Koe poikkesi kirjoittamisen kokeesta siinä, että puhekokeessa kaikkien oli suoritettava jotakuinkin samat tehtävät (taulukko 59), kun taas kirjoitelman aiheen saattoi itse valita. Kirjoittamisen tulos on sikäläkin yllättävä, että oppilaat pitivät kirjoittamista kieliopin jälkeen toiseksi vaikeimpana osa-alueena (Tuokko 2000:104) ja opettajat katsoivat harjoittavansa sitä vähemmän kuin muita osa-alueita (Tuokko 2000:111).

Kieliopillisen kompetenssin kokeessa jäi vain pieni määrä oppilaita alle A2-tason, mikä on myönteistä, koska oppilaat itse pitivät yleensä kielioppia vaikeana. Suomi menestyi juuri kieliopillisen kompetenssin kokeessa parhaiten kahdeksan Euroopan maan arvioinnissa, ja perinteisesti täällä on panostettu siihen johtuen suomen kielen erilaisuudesta moniin muihin kieliin verrattuna. Kommunikatiivisuuden painottaminen ei (onneksi) ole heikentänyt kieliopillista kompetenssia, vaikka tätä on pelätty, ja mikä Köhlmyrin (2003) tutkimuksen mukaan on tosiasia Ruotsissa. Kommunikatiivisen kompetenssin tavoittelu ei kuitenkaan merkitse kieliopillisen tarkkuuden tai laajaan, monipuoliseen sanavarastoon pyrkimisen hylkäämistä. Entisistä vahvuuksista tulee pitää kiinni ja uusia lisätä.

Heikommin onnistuneiden oppilaiden tekstin ymmärtämisen kokeen tulos on hieman yllättävä, koska sekä puheen ymmärtämisessä että kieliopillisen kompetenssin kokeessa menestytettiin paremmin. Myönteistä kokonaisuuden kannalta on, että taitotasolle A1 (1, 2, 3) jäi vain 5-10 % oppilaista; useinhan alisuoriutujia todetaan olevan noin 20 %.

Ymmärtämisen arviointia samoin kuin sanaston laajamittaisempaa arviointia vaikeuttaa se, että oppilaiden sanavarasto ei ole yhteismitallinen. Frekvenssisanalistojen tuntemusta ja korpusten, erityisesti elektronisten, käyttöä tulisi lisätä ja jollakin tavalla varmistaa, että oppimateriaalit tukeutuisivat niihin. Opittava keskeinen sanasto olisi silloin helpommin määriteltävissä, minkä johdosta myös sanastosta johtuvien vaikeuksien tutkiminen esimerkiksi tekstin ymmärtämisessä helpottuisi. Keskeisen sanaston jonkinasteinen määrittely loisi pohjaa arvioinnillekin. Sanaston kartuttamiseen tulisi ehkä panostaa systemaattisemmin.

Suullisen kielitaidon pedagogiikassa tulisi koulun tasolla lisätä puhumistehtäviä, jotta oppilaat sisäistäisivät tavallisimmat tehtävä- ja koekäytänteet ja suhtautuisivat puhumistilanteisiin vakavasti. Yleisimmät puutteet itse tuottamisessa olivat kohteliaisuuden, vuorovaikutuksen ja siirtymäsanontojen puuttuminen. Verbillä alkavan kysymyslauseen nouseva intonaatio puuttui monilta. Näissä suhteissa koulujen välillä oli suuria eroja. Koetta vastaanottavilla opettajilla saattoi olla ongelmia oppilaiden tukemisessa silloin, kun näiden keskuste-

luun tuli katko; useat ”autoivat” liikaa. Suorituksia, joissa opettaja puhuu oppilasta enemmän, on mahdoton arvioida oppilaan tason kannalta, minkä vuoksi tällainen tuotos on korvattava toisella. Tämä taas vääristää otantaa.

Myös äänitystekniikan parantamiseen tulisi kiinnittää huomiota. Luotettava otantamenettelyä oli vaikea noudattaa suoritusten kuuntelussa, jos valitun oppilasparin suorituksesta ei teknisistä syistä tai käyttäjän taitamattomuuden vuoksi nauhalta kuulu mitään. Ohje pyytää oppilaita sanomaan äänityksen alussa nimensä ja kirjain (esittikö oppilas puhujaa A vai B) oli myös usein unohnut antaa, mikä turhautti arvioijia, koska aikaa kului äänen perusteella oikean puhujan tunnistamiseen. Toisinaan tämä ei onnistunut lainkaan.

11.2 IRT -menetelmien käytön edut

Vaikka taitotasojen asettamisessa ei tässä tutkimuksessa onnistuttu kaikilta osin optimaalisesti, se antaa kuitenkin hyvän pohjan tuleville linkittämismenetelleyille. Luvussa 10 on yksilöity tämän tutkimuksen ongelmia, jotka poistamalla voidaan jo nyt saada aikaan vieläkin luotettavampi taitotasojen vertailu. Aikaan OPS 2004:n arvioinnissa kokeiden laadinta ja arviointi tulee perustumaan sen sisältämiin taitotasokuvauksiin. Todennäköinen on myös tarve linkittää silloin saatavat tulokset CEF -taitotasoihin kansainvälisen vertailun mahdollistamiseksi. Jotta kaikkien vieraiden kielten OPS 1994:n mukainen taso tulisi ilmaistuksi myös taitotasolla, olisi tarpeen käynnistää myös ruotsin, finskan, ranskan, saksan ja venäjän vuosina 2001-2003 saatujen oppimistulosten linkittäminen CEF -taitotasoasteikkoon, mikä varmistaisi sen, että eri kielten kansallisten arviointien tuloksia voidaan luotettavasti verrata toisiinsa.

Kriteeriviitteinen arviointi edellyttää kokeita, jotka vastaavat kriteerien määritelmiä. Nämä taas ovat oppimistavoitteita. Jotta tiedetään, onko tavoitteet saavutettu, oppimistulokset (koepisteet) on sijoitettava niitä vastaaville taitotasolle. Tämä edellyttää oppimistulosten osoittamista θ -arvoina, jotka ilmaisevat sekä oppilaan taidon että osion vaikeustason. Nämä arvot voidaan laskea vain IRT -menetelmillä. Arvojen perusteella voidaan suorittaa taitotasojen asettaminen, mikä myös edellyttää IRT -menetelmien käyttöä. Esim. OPLM -analyysin ja tarvittaessa lisäanalyysin avulla taitotasojen katkaisukohtat voidaan määrittää, jos käytävissä on sekä asiantuntija-arvio että empiirinen tieto osioiden vaikeustasosta. Näin päästään perusteltuun tulkintaan siitä, mitä eri pistemäärät tarkoittavat eri taitotasolla, joskin tähän linkittämistulokseen liittyy epävarmuustekijöitä, koska päättelyprosessi on tuskin koskaan täysin objektiivinen.

Käsillä olevassa tutkimuksessa todettiin yhdeksi IRT -menetelmien käytön kiistattomaksi eduksi se, että niiden avulla pystytään määrittämään sekä testattavan taito että osion vaikeustaso. Aineiston analyysi osoitti, että reseptiivisen kielitaidon kokeiden eri osioihin vastanneiden vastaajien pistemäärät ovat vertailukelpoisia. Tämä tarkoittaa sitä, että osioiden vaikeustasoa voidaan verrata,

vaikka niihin ovat vastanneet osittain eri vastaajat eli niiden avulla voidaan vertailla kielitaidon eri osa-alueiden tuloksia. Tieto testattavan taidosta ja osion vaikeustasosta on yleistettävissä perusjoukkoon (kaikki 9. -luokkalaiset ja kaikki samaa asiaa mittaavat osiot). IRT -menetelmät eivät siis anna tietoa pelkästään arviointiin osallistuneesta otospopulaatiosta kuten klassinen testiteoria. Parhaaseen tulokseen osioiden erottelukyvyyssä päästään silloin kun osiot on mahdollista valita varsinaiseen kokeeseen niin, että jo esikokeilutulokset analysoidaan IRT -menetelmällä. Useissa maissa IRT -analyysit ovat jo yksinomaisia kansallisten arviointien analyysimenetelmiä (esim. Ranskassa ja Espanjassa). Ruotsi ja Hollanti ovat niihin siirtymässä (ks. luku 3).

Toinen etu on, että kokeita voidaan näin räätälöidä tietyille henkilöille tai henkilöryhmille. Usealla tahoilla on jo siirrytty tai ollaan siirtymässä ainakin osittain tietokoneella suoritettaviin kokeisiin (esim. Hollanti ja Baijeri). IRT -menetelmiä käyttäen voidaan suunnitella tietokonekokeita, joissa osiot tai pikemminkin eritasoisista osioista koostuvat ”testletit”, joita kullekin testattavalle tarjotaan, valitaan sen perusteella, kuinka edellisiin osioihin on vastattu (ns. adaptiivinen testaus). Näin päästään luotettavaan arvioon käyttämällä minimimäärä osioita ja koeaika. Vaikka ei edettäisikään adaptiiviseen testaukseen asti, kalibroitu osiopankki mahdollistaa arvioinnin tehostamisen ja sen laadun parantamisen.

Taitotasojen asettamisesta on toistaiseksi melko vähän tutkimusta. Erityisesti kaivattaisiin tutkimustuloksia oppilaskeskeisten (*examinee-centered*) menetelmien käytöstä. Yleensäkin taitotasojen asettamisessa suositellaan käytettäväksi useampia kuin yhtä menetelmää. Käsillä olevassa tutkimuksessa se ei ollut mahdollista käytettävissä olleiden resurssien puitteissa. Tietämyksen lisäämiseksi tältä alueelta olisi suositeltavaa järjestää pilottitutkimus, jonka kokeessa olisi IRT -analyysiin ihanteellisesti sopiva määrä sekä osioita että testattavia, ja linkittää nämä tulokset tarpeellisiksi katsottuihin taitotasoasteikkoihin käyttämällä sekä testikeskeistä että oppilaskeskeistä menetelmää. Tässä käytetyn Angoff -menetelmän sovelluksen lisäksi on käytettävissä monia muita menetelmiä, joista osa voisi toimia jopa sitä paremmin ja vähäisemmällä työpanoksella. Tämä vaatisi kuitenkin käytännön kokeilua. Käsillä olevassa tutkimuksessa produktiivisen kielitaidon taitotasojen asettamista vaikeutti erityisesti puhumisen arvioinnissa otoksen pienuus. Normaalisti 50 henkilön otos olisi laajempaan osallistujajoukkoon ekstrapoloimiseksi riittävä, mutta kun taitotasoja, joille tuotokset on sijoitettava on esim. 10 (OPS 2004), jokaiselle tasolle ei saada riittävää määrää arvoja luotettavien johtopäätösten tekemiseksi. Tämä asia on jatkotutkimuksella korjattavissa, koska suullisen kokeen tuotokset ovat tiittävästi vielä Opetushallituksessa. C-kasettien laatu voi kuitenkin ajan mittaan heiketä ja kasettisoitintenkin saatavuus vähenee uuden tekniikan korvattessa ne. Joka tapauksessa on todettava, että mitä useampia taitotasoja käytetään, sitä useampia arvioitavia tuotoksia tarvitaan.

Kun kokeen laadinnan peruseriaate on, että kokeen tulee mitata sitä, mitä konstruktissa esitetään, sitä edustavan tehtävän vaikeustason tunnistaminen on ensiarvoisen tärkeitä. Kun nyt tehtävien todellinen vaikeustaso on IRT -analyysi-

sien johdosta tiedossa, ainakin osaa aineistosta voisi käyttää koulutus- tutkimusmateriaalina kokeen- ja erityisesti oppimateriaalin laatijoille, jotka viime kädessä ovat keskeisiä opetussuunnitelman ja taitotasojen havainnollistajia käytännössä.

11.3 Arvioinnin kehittäminen

Kansallinen englannin kielen arviointi perustui järjestämisvastuun osalta saman vuoden 1999 alussa voimaan astuneeseen perusopetuslakiin ja sisällöltään vuoden 1994 opetussuunnitelman perusteisiin. Kansallista arviointia on tämän jälkeen täsmennetty ainakin asetuksella 150/20.2.2003, jossa ainekohtaisten oppimistulosten arviointia toimeenpaneva Opetushallitus määrittellään osaksi koulutuksen arviointiasiantuntijoiden verkostoa. Tavoitteet määräytyvät yhä kansallisen oppimistulosten arviointijärjestelmän pohjalta.

Arviointi toteutettiin 1.9.1998 – 31.1.1999 välisenä aikana. Jo tätä ennen Opetushallituksessa oli käyty neuvotteluja arvioinnin organisoimisesta, ja raportin koostamisen jälkeen projektisihteeri ja toimistohenkilökunta vielä huolehtivat raportin painokuntoon saattamisesta. Käsillä olevan tutkimuksen tekijän kokemuksen mukaan arviointiprosessin aikataulu oli liian tiukka. Edellyttäen, että prosessi noudattaa jatkossa aikaisempaa kaavaa ja arviointi järjestetään samassa laajuudessa, luotettavasti toimeenpannun arviointiprosessin elinkaaren tulisi olla minimissään 2 ½ vuotta. Projektipäällikön ja ohjausryhmän tulisi voida alkaa työ sen vuoden alussa, jota seuraavana keväänä perusopetuksen päättövaiheen arviointi järjestetään. Projektipäällikölle tulisi varata vähintään kaksi kuukautta testispesifikaatioiden laatimiseen ja koekonstruktien suunnitteluun. Näin voidaan varmistaa, että kriteeriviitteiselle mittaukselle saadaan luotettava perusta. Kokeenlaatijat, jotka ovat yleensä muissa tehtävissä toimivia asiantuntijoita, tulisi nimittää niin, että he ovat valmiit arviointityöhön maaliskuun alussa, jotta he voivat osallistua ohjausryhmän suorittamaan alustavien koekonstruktien kommentointiin ja niistä päättämiseen työnsä pohjaksi. Koe ehdotukset tulisi luovuttaa toukokuun loppuun mennessä, jolloin projektipäälliköllä on mahdollisuus koota tehtäväsarjat esikokeilua varten elokuun loppuun mennessä. Kertaluonteisissa arvioinneissa, jollaisia vieraiden kielten arvioinnit ovat tähän mennessä olleet, olisi myös harkittava esikokeiluun osallistujien otostamista satunnaisesti samalla tavoin kuin varsinaiset osallistujakoulutkin valitaan. Kokeenlaatijoiden tulisi osallistua laatimiensa kokeiden muokkaukseen esikokeilun antaman palautteen pohjalta.

Varsinaisen raportin tuottamiseen voidaan ryhtyä vasta analyysien valmistumisen jälkeen. Jos kansallisia arviointeja toimeenpannaan kaksi tai kolme (esim. vuonna 1999 äidinkieli, modersmålet ja englandi) suunnilleen samaan aikaan, analyysit valmistuvat ehkä syys- lokakuun alussa. Tällöin ei ole realistista odottaa raportin valmistuvan ennen seuraavan vuoden toukokuun loppua. Useissa Euroopan maissa (liite 1) loppuraportointiin on varattu 1-2 vuotta.

Laajat kansalliset oppimistulosten ohjelma-arvioinnit ovat paikallaan ainakin kerran opetussuunnitelman perusteiden elinkaaren aikana. Tällöin tulisi olla käytettävissä myös mahdollisimman paljon kansainvälistä arviointi- ja trenditietoa, jotta kansallinen arviointi osattaisiin suunnata arvioimaan myös näitä. Jonkinlaiset edellytykset pienimuotoisempien arviointien järjestämiseksi tulisi kuitenkin myös sisällyttää arviointiohjelmaan. Tässä vieraiden kielten arviointi kaipaisi laajempaa organisatorista tukea.

Arviointiketjua tulisi pidentää niin, että opetusministeriössä olisi vieraiden kielten asiantuntija, joka vastaisi kielikoulutuksen kokonaisvaltaisen suunnittelun edistämisestä, vieraiden kielten arvioinnin systemaattisesta kehittämisestä ja sen liittymisestä kansainvälisiin kehityslinjoihin perusopetuksesta kolmannen asteen opetukseen ja aikuiskoulutukseen. Näin sekä kansallisesta arvioinnista että sitä tukevista osa-arvioinneista tulisi pitkäjänteisempää samaan tapaan kuin nyt äidinkielen ja matematiikan arvioinnissa perusopetuslain nojalla. Tätä tavoitetta on pyritty tuomaan esille mm. Säätytalolla marraskuussa 2004 järjestetyssä Kielikoulutuksen valtakunnallisessa ´pyörän pöydän` tapahtumassa. Vieraan kielen opetus on maassamme vankalla pohjalla ja korkealla tasolla. Vieraiden kielten arvioinnilla on myös pitkät perinteet. Nyt sille olisi annettava sen ansaitsema asema ja jänteveys.

11.4 Kansainvälinen arviointiyhteistyö

Nopean kansainvälistymisen myötä kielitaidosta on tullut yhä tärkeämpi työ- ja yhteydenpitoväline yhä useammalle. Kielitaidon kehittämistä tukee arviointi, jonka avulla voidaan pysyä selvillä siitä, mikä kansakunnan kielivaranto on ja miten kieltenopetusta tulee suunnata. Suomi ei ole ollut kovin aloitteellinen kansainvälisten vieraiden kielten arviointien järjestämisessä. Se ei ole myöskään hakeutunut kumppaniksi vireillä oleviin merkittäviin hankkeisiin. Tällainen on mm. EBAFLS¹-hanke, johon osallistuu kahdeksan eurooppalaista maata². Se on osa Euroopan unionin komission tukemaa Socrates Lingua 2 -ohjelmaa, joka kattaa vuodet 2004-7 ja jota koordinoi CITO Hollannissa. EBALFSin tavoitteena on koota eri maissa englannin, ranskan ja saksan kansallisissa kokeissa käytettyjä CEFiin linkitettyjä osioita uusien kansallisten kokeiden ankkuriosioina käytettäväksi. Suomella olisi ollut todennäköisesti varsin hyvä mahdollisuus osallistua hankkeeseen tarjoamalla ko. kielissä kansallisissa kokeissa käytettyjä koeosioita projektiin.

Toinen kiinnostava hanke, johon Suomi kuitenkin Euroopan unionin jäsenenä tulee osallistumaan, on The European Indicator of Language Competence, eurooppalainen kielitaidon indikaattori. Se perustuu Eurooppa -neuvoston Barcelonassa vuonna 2002 tekemään päätökseen selvittää Euroopan unionin kansalaisten vieraiden kielten taito ja kehittää luotettavia järjestelmiä mitta-

¹ European Bank of Anchor Items for Foreign Language Skills.

² Espanja, Hollanti, Luxemburg, Ranska, Ruotsi, Saksa, Skotlanti ja Unkari.

maan saavutettuja taitoja. Sen koordinointi kansalliseen arviointiin tulisi selvittää ja varmistaa.

11.5 Yhteenveto

Käsillä olevan tutkimuksen johtopäätöksissä tehdään koulua, koulutusta, tutkimusta ja arvioinnin organisointia koskevia ehdotuksia. Koulun tasolla arviointia tukisi keskeisen sanaston määrittely ja puhumisen arviointikäytänteisiin panostaminen. Molemmat tavoitteet ovat myös yhteydessä koulutukseen. Opettajille tulisi suunnata koulutusta puhumistilanteiden järjestämisessä ja tuotosten arvioinnissa. Oppimateriaalin laatijat taas ovat lähinnä vastuussa keskeisen sanaston harmonisoinnista. He, samoin kuin arvioijatkin, kaipaavat harjaantumista tekstien ja tehtävien vaikeustason arvioinnissa. Tutkimuskohteiksi ehdotetaan alakoulun englannin kielen kansallista arviointia, oppilaskeskeisen taitotasojen asettamisen projektia, ymmärtämis- ja tuottamisprosessien tutkimusta, vuoden 1999 englannin arvioinnin produktiivisten tuotosten suuremman otoksen arviointia sekä taitotasojen asettamisen että tulosten varmistamiseksi sekä muiden vieraiden kielten kansallisten arviointien tulosten linkittämistä CEF- ja OPS -2004 -taitotasoille. Arvioinnin organisointia koskevissa ehdotuksissa otetaan kantaa arviointiprosessin sekä vieraiden kielten arvioinnin kokonaisvaltaiseen kehittämiseen. Lopuksi korostetaan kansainvälisten yhteyksien ja projekteihin osallistumisen merkitystä myös aloitteellisena osapuolena vieraiden kielten opetukselle ja arvioinnille.

SUMMARY

Background

One of the fundamental principles of the European Union is to ease the movement of goods, services and people in its area. The movement of students and workforce has been limited, among other things, by the fact that examinations and marks/grades are not commensurable, not even easily comparable. One major instrument in the attempt to provide greater transparency in language education in general and in assessment in particular, was the development (by the Council of Europe) of the Common European Framework of Reference for Languages (CEF) in 2001. Its scales of language proficiency are proving an important tool in assessing language skills and expressing them on those scales.

