

Viljo Kohonen:

### Cloze-testien ongelmia testianalyysin valossa

1.0. Tässä artikkelissa tarkastellaan kahden suomalaisen koehenkilöjoukon (79 ja 48 opiskelijaa) suorittamien kolmen cloze-testin analyysituloksia. Testeistä oli yksi suomenkielinen ja kaksi englanninkielisiä. Ne arvosteltiin kahdella eri korjausmenetelmällä, sanatarkalla ja kontekstuaalisella. Jälkimmäisessä arvostelun suorittivat toisistaan riippumatta kaksi suomalaista ja syntyperäistä englantilaista puhujaa. Arvostelijoiden välistä yksimielisyyttä ja korjausmenetelmien antaman paremmuusjärjestyksen vertailukelpoisuutta selviteltiin korreloimalla cloze-pisteitä keskenään. Empiirisen validiteetin tutkimiseksi niitä korreloitiin myös samojen koehenkilöiden suorittamiin muihin, monivalintateknikalla mitattuihin englannin kielitaitotesteihin. Tulosten perusteella tarkastellaan eräitä cloze-testien käyttöön liittyviä ongelmia.

#### 1. Cloze kielitaidon mittarina: aikaisempia tutkimuksia

Cloze-testejä on toistaiseksi käytetty vielä melko vähän vieraiden kielten kielitaidon testauksessa. Suoritetuissa tutkimuksissa niiden reliabiliteetit ovat olleet suuruusluokkaa .80 ja sen ylikin. Ne ovat myös osoittautuneet kohtalaisen vaikeiksi ja erottelukykyisiksi. Cloze-pistemäärien on todettu korreloivan parhaiten kriteeritesteinä käytettyjen TOEFL- ja UCLA ESLPE -testistöjen integratiivista kielitaitoa mittaaviin kuullun ymmärtämis-, sanelu- ja luetun ymmärtämistesteihin (.70-.80 välillä olevia korrelaatiokertoimia). Niillä on myös ollut noin 70-80% yhteistä varianssia näiden testistöjen loppupistemäärien kanssa.

Arvostelumenetelminä käytetyistä sanatarkasta ja kontekstuaalisesta korjaamisesta on jälkimmäisen todettu toimivan paremmin vieraan kielen testauksessa: se on erottelukyvyltään ja reliabiliteetiltään parempi ja korreloi myös paremmin mainittuihin kriteeri-

testistöihin. Kontekstuaalisen korjaamisen haittapuolena sanatarkkaan verrattuna on se, että hyväksyttävistä vastauksista päättäminen on käytännössä melko suuritöistä. Eräissä viimeaikaisissa tutkimuksissa menetelmien välille on saatu huomattavan korkeitakin korrelaatioita (.94 ja .97). Tämä viittaisi siihen, että vieraankin kielen arvostelussa voitaisiin käyttää sanatarkkaa korjaustapaa testattavien paremmuusjärjestyksen säilyessä samana. (Oller ja Conrad 1971; Oller 1972; Oller ym. 1974; Stubbs ja Tucker 1974.) Cloze-testien korjaamisesta on myös esitetty informaatioteoriaan perustuva tietokonesovellutus ("clozentropia", Darnell 1970), jota on esitelty toisaalla tässä volyymissa (Enkvist ja Kohonen).

Umpimähkäiseen aukotukseen perustuvan "standarditeknikan" lisäksi cloze-testeistä on kokeiltu myös "modifioitua" versiota, jossa sanoja jätetään pois esimerkiksi jonkin lingvistisen periaatteen mukaan. Eräässä tällaisessa tutkimuksessa (Oller ja Inal 1971) aukotusperiaatteenä oli prepositioiden poistaminen yhtäjaksoisessa tekstissä. Tulosten todettiin korreloivan parhaiten kriteerinä käytetyn UCLA ESLPE-testistön kielioppitestiin, ja korrelaatio testistön kokonaispistemääräänkin oli yli .70. Testi oli myös vaikeustasoltaan sopiva, sillä ratkaisuprosentit olivat keskimäärin 65 prosentin suuruusluokkaa.