CEF has two dimensions: qualitative, referring to the taxonomy (a descriptive system) of language activities, and quantitative, referring to the six levels of proficiency. The descriptive system consists of categories for reception, production and interaction and their division into the modes “spoken” and “written” for each of them. These cells are then subdivided into more narrowly defined activities dependent on the language use situation. In addition, categories for aspects of the learner’s competences (e.g. spoken fluency, grammatical accuracy etc.) have been provided. The quantitative scale was formed by arranging the descriptors of language skills in the order of difficulty, and then splitting them into six consecutive intervals A1, A2, B1, B2, C1, C2.

For some time already, national scales have been developed or are being worked on in several EU countries. The purpose of most of them is to link them to the CEF scale to make it possible to show international correspondence. Since there has been a growing interest in defining the correspondence between national and CEF scales, the Council of Europe’s Language Policy Division launched a Manual¹ to help the providers of examinations to develop, apply and report procedures in order to situate their examinations in relation to the CEF. Among other things, more detailed scales to assess skills have also been included. The linkage is expected to contain the description and analysis of both the design of tests as well as the analyses of scoring and the results.

Standard setting, the process of setting and validating the cut-off score(s) for the different marks/grades reported from a test, is crucial to the linking process. It means that it is necessary to state that a certain score (the standard) on the test or examination corresponds to the minimum requirement to attain a certain level on a CEF scale. To support reliable linkage of tests and examinations, a complementary publication, the Reference Supplement, was

¹ Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR).

produced. It contains information on approaches to standard setting, classical test theory, item response theory etc.

A new national core curriculum for basic education was introduced in Finland in 2004. In it the language skills requirements are defined in terms of levels of proficiency (OPS 2004) based on the CEF levels. Unlike CEF, OPS 2004 has 10 levels² of language proficiency, and it only takes the categories reception and production as well as the modes “spoken” and “written” into consideration.

In 1999, 2001 and 2002 national evaluations were conducted in English, the second domestic languages³, German, French and Russian. They were based on the Framework Curriculum for Comprehensive School 1994, which did not include the concept of language proficiency levels. Furthermore, the tests of productive skills in the national assessment of English were assessed using a different scale from those of the other national tests.

For the Finnish results in those national language assessments to be comparable with European exams and marks/grades, it is important to link them to CEF scale. Another reason for the importance of linking is that it makes it possible to compare the results of future Finnish assessments in a reliable way with the earlier results, by providing an appropriate baseline.

Objective and research

There were two reasons for why the 1999 national assessment of the English language skills of Finnish 9th form pupils was chosen as the target of the present research. Firstly, its sample was larger than that of the others, in fact, the largest ever in Finland, and secondly, it is of potential considerable international interest, because English is the language taught in every EU country. The tests were developed according to the principles of criterion-referenced measurement.

The objective of the research was to link the results of the reading and listening as well as grammar tests to the CEF and OPS 2004 scales, and those of the productive skills to the CEF, OPS 2004 and A2001 scales (0 – 5). The most important objective, however, was to show what levels of English language proficiency can be reached in basic education.

The research data consisted of the 1999 national assessment receptive skills test results (n= 4854) and 111 listening, reading, and grammar test items. For productive skills two samples were assessed. After the larger sample had been assessed using the A2001 scales, another 52 oral and 50 writing tests were assessed using the scales CEF, OPS 2004 and A2001. The “written” test consisted of six test versions, three of which included the reading comprehension and writing tasks, and the other three those of listening and

² A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2, C1.1

³ Swedish for speakers of Finnish, Finnish for speakers of Swedish

grammar tasks. One of the comprehension and writing tasks was an anchor task, i.e. they were similar in each version. In the grammar test there were 25 anchor items on average in each version. The oral test consisted of six versions with several anchor tasks (see Table 59). Each version had two booklets, A and B, one for each of the partners doing pair work.

To design an evaluation project (like the one reported here) and carry it out call for familiarity with the concept of evaluation in general and its application to program evaluation, in this case to that of educational and curriculum evaluation, in particular. Ethical principles and program evaluation standards must also be taken into consideration. When assessing language competence, its construct has to be defined after which the tests themselves can be developed. In the theoretical part of the present research these aspects are dealt with as the conceptual background to the implemented evaluation process, and to the reporting of the research results.

Research questions, analyses and answers to questions

The first research question was: *What is the quality of the listening and reading comprehension and grammar test items when assessed using the methods of classical and new test theory?*

This research question is methodological in nature. It has to be dealt with, as a fundamental step towards other analyses. For standard setting, it is of added value to express the test results as θ -values. These can only be achieved by using IRT (the new test theory) methods. The use of IRT analyses calls for an analysis of the suitability of the test items to the chosen IRT model. For that, the measuring ability of test items must be known. They are expressed by calculating their difficulty and discrimination indices. In the present research this was done by using the OPLM statistical tool, which produces the indices in terms of the two test theories. The amount of proficiency to obtain a probability of $\frac{1}{2}$ for a correct answer is defined as the index of difficulty for the item in IRT. One more advantage of using an IRT model is that it is possible to compare test takers with respect to some proficiency even if they did not take the same test because in the two-parameter model, all test takers with the same weighted score have the same θ -estimate.

When using probabilistic models like the OPLM, the model itself is the statistical null hypothesis, which is not expected to be rejected. The purpose is to show that the observations fit well with the predictions of the model. The goodness-of-fit (model-fit) was tested using the R1c* analysis included in the OPLM. Mainly due to the very large sample, a model-fit could not be found. Therefore the pupil sample was divided randomly into two parts, which were calibrated separately. The relationships between the parts was studied through correlation and principal factors analyses. They showed the statistical unidimensionality of the test, meaning that all items were indicative jointly of

just one underlying latent variable. Thus it was appropriate to continue the IRT analyses.

To sum up, the items testing receptive skills proved to be good and the data statistically unidimensional, which means it was suitable to be used in IRT analyses, and the test versions were comparable enough. The listening comprehension test proved to be more difficult than the other tests. The reliabilities of the listening and reading tests were lower than that of the grammar test, mainly due to the fewer number of items.

The second research question was: *What is the correspondence between CEF and OPS 2004 scales when assessing receptive skills?*

The writer of the present research and an expert on the field judged the difficulty of the listening and reading texts and test items using the "Grids", designed for this purpose by the Dutch Construct Project Group, and applying both the CEF scales and those provided by the *Manual Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF)* (2005). In addition to the CEF scales, the grammar items were judged using validated YKI⁴ scales, which was supported by reference to the OPS 2004 grammar descriptions (included in the descriptors of receptive and productive skills). Analysis of the relationships between these judgements and the empirical results showed that the correlations were rather low: 0,348** referring to CEF and 0,294** referring to the finer OPS 2004 scales respectively.

The judgements were made by two people in consensus, whereas the most common recommendation is to have judgements made by (preferably more than) ten judges. Practical experience in Finland and abroad has shown, however, that their average correlations have not been higher than those of the present research, which shows how difficult it is to judge the difficulty of tests of receptive skills.

The consensus judgements were, thus, considered of sufficient quality, and they were then combined with the empirical results. When placing the results in the order of their average difficulty on the CEF scale, it was noticed that the scale was properly ascending but still did not show a clear enough level difference. Hence another standard setting round was needed. This was done by calculating the sum of raw scores on each level and converting the raw score into θ -values. After this measure, the division of the items on the CEF scale was satisfactory (Figure 49). The same measures were taken with the OPS 2004 scale. However, even after two standard setting rounds the result was not quite satisfactory enough.

The answer to the second research question is, then, that the correlations between the expert judgements and the empirical results in the tests of receptive skills were rather low. Still, the second standard setting round produced adequate cut-off points on the CEF scale. By contrast, those on the OPS 2004 scale are only indicative/suggestive. In the end, the two scales proved to be more or less commensurable as can be seen in Table 72.

⁴ The Finnish National Certificate Scale (National Board of Education 2002)

The third research question was: *What was the average level of the receptive language skills expressed on the CEF and OPS 2004 scales?*

In the OPLM analysis program, the θ -values may vary from 0 in the positive direction (+3) and negative (-3) direction, the former indicating difficult items and high ability and the latter easy items and low ability. The average θ -value of the whole test as well as that of each subtest in the receptive skills test was 0,27, which is a good result. Figures 58-61 also show a close resemblance to the normal distribution. When placed on the CEF and OPS 2004 scales, the average results are B1 and B1.1 respectively. The latter is the average target level of the Finnish national core curriculum.

The fourth research question seeks for an answer to the question: *How do the CEF, OPS 2004 and A2001 productive skills scales correspond to each other?*

The writer of the present research and an expert on the field applied consensus judgement when placing 52 speaking and 50 writing products on CEF and OPS 2004 scales. An additional judgement was made by a third expert. The same procedure had been followed before when another sample had been judged by two judges⁵ on the A2001 scale. The linking of the productive skills results was done by cross-tabulating the frequencies and criteria of different skills scales. The results were illustrated by correspondence analysis. Figure 73 is an example of that analysis between the CEF (*overall*) and OPS 2004 scales. The correspondence between the scales OPS 2004 and A2001 was rather low while that between CEF and OPS 2004 was relatively high. The correlations of the writing criteria were better than those of speaking on all levels. Table 81 shows the distribution of the speaking skills levels and Table 82 that of the writing skills levels. To reach a reliable linkage between OPS 2004 and A2001 was rather hard due to the big difference in the number of levels. On the other hand, there were too few cases to be placed on the OPS 2004 levels. Still, the linkage between CEF and OPS 2004 succeeded quite well. A bigger number of sample cases would have made the comparison more reliable.

In the last, fifth, research question it was inquired: *What was the level of the productive skills expressed in the CEF, OPS 2004 and A2001 scales?*

The median in the speaking test was B1.1 on the OPS 2004 scale, which is one level higher than the Finnish core curriculum target level A2.2. On the CEF scale the level was B1, and on the A2001 scale 4. The median of the writing test also exceeded that of the core curriculum target of A2.2 by about one level, being B1.1. The corresponding CEF level result was B1 and 4 on the 2001 scale. This raises the question whether the oral test was too easy, whether the proficiency requirements on the levels B1 – B2 and B1.1 – B2.1 are too modest, and whether the Matriculation Examination writing tests has some washback on teaching in basic education contrary to what the teachers maintained when answering the teacher questionnaire about their teaching priorities.

⁵ The writer of the present research did not participate in the previous judgement.

Conclusion

In addition to its potential EU-linkage, the research is important from the Finnish perspective. The results are expressed in θ -values (IRT), which show simultaneously the difficulty of the item and the skill of the examinee. In this way the results of different tests at different times and places, can be compared with each other providing the items are calibrated at the same time. That made it possible to note in the present research that the listening comprehension test was more difficult than the reading comprehension test. The use of IRT analyses is also a prerequisite to effective standard setting, i.e. defining the cut-off points of language proficiency levels. Standard setting and linking test results to scales levels are new techniques, which have not been applied much so far. The present research aims at showing that that kind of linking is possible but that a more reliable result might be reached if more studies/research were conducted. It also demonstrates the advantages of applying criterion-referenced measurement in curriculum evaluation, as well as the importance of defining the constructs and test specifications when developing tests.

To ensure the reliability of standard setting, research in the use of other methods in addition to the test-centered continuum method used here, are needed. One of them is the examinee-centered method, which could be used if a properly representative sample of pupils (which could be much smaller than the present sample) were administered these tests and teachers provided estimates of what level each pupil had reached. It is advisable to apply this procedure in the next round of national language assessments. It is also recommended that a bigger sample of productive skills cases be assessed/judged, although the CEF and OPS 2004 scales proved to be fairly comparable in the present research. The correspondence of the other national language assessments to CEF should also be defined (in retrospect) both to compare the results with those of the present research and those of other EU evaluations as well as to have a baseline on which to build the future assessments.

Judging the difficulty of the lowest language proficiency levels (A1; A1.1-A1.3) proved to be somewhat unstable. Therefore it is suggested that a national test in English be conducted in primary school. The assessment should be criterion-referenced, and be analysed with IRT methods to make it possible to compare the difficulty. More experience is needed in judging the difficulty of texts and items of tests measuring receptive skills, both in teacher education and in-service training but training should be targeted to raters and material writers in particular. It is of prime importance if we want to assess these skills in a reliable way.

If we want to cover a wider area of the curriculum, the number of test versions should be increased, say, from three to five, which was the objective of the original national evaluation in English. A solution is to compose booklets with more items and rotate them between examinees or alternatively, to

accelerate the testing cycle. That would result in the opportunity to increase the number of listening and reading comprehension items, which would add to the construct representation and improve the reliability of the results of these skills. However, all skills should be assessed also in the future.

SAMMANFATTNING

I många EU länder pågår revidering av läroplaner i moderna språk i syfte att beskriva språkfärdigheter genom nivåbeskrivningar som anknyts till CEF¹. Den finska läroplanen för den obligatoriska utbildningen (grundskolan) 2004 är redan baserad på dessa nivåer. Standardbeskrivningar av detta slag möjliggör jämförelse av provresultat såväl nationellt som internationellt och bidrar därmed till ökad transparens vad gäller inlärningsresultat, till exempel vid betygsättning. För att jämförelseresultaten skall kunna anses tillförlitliga har Europarådets språkpolitiska enhet utarbetat instruktioner, vilka bland annat förutsätter en detaljerad dokumentation av utvärderingsprocessen. Trots att andra metoder också kan användas, är dessa anvisningar i allmänt bruk i dag.

Fokus i denna avhandling är utvärdering av målen i engelska som de uttrycks i den finska grundskolans läroplan 1994. Syftet var att relatera resultaten av den nationella utvärderingen i engelska i grundskolans nionde klass 1999 till CEF -skalan (A1, A2, B1, B2, C1 och C2) och OPS 2004 -skalan² med nivåerna A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2 och C1.1 från lägst till högst. För att bedöma de produktiva kompetenserna användes ytterligare en skala, A2001³ (0 - 5). För relatering användes en provrelaterad metod för standardsättning. Standardsättningen av de receptiva kompetenserna förutsätter användning av IRT -teknik vid analys av empiriska data, vars resultat anges i theta-värden (θ). Tekniken gör det möjligt att bestämma såväl uppgifters svårighetsgrad som personers färdighetsnivå så att svårigheten av items i olika prov kan jämföras. Här användes OPLM -teknik, där θ -skalan kan variera mellan -3 och +3, där minus tyder på lätta uppgifter och svag färdighet och plus på svåra uppgifter och utmärkt färdighet. Uppgiftens kvalitet visas närmast av diskriminationsförmåga, som kan variera från 1-15, men som vanligen, liksom i denna undersökning, faller mellan värdena 2 och 3.

Undersökningens resultat visade, att θ -medelvärdena både i hela det receptiva provet samt i delproven i hör- och läsförståelse och grammatikproven var 0,27, vilket indikerar goda kunskaper och färdigheter. Vid standardsättningen kombinerades bedömarnas värderingar av items och produktioner och de empiriska resultaten för att etablera en länk mellan provresultaten och färdighetsnivåerna. På CEF -skalan var medianen B1 och på OPS 2004 -skalan B1.1, vilket betyder att den finska målnivån B1.1 uppnåddes. Nivåerna i proven på produktiva kompetenser bedömdes i enlighet med de tre ovannämnda skalorna. Medianvärden på muntlig och fri skriftlig produktion var B1 på CEF -, B1.1 på OPS 2004 - och 4 på A2001 -skalan, vilket betyder att resultaten var cirka en nivå högre än den finska målnivån, A2.2.

Nyckelord: utvärdering, bedömning, engelska, färdighetsnivå, standardsättning

1 Common European Framework of Reference for Languages.
 2 Finska grunderna för den grundläggande utbildningen 2004.
 3 Utbildningsstyrelsens skala för produktiva kompetenser 2001.

TIIVISTELMÄ

Useissa Euroopan unionin maissa uusitaan opetussuunnitelman perusteita, joissa tavoitteeksi asetettua kielitaitoa kuvataan yleiseurooppalaiseen CEF -taitotasasteikkoon¹ perustuvien taitotasoin. Näin on menetelty myös Suomen vuoden 2004 opetussuunnitelman perusteita laadittaessa. Kielitaidon kuvaus taitotasoin tekee mahdolliseksi verrata saavutettuja koetuloksia toisiinsa sekä kansallisesti että kansainvälisesti ja näin lisätä esim. arvosanojen läpinäkyvyyttä. Luotettavan linkityksen takaamiseksi Euroopan neuvoston kielipolitiikan jaos on laatinut ohjeet, jotka edellyttävät mm. kattavaa koko arviointiprosessin kuvaamista. Vaikka muitakin menetelmiä voidaan käyttää, tämä ohjeisto on jo nyt yleisesti käytössä.

Tämän tutkimuksen tavoitteena on vuoden 1994 peruskoulun opetussuunnitelman perusteiden määrittämisen, vuonna 1999 toimeenpannun 9. luokan englannin kielen kansallisen arvioinnin oppimistulosten vastaavuuden ilmaiseminen kolmella taitotasolla. Vastaavuuden löytämiseen käytettiin testikeskeistä taitotasojen asettamisenmenettelyä. Reseptiivisen kielitaidon kokeet linkitettiin sekä CEF -taitotasasteikkoon, joka käsittää alimmasta ylimpään lueteltuna tasot A1, A2, B1, B2, C1 ja C2, ja Suomen nykyiseen OPS 2004 -taitotasasteikkoon, joka käsittää tasot A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2 ja C1. Produktiivisen kielitaidon² koetulokset linkitettiin lisäksi Opetushallituksessa laadittuun A2001 -taitotasasteikkoon (0-5). Reseptiivisen kielitaidon kokeiden taitotasojen asettaminen edellytti IRT -analyysien käyttöä, jolloin kokeiden vaikeustaso ilmaistaan θ -arvoina. Ne ilmaisevat sekä oppilaan taidon että osion vaikeuden, mikä mahdollistaa kielitaidon eri osa-alueita mittaavien osioiden vaikeustason keskinäisen vertailun. Tähän ei muilla menetelmillä pystytä. Käytetyssä OPLM -ohjelmassa θ -arvot voivat vaihdella 0-arvosta negatiiviseen ja positiiviseen suuntaan välillä -3 (helppo osio/heikko kielitaito) ja +3 (vaikea osio/erinomainen kielitaito). Osioden erottelukyky ilmaistaan OPLM -ohjelmassa asteikolla 1-15, ja yleensä tulokseksi saadaan arvoja 2-3, kuten tässäkin arvioinnissa.

Tutkimustulokset osoittivat, että sekä koko reseptiivisen kielitaidon kokeen että sen osakokeiden keskimääräinen θ -arvo oli 0,27, mikä on hyvä tulos. Taitotasorajojen asettamisessa yhdistettiin asiantuntija-arviot ja empiiriset tulokset. Tämän perusteella tulokset voitiin muuntaa taitotasokuvauksiksi. Tällöin CEF -asteikolla keskimääräiseksi tulokseksi saatiin B1 ja OPS 2004 -asteikolla B1.1. Suomen vuoden 2004 perusopetuksen opetussuunnitelman perusteiden taitotaso-odotus on B1.1, joten tavoite saavutettiin. Produktiivisen kielitaidon kokeet arvioitiin suoraan taitotasasteikoilla. Sekä puhumisen että kirjoittamisen kokeen mediaani oli CEF -asteikolla B1, OPS 2004 -asteikolla B1.1 ja A2001 -taitotasasteikolla 4. Suomen taitotasotavoite on A2.2. Produktiivisen kielitaidon taso oli siis yhden tason tavoite-tasoa korkeampi.

Avainsanat: arviointi, englannin kieli, taitotasasteikko, taitotasojen asettaminen

¹ Common European Framework of Reference for Languages.

² Puheen ja tekstin ymmärtäminen sekä kieliopillinen kompetenssi.

LÄHTEET

- Alhlström, K-G. 1971. *Korkeamman opetuksen käsikirja*. Helsinki: Otava.
- Alderson, J. C., 1992. *Guidelines for the evaluation of language education*. Alderson J.C. & Beretta, A. (toim.) *Evaluating Second Language Education*. Cambridge: Cambridge University Press. 274-304.
- Alderson, J. C. 1993. *The state of language testing in the 1990s*. Huhta A, Sajavaara K., Takala S. (toim.) *Language Testing: New Openings*. Jyväskylä: Oy Sisäsuomi.
- Alderson, J. C. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. 2005. *The challenge of (diagnostic) testing: do we know what we are measuring?* Plenary presentation at the 2005 LTRC, Ottawa, Canada.
- Alderson J. C. & Banerjee J. 2002. *Language testing and assessment (Part 2)*. State-of-the-Art Review. *Language Teaching* 35. Cambridge University Press. 79-113.
- Alderson, J. C., Clapham, C. and Wall, D. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., Tardieu, C. 2005. *Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct Project (Grid)*. Moniste.
- AMKKIA. Ks. Juurakko, T.
- Austin, J. L. 1962. *How to Do Things with Words?* Oxford: Clarendon Press.
- Bachman, L. F. 1990. *Fundamental Considerations on Language Testing*. Hong Kong: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. 1984. *Some comments on the terminology of language testing*. Rivera, C. (toim.) *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application*. Multilingual Matters. Ltd. Bodmin: Robert Hartnoll Ltd. 34-43.
- Bachman, L. F. & Palmer, A. S. 1996. *Language Teaching in Practice: Designing and Developing Useful Language Tests*. 1. ja 5. painos. Oxford: Oxford University Press.
- Baker, F. 2001. *The Item Characteristic Curve*. Baker, F. *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Education. University of Maryland, College Park, MD. 5-11.
- Barrett, T. C. 1974. *Taxonomy of reading comprehension*. Smith R. C. & Barrett, T. C. (toim.) *Teaching reading in the middle grades*. Reading, MA: Addison-Wesley.
- Bernhardt, E. B. & Kamil, M. L. 1995. *Interpreting relationships between L1 and L2 reading: consolidating the linguistic threshold and the linguistic independence hypotheses*. *Applied Linguistics* 16 (1), 15-34.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krarwohl, D. R. (toim.). 1956. *Taxonomy of educational objectives: cognitive domain*. New York: David McKay.

- Bonnét, G. (toim.). 2004. *The assessment of pupils' skills in English in eight European countries in 2002*. A European project. Commissioned by the European network of policy makers for the evaluation of educational systems. <http://cisad.adc.education.fr/rev/>
- Brown, H. D. 2001. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Second Edition. NY: Longman.
- Brown, J. D. 1995. *The Elements of Language Curriculum. A Systematic Approach to Program Development*. Boston: Heinle & Heinle.
- Buck, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press.
- Bygate, M. 1987. *Speaking*. Oxford: Oxford University Press.
- Canale, M. & Swain, M. 1980. *Theoretical bases of communicative approaches to second language teaching and testing*. *Applied Linguistics* 1 (1), 1-47.
- Carroll, J. B. 1958. *A Factor Analysis of Two Language Aptitude Batteries*. *The Journal of General Psychology*. 59. 3-19.
- Carroll, J. B. 1972. *Defining language comprehension*. Freedle, R.O & Carroll, J.B. (toim.) *Language comprehension and the acquisition of knowledge*. New York: John Wiley and Sons.
- Carroll, J. B. 1993. *Human Cognitive Abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- CEF. Ks. *Common European Framework of Reference for Languages*.
- CEF-käsikirja. Ks. *Relating Language Examinations*.
- Chalhoub-Deville, M. 2003. *Second language interaction: current perspectives and future trends*. *Language Testing* 20 (4). Arnold. 369-383.
- Chalhoub-Deville, M., Alcaya, C. and McCollum Lozier V. 1997. *Language and measurement issues in developing computer-adaptive tests of reading ability: The University of Minnesota approach*. Huhta, A., Kohonen, V., Kurki-Suonio L, & Luoma, S. (toim.). *Current developments and alternatives in language assessment*. *Proceedings of LTRC* 96. 557-
- Chelmsky, E. 1997. *The Coming Transformation in Evaluation*. Chelmsky, Eleanor & Shadish, William R. (toim.). *Evaluation for the 21st century. A Handbook*. Thousand Oaks, CA. Sage. 1-25.
- Chomsky, N. 1972. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. 1980. *Rules and Representations*. Oxford: Basil Blackwell.
- Chun, D. M. & Plass, J. L. 1997. *Research on Text Comprehension in Multimedia Environments*. *Language Learning and Technology*. Vol. 1, No. 1 July 1997. 60-81.
- Clark, H. H. 1996. *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H. & Clark E.V. 1977. *Psychology and language: an introduction to psycholinguistics*. New York: Harcourt Bruce Jovanovich.
- Common European Framework of Reference for Languages: Learning, teaching, assessment (CEF)*. 2001. Council of Europe. Modern Languages Division. Strasbourg. Cambridge: Cambridge University Press.
- Cook, T. D. 1997. *Lessons Learned in Evaluation Over the Past 25 Years*. Chelmsky, Eleanor & Shadish, William R. (toim.). *Evaluation for the 21st century. A Handbook*. Thousand Oaks, CA. Sage. 30-52.

- Cousins, J. B. 2003. *Utilization Effects on Participating Evaluation*. Kellaghan, T., Stufflebeam, D. L. & Wingate Lori A. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 245-65.
- Cronbach, L. J. and Associates. 1981. *Toward Reform of Program Evaluation*. San Francisco CA: Jossey-Bass.
- Cronbach, L. J. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey-Bass.
- Cronbach L. J. 1987. *Issues in Planning Evaluations*. R. Murphy & H. Torrance (toim.) *Evaluating education: Issues and methods*. London: Harper & Row. 4-35.
- Cutler, A. & Clifton, C. Jr. 1999. *Comprehending spoken language: a blueprint of the listener*. Brown, Colin M. & Hagoort Peter (toim.) *The Neurocognition of Language*. Oxford: Oxford University Press. 122-155.
- Davies, A. 1978. *Language Testing*. Survey Article. *Language Testing and Linguistics*. Abstracts. Vol. 11 No 3 24, 1978. 145-159.
- Davies, A. 1997. *Introduction: The units of ethics in language testing*. *Language Testing*. 14, 3. 235-241
- Davidson, F. & Lynch, B. 1993. *Criterion-Referenced Language Test Development: a Prolegomenon*. Huhta, A, Sajavaara, K., Takala, S. (toim.) *Language Testing: New Openings*. Institute for Educational Research. Jyväskylä. 73-89.
- De Jong, J. H. A. L. 1987. *Defining tests for proficiency in listening comprehension: A response to Don Douglas's "Testing Listening Comprehension"*. Teoksessa A. Valdman (toim.) *Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency*. Committee for Research and Development in Language Instruction. Bloomington: Indiana University. 115-124.
- De Jong, J. H. A. L. 2004. *What is the role of the Common European Framework of Reference for Languages: learning, teaching, assessment?* Invited lecture at the first Annual Meeting of the European Association for Language Testing and Assessment (EALTA), Kranjska Gora, Slovenia.
- Douglas, D. 2000. *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Ebel, R.L. 1969. *The relation of scale fineness to grade accuracy*. *Journal of Educational Measurement*. Vol 6 4 s. 217-221.
- Ebel, Robert L. 1979. *Essentials of Educational Measurement*. Englewood Cliffs: Prentice Hall.
- Erickson, Gudrun. 1999. *Från Sp 8 till Äp 9. Om utvecklingen av ett nytt nationellt prov i engelska i grundskolan*. Papers on Language, Learning, Teaching, Assessment. Carlsson, I. & af Ekenstam, N. (toim.). Festskrift till Torsten Lindblad. IPD-rapporter nr 1999:02. Göteborgs universitet. Institutionen för pedagogik och didaktik. 202-226.
- Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys*. 2003. Porvoo: WSOY
- Evensen, L. S, Berge, K. & Vagle W. 2004. *Kirjoittamisprosessi -kaavio*. Julkaisematon käsikirjoitus.
- Faerch, C. & Kasper, G. 1986. *The role of comprehension in second language learning*. *Applied Linguistics* 7. 257-274.

- Figueras, N., North, B., Takala, S., Verhelst, N., Van Avermaet, P. 2005. *Relating examinations to the Common European Framework: a manual*. Language Testing. 20(3) Edward Arnold Publishers Ltd. 261-279.
- Finocciaro, M. & Sacks, S. 1983. *Foreign Language Testing. A Practical Approach*. New York: Regents.
- Franke-Wikberg, S. & Lundgren, U. B. 1982. *Att värdera utbildning*. Del 1. En introduktion till pedagogisk utvärdering. ALMAserien 103. Stockholm: Wahlström & Widstrand.
- Freedle, R. & Kostin, I. 1993. *The Prediction of TOEFL Reading Comprehension Item Difficulty for Expository Prose Passages from Three Item Types: Main Idea, Inference, and Supporting Idea Items*. TOEFL Research Reports 44. Princeton, NJ: ETS.
- Fulcher, G. 2003. *Testing Second Language Speaking*. Pearson ESL
- Gagné, R. M. 1970. *Instructional Variables and Learning Outcomes*. Wittrock, M. C. & Wiley, D. E. *The Evaluation of Instruction. Issues and Problems*. USA: Holt, Rinehart and Winston.
- Greenacre, M. J. 1993. *Correspondence Analysis in Practice*. London: Academic Press.
- Grid. 2005. Ks. Alderson, J. C., Figueras, N. ym.
- Halliday, M. A. K. 2003. *On Language and Linguistics*. Coleston, Jonathan (toim.) Collected works of M A K Halliday. Volume 3. London: Continuum.
- Hamp-Lyons, L. 1991. *Basic Concepts*. Hamp-Lyons, L. (toim.) *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex.
- Harste, J. C. 1985. *Portrait of a new paradigm: Reading comprehension research*. Crismore, A (toim.) *Landscapes: A state-of-the-art assessment of reading comprehension research*. Indiana University, School of Education. Final Report Project USDE-C-300-83-0130, luku 12.
- Havola-Pitkänen, L. 1974. *Englanninkielen koulusaavutukset peruskoulun ala-asteen ja yläasteen päätyessä*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitoksen julkaisuja 231.
- Havola, L. & Saari, H. 1993. *Kaksikymmentä vuotta englannin kuullun ymmärtämistä - Onko taidoissa edistytty? Linnakylä, P. & Saari, H. (toim.) Oppiiko oppilas peruskoulussa? Peruskoulun arviointi 90-tutkimuksen tuloksia*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos. 157-173.
- Heinonen, V. 1961. *Koulusaavutustestit*. Jyväskylä: K.J. Gummeruksen kirjapaino.
- Hellekjaer, G. 2005. *The Acid Test: Does Upper Secondary EFL Instruction Effectively Prepare Norwegian Students for the Reading of English Textbooks at Colleges and Universities?* Oslo: University of Oslo.
- Hildén, R. 2000. *Att tala bra, bättre och bäst*. Suomenkielisten abiturienttien ruotsin kielen suullinen taito testisuoritusten valossa. Helsingin yliopiston opettajankoulutuslaitos. Helsinki: Hakapaino.
- House, E. R. 2003. *Introduction*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 9-14.

- House, E. R. & Howe, K. 1999. *Values in evaluation and social research*. Thousand Oaks, CA: Sage.
- House, E. R. & Howe, K. 2003. *Deliberative Democratic Evaluation*. Kellaghan, T, Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 72-100.
- Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: Cambridge University Press. 4. painos.
- Huhta, A. 1990. *Suullisen kielitaidon arviointi ja arviointikriteerit*. Moniste. Korkeakoulujen kielikeskus.
- Huhta, A. 1993. *Teorioita kielitaidosta: Onko niistä hyötyä testaukselle?* Takala, S. (toim.) *Suullinen kielitaito ja sen arviointi*. Kasvatustieteiden tutkimuslaitos. Jyväskylä. 77-131.
- Huhta, A. 1997. *Yleisten kielitutkintojen arvosteluperusteet*. Huhta, A, Määttä, A., Takala, S. & Tarnanen, M. (toim) *Yleiset kielitutkinnot ja kielenopetus*. Opetushallitus. Helsinki: Hakapaino Oy. 21-43.
- Huhta, A. & Takala, S. 1999. *Kielitaidon arviointi*. Sajavaara, K. & Piirainen-Marsh, A. (toim.) *Kielenoppimisen kysymyksiä. Soveltavan kielentutkimuksen teoriaa ja käytäntöä*. Soveltavan kielentutkimuksen keskus. Jyväskylän yliopisto. 179-218.
- Hymes, D. 1972. *On Communicative Competence*. J. B. Pride and J. Holmes (toim.) *Sociolinguistics*. Harmondsworth: Penguin, 269-93.
- Jaakkola, H. 1997. *Kielitieto kielitaitoon pyrittäessä*. Vieraiden kielten opettajien käsityksiä kieliopin oppimisesta ja opettamisesta. Kasvatustieteiden laitos. Jyväskylän yliopisto.
- Jakku-Sihvonen, R. & Heinonen, S. 2001. *Johdatus koulutuksen uudistuvaan arviointikulttuuriin*. Opetushallitus. Arviointi 2/2001. Helsinki: Yliopistopaino.
- Joint Committee on Standards for Educational Evaluation. 1981. *Standards for Evaluations for Educational Programs, projects and materials*. New York: McGraww-Hill.
- Jones, L. W. 2003. *National Assessment in the United States: The Evolution of a Nation's Report Card*. Kellaghan, T. Stufflebeam, D. & Wingate, L. A. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 882-904.
- Juurakko, T. (toim). 2002. *Kielitaidon arvioinnin yhdenmukaistaminen ammattikorkeakouluissa (AMKKIA)*. Hämeenlinna: Hämeen ammattikorkeakoulu. A:16.
- Kaftandjieva, F. 1999. *Item response modelling*. Pietilä, P. ja Salo, O-P. (toim.). *Multiple languages - multiple perspectives: Texts on Language Teaching and Linguistic Research*. AFinLA Series No. 57. Jyväskylä: Jyväskylän yliopistopaino. 94-99.
- Kaftandjieva, F. 2004. *Standard Setting*. Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. Section B. Language Policy Division, Strasbourg.

- Karlsson, O. 2003. *Program Evaluation in Europe: Between Democratic and New Public Management Evaluation*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 429-439.
- Kellaghan, T. 2003. *National and International Levels of System Evaluation. Introduction*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 873-882.
- King, J. 2003. *Evaluating Educational Programs and Projects in the USA*. Kellaghan, T., Stufflebeam, D. & Wingate, L.A. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 721-732.
- Kintsch, W. 1998. *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Knapp, M. 1979. *Tinkering with Open Systems: organizational Theory Perspectives on Educational Program Evaluation*. Datta, L-E & Perloff, R. (toim.) *Improving Evaluations*. Beverly Hills, CA: Sage. 226-244.
- Kouluhallitus. 1976. *Oppilaalle vieraat kielet. Ehdotus perustavoitteiksi ja perusoppiainekseksi peruskoulussa. 3 d.* (Ks. VIKKE) Valtion Painatuskeskus.
- Kouluhallitus 1985. *Peruskoulun opetussuunnitelman perusteet 1985*. Helsinki: Valtion painatuskeskus.
- Koulutuksen arviointistrategia. 1996. <http://kuntaliitto.fi/opetus/arvioin/htm>.
- Kynnystaitoluettelo. 1990. Ks. van Ek, J. A. & Trim, J. L. M.
- Kärkkäinen, K. & Takala, S. 1977. *Vieraiden kielten rakenteiden hallintaa mittaavia koetyyppejä*. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 95/1977.
- Köhlmyr, P. 2003. *To Err Is Human. An investigation of grammatical errors in Swedish 16-year-old learners' written production in English*. Göteborg University: Department of English.
- Lado, R. 1961. *Language Testing. The Construction and Use of Foreign Language Tests*. Bristol: Western Printing Services Ltd.
- Laki 476/1983. Peruskoululaki.
- Laki 182/1991. Laki Opetushallituksesta.
- Laki 705/1992. Laki opetus- ja kulttuuritoimen rahoituksesta.
- Laki 628/1998. Perusopetuslaki.
- Laukkanen, R. 1994. *Koulutuksen tuloksellisuuden arvioinnin keskeiset kysymykset ja trendit keskushallinnon näkökulmasta*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja A. Tutkimuksia 59.
- Laukkanen, R. 1998. *Opetustoimen keskushallinnon evaluaatioajattelun kehitys Suomessa 1970-luvulta 1990-luvulle*. Koulutuksen tutkimuslaitos. Jyväskylän yliopisto. Tutkimuksia 5. Jyväskylä.
- Lehtisalo, L. & Raivola, R. 1999. *Koulutus ja koulutuspolitiikka 2000 -luvulle*. Juva: WSOY.
- Leimu, K. 2004a. *Kansainvälisyys koulutuksen arviointitutkimuksessa*. Leimu, K. (toim.) *Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Yliopistopaino. 7-26.

- Leimu, K. 2004b. *Vertailun käsitteestä koulutustutkimuksissa..* Leimu, K. (toim.) Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa. Jyväskylä: Yliopistopaino. 27-42
- Leimu, K. 2004c. *IEA -tutkimusyhteistyön taustasta ja luonteesta.* Leimu, K. (toim.) Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa Jyväskylä: Yliopistopaino. 43-76.
- Leimu, K. & Saari, H. 1976. *Arviointi- ja koulukoetoiminnan kehittämisesstä Suomessa.* Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita. 77/1976.
- Levelt, W. J. M. 1989. *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.
- Levelt, W. J. M, Colin M. & Hagoort P. (toim.) *The Neurocognition of Language.* Oxford: Oxford University Press. 84-117.
- Lincoln, Y. S. 2003. *Constructivist Knowing, Participatory Ethics and Responsive Evaluation.* A Model for the 21st Century. Kellaghan, T., Stufflebeam, D. & Wingate, L (toim.) *International Handbook of Educational Evaluation.* London: Kluwer. 69-78.
- Lindquist, E. F. 1961. *Preliminary Considerations in Objective Test Construction.* Lindquist, E. F. (toim.) *Educational Measurement.* Washington D. C.: American Council on Education.
- Linn, R.L. & Gronlund, N. E. 1995. *Measurement and Assessment in Teaching.* Columbus: Prentice-Hall.
- Linnakylä, P. 1995. *Lukutaidolla maailmankartalle.* Kansainvälinen lukutaitotutkimus Suomessa. Kasvatustieteiden tutkimuslaitos.
- Linnakylä, P. 2004. *Koulutuksen arvioinnin eettiset ehdot.* Virkaanastujaisesitys. Koulutuksen tutkimuslaitos. Jyväskylän yliopisto.
- Linnakylä, P. & Saari, H. 1993. *Miten peruskoulua arvioitiin?* Linnakylä, P. & Saari, H. (toim.) *Oppiiko oppilas peruskoulussa? Peruskoulun arviointi 90- tutkimuksen tuloksia.* Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos. 1-4.
- Lloyd-Jones, R. 1977. *Primary trait scoring.* Cooper, C. R & Odell, C. (toim.) *Evaluating Writing.* New York: National Council of Teachers of English. 33-69.
- Luoma, S. 2001. *What does your test measure? Construct definition in language test development and validation.* Centre for Applied Language Studies. Jyväskylä: Jyväskylä University Printing House.
- Luoma, S. 2004. *Assessing Speaking.* Cambridge Language Assessment Series. Toim. J. C. Alderson & Lyle F. Bachman. Cambridge: Cambridge University Press.
- Lynch, B. 1996. *Evaluating a program inside and out.* J. C. Alderson & A. Beretta (toim.) *Evaluating Second Language Education.* Cambridge Applied Linguistics. Cambridge: Cambridge University Press. 61-96
- Lynch, B. 2003. *Language Assessment and Programme Evaluation.* Edinburgh: Edinburgh University Press Ltd.
- Mager, R. F. 1971. *Opetustavoitteiden määrittäminen.* Helsinki: Otava.

- Mauranen, A. 1986. *Monivalintacloze englannin tekstinyymmärtämisen mittarina*. Laudaturtyö. Helsingin yliopisto. Kasvatustieteen laitos. 25.11.1986.
- McNamara, T. 1996. *Measuring Second Language Performance*. London and New York: Longman.
- McNamara, T. 2000. *Language Testing*. Oxford: Oxford University Press.
- Messick, S. 1998. *The once and future issues of validity*. Assessing the meaning and consequences of measurement. Wainer H. & Brown, H. (toim.) *Test Validity*. Hillsdale, NJ: Erlbaum. 33-45.
- Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference. Draft 2 of Framework Proposed*. 1996. Strasbourg, Council of Cultural Co-operation. CC_Lang (95). 5 rev. 10-17.
- Morris, M. 2003. *Ethical Considerations in Evaluation*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer .303-327.
- Morrow, K. 1979. *Communicative Language Testing: revolution or evolution?* Brumfit C. J. & Johnson K. (toim.) *The Communicative Approach to Language Teaching*. Oxford: Oxford University Press. 143-157.
- Moss, P. A. 1994. *Can there be validity without reliability?* *Educational Researcher* 23(2). 5-12.
- Nation, P. 1990. *Teaching and Learning Vocabulary*. New York: Heinle and Heinle.
- Nitko, A. J. 2001. *Educational Assessment of Students*. 3 painos. New Jersey: Prentice-Hall, Inc.
- Norris, N. 1990. *Understanding Educational Evaluation*. Published in association with CARE, School of Education, University of East Anglia. London: Kogan Page.
- Nummenmaa, L. 2004. *Käyttätymistieteiden tilastolliset menetelmät*. Helsinki: Tammi.
- Nunan, D. 1991. *Language Teaching Methodology. A Textbook for Teachers*. New York: Prentice-Hall.
- Nunan, D. 2001. *The Learner-Centred Curriculum. A study in second language teaching*. Long, M. H. & Richards, J. C. (toim.) *Cambridge Applied Linguistics*. Cambridge: Cambridge University Press.
- Oller, J. W. Jr. 1979. *Language Tests at School*. London: Longman.
- OPH/Opetushallitus. 1993. *Suullisen kielitaidon kehittämisen kokeilun väliraportit*. Moniste.
- OPH/Opetushallitus. 1998a. *Kansallinen oppimistulosten arviointijärjestelmä*. Arviointi 4/1998.
- OPH/Opetushallitus. 1998b. *Koulutuksen tuloksellisuuden arviointimalli*. Arviointi 7/1998. Helsinki: Yliopistopaino.
- OPH/Opetushallitus. 1999. *Perusopetuksen päättöarvioinnin kriteerit*. Arvosanan hyvä (8) kriteerit yhteisissä oppiaineissa. Helsinki: Yliopistopaino.
- OPH/Opetushallitus. 2000. *Peruskoulun opetussuunnitelman perusteet 1994*. 4. korjattu painos. Helsinki: Edita Oy.
- OPH/Opetushallitus. 2004. *Perusopetuksen opetussuunnitelman perusteet 2004*. Vammala: Vammalan kirjapaino.