Cloze-testien kuullun ymmärtämistä mittaavassa vastineessa redundanssia vähennetään äänittämällä nauhalle yhtäjaksoinen äänisignaali ("white noise"), joka tekee kuulemisen paikatellen vaikeaksi tai mahdottomaksi (Spolsky 1971 ja 1973). Tällainen melusignaali saattaa kuitenkin ärsyttää joitakin testattavia ja siten tuoda tuloksiin kielitaidosta riippumatonta virhevarianssia. Eräässä tätä koskeneessa tutkimuksessa (Johansson 1973) todettiin myös, että osa melutestissä heikosti suoriutuneista opiskelijoista menestyi hyvin tavanomaisissa kuullunymmärtämistesteissä. Testin keinotekoisuutta voidaan ilmeisesti vähentää äänittämällä nauhalle esimerkiksi liikenteen melua tai puheensorinaa, jolloin testattavat kokevat testaustilanteen luonnollisempina. Melutestiä on kokeiltu myös monivalintaversiona, jossa nauhalta kuultaville lauseille esitetään tehtävämönisteissa vaihtoehtoja, ja testattava valitsee näistä sen, jonka hän mielestään on kuullut nau-

halta. Tällaisen testin todettiin korreloivan hyvin tuottamista mittavaan haastattelutestiin (.79) sekä toisena kriteerinä käytetyn TOEFL-testistön kuullun ymmärtämistestiin (.75) ja kielipillisten rakenteiden testiin (.70) (Gradman 1973). Nämä korrelaatiot perustuvat kuitenkin vain 25 koehenkilön suoritukseen, ja käytetyn melutestin reliabiliteetti oli myös suhteellisen alhainen (.68). Cloze-tekniikalla näyttää siis olevan auraalisessakin versiossa useita variaatiomahdollisuuksia.

## 2. Cloze-kokeilu suomalaisella koehenkilöjoukolla

2.1. Koejärjestelyt ja testit. Syyskuussa 1975 suoritettuun kokeiluun osallistui 79 Jyväskylän yliopiston sekä 48 Turun yliopiston ja Joensuun korkeakoulun ensimmäisen vuoden englannin opiskelijaa. Nämä suorittivat kolme testiä, joista yksi oli suomenkielinen ja kaksi englanninkielisiä (Jyväskylän ryhmässä käytetyt testit on esitetty liitteessä 1). Suomenkielinen tekstiote ja toinen englanninkielinen teksti pyrittiin saamaan vertailukelpoisiksi ottamalla ne sanoma- ja aikakauslehdestä (Helsingin Sanomat, Time). Ne käsittelivät molemmat päivän politiikkaa, tarkemmin sanottuna sen hetkisiä Lähi-Idän rauhanneuvotteluja ja presidentti Fordin toiminnan arviointia; kummassakin oli 42 osiota. Toinen englanninkielinen tekstiote oli asiallisesti ja kielellisesti vaikeampaa popularisoitua tiedettä (Penguin Science Library -sarjasta) ja käsitteli propagandan teoriaa, osiomäärä 39.

Ford-testi oli täsmälleen sama molemmissa ryhmissä, ja siinä oli aukotuseriaatteena joka 6. sanan poisto. Toisen englanninkielisen testin aukotuseriaate oli samoin joka 6. sana, mutta sitä varioitiin siten, että Turun ja Joensuun ryhmän testissä alkupeätäisen tekstin normaali, sidosteinen lauseiden järjestys muutettiin umpimähkäiseksi. Poisjätetyt sanat olivat kuitenkin molemmissa versioissa samat. Suomenkielisessä tekstissä kokeiltiin kahden aukotuseriaatetta. Jyväskylän ryhmän testissä jätettiin pois joka 5. sana, mutta kontekstia lisättiin hiukan Turun ja Joensuun ryhmissä muuttamalla aukotuseriaatetta joka 6. sanan poistoksi. Koehenkilöryhmiä voidaan pitää vertailukelpoisina, sillä ne edustavat sitä 20 prosentin parhaimmistoa, joka voitiin hyväksyä näihin korkeakouluihin kesällä 1975 suoritettussa kaikille pakollisessa valintakokeessa.

## 2.2. Sanatarkan ja kontekstuaalisen korjauksen vertailu.<sup>1</sup>

Testianalyysi suoritettiin tietokoneella käyttäen OPSAM-ohjelmaa. Testiversioiden analyysin tärkeimmät tulokset olivat sanatarkalla arvostelumenetelmällä korjattuna seuraavat.

Taulu 1. Sanatarkalla menetelmällä arvostellut analyysitulokset koehenkilöryhmissä (testit: 1 "Lähi-Itä", 2 "Ford" ja 3 "Propaganda").