- Owen, J. M. & Rogers, P. J. 1999. *Program Evaluation. Forms and Approaches*. Thousand Oaks, CA.: Sage.
- Pasanen, U-M. 1996. *Englannin ja ruotsin kielten taidoista peruskoulun päättövaiheessa*. Jakku-Sihvonen, R., Lindström, A. & Lipsanen, S. (toim.). Toteuttaako peruskoulu tasa-arvoa? Opetushallitus. Arviointiraportteja 1/96, 352-362.
- Patton, M. Q. 1981. *Creative Evaluation*. Beverly Hills CA: Sage.
- Patton, M. Q. 1997. *Utilization-Focused Evaluation*. The New Century Text. Thousand Oaks, CA: Sage.
- Patton, M.Q. 2003. *Utilization-Focused Evaluation*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.). *International Handbook of Educational Evaluation*. London: Kluwer 223-243.
- Pearson, P. D. & Johnson, D. D. 1978. *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.
- Perfetti, C. A. 1999. *Comprehending written language: a blueprint of the reader*. Brown, Colin M. & Hagoort Peter (toim.) *The Neurocognition of Language*. Oxford: Oxford University Press. 167-198.
- Peruskoulun opetussuunnitelmakomitean mietintö I. (POPS I) 1970*. Komiteamietintö 1970:A4. Helsinki: Valtion painatuskeskus
- Phillips, Dianne. 1995. *Correspondence analysis*. Social Research Update. Department of Sociology, University of Surrey, UK.
- Pohjoismaiset mietinnöt. 1973. *Harmonisering av den obligatoriske engelskundervisning i de nordiske land*. Nordisk samarbeid på undervisningsområdet. Nordisk utredningsserie 14/73. Nordisk råd og sekretariate for nordisk kulturelt samarbeid.
- Popham, W. J. 1978. *Criterion-referenced measurement*. Englewood Cliffs, New Jersey: Prentice Hall.
- Popham, W. J. 1999. *Classroom Assessment. What Teachers Need to Know*. Boston: Allyn and Bacon.
- POPS I. 1970. Ks. Peruskoulun opetussuunnitelmakomitean mietintö I.
- Postlethwaite, T. N. 1975. *The Surveys of the International Association for the Evaluation of Educational Achievement (IEA)*. Purves, Alan C. & Levine, Daniel U. (toim.) *Educational Policy and International Assessment*. Berkeley: McCutchan Publishing Corporation. 1-32.
- Purpura, J. E. 2004. *Assessing Grammar*. Cambridge Language Assessment Series. J. Charles Alderson & Lyle F. Bachman (toim.) Cambridge: Cambridge University Press.
- Putnam, H. 2002. *The collapse of the fact/value dichotomy and other essays*. Cambridge, Mass.: Harvard University Press.
- Raivola, R. 2000. *Tehoa vai laatua koulutukseen?* Juva: WSOY
- Rea-Dickins, P. 1991. *What makes a grammar test communicative?* Alderson J. C. & North, B. (toim.) *Language Testing in the 1990's: the Communicative Legacy*. New York: HarperCollins. 112-135.
- Rea-Dickins, P. & Germaine, K. 1992. *Evaluation*. Oxford: OUP.
- Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Relating Language Examinations to the Common European Framework of Reference for*

- Languages: Learning, Teaching, Assessment (CEF)*. 2003. Manual. Preliminary Pilot Version. Strasbourg. Language Policy Division of the Council of Europe.
- Rich, R.F. 1979. *Emerging Issues for Evaluators and Evaluation Users*. Datta, L.- E. & Perloff, R. (toim.). *Improving Evaluations*. Beverly Hills, CA: Sage. 245-256.
- Richards, J. C. 1985. *The Context of Language Teaching*. Cambridge: Cambridge University Press.
- Rinne, R. 2001. *Suomalainen yhtenäiskoulu ja ylikansallinen koulutuspolitiikka*. Esitelmä Opetushallituksen juhlaseminaarissa 26.4.2001.
- Rinne, R. & Vuorio-Lehti, M. 1996. *Tasa-arvon suuri masuuni: muistutuksia 1960- ja 70 - lukujen peruskoulukeskusteluista*. Jakku-Sihvonen, R., Lindström, A. & Lipsanen, S. (toim.) *Toteuttaako peruskoulu tasa-arvoa? Arviointi 1/96*. Opetushallitus ja kirjoittajat. Helsinki: Yliopistopaino. 36-65.
- Rivers, W. M. 1972. *Vieraan kielen opettamisen psykologiaa*. Jyväskylä: K. J. Gummeruksen kirjapaino.
- Rost, M. 1990. *Listening in language learning*. London: Longman.
- Saari, H. 1996. *Englannin kielitaidot, puheen ymmärtäminen peruskoulun 9. luokalla v. 1995*. Jakku-Sihvonen, R., Lindström, A. & Lipsanen, S. (toim.) *Toteuttaako peruskoulu tasa-arvoa? Arviointi 1/96*. Opetushallitus ja kirjoittajat. Helsinki: Yliopistopaino. 451-463.
- Saleva, M. 1997. *Now They're Talking*. Testing Oral Proficiency in a Language Laboratory. Faculty of Humanities. Jyväskylä University.
- Scriven, M. 1980. *The Logic of Evaluation*. Pt. Reyes CA: Edgepress.
- Scriven, M. 1981. *Evaluation Thesaurus*. Third edition. California: Edgepress.
- Scriven, M. 1991. *The Science of Valuing*. Shadish, William R. Jr, Cook, Thomas D. & Leviton, Laura C.. *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, CA: Sage. 73-118.
- Shadish, W. R. Jr. 1991. *Social Program Evaluation: Its History, Tasks, and Theory*. Shadish W. R. Jr, Cook, T. D. & Leviton L. C. (toim.). 1991. *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, CA: Sage. 19-35.
- Shohamy, E. 1993. *The Exercise of Power and Control in the Rhetorics of Testing*. Huhta, A., Sajavaara, K. & Takala, S. (toim.) *Language Testing: New Openings*. Jyväskylä. 23-38.
- Skolverket, 2004. *Engelska i åtta europeiska länder*. En undersökning av ungdomars kunskaper och uppfattningar. Svensk huvudrapport. Rapport 242. Kalmar: Lenanders Grafiska.
- Smith, E. R., Tyler, R. N. & the Evaluation Staff. 1942. *Appraising and recording student progress. Adventure in American Education*. Volume III. New York: Harper & Brothers.
- Smith, M. F. 1989. *Evaluability Assessment. A Practical Approach*. Norwell: MA Kluwer.

- Smith, M. L. & Glass, G. V. 1987. *Research and evaluation in education and the social sciences*. Englewood Cliffs, NJ: Prentice-Hall.
- Spolsky, B. 1993. *Institutional and political aspects of language testing*. Huhta A., Sajavaara K. & Takala, S. (toim.) *Language testing: New Openings*. Jyväskylä. 38-45.
- Stake, R.E. 1991. *Responsive Evaluation and Qualitative Methods*. Shadish, W.R. Jr, Cook, T.D. & Leviton, L.C. 1991. *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, CA: Sage. 270-314.
- Stake R. E. 2003. *Responsive Evaluation*. Kellaghan, T., Stufflebeam, D. & Wingate, L.(toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 63-68.
- Stufflebeam, D. 1994. *Empowerment Evaluation*. *Objectivist Evaluation and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go*. *Evaluation Practice*. 15(3):321-328.
- Stufflebeam, D. 2003. *The CIPP Model for Evaluation*. Kellaghan, T., Stufflebeam, D. & Wingate, L.(toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 30-62.
- Stufflebeam, D. 2003. *Professional Standards and Principles for Evaluations*. Kellaghan, T., Stufflebeam, D. & Wingate, L. (toim.) *International Handbook of Educational Evaluation*. London: Kluwer. 279-302.
- Takala, S. 1973. *Kielitaidon mittaaminen*. Opas 3b. POPS-70. Kouluhallitus. Kokeilu- ja tutkimustoimisto. Helsinki: Valtion painatuskeskus.
- Takala, S. 1980. *Kriteerimittamisen käsitteestä ja käytännön sovelluksista*. Kasvatustieteiden tutkimuslaitos. *Selosteita ja tiedotteita* 146/1980.
- Takala, S. 1982a. *First national assessment of teaching in the comprehensive school 1979. English as a Foreign Language, Grade 9: Data on Listening Comprehension, Reading Comprehension and Grammar Test Items. Part III*. Kasvatustieteiden tutkimuslaitos. *Selosteita ja tiedotteita* 205/1982.
- Takala, S. 1982b. *On the origins, communicative parameters and processes of writing..* Choppin B. H. & Postlethwaite T. N. (toim.) *Evaluation in Education*. 1982. Vol. 5, nro 3. *An International Review Series*. Great Britain: Pergamon Press. 209-230.
- Takala, S. 1984. *Evaluation of Students' Knowledge of English Vocabulary in the Finnish Comprehensive School*. Kasvatustieteiden tutkimuslaitoksen julkaisuja 350.
- Takala, S. (toim.). 1987. *Katsauksia kirjoittamiseen*. KTL:n julkaisusarja B. Teoriaa ja käytäntöä 5.
- Takala, S. 1992. *Virikkeitä uutta luovaan koulutyöhön*. 7 40. Kasvatustieteiden tutkimuslaitos.
- Takala, S. 1993. *Suullisen kielitaidon testaaminen peruskoulun kokeiluvaiheessa*. Takala, S. (toim.) *Suullinen kielitaito ja sen arviointi*. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 77. 15-26.

- Takala, S. 1997. *Vieraan kielen kehittymisen arviointiperusteita*. Jakku-Sihvonen, Ritva (toim.) *Onnistuuko oppiminen – oppimistuloksien ja oppimisen laadun arviointiperusteita peruskoulussa ja lukiossa*. Opetushallitus. Arviointi 3/1997. Helsinki: Yliopistopaino. 87-114.
- Takala, S. 1998. *Englannin kielen taso Suomessa – hyvää kehitystä*. Takala, S. & Sajavaara, K. (toim.) *Kielikoulutus Suomessa*. Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus. Jyväskylä: Yliopistopaino. 73-89.
- Takala, S. 2004. *Englannin kielitaidon tasosta Suomessa*. Leimu, K. (toim.) *Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Yliopistopaino. 255-276.
- Takala, S. 2005a. *Ylioppilastutkinnon B-ruotsin arvosanojen ja kielten eurooppalaisen viitekehyksen taitotasojen vastaavuutta koskeva tutkimus*. Muistio 1.2.2005.
- Takala, S. 2005b. *Reflections on the development and impact of language education in Finland*. Annikki Koskensalon juhlaKirja J. Smeds, K. Sarmavuori, E. Laakkonen & R. de Cillia (toim.) *Multicultural communities, multilingual practice: monikulttuuriset yhteisöt, monikielinen käytäntö*. Festschrift für Annikki Koskensalo zum 60. Geburtstag. Annales Universitatis Turkuensis B 285.
- Takala, S. & Havola, L. 1983. *Kielitieteen, psykologian ja filosofian väliset yhteydet kielen tutkimuksen näkökulmasta*. Kasvatus, Vol. 14, No 4. 319-326.
- Thompson, B. 2003. *Score Reliability*. Texas: A&M University.
- Threshold Level*. 1990. Kynnystaitoluettelo. Ks. van Ek, J. A. & Trim, J. L. M.
- Tiittula, L. 1993. *Suullinen kielitaito: puhutun kielen ja suullisen viestinnän ominaispiirteitä*. Takala, S. (toim.) *Suullinen kielitaito ja sen arviointi*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 77. 63-76.
- Tuokko, E. 2000. *Peruskoulun 9.vuosisluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. Opetushallitus. Oppimistulosten arviointi 3/2000. Helsinki: Yliopistopaino.
- Tuokko, E. 2002. *Perusopetuksen päättövaiheen ruotsin kielen oppimistulosten kansallinen arviointi 2001*. Opetushallitus. Oppimistulosten arviointi 3/2002. Helsinki: Yliopistopaino.
- Tuokko, E. 2003. *Perusopetuksen päättövaiheen englannin kielen oppimistulosten kansainvälinen arviointi 2002*. Opetushallitus. Oppimistulosten arviointi 3/2003. Helsinki: Yliopistopaino.
- Tuokko, E. 2004. *Perusopetuksen kansallinen arviointi: englannin ja ruotsin osaamistaso*. Sajavaara, K. & Takala, S. (toim.) *Kielikoulutus tienhaarassa*. Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus. Jyväskylä: Yliopistopaino. 149-165.
- Törmäkangas, K. 2004. *Kansainvälisten koulusaavutustutkimusten otannat Suomessa 1960-2000*. Leimu, K. (toim.) *Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Yliopistopaino. 77-98.
- Töttö, P. 2005. *PISA -testi mittasikin ehkä älykkyyttä*. Tieteessä tapahtuu 1/2005.

- Underhill, N. 1987. *Testing Spoken Language*. A handbook of oral testing techniques. Cambridge: Cambridge University Press.
- Vahervuo, T. 1958. *Arvosanojen antaminen*. Helsinki: Otava.
- Valette, R. 1977. *Modern language testing*. New York: Harcourt Brace Jovanovich.
- Van Ek, J. A. & Trim J. L.M. 1990. *Threshold Level 1990*. Council for Cultural Co-operation. Strasbourg: Council of Europe Press.
- Vaurio, Leena. 1998. *Lexical Inferencing in Reading in English on the Secondary Level*. Jyväskylä: Jyväskylä University.
- Verhelst, N.D. 2004a. *Classical Test Theory*. Section C. Reference Supplement to the Preliminary Pilot Version of the Manual. Council of Europe. Language Division. Strasbourg.
- Verhelst, N.D. 2004b. *Item Response Theory*. Section G. Reference Supplement to the Preliminary Pilot Version of the Manual. Council of Europe. Language Division. Strasbourg.
- VIKKE. 1976. Ks. Kouluhallitus 1976.
- Vähäpassi, A. 1986/2004. *Kansainvälisten kirjoitelmatutkimuksen lähtökohtia*. Leimu, K. (toim.) *Kansainväliset IEA -tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Yliopistopaino. 227-236.
- Vähäpassi, A. 1987a. *Kansainvälisen kirjoitelmatutkimuksen taustaa ja tehtävien valinnan perusteet*. Vähäpassi, A. *Todellisuuden kuvausta, pohdintaa ja arviointia edellyttäviä kirjoitustehtäviä ja niiden arviointi*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 6. 3-13.
- Vähäpassi, A. 1987b. *Argumentoivan, taivuttelevan kirjoitustehtävään lähtökohdistista ja kehittelystä*. Vähäpassi, A. (toim.) *Todellisuuden kuvausta, pohdintaa ja arviointia edellyttäviä kirjoitustehtäviä ja niiden arviointi*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 6. 111-123.
- Vähäpassi, A. 1987c. A Vähäpassi (toim.) *Kansainvälisen kirjoitelmatutkimuksen taustaa, tehtävien ja arviointimenetelmien valinnan perusteet*. Vähäpassi, A. (toim.) *Käytännöllisluonteisia ja opiskelussa tarvittavia kirjoitustehtäviä ja niiden arviointi*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 7. 3-44.
- Weigle, S. 2002. *Assessing Writing*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.
- Weir, C. 1993. *Understanding and Developing Language Tests*. Hemel Hempstead: Prentice Hall.
- Weir, C. & Roberts, J. 1994. *Evaluation in ELT*. Oxford: Blackwell Publishers.
- White, E.M. 1985. *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.
- Widdowson, H. 1978. *Exploration in Applied Linguistics. The process and purpose of reading*. Esitys TESOLin kokouksessa Miamiassa heinäkuussa 1977. oup.com/elt/global/teachersclub/teaching/library (23.10.2002).
- Yrjölä, P. 1998. *Nuorten arviointivirkamiesten perehdyttämiskoulutus*. Opetushallitus. Aineistoa arvioinnin: ymmärtämiseen. Moniste.

LIITE 1 Eräissä Euroopan maissa järjestettävät kansalliset kokeet tai arvioinnit.

MAA	ESPANJA	NORJA	RANSKA	SAKSA (Baijeri)	SLOVENIA	SVEITSI ¹	UNKARI
Tietoja							
Luokka-aste	6.	4.	(5.)	6.	9.	3. - 5.	6.
	8.	7.					8.
	10.	10. 11.	9.	10.			10. 12.
Oppilaiden ikä	12	9-10	(9)	12	14-15	9 - 11	12
	14	12-13					14
	16	15-16 16-17	15	16			16 18
Järjestämistiheys	1999	2005	Joka vuosi	Joka vuosi	Kaksi kertaa vuodessa toukokuussa	Joka vuosi	2000, 2002, 2003
	2001	2005					2000, 2002
	2001	2004, 2005					2002, 2003
Arvioitavat osa-alueet	PY, TY, K	TY	4-5 vuoden välein arvioidaan eri taitoja: TY, PY, K	TY, PY, K englannin käyttö (puhumista ei testata keskitetysti)	TY, K, P, kielioppi ja sanasto	PY, TY, K, puhuminen, johon sisältyy diskurssikompetenssi	TY, PY, K
		TY, K					
		TY, K					
Kokeen pituus	à 20 min.	45-60 min.	90 min.	45 min.	1. W: 45 min. 2. P: 15 min (yleensä 6-10 min.)	PY, TY: 45 min. Suullinen vuorovaikutus 45 min.	TY 45 min. PY + K 45 min.
		60 min.	120 min.				
		90 min. 120 min.					
Järjestämisaikankohta	Yhtenä päivänä	TY: jokin päivä 3 viikon aikana; K tietynä päivänä	Yhden viikon aikana 2-3 englannin oppitunnilla	-	1. Yhtenä päivänä	Huhti - toukokuu	Yhtenä päivänä
Arviointien alkamisen aikankohta	1999	2004	Järjestetty useita vuosia	2004	2. 15.3. - 15.4.	2006-2008 (pitkittäistutkimus)	-

¹ Ei kansallinen arviointi, vaan projekti keski-Sveitsin kolmessa valtiossa: Obwalden, Schwyz, Lucern)

Koeversioiden lkm	Useita; rotatoitu	?	Useita; matriisiotanta	1	1	1	kaksi versiota, samat tehtävät eri järjestyksessä
Otostaminen	pps; kaksivaiheinen klusteriotos	Kaikki osallistuvat	Ositettu otos (koulut, kokonaiset luokat)	Kaikki osallistuvat		Ositettu satunnaisotos	Koulumuodon, maantieteellisen aseman ja asutustyypin mukaan
Osiopankki	On	On; TY	On	Ei	On	Ei	Ei
Järjestämistä päättävä taho	INECSE ²	NEB ³	Opetusministeriö	Opetusministeriö	National Testing Centre	⁴	-
Testausorganisaatio	INECSE; pysyvä	Tilapäinen, nimitetään toistaiseksi vuodeksi	OM:n osasto	The State Institute for School Quality and Educational Research	Ks. Ed.	University for Teacher Education, Institute for Teaching and Learning	-
Kokeenlaatijat	Ulkop. asiantunt.	Org. jäsenet + ulkop. asiantunt.	OM:n johdolla opettajia, yliopistonopett., tarkastajia	Virkamiehet ja kokeneet opettajat; laaditaan virastossa	Opettajat, yliopistonop., OPH:n edustaja	Cambridge Young Learners English ym. kokeita	Ulkopuoliset asiantuntijat
Ohjausryhmä	INECSE; asiantunt. eri alueilta	Opettajia ja opett.kouluttajia	Ks. ed.	Joskus professorit	On	Professoreita	On
Oppimistulosten julkistaminen	-	Kyllä.	Ei	OM julkistaa	Kyllä	Kyllä, suunnitelmissa	-
Tehtävien julkistaminen	-	K,;: kyllä. Esimerkkityyppejä TY:stä	Ei	Kyllä	Osittain	Ks. ed.	-

² INECSE = National Institute for Evaluation and Quality of the Education System

³ The National Examination Board

⁴ University for Teacher Education of Central Switzerland, Lucerne, Department of Research and Development, Institute for Teaching and Learning, financed by the Swiss National Science Foundation

Analyysit	INECSE; IRT	Ulkop. asiantuntija	OM:n osasto		Opettajat, yliopistonop., OPH:n edustaja. OPH analysoi puhekokeen.	Tri. Andrea Haenni, Fil. lis.. Sybille Heinzmann	Ulkopuoliset asiantuntijat
Raportointi	INECSE: päätulokset 4 kk; raportti 1 v	-	6-siv. raportti; myöh. koko raportti	OM	Laaditaan.	Tri Andrea Haenni ja prof. Erika Werlen	Opetusministeriö

LIITE 2 Kielellisen viestintätaidon taitotasosteikot.

Pääkomponentit	Kategoriat ja arviointia varten laaditut taitotasosteikot	Esimerkkejä	
Kielelliset tiedot ja taidot (Linguistic competences)	Sanastolliset tiedot ja taidot (Lexical competence) - Sanaston laajuus (s. 159) - Sanaston hallinta (s. 159)	- vakiintuneet sanonnat - yksittäiset sanat - kieliopilliset umpiluokat	
	Kieliopilliset tiedot ja taidot (Grammatical competence) - Kielio pillinen virheettömyys (s. 161)	- elementit - kategoriat - luokat - rakenteet - prosessit - suhteet	
	Semanttiset tiedot ja taidot (Semantic competence)	- leksikaalinen semantiikka - kieliopillinen semantiikka - pragmaattinen semantiikka	
	Fonologiset tiedot ja taidot (Phonological competence) - Ääntäminen (s. 166)	- kielen äänneyksiköt - distinktiiviset piirteet - sanojen foneettinen koostumus - lauseen fonetiikka (prosodia) - foneettinen redusoituminen	
	Oikeinkirjoitustaito (Orthographic competence) - Oikeinkirjoitus (s. 167)	- isot ja pienet kirjaimet - sanojen oikeinkirjoitus - välimerkit - typografiset käytänteet - yleisesti käytetyt logografiset merkit	
	Kirjoituksen luonteva ääneenlukutaito (Orthoepic competence)	- oikeinkirjoituksen tuntemus - sanakirjan käyttö - kirjoitustavan ja esim. välimerkkien vaikutus jäsentämiseen ja intonaatioon - kyky ratkaista monitulkintaisuus	
	Sosiolingvistiset tiedot ja taidot (Sociolinguistic competences)	Sosiaalisten suhteiden kielellinen osoittaminen - Sosiolingvistinen tarkoituksenmukaisuus (A1-B1) (s. 171) - Vuorovaikutustaidot: Puheenvuoron vaihtuminen (s. 175)	- tervehdysten käyttö ja valinta - puhuttelumuotojen käyttö ja valinta - puheenvuoron vaihtumissäännöt - voimasanojen käyttö ja valinta
		Kohteliaisuussäännöt - Sosiolingvistinen tarkoituksenmukaisuus (A1-B1) (s. 171)	- ”positiivinen” kohteliaisuus - ”negatiivinen” kohteliaisuus - kohteliaisuussanojen käyttö - epäkohteliaisuus
		Kansanviisauden ilmaukset - Sosiolingvistinen tarkoituksenmukaisuus B2- (s. 171)	- sananlaskut - idiomit - sitaatit - uskomuksiin, asenteisiin ja arvoihin liittyvät ilmaukset
		Rekisterien erot - Sosiolingvistinen tarkoituksenmukaisuus B2- (s. 171)	- kaavoittunut - muodollinen - neutraali - epämuodollinen - tuttavallinen - läheinen

	Kielimuodon vaihtelu - Sosiolingvistinen tarkoituksenmukaisuus B2- (s. 171)	- sosiaaliluokka - alueellinen tausta - kansallinen tausta - etninen tausta - ammatillinen tausta
Pragmaattiset tiedot ja taidot (Pragmatic competences)	Diskurssikompetenssi - Teeman käsittely (s. 175) - Joustavuus (s. 174) - Tekstin sidosteisuus (s. 174)	Lauseiden järjestäminen tekstiksi ottaen huomioon - topiikki / fokus - tuttu / uusi informaatio - ”luonnollinen”, esim. aikajärjestys - syy / seuraus - temaattisuus - johdonmukaisuus ja sidosteisuus - looginen järjestys - tyyli ja rekisteri -retorinen tehokkuus - yhteistyön periaate (l. Gricen maksimit) Tekstin muotoilu tietäen, kuinka - informaatio jäsennetään eri tekstityyppejä varten - tarina kerrotaan - perusteluketju rakennetaan - kirjoitettu teksti asetellaan typografisesti, varustetaan siirtymäsignaaleilla ja jaksotetaan Funktionaalinen kompetenssi - Puheen sujuvuus (s. 180) - Ilmaisun tarkkuus (s. 181)

LIITE 3 Eurooppalaisen viitekehyksen kommunikatiivisten toimintojen ja strategioiden taitotasoasteikot.

Kielelliset toiminnot	Taitotasoasteikot	Sivu
<i>Tuottaminen</i>		
Puhuminen	Yleiset taitotasot	93
	Kokemuksen kuvaaminen	93
	Asian esittäminen (esim. väittelyssä)	94
	Julkinen tiedottaminen	94
	Yleisölle puhuminen	95
Kirjoittaminen	Yleiset taitotasot	96
	Luova kirjoittaminen	97
	Raportit ja kirjoitelmat	92
Tuottamisstrategiat	Suunnittelu	100
	Kompensointi	100
	Tarkkailu ja korjaaminen	100
<i>Vastaanottaminen</i>		
Kuuntelu	Yleiset taitotasot: Kuullun ymmärtäminen	102
	Syntyperäisten kielenpuhujien välisen keskustelun ymmärtäminen	103
	Paikalla olevana yleisönä kuunteleminen ja ymmärtäminen	103
	Kuulutusten ja ohjeiden ymmärtäminen	104
	Radion ja äänitteiden ymmärtäminen	104
Lukeminen	Yleiset taitotasot: Luetun ymmärtäminen	106
	Kirjeenvaihdon lukeminen ja ymmärtäminen	106
	Lukeminen asiaan tutustumista varten	107
	Lukeminen tiedonhankintaa ja argumentointia varten	107
	Ohjeiden lukeminen ja ymmärtäminen	108
Audiovisuaalinen vastaanottaminen	Television ja elokuvan katseleminen	109
Vastaanottamisstrategiat	Vihjeiden tunnistaminen ja johtopäätösten teko puhutussa ja kirjoitetussa tekstissä	110
<i>Vuorovaikutus</i>		
Suullinen vuorovaikutus	Yleiset taitotasot: Suullinen vuorovaikutus	112
	Syntyperäisen puhekumppanin ymmärtäminen	113
	Keskustelu	114
	Vapaamuotoinen keskustelu: Juttelu ystävien kanssa	115
	Virallinen keskustelu: Neuvottelut ja kokoukset	116
	Tavoitteellinen yhteistyö (esimerkiksi auton korjaaminen, dokumentista keskusteleminen, tapahtuman järjestäminen)	117
	Tavaroiden ja palvelujen hankkiminen	118
	Tiedonvaihto	119
	Haastatteleminen ja haastateltavana oleminen	120
	Kirjallinen vuorovaikutus	Yleiset taitotasot: Kirjallinen vuorovaikutus
	Kirjeenvaihto	121
	Kirjelaput, viestit, lomakkeet	122
Vuorovaikutusstrategiat	Puheenvuoron vaihtaminen	125
	Yhteistyö	125
	Tarkennuksen pyytäminen	126
Kirjallinen reagoiminen puhuttuun tai kirjoitettuun syötökseen	Muistiinpanojen tekeminen esimerkiksi luennoilla tai seminaareissa	138
	Tekstin työstäminen	139

LIITE 4 Puheen ymmärtämisen kokeen sisältö ja vaikeustaso Grideillä arvioituna.

TEHTÄVÄ 1.

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Laadittu teksti	1.	MV	3.1.6	A2/B1	B1
2. Autenttisuus	Pedagoginen	2.	MV	3.1.6	A2	B1
3. & 4. Diskurssityyppi	Pääasiassa kertovaa	3.	MV	3.1.6	A2	A2
5. Kielenkäyttöalue	Henkilökohtainen	4.	MV	3.1.6	A2	A2
6. Aihe	Suhteet muihin ihmisiin	5.	MV	3.1.6	B1	A2
7. Sisältö	Pääasiassa konkreettinen	6.	MV	3.1.6	B1	B1
8. Kesto	1'41''	7.	MV	3.1.6	A2	A2
9. Sanasto	Enimmäkseen frekventtiä	8.	MV	3.1.6	B1	B1
10. Kielioppi	Pääasiassa yksinkertaista	9.	MV	3.1.6	B1	B1
11. Nopeus	Nopea (osioissa 1-2 normaali)					
12. Osallistujat	1 (1-6), 2 (7-9)					
13. Aksentti	Standardi					
14. Artikulaatio	Normaali					
15. Kuuntelukerrat	2					
16. Taitotaso					A2/B1	A2/B1

TEHTÄVÄ 2

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Aito: uutinen	1.	O/V	1.1.2	B1/B2	B1
2. Autenttisuus	Autenttinen	2.	O/V	1.1.2	A2/B1	B1
3. & 4. Diskurssityyppi	Kertova, raportti					
5. Kielenkäyttöalue	Julkinen					
6. Aihe	Vapaa-aika/Viihde/Urheilu					
7. Sisältö	Melko laaja, abstrakti					
8. Kesto	1'9''					
9. Sanasto	Melko laaja					
10. Kielioppi	Jossakin määrin kompleksista					

11. Nopeus	Normaali				
12. Osallistujat	1				
13. Aksentti	Standardi				
14. Artikulaatio	Normaali				
15. Kuuntelukerrat	2				
16. Taitotaso				B1/B2	B1/B2

TEHTÄVÄ 3

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Aito: turisti-informaatio	1.	AVO	1.1.2	B1	A2
2. Autenttisuus	Autenttinen	2.	AVO	1.1.2	B1	B1
3. & 4. Diskurssityyppi	Kertova, opastava	3.	AVO	1.1.2	B1	A2
5. Kielenkäyttöalue	Julkinen	4.	AVO	1.1.2	B1	A2
6. Aihe	Vapaa-aika/Viihde/Matkailu					
7. Sisältö	Melko laaja					
8. Kesto	1' 38''					
9. Sanasto	Melko laaja					
10. Kielioppi	Jossakin määrin kompleksista					
11. Nopeus	Normaali					
12. Osallistujat	1					
13. Aksentti	Standardi					
14. Artikulaatio	Normaali					
15. Kuuntelukerrat	2					
16. Taitotaso					B1	A2

TEHTÄVÄ 4

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Aito: Radio-/tv-dokumentti	1.	MV	1.1.2	B1	A2/B1
2. Autenttisuus	Autenttinen	2.	MV	1.1.2	B1	B1
3. & 4. Diskurssityyppi	Kertova, raportti	3.	MV	1.1.2	B1	A2/B1
5. Kielenkäyttöalue	Julkinen, ammatillinen					
6. Aihe	Vapaa-aika/Viihde/Urheilu					
7. Sisältö	Melko laaja, abstrakti					
8. Kesto	1' 33''					
9. Sanasto	Melko laaja					
10. Kielioppi	Jossakin määrin kompleksista					
11. Nopeus	Normaali					
12. Osallistujat	1					
13. Aksentti	Standardi					
14. Artikulaatio	Normaali					
15. Kuuntelukerrat	2					
16. Taitotaso					B1	A2/B1

TEHTÄVÄ 5

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Aito: Ääneen luettuja kirjoitettuja tekstejä	1.	AVO	1.1.2	A2/B1	B1
2. Autenttisuus	Autenttinen	2.	AVO	1.1.2	A2/B1	A2/B1
3. & 4. Diskurssityyppi	Kertova					
5. Kielenkäyttöalue	Henkilökohtainen					
6. Aihe	Suhteet muihin ihmisiin					
7. Sisältö	Melko laaja, abstrakti					
8. Kesto	1' 20''					
9. Sanasto	Melko laaja					
10. Kielioppi	Jossakin määrin kompleksista					
11. Nopeus	Normaali					

12. Osallistujat	3		
13. Aksentti	Standardi		
14. Artikulaatio	Normaali		
15. Kuuntelukerrat	2		
16. Taitotaso		A2/B1	A2/B1

TEHTÄVÄ 6

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Aito: Radio-/tv-dokumentti	1.	MV	2.2.2	B1	B1
2. Autenttisuus	Autenttinen	2.	MV	2.1.2	B1	A2/B1
3. & 4. Diskurssityyppi	Selittävä/Selostava	3.	MV	1.1.2	B1	A2/B1
5. Kielenkäyttöalue	Henkilökohtainen					
6. Aihe	Muu: trendit					
7. Sisältö	Melko laaja, abstrakti					
8. Kesto	1' 45''					
9. Sanasto	Melko laaja					
10. Kielioppi	Jossakin määrin kompleksista					
11. Nopeus	Normaali					
12. Osallistujat	3					
13. Aksentti	Standardi					
14. Artikulaatio	Normaali					
15. Kuuntelukerrat	2					
16. Taitotaso					B1/B2	A2/B1

TEHTÄVÄ 7

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso	
					T	O
1. Lähdetyyppi	Laadittu: Ääneen luettu kirjoitettu teksti	1.	Täyd.	1.1.2	A2	A2
2. Autenttisuus	Autenttinen	2.	Täyd.	1.1.2	A2/B1	A2/B1
3. & 4. Diskurssityyppi	Opastava	3.	Täyd.	1.1.2	A2	A2
5. Kielenkäyttöalue	Henkilökohtainen	4.	Täyd.	1.1.2	B2	B2
6. Aihe	Ruoka ja juoma	5.	Täyd.	1.1.2	B1	B1
7. Sisältö	Pääasiassa konkreettinen	6.	Täyd.	1.1.2	A2	A2
8. Kesto	¹	7.	Täyd.	1.1.2	B1/B2	B1/B2
9. Sanasto	Enimmäkseen frekventtiä	8.	Täyd.	1.1.2	A2	A2
10. Kielioppi	Pääasiassa yksinkertaista					
11. Nopeus	Normaali					
12. Osallistujat	1					
13. Aksentti	Standardi					
14. Artikulaatio	Normaali					
15. Kuuntelukerrat	2					
16. Taitotaso					A2/B1	A2/B1

¹ 1. lukemiskerta 14'' (johdanto) + 38'' = 52''; 2. lukemiskerta hidastettuna ilman johdantoa: 1'32''.

LIITE 5 Puheen ymmärtämisen kokeen sisällön arvioinnin tulkinta-avain

Lähdetyyppi	Tekstilähde	
	Aito	Laadittu
	1.1 Kuulutukset ja ohjeet	
	1.2 Julkiset puheet, luennot, esitykset	
	1.3 Viihde (esim. draama)	
	1.4 Uutislähettykset	
	1.5 Keskustelut	
	1.6 Puhelinkeskustelut	
	1.7 Haastattelut	
	1.8 Puhelintiedotukset	
	1.9 Liikennetiedotteet	
	1.10 Sää tiedotukset	
	1.11 Opastus (esim. nauhoitus)	
	1.12 Mainokset	
	1.13 Rutiinikehotukset (esim. lentokoneessa)	
	1.14 Radio- ja tv-dokumentit	
	1.15 Radion soitto-ohjelmat (Phone-in)	
2. Autenttisuus	2.1 Autenttinen 2.2 Lyhennetty/Muokattu 2.3 Pedagoginen 2.4 Yksinkertaistettu	
3 & 4. Diskurssityyppi	3.1 Pääasiassa kuvaileva 3.2 Pääasiassa kertova 3.3 Pääasiassa selittävä 3.4 Pääasiassa argumentoiva 3.5 Pääasiassa ohjaava/opastava	
5. Kielen käyttöalue	5.1 Henkilökohtainen 5.2 Julkinen 5.3 Ammatillinen 5.4 Koulutukseen liittyvä	
6. Aihe (ks. CEF s. 52 + Threshold Level.	6.1 Suhteet muihin ihmisiin 6.2 Vapaa-aika, viihde 6.2.1 Urheilu 6.3 Syöminen ja juominen 6.4 Muu: Jokapäiväinen elämä, trendit	
7. Sisältö	7.1 Pelkästään konkreettinen 7.2 Pääasiassa konkreettinen 7.3 Melko laaja, abstrakti 7.4 Pääasiassa abstrakti	
8. Pituus		
9. Sanasto	9.1 Vain frekventtiä 9.2 Enimmäkseen frekventtiä 9.3 Melko laaja 9.4 Laaja	
10. Kielioppi	10.1 Yksinkertaista 10.2 Pääasiassa yksinkertaista 10.3 Jossakin määrin kompleksista 10.3 Erittäin kompleksista	
11. Nopeus		
12. Osallistujien lukumäärä		
13. Aksentti	13.1 Standardi 13.2 Lievästi murteellinen 13.3 Vahvasti murteellinen	

14. Artikuloinnin selkeys	14.1 Teennäinen		
	14.2 Selkeä		
	14.3 Normaali		
	14.4 Ajoittain epäselvä		
15. Kuuntelukerrat			
16. CEF -taitotaso			
17. Tehtävätyyppi	17.1 Monivalinta		
	17.2 Oikein / Väärin		
	17.3 Lyhyt avovastaus		
	17.4 Muu		
	17.4.1 Käännös (E_S)		
18. Operaatio	Tunnistaa ja esittää	Eksplisiittinen	Pääajatus
	Tehdä johtopäätöksiä		Yksityiskohta
	Arvioida		Mielipide
		Implisiittinen	Asenne
			Johtopäätös
			Kommunikatiivinen tarkoitus
		Tekstin osien väliset yhteydet	
19. CEF -taitotaso			

LIITE 6 Tekstin ymmärtämisen kokeen sisältö ja vaikeustaso Grideillä arvioituna.

TEHTÄVÄ 1.

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Sanomalehti	1.	MV	1.1.2	A2
2. Autenttisuus	Autenttinen	2.	MV	1.1.2	A2
3 & 4. Diskurssityyppi	Pääasiassa ohjaava/opastava	3.	MV	1.1.2	A2
5. Kielenkäyttöalue	Henkilökohtainen	4.	MV	1.1.2	A2
6. Aihe	Vapaa-aika/Viihde/Harrastukset	5.	MV	1.1.2	A2
7. Sisältö	Pääasiassa konkreettinen				
8. Pituus	400 sanaa				
9. Sanasto	Enimmäkseen frekventtiä				
10. Kielioppi	Jossakin määrin kompleksista				
11. Taitotaso	A2/B1				A2

TEHTÄVÄ 2

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Sanomalehti	1.	MV	2.1.5/7	B1/2
2. Autenttisuus	Autenttinen	2.	MV	2.1.5/7	B1
3 & 4. Diskurssityyppi	Pääasiassa kertova	3.	MV	2.1.5/7	B1
5. Kielenkäyttöalue	Henkilökohtainen / Ammatillinen	4.	MV	2.2.5	B1/2
6. Aihe	Muu: Työelämä/Vapaa-aika/Urheilu				
7. Sisältö	Pääasiassa konkreettinen				
8. Pituus	394 sanaa				
9. Sanasto	Enimmäkseen frekventtiä				
10. Kielioppi	Yksinkertaista				
11. Taitotaso	A2/B1				B1/2

TEHTÄVÄ 3

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Matkaliput/Aikataulut	1.	AVO	1.1.2	A1/2
2. Autenttisuus	Autenttinen	2.	AVO	2.2.2	A2
3 & 4. Diskurssityyppi	Pääasiassa selittävä	3.	AVO	2.2.2	A2
5. Kielenkäyttöalue	Julkinen	4.	AVO	2.2.2	A2
6. Aihe	Matkustaminen	5.	AVO	2.2.2	A2
7. Sisältö	Pääasiassa konkreettinen	6.	AVO	1.1.2	A1/2
8. Pituus	578 sanaa (joista 94 numeroa)				
9. Sanasto	Enimmäkseen frekventtiä				
10. Kielioppi	Pääasiassa yksinkertaista				
11. Taitotaso	A2				A2

TEHTÄVÄ 4

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Lähdeteos	1.	MV	1.1.1	A2
2. Autenttisuus	Autenttinen	2.	MV	1.1.1	A2
3 & 4. Diskurssityyppi	Pääasiassa selittävä (lyhyt määritelmä)	3.	MV	1.1.1	A2
5. Kielenkäyttöalue	Julkinen				
6. Aihe	Koti/Ympäristö				
7. Sisältö	Pääasiassa konkreettinen				
8. Pituus	120 sanaa				
9. Sanasto	Melko laaja				
10. Kielioppi	Jossakin määrin kompleksista				
11. Taitotaso	B1				A2

TEHTÄVÄ 5

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Aikakauslehti	1.	O/V	1.1.2	A1/2
2. Autenttisuus	Autenttinen	2.	O/V	1.1.2	A2
3 & 4. Diskurssityyppi	Kertova/Kuvaileva	3.	O/V	2.2.5	A2/B1
5. Kielenkäyttöalue	Henkilökohtainen/Julkinen	4.	Käännös	1.1.2	A2
6. Aihe	Ostokset/Vaatteet	5.	Käännös	1.1.2	A2
7. Sisältö	Pääasiassa konkreettinen	6.	Käännös	1.1.2	A2
8. Pituus	189 sanaa	7.	Käännös	1.1.2	A2
9. Sanasto	Enimmäkseen frekventtiä				
10. Kielioppi	Jossakin määrin kompleksista				
11. Taitotaso	B1				A2

TEHTÄVÄ 6

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Sanomalehti	1.	MV	3.1.1	A2/B1
2. Autenttisuus	Autenttinen	2.	MV	3.1.1	B1
3 & 4. Diskurssityyppi	Kertova / Kuvaileva/Selittävä	3.	MV	3.1.1	B1
5. Kielenkäyttöalue	Julkinen	4.	MV	3.1.1	B1
6. Aihe	Muu: Kulttuurien välistä vertailua				
7. Sisältö	Melko laaja, abstrakti				
8. Pituus	350 sanaa				
9. Sanasto	Melko laaja				
10. Kielioppi	Jossakin määrin kompleksista				
11. Taitotaso	B1/2				B1

TEHTÄVÄ 7

Tekstin piirteet	Tekstin kuvaus	Osio	Tyyppi	Operaatio	Taitotaso
1. Lähdetyyppi	Aikakauslehti	1.	Käännös	1.2.2	A2
2. Autenttisuus	Autenttinen	2.	Käännös	1.2.2	A2
3 & 4. Diskurssityyppi	Kuvaileva	3.	Käännös	1.2.2	A2/B1
5. Kielenkäyttöalue	Julkinen / Henkilökohtainen	4.	Käännös	1.2.2	A2/B1
6. Aihe	Terveys: Henkilökohtainen hygienia	5.	AVO	1.2.2	A1/2
7. Sisältö	Pääasiassa konkreettinen	6.	AVO	2.1.2	A2
8. Pituus	202 sanaa	7.	AVO	2.2.2	B1
9. Sanasto	Enimmäkseen frekventtiä				
10. Kielioppi	Pääasiassa yksinkertaista				
11. Taitotaso	A2+				A2

- 1.1.1 = Tunnistaa eksplisiittisesti ilmaistun pääajatus
1.1.2 = Tunnistaa eksplisiittisesti ilmaistu yksityiskohta
1.2.2 = Tunnistaa implisiittisesti ilmaistu yksityiskohta
2.1.7 = Tehdä johtopäätöksiä eksplisiittisesti ilmaistusta yksityiskohdasta
2.2.2 = Tehdä johtopäätöksiä implisiittisesti ilmaistusta yksityiskohdasta
3.1.1 = Arvioida eksplisiittisesti ilmaistuja tekstin osien välisiä yhteyksiä.

LIITE 7 Tekstin ymmärtämisen kokeen sisällön arvioinnin tulkinta-avain

1. Lähdetyyppi	1.1 Aikakauslehti		
	1.2 Matkaliput, aikataulut		
	1.3 Sanomalehti		
	1.4 Lähdeteos		
	jne.		
2. Autenttisuus	2.1 Autenttinen		
	2.2 Lyhennetty/Muokattu		
	2.3 Pedagoginen		
	2.4 Yksinkertaistettu		
3 & 4. Diskurssityyppi	3.1 Pääasiassa kuvaileva		
	3.2 Pääasiassa kertova		
	3.3 Pääasiassa selittävä		
	3.4 Pääasiassa argumentatiivinen		
	3.5 Pääasiassa ohjaava/opastava		
5. Kielen käyttöalue	5.1 Henkilökohtainen		
	5.2 Julkinen		
	5.3 Ammatillinen		
	5.4 Koulutukseen liittyvä		
6. Aihe (CEF s. 52 ja Threshold: 1995:.65-92)	6.1 Koti, ympäristö		
	6.2 Matkustaminen		
	6.3 Ostokset: Vaatteet		
	6.4 Viihde, vapaa-aika		
	6.5 Terveys/Henkilökohtainen hygienia		
	6.6 Muu:		
	6.6.1 Työ		
	6.6.2 Kulttuurien väliset erot		
7. Sisältö	7.1 Pelkästään konkreettinen		
	7.2 Pääasiassa konkreettinen		
	7.3 Melko laaja, abstrakti		
	7.4 Pääasiassa abstrakti		
8. Pituus			
9. Sanasto	9.1 Vain frekventtiä		
	9.2 Enimmäkseen frekventtiä		
	9.3 Melko laaja		
	9.4 Laaja		
10. Kielioppi	10.1 Yksinkertaista		
	10.2 Pääasiassa yksinkertaista		
	10.3 Jossakin määrin kompleksista		
	10.4 Erittäin kompleksista		
11. Tekstin taitotaso			
12. Tehtävätyyppi	12.1 Monivalinta		
	12.2 Oikein / Väärin		
	12.3 Lyhyt avovastaus		
	12.4 Muu:		
	12.4.1 Käännös (E_S)		
13. Operaatio	Tunnistaa ja esittää	Eksplisiittinen	Pääajatus
	Tehdä johtopäätöksiä		Yksityiskohta
	Arvioida		Mielipide
			Asenne
		Implisiittinen	Johtopäätös
			Kommunikatiivinen tarkoitus
			Tekstin osien väliset yhteydet
14. Taitotaso			

LIITE 8 Kieliopin hallinnan kriteerit

CEF taitotaso / YKI	YKI	General Linguistic Range, CEF p. 110	Grammatical Accuracy, CEF p. 114	Suomen taitotasot	Suomen taitotasojen kuvaukset S= puhuminen, W = kirjoittaminen
A1	Peruskieliopin hallinnassa on suuria puutteita. Osaa tuottaa lähinnä lyhyitä päälauseita. Rakenteiden käyttö on epäjohdonmukaista ja rakenneosia jää pois. Käyttää sanoja usein sattumanvaraisessa muodossa.	Has a very basic range of simple expressions about personal details and needs of a concrete type.	Shows only a limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.	A1.1	S: Osaa joitakin opeteltuja vakioilmaisuja. W: Osaa joukon erillisiä sanoja ja sanontoja.
				A.1.2	S: Osaa joitakin tilannesidonnaisia ilmaisuja ja peruskieliopin aineksia. W: Pystyy kirjoittamaan hyvin yksinkertaisia päälauseita.
				A1.3	S: Osaa rajallisen joukon perustason lauserakenteita. Alkeellisessakin puheessa esiintyy paljon peruskielioppivirheitä. W: Osaa kirjoittaa muutamia yksilauseisia virkkeitä.
A2	Hallitsee yksinkertaisimman peruskieliopin, mutta rakenteiden käyttö on vielä horjuvaa. Osaa muodostaa yksinkertaisia lauseita ja lyhyitä hieman monimutkaisempia virkkeitä. Jättää pois esimerkiksi sanojen päätteitä, artikkeleita ja prepositioita	Can produce brief everyday expressions in order to satisfy simple needs of a concrete type: personal details, daily routines, wants and needs, requests for information. Can use basic sentence patterns and communicate with memorised phrases, groups of a few words and formulae about themselves and other people, what they do, places possessions etc. Has a limited repertoire of short memorised phrases covering	Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement, nevertheless, it is usually clear what he/she is trying to say.	A2.1	S: Osaa monia keskeisimpiä rakenteita (kuten menneen ajan muotoja ja konjunktioita). Hallitsee kaikkein yksinkertaisimman kieliopin alkeellisessä vapaassa puheessa, mutta virheitä esiintyy yhä paljon perusrakenteissakin. W: Osaa perusaikamuotoja sekä yksinkertaisin sidossanoin (ja, mutta) liitetyjä rinnasteisia lauseita. ... tekee toistuvasti virheitä perusasioissa (aika-muodot, taivutus).

		<p>predictable survival situations;; frequent breakdowns and misunderstandings occur in non-routine situations.</p> <hr/> <p>Has a repertoire of basic language which enables him/her to deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words.</p>			
				A2.2	<p>S: Osaa useita yksinkertaisia ja myös joitakin vaativampia rakenteita. Laajemmassa vapaassa puheessa esiintyy paljon virheitä perusasioissa (esim. verbien aikamuodoissa) ja ne voivat joskus haitata ymmärrettävyyttä. W: Osaa perusrakenteet ja tavallisimmat sidoskeinot. ... tekee virheitä harvinaisemmissa rakenteissa ja muodoissa ja tuottaa kömpelöjä ilmaisuja</p>
B1	<p>Hallitsee melko hyvin peruskieliopin keskeiset rakenteet. Osaa muodostaa yleisimmät lausetyypit ja käyttää myös pidempiä ilmauksia. Vaikka rakenteissa saattaa esiintyä epätarkkuutta ja kieliopin hallinta on vielä ajoittain epävarmaa, virheet eivät estä kommunikaatiota.</p>	<p>Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work travel and current events but lexical limitations cause repetition and even difficulty with formulation at times.</p> <hr/> <p>Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics such and music and films.</p>	<p>Uses reasonably accurately a repertoire of frequently used `routines` and patterns associated with more predictable situations.</p> <hr/> <p>Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.</p>	B1.1	<p>S: Käyttää useita erilaisia rakenteita. Laajemmassa vapaassa puheessa kielioppivirheet ovat tavallisia (esim. artikkeleita ja päätteitä puuttuu), mutta ne haittaavat harvoin ymmärrettävyyttä. W: Osaa useimpien tutuissa tilanteissa tarvittavien tekstien laadintaan riittävät rakenteet. Rutiininomainen kieliaines ja perusrakenteet ovat jo suhteellisen virheettömiä, mutta jotkut vaativammat rakenteet ja sanaliitot tuottavat ongelmia.</p>

				B1.2	<p>S: Käyttää monenlaisia rakenteita ja mutkikkaitakin lauseita. Kielioppivirheitä esiintyy jonkin verran, mutta ne häiritsevät harvoin laajempaan viestintään.</p> <p>W: Hallitsee melko monenlaiseen kirjoittamiseen tarvittavia lauseenrakenteita. Osaa ilmaista rinnastisuutta ja alisteisuutta. Pystyy kirjoittamaan ymmärrettävää ja kohtuullisen virheetöntä kieltä, vaikka virheitä esiintyy vaativissa rakenteissa ja vaikka äidinkielen tai jonkin muun kielen vaikutus on ilmeinen</p>
B2	<p>Hallitsee jokseenkin kaikki peruskieliopin rakenteet hyvin. Tuntee myös joitakin monimutkaisia rakenteita, mutta niiden käyttö on vielä kaavamaisista. Kieliopillisia virheitä ja epätarkkuutta esiintyy vielä ajoittain harvinaisten ja monimutkaisten rakenteiden käytössä</p>	<p>Has a sufficient range of language to be able to give clear descriptions, express viewpoints and develop arguments without much conspicuous searching for words, using some complex sentence forms to do so.</p> <p>Can express him/herself clearly and without much sign of having to restrict what he/she wants to say</p>	<p>Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.</p> <p>Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.</p>	B2.1	<p>S: Osaa käyttää monipuolisesti kielen rakenteita. Kieliopin hallinta on melko hyvää eivätkä satunnaiset virheet yleensä häiritsevät ymmärrettävyyttä.</p> <p>W: Hallitsee melko hyvin ... kieliopin käytön eivätkä virheet johda väärinkäsityksiin, vaikka tuotoksessa saattaa näkyä äidinkielen vaikutusta. Vaativat rakenteet tuottavat ongelmia..</p>
				B2.2	<p>S: Hallitsee laajasti kielelliset keinot ilmaista konkreetteja ja</p>

					<p>käsitteellisiä, tuttuja ja tuntemattomia aiheita varmasti, selkeästi ja tilanteen vaatimaan muodollisuusastetta noudattaen.</p> <p>Kielelliset syyt rajoittavat ilmaisua erittäin harvoin. Kieliopin hallinta on hyvää. Usein puhuja korjaa virheensä itse, eivätkä virheet haittaa ymmärrettävyyttä.</p> <p>W: Kielellinen ilmaisuvaramo ei rajoita havaittavasti kirjoittamista. Hallitsee hyvin kieliopin. Virheitä voi esiintyä harvinaisissa rakenteissa ja idiomaattisissa ilmauksissa sekä tyyliseikoissa</p>
C1	<p>Hallitsee käytännöllisesti katsoen kaikki kieliopilliset rakenteet hyvin ja osaa myös varioida ilmaisuaan. Harvinaisissa rakenteissa saattaa vielä esiintyä satunnaista epätarkkuutta. Vähäinen määrä virheitä saattaa esiintyä, mutta kielenkäyttöä pystyy korjaamaan lähes aina itse.</p>	<p>Can select an appropriate formulation from a broad range of language to express him/herself clearly, without having to restrict what he/she wants to say.</p>	<p>Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot</p>	C1	<p>S: Rakenteisto on hyvin laaja ja rajoittaa ilmaisua erittäin harvoin. Kieliopin hallinta on hyvää. Satunnaiset virheet eivät hankaloita ymmärtämistä ja puhuja osaa korjata ne itse.</p> <p>W: Kielellinen ilmaisuvaramo on hyvin laaja. Hallitsee erittäin hyvin kieliopin. Virheitä voi esiintyä satunnaisesti idiomaattisissa ilmauksissa sekä tyyliseikoissa.</p>
C2	<p>Osoittaa varmaa, joustavaa ja intuitiivista kielen rakenteiden hallintaa. Käyttää monipuolisesti monimutkaisia lause- ja diskurssirakenteita, jotka hallitsee johdonmukaisesti. Pienetkin virheet ovat harvinaisia.</p>	<p>Can exploit a comprehensive and reliable mastery of a very wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity.. No signs of having to restrict what he/she wants to say.</p>	<p>Maintains systematic grammatical control of complex language, even when attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).</p>		

LIITE 9 A2001 -taitotasoasteikon puhumisen kriteerit. (Lähde: Tuokko: 2002:141.)

Asteikko	Tehtävän suorittaminen ja sujuvuus	Ilmaisun laajuus	Virheettömyys	Vuorovaikutus
5	Viesti välittyy selkeästi ja vaivattomasti. Oppilas suoriutuu melko komplekseista puheenvuoroista takeltematta.	Oppilaan sanavarasto on runsas. Hän käyttää vaihtelevia lauseenrakenteita. Käyttää luontevasti kohdekielen idiomeja ja sanontoja.	Oppilas hallitsee käyttämänsä kieliopilliset muodot ja osaa korjata satunnaiset virheet myös ääntämisessään. Kiinnittää huomiota myös intonaatioon.	Oppilas viestii luontevasti ja aloitteellisesti vaativammissakin tilanteissa. Interaktio toimii hyvin.
4	Viesti välittyy tarkoitettulla tavalla. Oppilas suoriutuu tehtävästä suhteellisen vaivattomasti, vaikka toisinaan epäröiden.	Oppilaalla on laajahko sanavarasto. Hän käyttää vaihtelevia lauseenrakenteita ja pystyy sisällyttämään tuotukseensa joitakin kohdekielen idiomeja ja sanontoja.	Oppilas käyttää kieliopillisiä muotoja jokseenkin virheettömästi. Ääntäminen on lähes virheetöntä.	Oppilas viestii luontevasti ja on melko aloitteellinen.
3	Viesti välittyy kohtuullisesti. Puhe saattaa olla epäröivää, mutta on paikoitellen kuitenkin sujuvaa ja luonnollista.	Oppilas käyttää perustason sanastoa ja lauseenrakenteita.	Oppilas käyttää perustason kieliopillisiä muotoja melko virheettömästi. Ääntämisvirheet eivät vaikeuta puheen ymmärrettävyyttä.	Oppilas ottaa huomioon kuulijan pyytäen myös tarvittaessa toistoa tai tarkennusta. On jonkin verran aloitteellinen.
2	Oppilas suoriutuu jotakuinkin viestin välittämisestä varsin ilmeisestä epäröinnistä huolimatta.	Sanavarastossa ja lauseenrakenteissa on puutteita.	Oppilas käyttää joitakin perustason kieliopillisiä muotoja lähes oikein. Ääntämisvirheet vaikeuttavat joskus ymmärrettävyyttä.	Oppilas pyrkii ottamaan kuulijan jossakin määrin huomioon, mutta ei ole kovin aloitteellinen.
1	Viesti välittyy vain osittain. Tuotos on katkonaista. Oppilas joutuu usein etsimään ilmauksia ja paikkailemaan tuotostaan.	Sanavarasto on suppea ja virkkeet lyhyitä tai epätäydellisiä.	Kieliopillisten rakenteiden hallinnassa suuria puutteita. Ääntämisvirheet vaikeuttavat suuresti ymmärrettävyyttä.	Oppilas ei ota juuri lainkaan huomioon kuulijaansa eikä ole aloitteellinen.
0	Tehtävä suorittamatta tai täysin käsittämätön tuotos.			

LIITE 10 A2001 -taitotasoasteikon kirjoittamisen kriteerit. (Lähde: Tuokko: 2002:142.)

Asteikko	Luettavuus	Ilmaisuvarasto	Asian käsittely	Kielelliset puutteet
5	Miellyttävää lukea. Aidon tuntuista ja sujuvaa.	Monipuolinen ja suhteellisen idiomaattinen.	Omintakeista ja monipuolista.	Joitakin erehdyksiä.
4	Helppolukuista ja sujuvaa.	Sopiva, mutta ei kovin monipuolinen sanasto. Melko vaihtelevia rakenteita.	Selkeää, mutta melko tavanomaista.	Joitakin virheitä ja epäaitoja ilmauksia.
3	Jokin kohta hankaloittaa lukemista, mutta pääsisältö välittyy.	Rakenteet ja sanasto rajoittuneita.	Aiheen käsittely suppeaa.	Joitakin virheitä perusrakenteissa tai interferenssiä
2	Verrattain hankalaa luettavaa. Kielen hallinta puutteellista. Joskus merkitys jää epäselväksi.	Vaatimatonta ja yksipuolista.	Yksipuolista.	Melko paljon virheitä perusrakenteissa tai häiritsevää interferenssiä.
1	Erittäin vaikeata lukea. Useissa kohdin sanoma ei välity.	Ilmaisuvarasto alkeellista.	Hyvin puutteellista. Irrallisia lauseita.	Paljon virheitä.
0	Tehtävä suorittamatta tai täysin käsittämätön tuotos.	Lähes olematon.	Lähinnä irrallisia lauseita.	Ei juuri mitään oikeaa.

LIITE 11 Puhumisen kokeen laadullisten kriteerien yhdistelmä.

Scale	CEF			Manual		
	Overall Spoken interaction	Phonological Control	Turntaking	Linguistic Range edited from General Linguistic Range, Vocabulary Range, Flexibility	Linguistic Accuracy Edited from Grammatical Accuracy and Vocabulary Control	Turntaking Edited from Turntaking, Cooperating, Asking for Clarification
C2	Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.	As C1	As C1	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Maintains consistent grammatical control of complex language, even when attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly. Can interweave his/her contributions into the joint discourse with fully natural turntaking, referencing, allusion making etc.
C1	Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.	Can select a suitable phrase from a readily available range of discourse functions to preface his/her remarks appropriately in order to get the floor, or to gain time and keep the floor whilst thinking.	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say.	Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.	Can select a suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/her own contributions skilfully to those of other speakers.
B2	Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational and leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality	Has acquired a clear, natural pronunciation and intonation.	Can intervene appropriately in discussion, exploiting appropriate language to do so. Can initiate, maintain and end discourse appropriately with effective turntaking. Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly.	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting

	appropriate to the circumstances. _____ Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments.		Can use stock phrases (e.g. "That's a difficult questions to answer") to gain time and keep the turn whilst formulating what to say.			others in, etc.
B1	Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thought on more abstract, cultural topics such as films, books, music etc. Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. Can enter unprepared into conversation on familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.	Can intervene in a discussion on a familiar topic, using a suitable phrase to get the floor. Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest.	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.
A2	Can interact with reasonable ease in structural situations and short conversations provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.	Can use simple techniques to start, maintain, or end a short conversation Can initiate, maintain and close simple, face-to-face conversation. Can ask for attention.	Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Uses some simple structures correctly, but still systematically makes basic mistakes.	Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord. Can ask for attention.

	<p>information on familiar topics in predictable everyday situations. Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of his/her own accord.</p>					
A1	<p>Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech, rephrasing and repair. Can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.</p>	<p>Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.</p>	<p>No descriptor available.</p>	<p>Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.</p>	<p>Shows very limited grammatical control of a few simple grammatical structures and sentence patterns in a memorised repertoire.</p>	<p>Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.</p>

LIITE 12 Written Assessment Criteria Grid. (Lähde: Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. 2003.)

82

Table 5.8: WRITTEN ASSESSMENT CRITERIA GRID

	Overall	Range	Coherence	Accuracy	Description	Argument
C2	Can write clear, highly accurate and smoothly flowing complex texts in an appropriate and effective personal style conveying finer shades of meaning. Can use a logical structure which helps the reader to find significant points.	Shows great flexibility in formulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Can create coherent and cohesive texts making full and appropriate use of a variety of organizational patterns and a wide range of connectors and other cohesive devices.	Maintains consistent and highly accurate grammatical control of even the most complex language forms. Errors are rare and concern rarely used forms.	Can write clear, smoothly flowing and fully engaging stories and descriptions of experience in a style appropriate to the genre adopted.	Can produce clear, smoothly flowing, complex reports, articles and essays which present a case or give critical appreciation of proposals or literary works. Can provide an appropriate and effective logical structure which helps the reader to find significant points.
C1	Can write clear, well-structured and mostly accurate texts of complex subjects. Can underline the relevant salient issues, expand and support points of view at some length with subsidiary points, reasons and relevant examples, and round off with an appropriate conclusion.	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say. The flexibility in style and tone is somewhat limited.	Can produce clear, smoothly flowing, well-structured text, showing controlled use of organizational patterns, connectors and cohesive devices.	Consistently maintains a high degree of grammatical accuracy, occasional errors in grammar, collocations and idioms.	Can write clear, detailed, well-structured and developed descriptions and imaginative texts in an assured, personal, natural style appropriate to the reader in mind.	Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. Can expand and support point of view with subsidiary points, reasons and relevant examples.
B2	Can write clear, detailed official and semi-official texts on a variety of subjects related to his field of interest, synthesising and evaluating information and arguments from a number of sources. Can make a distinction between formal and informal language with occasional less appropriate expressions.	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, using some complex sentence forms to do so. Language lacks, however, expressiveness and idiomatity and use of more complex forms is still stereotypic.	Can use a limited number of cohesive devices to link his/her sentences into clear, coherent text, though there may be some 'jumpiness' in a longer text.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstandings.	Can write clear, detailed descriptions of real or imaginary events and experiences marking the relationship between ideas in clear connected text, and following established conventions of the genre concerned. Can write clear, detailed descriptions on a variety of subjects related to his/her field of interest. Can write a review of a film, book or play.	Can write an essay or report that develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail. Can evaluate different ideas or solutions to a problem. Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. Can synthesise information and arguments from a number of sources.
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence. The texts are understandable but occasional unclear expressions and/or inconsistencies may cause a break-up in reading.	Has enough language to get by, with sufficient vocabulary to express him/herself with some circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Can link a series of shorter discrete elements into a connected, linear text.	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more common situations. Occasionally makes errors that the reader usually can interpret correctly on the basis of the context.	Can write accounts of experiences, describing feelings and reactions in simple connected text. Can write a description of an event, a recent trip – real or imagined. Can narrate a story. Can write straightforward, detailed descriptions on a range of familiar subjects within his field of interest.	Can write short, simple essays on topics of interest. Can summarise, report and give his/her opinion about accumulated factual information on a familiar routine and non-routine matters, within his field with some confidence. Can write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions.
A2	Can write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'. Longer texts may contain expressions and show coherence problems which makes the text hard to understand.	Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Can link groups of words with simple connectors like 'and', 'but' and 'because'.	Uses simple structures correctly, but still systematically makes basic mistakes. Errors may sometimes cause misunderstandings.	Can write short simple imaginary biographies and simple poems about people. Can write very short, basic descriptions of events, past activities and personal experiences.	
A1	Can write simple isolated phrases and sentences. Longer texts contain expressions and show coherence problems which makes the text very hard or impossible to understand.	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Can link words or groups of words with very basic linear connectors like 'and' and 'then'.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire. Errors may cause misunderstandings.	Can write simple phrases and sentences about themselves and imaginary people, where they live and what they do.	

LIITE 13 Arviointiprosessi

Ajankohta	Prosessin vaiheet	Käytetyt menetit
Elokuu 1998	1. Kokeen sisällöstä päättäminen	Kokeenlaatijoiden ja Opetushallituksen arviointivirkamiehien edustajien kokous
	2. Testispesifikaatiot	Kokeenlaatijoiden, asiantuntijaryhmän sekä erikoissuunnittelijan kokous
Syyskuu 1998		LEHDISTÖLLE TIEDOTE ARVIOINNIN JÄRJESTÄMISESTÄ
Syyskuu- lokakuu 1998	3. Tehtävien laadinta	<ul style="list-style-type: none"> - kokeenlaatijaparit kommentoivat työtään - kokeenlaatijaparit kommentoivat toisten parien työtä - asiantuntijaryhmä kommentoi tuotettua aineistoa
Marraskuu 1998		OTANNAN SUORITTAMINEN
Lokakuu – marraskuu 1998	4. Kirjallisen kokeen esikokeiluaineiston kokoaminen ja osittainen esikokeilu	<ul style="list-style-type: none"> - asiantuntijaryhmän ja erikoissuunnittelijan kokous - asiantuntijoiden ja erikoissuunnittelijoiden neuvottelut - esikokeilun tehtävävihkojen (5 + 6 kpl) laadinta - esikokeilun arvostelulomakkeiden laadinta - pisteitysehdotusten suunnittelu
Joulukuun 1998	5. Suullisen kokeen esitestausaineiston kokoaminen, kirjallisen kokeen osittainen esikokeilu ja palautteen käsittely	<ul style="list-style-type: none"> - suullisen kokeen laatijoiden, asiantuntijaryhmän ja erikoissuunnittelijan kokous - asiantuntijoiden ja erikoissuunnittelijan neuvottelut - esikokeilun äänitteiden teettäminen - esikokeilun äänityskäsikirjoituksen laadinta - esikokeilun tehtävävihkojen (6x2) kokoaminen
Tammikuu 1999	6. Suullisen kokeen esikokeilu, kirjallisen kokeen esikokeilu ja palautteen käsittely	<ul style="list-style-type: none"> - ohjeet järjestäjille ja osallistujille - palaute järjestäjiltä - koetulokset kouluilta
		VARAKOULUJEN VALINTA
Helmikuu 1999	7. Suullisen kokeen palaute, kirjallisen kokeen tehtävistä päättäminen, suullisen kokeen tehtävistä päättäminen	<ul style="list-style-type: none"> - tehtävien ja osioiden valinta osioiden toimivuuden tarkastelun pohjalta: <ul style="list-style-type: none"> • osion korrelaatio mittarin summaan - kirjallinen tuottaminen: <ul style="list-style-type: none"> eniten valitut kirjoittamisen eri lajeja edustavat aiheet
		- TEHTÄVIEN VALINTA PALAUTTEEN POHJALTA
Helmikuu – maaliskuu 1999	9. Kokeiden laatiminen	<ul style="list-style-type: none"> - äänitteiden teettäminen - äänityskäsikirjoitusten laadinta - tehtävävihkojen (3 + 12) suunnittelu ja painattaminen - vastauskaavakkeiden suunnittelu ja painattaminen - optisen lomakkeen suunnittelu ja painattaminen - opettajan ohjeiden laatiminen - rehtorin ohjeiden laatiminen - valvovan opettajan ohjeiden laatiminen - oikoluku - kokeiden postitus
Huhtikuu 1999	10 Arvioinnin järjestäminen	<ul style="list-style-type: none"> - järjestelyt rehtorin vastuulla - otoskoulujen opettajat korjasivat kokeet yhdessä
Toukokuu 1999		PIKAPALAUDE KOULUILLE
Kesäkuu – marraskuu 1999	11. Tulokset	<ul style="list-style-type: none"> - suullisen ja kirjallisen tuottamisen kokeiden osittainen uudelleen arviointi - kvantitatiivinen analyysi - raportin kirjoittaminen
Joulukuun 1999 – tammikuu 2000	12. Raportti	<ul style="list-style-type: none"> - raportin painattaminen ja julkaiseminen - tiedotustilaisuus medialle

LIITE 14 Perusopetuksen opetussuunnitelman perusteiden 1994 ja Perusopetuksen päättöarvioinnin hyvän arvosanan (8) kriteerien englannin kielen taidolle asettamat tavoitteet perusopetuksen päättyessä.

Kielitaidon osa-alue	Opetussuunnitelman perusteet 1994	Päättöarvioinnin kriteerit (arvosana 8)
	Oppilas	Oppilas
<i>Tekstin ymmärtäminen</i>	- ymmärtää helpohkoa yleisiä asioita käsittelevää kirjoitettua kieltä.	- ymmärtää pääkohdittain lyhyitä, selkeitä, tuttuja aiheita käsitteleviä tarinoita ja asiatekstejä, myös lehti uutisia. Vaikeampien ja laajempien tekstikokonaisuuksien hahmottaminen tuottaa ongelmia, mutta hän selviää niistä apuneuvoja käyttäen.
	- on omaksunut kielen keskeistä sanastoa, keskeisiä sanontoja, ...	- pystyy löytämään ja ymmärtämään jokapäiväisissä tilanteissa tarvitsemiaan tietoja, esimerkiksi aikatauluista, ruokalistoista ja mainoksista.
	- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja.	- ymmärtää hänelle osoitettuja kirjallisia viestejä sekä tapahtumien, tunteiden ja toiveiden kuvauksia hänelle osoitetuissa kirjeissä, korteissa ja tietokonevälitteisessä viestinnässä, esimerkiksi sähköpostissa.
	- ymmärtää hänelle osoitettuja kirjallisia viestejä sekä tapahtumien, tunteiden ja toiveiden kuvauksia hänelle osoitetuissa kirjeissä, korteissa ja tietokonevälitteisessä viestinnässä, esimerkiksi sähköpostissa.	- ymmärtää pääkohdittain lyhyitä, selkeitä, tuttuja aiheita käsitteleviä tarinoita ja asiatekstejä, myös lehti uutisia. Vaikeampien ja laajempien asiakokonaisuuksien hahmottaminen tuottaa ongelmia, mutta hän selviää niistä apuneuvoja käyttäen.
<i>Puheen ymmärtäminen</i>	- ymmärtää jokapäiväisiä asioita käsittelevää normaalitempoista puhetta.	- ymmärtää itselleen esitettyjä henkilökohtaisia kysymyksiä, käskyjä ja kehotuksia.
	- on omaksunut kielen keskeistä sanastoa, keskeisiä sanontoja ...	- ymmärtää puhetta riittävästi selviytyäkseen tavallisimmista arkipäivän asiointi- ja vuorovaikutustilanteista, esimerkiksi kaupassa, lippumyymälässä, postissa, pankissa ja ruokailu- ja opastustilanteissa.
	- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja.	- ymmärtää syntyperäisen puhujan hänelle suuntaamaa selkeää yleiskielistä puhetta tutuista aiheista, kunhan puhekumppania voi aika ajoin pyytää toistamaan tai muotoilemaan uudestaan sanomaansa.
		- ymmärtää pääkohdittain tuttuja ja läheisiä asioita käsittelevää selkeää yleiskielistä puhetta, myös tiedotusvälineistä.

		- tunnistaa kielen erilaisia variantteja ja tuntee niiden tärkeimpiä arkielämän sanastoeroja.
<i>Kieliopilliset rakenteet</i>	- on omaksunut kielen ... perusrakenteita.	- pystyy kirjoittamaan virheistä huolimatta ymmärrettävästi henkilökohtaisia kirjeitä tai sähköpostiviestejä ja ilmaisemaan niissä mielipiteitään tai tunteitaan sekä kuvailemaan tapahtumia.
		- pystyy kirjoittamaan virheistä huolimatta ymmärrettävästi yhtenäisiä, yksinkertaisia kertovia tekstejä oman elinympäristönsä arkipäiväisistä aiheista ja ymmärrettäviä tarinoita tarvittaessa apuneuvoja käyttäen.
<i>Puhuminen</i>		- ääntää ymmärrettävästi, joitakin virheitä esiintyy aika ajoin. Intonaatio noudattaa enimmäkseen puhetilanteen vaatimuksia. Puheessa saattaa esiintyä myös hapuilua, mutta viesti kuitenkin välittyy.
	- pystyy osallistumaan tavallisia asioita käsittelevään keskusteluun soveltaen luonnollista ja sujuvaa ääntämistä, painotusta, rytmisiä ja intonaatiota.	- pystyy osallistumaan tutuista aiheista käytävään vapaamuotoiseen keskusteluun, kunhan puhutaan selkeästi artikuloitua puhekieltä ja puhekumppania voi aika ajoin pyytää toistamaan ja muotoilemaan sanomansa uudelleen.
	- pystyy melko sujuvasti kertomaan itsestään ja elämäntilanteestaan sekä pitämään yksinkertaisen, ennalta valmistellun esityksen jostain itseään kiinnostavasta aiheesta.	
	- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja.	- käyttää kohdekulttuurille ominaisia ilmauksia tavallisimmissa lähtilanteissa, esimerkiksi tervehdittäessä, esiteltäessä, kiitettäessä ja hyvästeltäessä.
<i>Vuorovaikutusstrategiat</i>	- pystyy osallistumaan tavallisia asioita käsittelevään keskusteluun.	pystyy keskustelussa kohteliaasti ilmaisemaan henkilökohtaisia mielipiteitään, yksimielisyyttä ja erimielisyyttä sekä tekemään ehdotuksia siitä, mitä tehdä, minne mennä, kuinka järjestää jokin tapahtuma yms.
	- on omaksunut kielen keskeistä sanastoa, keskeisiä sanontoja ja ...	
	- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja.	- pystyy (myös) pyytämään toistoa tai tarkennusta ... sekä käyttämään yksinkertaisempaa ilmaisuja. Osaa muotoilla sanottavansa toisin silloin, kun syntyy kommunikaatiokatkos.

Kirjoittaminen	- kykenee tuottamaan lyhyen, esimerkiksi kertovan tai kuvailevan kirjoitetun tekstin tarvittaessa apuneuvoja käyttäen.	- pystyy kirjoittamaan virheistä huolimatta ymmärrettävästi yksinkertaisia henkilökohtaisia kirjeitä tai sähköpostiviestejä ja ilmaisemaan niissä mielipiteitään tai tunteitaan sekä kuvailemaan tapahtumia.
	- on omaksunut kielen keskeistä sanastoa, keskeisiä sanontoja ja perusrakenteita.	- pystyy täyttämään henkilötietolomakkeen esimerkiksi hotelliin kirjoittauduttaessa, sekä kirjoittamaan arkipäivän viestintätilanteissa tarpeellisia lyhyitä puhelin-, faksi- tai sähköpostiviestejä ja kortteja.
Kulttuurikompetenssi	- tuntee kohdekielelle ja sen kulttuurille ominaisia viestintätapoja.	- ymmärtää kielen ja kulttuurin (laajassa merkityksessä) yhteyden, tuntee keskeisiä kohdekulttuuriin kuuluvia normeja, arvoja, uskomuksia ja tapoja, tietää, miten ne heijastuvat viestintään ja sopeuttaa oman viestintänsä kohdekulttuuriin.
	- on omaksunut tietoja kielialueen maista, kansoista ja kulttuureista.	- tietää keskeisiä asioita kielialueen maantieteestä, elinkeinoista ja kulttuureista.
		- käyttää kohdekulttuurille ominaisia ilmauksia tavallisimmissa lähitilanteissa, esimerkiksi tervehdittäessä, esiteltäessä, kiitettäessä, anteeksipyydettyä ja hyvästeltäessä.

LIITE 15 Esimerkki eräiden tekstin ymmärtämisen esikokeilutehtävien testi-
spesifikaatioista.

Aihepiiri/Otsikko	Vaikeustaso	Funktio	Testattava taito	Tehtävätyyppi	Kieli	
					Tehtävä	Vastaus
Joukkotiedotus	1	tiedon välittäminen.	1. sel. yks.	avo	äk	äk
			2. k			
			3a. p			
			3b. p			
			3c. p			
			4. sel. yks.			
5. sel. yks						
Kierrätys	3	tiedon välittäminen.	1. kesk.	mv	vk	vk
			2. kesk.			
			3. kesk..			
Luonto	2	kuvaileva	1. ymm.	mv	äk	äk
			2. ymm.			
			3. ymm.			
			4. yks.			
			5. p	parafrasia vastaavan sanan tunnistaminen	vk	vk
			6. p			
			7. p			
			8. p			
			9.p			

Lyhenteiden selitykset

sel. yks.	= Kysytyn implisiittisesti ilmaistun tiedon paikallistaminen / selaaminen
k	= Käännös
p	= Kysytyn eksplisiittisesti ilmaistun tiedon paikallistaminen
kesk.	= Keskeisen / pääasian ymmärtäminen
ymm.	= Syy -seuraussuhteen ymmärtäminen
avo	= kysymys, johon odotetaan lyhyttä vastausta
mv	= monivalintatehtävä

LIITE 16 Vuoden 1999 englannin kielen arvioinnin Opettajan ohjeiden sisällysluettelo. (Lähde: Opetushallitus 1999. perusopetuksen 9. vuosiluokan englannin koe 1999. Opettajan ohjeet. Moniste.)

2

Ti 30.3.1999	Suullinen koe	15 + 15 min. /oppilas
Ke 31.3.1999	Koe A: Tekstin ymmärtäminen ja kirjoittaminen	45 min.
	Oppilaskyselyyn vastaaminen	15 min.
	Koe B: Kuullun ymmärtäminen ja rakenteet	45 min.

Sisällys

1	Suullinen koe
1.1	Suullisen kokeen toimeenpaneminen
1.2	Arviointiohjeet
2	Kirjallisen kokeen ohjeiden antaminen oppilaille
2.1	Tehtävätyypit ja koetehtäviin vastaaminen
2.2	Optisen lomakkeen täyttäminen
3	Tekstin ymmärtämisen koe
3.1	Korjaus- ja pisteytysohjeet
4	Kirjoittamisen koe
4.1	Arviointiohjeet
5	Kuullun ymmärtämisen koe
5.1	Korjaus- ja pisteytysohjeet
6	Rakennekoe
6.1	Oikeat vastaukset
Liitteet	1 Suullisen kokeen ajanvarauslista
	2 Oppilaan ohjeet suulliseen kokeeseen osallistumiseksi
	3 Opettajan muistilista ja vihjeet suullisen kokeen toimeenpanemiseksi
	4 Suullisen kokeen arviointikaavake
	5 Kuullun ymmärtämisen kokeiden käsikirjoitukset
	6 Valvojan ohjeet

LIITE 17 Vuoden 1999 englannin kielen arvioinnin puhumisen kokeen oppilaan ohjeet. (Lähde: Opetushallitus 1999. perusopetuksen 9. vuosiluokan englannin koe 1999. Opettajan ohjeet. Moniste.)

Englannin suullinen koe tiistaina 30.3.1999

Oppilaan ohjeet

Kokeessa on kolmenlaisia tehtäviä:

- lyhyitä vuoropuheluita
- lukemistehtäviä
- keskusteluita(ja).
- Ensimmäisen tehtävän arvostelussa kiinnitetään huomiota tilanteesta selviytymiseen ja luonteviin sanontoihin.
- Toisessa tarkkaillaan äänneitä ja painotusta.
- Kolmannessa huomiota kiinnitetään siihen, miten paljon ja monipuolisesti pystyt tuottamaan puhetta annetusta aiheesta.
- Opettaja keskeyttää tarvittaessa saatuaan riittävästi näyttöä kielitaidostasi!

Valmistautumiseen on ehdottoman tärkeää tulla sovitun aikataulun mukaisesti.

Kokeessa tulisi pyrkiä antamaan itsestään mahdollisimman hyvä ja monipuolinen kuva englannin puhujana.

Molempien keskustelijoiden tulisi tuottaa suunnilleen saman verran puhetta.

Muistiinpanoihin tulisi turvautua mahdollisimman vähän.

Kielivirheisiin ei tule kiinnittää tarpeettoman paljon huomiota.

Tehtävissä ei ole tarpeen kiirehtiä; nopeus ei ole kielitaidon mitta!

Onnea kokeeseen!

LIITE 18 Vuoden 1999 englannin arvioinnin puhumisen kokeen opettajan muistilista. (Lähde: Opetushallitus 1999. perusopetuksen 9. vuosiluokan englannin koe 1999. Opettajan ohjeet. Moniste.)

Englannin suullinen koe tiistaina 30.3.1999

Opettajan muistilista

Koejärjestelyt

- sovitaan koetta vastaanottavat opettajat
- varataan rauhallinen kokeeseen valmistutumistila
- varataan rauhallinen koetila
- laaditaan oppilasaikataulu, johon merkitään valmistautumis- ja koetila, ja sijoitetaan aikataulu näkyvään paikkaan, jotta oppilaat voivat tarkistaa koeaikansa
- muodostetaan oppilasparit
- tiedotetaan oppilaille kokeesta ja aikataulusta ja jaetaan otosoppilaille kirjalliset koeohjeet
- varataan nauhuri (mieluummin verkkovirralla toimiva), mikrofoni ja mikrofonin jalusta sekä OPH:n toimittama tyhjä kasetti ja varmistetaan, että laitteet toimivat (esim. että tarvittavissa paristoissa on riittävästi virtaa)
- varataan koetilan pöydälle kello, josta ”puheenjohtaja” voi tarkkailla ajan kulku oppilaan havaitsematta
- monistetaan arviointikaavakkeet kokeen vastaanottajille ja merkitään niihin oppilaiden nimet pareittain siinä järjestyksessä, jossa he tulevat kokeeseen samoin kuin koeversion numero
- tarkistetaan, että tehtävävihkot ovat Opetushallituksesta tullessa järjestyksessä ja merkitään oppilaiden nimet tehtävävihkojen kansilehdelle
- varataan valmistautumistilaan ensimmäistä oppilasparia varten tehtävät
- ennen keskustelun alkua oppilaita pyydetään sanomaan nimensä nauhalle
- parin lopetellessa koetta toinen testaa ja noutaa seuraavan parin kokeeseen ja jättää tätä seuraavalle parille tehtävävihot
- tehtävävihot kootaan pois kunkin parin suorituksen jälkeen
- koetilaisuuden jälkeen sovitaan oppilaiden arvostelusta ja merkitään lopulliset pisteet arviointikaavakkeelle ja optiselle lomakkeelle
- kirjoitetaan kasettikoteloon koulun nimi

Vihjeitä testaajille

- aloita koe lyhyellä lämmittelykeskustelulla saadaksesi oppilaan vapautumaan – kysele kuulumisia tms.
- luo katsekontakti oppilaaseen – sillä on rauhoittava vaikutus
- siedä taukoja ja hiljaisuutta
- anna oppilaan puhua mahdollisimman paljon
- älä keskeytä – ainakaan korjataksesi virheitä
- ole kiinnostunut
- muotoile mahdolliset kysymyksesi selkeästi ja esitä mieluummin sellaisia, jotka vaativat useamman kuin yhden virkkeen tuottamista
- älä auta sanastollisesti; tarvittaessa voit tulkita
- pidä kiinni aikataulusta
- päättää koe positiiviseen kommenttiin oppilaiden suorituksesta
- tarkista, että saat tehtäväpaperit takaisin

LIITE 19 Vuoden 1999 englannin kielen arvioinnin valvojan ohjeet. (Lähde: Opetushallitus 1999. perusopetuksen 9. vuosiluokan englannin koe 1999. Opettajan ohjeet. Moniste.)

Perusopetuksen 9. vuosiluokan englannin koe keskiviikkona 31.3.1999

Valvojan ohjeet

1. tunti

1. Jaa oppilaille optiset lomakkeet ja konseptiarkin puolikas.
2. Pyydä oppilaita kirjoittamaan sekä optiselle lomakkeelle että konseptiarkin puolikkaan oikeaan yläkulmaan koulun täydellinen nimi ja oma nimensä.
3. Sano, että optisen lomakkeen oppilaskyselyyn vastataan vasta kokeen jälkeen.
4. Jaa oppilaille kokeen A tehtävävihot, sano, että niitä ei saa vielä avata ja kehota oppilaita kirjoittamaan optiseen lomakkeeseen ja konseptiarkille oman tehtävävihkonsa kannessa oleva **tehtäväsarjan numero**.
5. Selitä, että optista lomaketta käytetään vastattaessa monivalinta- ja oikein-/vääriin-kysymyksiin. Näissä on vain yksi oikea vaihtoehto. Lomakkeen kohdasta **Oppilas mustaa** mustataan osion numeroa vastaava kirjain, jota oppilas pitää oikeana vaihtoehtona. Oikein-/vääriin-väittämässä A on oikea vaihtoehto, B väärä. Rastit tai ympyröinnit tulkitaan vääriksi vastauksiksi.
6. Konseptiarkille kirjoitetaan tuottamistehtävien vastaukset selkeällä käsialalla. Toiselle puolelle arkkia kirjoitetaan kirjoitelma joka toiselle riville. Tuottamistehtäviin vastataan samalla kielellä, jolla kysymys on tehty. Kirjoitelma kirjoitetaan englanniksi. Väärällä kielellä annetuista vastauksista tulos on 0 p.
7. Anna oppilaille lupa avata koevihat.
8. Anna tarvittaessa tarkentavia vastausohjeita.
9. Muistuta oppilaita hyvissä ajoin ennen tunnin päättymistä kirjoitelman kirjoittamisesta.
10. Kokoa tunnin päättyessä koevihat ja konseptiarkit.

Välitunti:

11. Kehota oppilaita täyttämään optisella lomakkeella oleva oppilaskysely. Kiellä tarvittaessa heitä keskustelemasta keskenään.
11. Aseta tällä välin kuuntelukasetti nauhuriin.

2. tunti

13. Jaa oppilaille konseptiarkin puolikas ja kehota heitä kirjoittamaan sen oikeaan yläkulmaan jälleen koulun täydellinen nimi ja oma nimensä.
14. Jaa oppilaille kokeen B tehtävävihot. Sano, että vihkoa ei saa vielä avata, ja kehota oppilaita kirjoittamaan optiselle lomakkeelle ja konseptiarkille vihon kannessa oleva tehtäväsarjan numero.
15. Kerro, että myös kuuntelukokeessa voi olla monivalinta-, oikein-/vääriin- ja tuottamistehtäviä. Näihin vastataan samalla tavalla kuin kokeessa A joko optiselle lomakkeelle tai konseptiarkille.
16. Kehota oppilaita kuuntelemaan huolellisesti nauhalta kuuluvia ohjeita ja sano, että ohjelmat ovat valmiiksi tauotettuja, joten kun kuuntelu alkaa, he eivät voi enää esittää kysymyksiä.
17. Kehota oppilaita avaamaan tehtävävihat.
18. Käynnistä nauhuri.
19. Kuuntelun päätyttyä kehota oppilaita jatkamaan rakennekokeen kysymyksiin vastaamista.
20. Kokeesta ei saa poistua ennen oppitunnin päättymistä.
21. Varmista, että tarvittavat vastaukset on siirretty optiselle lomakkeelle. Kokoa koevihat, konseptiarkit ja optiset lomakkeet.

LIITE 20 Reseptiivisen kielitaidon kokeen osioanalyysit tehtäväsarjoittain

Taulukon selitys:

Sarake 1. Osion numero

Sarake 2. Ratkaisuprosentti (p-arvo)

Sarake 3. Painottamaton osiotestikorrelaatio

Sarake 4. Painotettu osiotestikorrelaatio

1	.788	.264	.238	1	.752	.298	.278	1	.762	.348	.325
2	.609	.511	.515	2	.635	.500	.502	2	.601	.505	.507
3	.961	.232	.228	3	.953	.320	.314	3	.952	.259	.248
4	.874	.257	.243	4	.866	.324	.312	4	.862	.307	.296
5	.856	.314	.308	5	.845	.387	.380	5	.853	.374	.369
6	.588	.463	.448	15	.821	.435	.436	25	.797	.509	.507
7	.529	.415	.397	16	.635	.501	.501	26	.764	.464	.460
8	.531	.527	.519	17	.506	.491	.488	27	.632	.327	.299
9	.490	.244	.212	18	.785	.494	.496	28	.704	.431	.418
10	.920	.259	.253	19	.908	.341	.333	29	.673	.633	.641
11	.526	.352	.327	20	.809	.346	.332	30	.747	.540	.547
12	.490	.303	.274	21	.582	.670	.684	31	.658	.712	.730
13	.567	.517	.517	22	.536	.514	.519	32	.462	.582	.587
14	.519	.494	.493	23	.334	.544	.542	33	.563	.435	.421
36	.445	.435	.425	24	.750	.560	.568	34	.503	.468	.454
37	.762	.451	.453	36	.454	.427	.414	35	.436	.411	.397
39	.800	.350	.335	37	.739	.471	.470	36	.431	.444	.437
40	.943	.390	.414	39	.756	.467	.460	37	.734	.477	.480
41	.690	.455	.454	40	.940	.424	.434	39	.775	.375	.374
42	.308	.337	.320	41	.686	.418	.411	40	.931	.451	.461
43	.427	.554	.555	42	.294	.314	.308	41	.706	.445	.442
44	.782	.549	.565	43	.422	.590	.598	42	.283	.335	.330
45	.650	.318	.279	44	.779	.538	.552	43	.405	.562	.570
46	.436	.241	.200	51	.780	.545	.549	44	.772	.529	.545
48	.233	.368	.356	52	.752	.337	.325	56	.517	.260	.235
49	.546	.330	.302	53	.720	.360	.346	57	.652	.438	.426
50	.475	.535	.538	54	.187	.354	.344	58	.502	.532	.534
67	.776	.469	.477	55	.592	.511	.504	59	.497	.530	.535
68	.755	.381	.371	67	.745	.555	.563	60	.199	.516	.529
69	.540	.478	.483	68	.749	.414	.408	61	.243	.511	.523
70	.646	.387	.381	69	.515	.482	.482	62	.302	.494	.498
72	.805	.487	.501	70	.649	.428	.416	63	.203	.482	.490
73	.786	.533	.557	72	.770	.483	.484	64	.448	.525	.528

92	.847	.470	.481	91	.572	.510	.512	84	.661	.333	.311
93	.617	.361	.352	101	.818	.420	.409	85	.798	.592	.603
94	.730	.519	.530	102	.473	.366	.356	86	.783	.582	.601
95	.785	.567	.590	103	.785	.583	.596	92	.773	.522	.531
96	.697	.640	.669	104	.602	.361	.357	93	.631	.410	.395
97	.460	.630	.641	105	.611	.636	.646	94	.722	.528	.532
98	.397	.543	.543	106	.588	.532	.536	95	.796	.542	.558
99	.392	.579	.587	107	.594	.493	.493	101	.820	.410	.404
100	.338	.530	.524	108	.179	.342	.330	102	.507	.420	.408
				109	.115	.312	.295	103	.773	.561	.571
								104	.606	.339	.334
								105	.624	.595	.605
								110	.878	.453	.460
								111	.561	.377	.365
								112	.789	.467	.468
								113	.338	.411	.399
								114	.653	.364	.355

LIITE 21 Koko reseptiivisen kielitaidon kokeen oppilasvastausten osiokoh-
taiset jakaumat koko reseptiivisessä kokeessa.

Osion numero	Osion kuvaus	a	N	P	0	1
1	R01_abc	1	4854	.768	1128	3726
2	R02_abc	3	4854	.615	1868	2986
3	R03_abc	2	4854	.956	216	4638
4	R04_abc	2	4854	.868	643	4211
5	R05_abc	2	4854	.851	722	4132
6	R06_a00	2	1699	.588	700	999
7	R07_a00	2	1699	.529	801	898
8	R08_a00	3	1699	.531	797	902
9	R09_a00	1	1699	.490	867	832
10	R10_a00	2	1699	.920	136	1563
11	R11_a00	2	1699	.526	806	893
12	R12_a00	1	1699	.490	867	832
13	R13_a00	3	1699	.567	735	964
14	R14_a00	3	1699	.519	818	881
15	R15_ob0	2	1585	.821	284	1301
16	R16_ob0	3	1585	.635	578	1007
17	R17_ob0	3	1585	.506	783	802
18	R18_ob0	3	1585	.785	341	1244
19	R19_ob0	2	1585	.908	146	1439
20	R20_ob0	2	1585	.809	303	1282
21	R21_ob0	4	1585	.582	662	923
22	R22_ob0	3	1585	.536	735	850
23	R23_ob0	3	1585	.334	1055	530
24	R24_ob0	3	1585	.750	397	1188
25	R25_0c0	3	1570	.797	318	1252
26	R26_0c0	2	1570	.764	371	1199
27	R27_0c0	1	1570	.632	577	993
28	R28_0c0	2	1570	.704	464	1106
29	R29_0c0	4	1570	.673	514	1056
30	R30_0c0	3	1570	.747	397	1173
31	R31_0c0	5	1570	.658	537	1033
32	R32_0c0	3	1570	.462	845	725
33	R33_0c0	2	1570	.563	686	884
34	R34_0c0	2	1570	.503	781	789
35	R35_0c0	2	1570	.436	885	685
36	L01_abc	2	4854	.443	2702	2152
37	L02_abc	3	4854	.745	1236	3618
39	L04_abc	2	4854	.778	1080	3774
40	L05_abc	5	4854	.938	301	4553
41	L06_abc	2	4854	.694	1485	3369

42	L07_abc	2	4854	.295	3420	1434
43	L08_abc	4	4854	.418	2824	2030
44	L09_abc	4	4854	.778	1079	3775
45	L10_a00	1	1699	.650	594	1105
46	L11_a00	1	1699	.436	958	741
48	L13_a00	2	1699	.233	1303	396
49	L14_a00	1	1699	.546	772	927
50	L15_a00	3	1699	.475	892	807
51	L16_0b0	3	1585	.780	349	1236
52	L17_0b0	2	1585	.752	393	1192
53	L18_0b0	2	1585	.720	444	1141
54	L19_0b0	2	1585	.187	1288	297
55	L20_0b0	2	1585	.592	646	939
56	L21_0c0	1	1570	.517	758	812
57	L22_0c0	2	1570	.652	546	1024
58	L23_0c0	3	1570	.502	782	788
59	L24_0c0	3	1570	.497	790	780
60	L25_0c0	5	1570	.199	1257	313
61	L26_0c0	4	1570	.243	1188	382
62	L27_0c0	3	1570	.302	1096	474
63	L28_0c0	4	1570	.203	1252	318
64	L29_0c0	3	1570	.448	866	704
65	L30_0c0	2	1570	.115	1389	181
66	L31_0c0	3	1570	.601	627	943
67	G01_abc	3	4854	.748	1222	3632
68	G02_abc	2	4854	.746	1233	3621
69	G03_abc	3	4854	.529	2287	2567
70	G04_abc	2	4854	.637	1761	3093
72	G06_abc	3	4854	.785	1046	3808
73	G07_abc	4	4854	.775	1094	3760
74	G08_abc	3	4854	.502	2419	2435
75	G09_abc	2	4854	.704	1437	3417
76	G10_abc	4	4854	.896	505	4349
77	G11_abc	3	4854	.838	786	4068
78	G12_abc	4	4854	.776	1089	3765
79	G13_abc	2	4854	.778	1077	3777
80	G14_abc	1	4854	.566	2108	2746
81	G15_abc	2	4854	.636	1768	3086
82	G16_abc	3	4854	.611	1890	2964
83	G17_abc	2	4854	.839	782	4072
84	G18_abc	1	4854	.684	1534	3320
85	G19_abc	4	4854	.813	907	3947
86	G20_abc	5	4854	.790	1020	3834
87	G21_ab0	2	3284	.589	1349	1935
88	G22_ab0	5	3284	.754	807	2477

89	G23_ab0	1	3284	.575	1396	1888
90	G24_ab0	5	3284	.822	583	2701
91	G25_ab0	3	3284	.562	1439	1845
92	G26_a0c	3	3269	.812	616	2653
93	G27_a0c	2	3269	.624	1229	2040
94	G28_a0c	3	3269	.727	894	2375
95	G29_a0c	4	3269	.790	687	2582
96	G30_a00	5	1699	.697	514	1185
97	G31_a00	4	1699	.460	917	782
98	G32_a00	3	1699	.397	1025	674
99	G33_a00	4	1699	.392	1033	666
100	G34_a00	3	1699	.338	1125	574
101	G35_0bc	2	3155	.819	571	2584
102	G36_0bc	2	3155	.490	1610	1545
103	G37_0bc	4	3155	.779	696	2459
104	G38_0bc	2	3155	.604	1250	1905
105	G39_0bc	4	3155	.617	1208	1947
106	G40_0b0	3	1585	.588	653	932
107	G41_0b0	3	1585	.594	643	942
108	G42_0b0	2	1585	.179	1301	284
109	G43_0b0	2	1585	.115	1403	182
110	G44_0c0	3	1570	.878	191	1379
111	G45_0c0	2	1570	.561	690	880
112	G46_0c0	3	1570	.789	331	1239
113	G47_0c0	2	1570	.338	1040	530
114	G48_0c0	2	1570	.653	545	1025

LIITE 22 Produktiivisen kielitaidon kokeiden suppeamman aineiston CEF-, OPS 2004- ja A2001 -taitotasoille sijoittuneiden oppilaiden frekvenssijakaumat ja prosentuaaliset osuudet kriteereittäin.

Puhumisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet CEF-taitotasoasteikolla

Taitotasoasteikko	Yleistaso		Laajuus		Virheettömyys		Vuorovaikutus		Fonologia	
	%	Fr	%	Fr	%	Fr	%	Fr	%	Fr
0	4	2	4	2	4	2	4	2	-	-
A1	2	1	2	1	2	1	-	-	6	3
A2	17	9	7	4	9	5	9	5	15	8
A2+*	19	10	-	-	-	-	11	6	-	-
B1	30	16	85	45	83	45	32	17	59	32
B1+*	26	14	-	-	-	-	26	14	-	-
B2	4	2	2	1	2	1	19	10	19	10
C1	-	-	-	-	-	-	-	-	2	1
Yht.	100	54	100	54	100	54	100	54	100	54

* Jokaista tasoa ei ollut jaettu kahtia.

Puhumisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet OPS 2004-taitotasoasteikolla

Taitotasoasteikko	%	Fr
0	4	2
A1.1	2	1
A1.2	-	-
A1.3	4	2
A2.1	7	4
A2.2	20	11
B1.1	30	16
B1.2	28	15
B2.1	6	3
Yht.	100	54

Puhumisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet A2001-taitotasoasteikolla

Taitotasoasteikko	Sujuvuus		Laajuus		Tarkkuus		Vuorovaikutus	
	%	Fr	%	Fr	%	Fr	%	Fr
1	8	4	6	3	4	2	4	2
2	8	4	6	3	9	5	4	2
3	22	12	19	10	37	20	13	7
4	45	24	50	27	37	20	44	24
5	18	10	20	11	13	7	35	19
Yht.	100	54	100	54	100	54	100	54

Kirjoittamisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet CEF-taitotasoasteikolla

Taitotaso-asteikko	Yleistaso		Laajuus		Sidosteisuus		Virheettömyys		Kuvaileva		Argumentatiivinen	
	%	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%	Fr
A1	14	7	16	8	16	8	24	12	14	6	13	1
A2	34	17	26	13	28	14	26	13	36	15	-	-
B1	42	21	42	21	32	16	34	17	43	18	38	3
B2	4	2	10	5	18	9	10	5	7	3	25	2
C1	2	1	2	1	4	2	2	1	-	-		
C2	4	2	4	2	2	1	4	2	-	-	25	2
Yht.	100	50	100	50	100	50	100	50	100	42	100	8

Kirjoittamisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet OPS 2004-taitotasoasteikolla

Taitotaso-asteikko	%	Fr
0	-	-
A1.1	-	-
A1.2	14	7
A1.3	12	6
A2.1	18	9
A2.2	12	6
B1.1	24	12
B1.2	12	6
B2.1	4	2
B2.2	-	-
C1.1	4	2
Yht.	100	50

Kirjoittamisen kokeen kriteerien frekvenssit ja prosentuaaliset osuudet A2001-taitotasoasteikolla

Taitotaso-asteikko	Sujuvuus		Laajuus		Sidosteisuus		Virheettömyys	
	%	Fr	%	Fr	%	Fr	%	Fr
0	4	2	2	1	2	1	2	1
1	8	4	10	5	12	6	12	6
2	16	8	14	7	12	6	34	173
3	32	16	30	15	26	13	24	12
4	22	11	28	14	28	14	18	9
5	18	9	16	8	20	10	10	5
Yht.	100	50	100	50	100	50	100	50

LIITE 23 Produktiivisen kielitaidon kokeiden suuremman aineiston CEF-, OPS 2004- ja A2001 -taitotasoille sijoittuneiden oppilaiden ekstrapoloidut frekvenssijakaumat ja prosentuaaliset osuudet kriteereittäin.

Puhumisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet CEF -taitotasoasteikolla

Taitotasoasteikko	Sujuvuus		Laajuus		Virheettömyys		Vuorovaikutus		Keskiarvo	
	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%
<=A1	10	5	12	5	15	7	11	5	11	5
A2	89	40	78	35	108	48	73	32	76	34
B1	82	36	85	38	71	31	65	29	87	39
>=B2	45	20	51	23	32	14	77	34	52	23
Yht.	226	100	226	100	226	100	226	100	226	100

Puhumisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet OPS 2004 -taitotasoasteikolla

Taitotasoasteikko	Fr	%
A1.2	11	5
A1.3	19	8
A2.1 - A2.2	57	25
B1.1 - B1.2	87	39
B2.1 C1.1	52	23
Yht.	226	100

Puhumisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet A2001 -taitotasoasteikolla

Taitotasoasteikko	Sujuvuus		Laajuus		Virheettömyys		Vuorovaikutus	
	Fr	%	Fr	%	Fr	%	Fr	%
0	-	-	-	-	-	-	1	-
1	10	4	12	5	15	7	10	4
2	21	9	20	9	26	12	24	11
3	68	30	58	26	82	36	49	22
4	82	36	85	38	71	31	65	29
5	45	20	51	23	32	14	77	34
Yht.	226	100	226	100	226	100	226	100

Kirjoittamisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet CEF -taitotasoasteikolla

Taitotasoasteikko	Sujuvuus		Laajuus		Koherenssi		Virheettömyys		Keskiarvo	
	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%
<=A1	25	5	27	5	21	4	52	10	22	4
A2	226	43	208	39	190	35	280	54	212	40
B1	155	29	185	35	169	32	113	21	171	32
>=B2	126	24	117	22	157	29	81	15	132	25
Kato	5	1	-	-	-	-	2	-	-	-
Yht.	537	100	537	100	537	100	537	100	537	100

Kirjoittamisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet OPS 2004 -taitotasoasteikolla

Taitotasoasteikko	Fr	%
A1.1	5	1
A1.2	17	3
A1.3	62	12
A2.1	150	28
A2.2 - B1.2	171	32
B2.1 C1.1	132	25
Yht.	537	100

Kirjoittamisen kokeen kriteerien ekstrapoloidut frekvenssit ja prosentuaaliset osuudet A2001 -taitotasoasteikolla

Taitotasoasteikko	Sujuvuus		Laajuus		Koherenssi		Virheettömyys		Pyöristetty keskiarvo	
	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%
0	4	1	6	1	6	1	5	1	5	1
1	21	4	21	4	15	3	47	9	17	4
2	72	14	58	11	58	11	130	24	62	12
3	154	29	150	28	132	25	159	30	150	28
4	155	29	185	35	169	32	113	21	171	32
5	126	24	117	22	157	29	81	15	132	25
Kato	5	1	-	-	-	-	2	-	-	-
Yht.	537	100	537	100	537	100	537	100	537	100