Turku & Joensuu (N=48)					
testi	osio- määrä	keski- arvo	haj.	ratk.- pros.	rel.
1.	35	15.6	2.6	44.6%	.35
2.	42	21.6	3.9	51.4%	.63
3.	39	12.7	4.7	31.5%	.77
koko testi	116	49.9	7.8		.75
Jyväskylän ryhmä (N=79)					
1.	42	20.2	2.8	48.2%	.31
2.	42	20.5	3.9	48.8%	.64
3.	39	10.2	4.3	26.1%	.73
koko testi	123	50.9	8.6		.79

Kuten molemmille ryhmille yhteinen testi 2 ("Ford") osoittaa, ryhmien suoritukset ovat hyvin samanlaiset. Huomiota kiinnittää se, että äidinkielen testi 1 osoittautui jopa hieman vaikeammaksi sanatarkalla menetelmällä korjattuna kuin aihepiiriltään vastaava testi 2. Kontekstin lisääminen Turun ja Joensuun ryhmässä (joka 6. sanan poistoksi) ei myöskään tehnyt testiä sen helpommaksi. Aihepiiriltään ja sanastoltaan hankalampi testi 3 osoittautui huomattavasti vaikeammaksi kuin testit 1 ja 2. Tämä johtui osittain myös siitä, että kaikki testattavat eivät olleet ehtineet suorittaa sitä loppuun; tyhjätkä aukot arvosteltiin nolliksi. On mielenkiintoista todeta, että lauseiden alkuperäisen järjestyksen muuttaminen umpimähkäiseksi (Turun ja Joensuun ryhmässä) ei tehnyt testiä 3 vaikeammaksi. Ilmeisesti vain 9 virkkeestä koostuva teksti oli vielä sittenkin riittävän sidosteinen, ja sen suhteellisen pitkien lauseiden sisäinen konteksti oli riittävä aukkojen täyttämiseksi.

<sup>1</sup> Tässä artikkelissa esitetään vain tilastollisen analyysin päätuloksia. Testien vastausjakautumia tarkastellaan esimerkkien valossa Tommolan artikkelissa, jossa pohditaan myös mahdollisia syitä siihen, miksi äidinkielen testi 1 osoittautui hieman yllättäen yhtä vaikeaksi kuin testi 2.

Ratkaisuprosentteja tarkastellen kaikki testit ovat hyvin erottelukykyisiä, eli ne ovat riittävän vaikeita. Englanninkielisten testien reliabiliteetteja voidaan myös pitää tyydyttävinä. Ne osoittavat kohtalaisen hyvää erottelukyvyn sisäistä johdonmukaisuutta (laskukaavat on esitetty ohjelman käsikirjassa, Mikkonen ja Mikkonen 1971). Äidinkielen testin 1 reliabiliteetti on sen sijaan huomattavan alhainen. Tämä viittaa siihen, että testi ei erotellut johdonmukaisesti, eli oikeisiin sanatarkkoihin vastauksiin olivat päätyneet pikemminkin sellaiset opiskelijat, jotka olivat menestyneet koko testissä 1 keskimääräistä heikommin. Erottelukyvyn johdonmukaisuushan edellyttää, että koko testistä parhaiten suoriutuneet yksilöt tuottavat myös keskimääräistä enemmän oikeita osiovastauksia; alhainen tai negatiivinen osiokorrelaatio osoittaa, että näin ei ollut asianlaita. Tämän johdosta sanatarkka korjausmenetelmä ei näytä perustellulta ainakaan äidinkielisillä suomalaisilla puhujilla, vaikka sitä onkin yleisesti käytetty äidinkielisten englantilaisten koehenkilöiden suoritusten korjaamisessa. Tämä saattaa johtua kielten välisistä rakenteellisista eroista (vrt. Tommola).

Kontekstuaalinen korjaus suoritettiin vain Jyväskylän ryhmässä (N=79). Tärkeimmät analyysitulokset olivat:

testi	osio- määrä	keski- arvo	haj.	ratk.- pros.	rel.
1.	42	33.2	3.1	79.1%	.53
2.	42	28.1	4.8	66.9%	.72
3.	39	24.0	6.7	61.5%	(ei laskettu)

Tauluun 1 verrattuna havaitaan, että testit ovat kontekstuaalisesti korjattuina tilastollisesti parempia: osioiden ratkaisuprosentit ja osiokorrelaatiot ovat kohonneet (ks. Liite 2.), ja testeistä 1 ja 2 lasketut reliabiliteetit viittaavat niiden sisäisen johdonmukaisuuden parantumiseen. Tämä parannus on hyvin selvä suomenkielisen testin 1 kohdalla, joskin sen erottelukyky jää vielä silttenkin heikoksi. Joka tapauksessa tämä reliabiliteetin kohoaminen merkitsee sitä, että testissä 1 paremmin menestyneiden koehenkilöiden vastaukset olivat usein kontekstuaalisesti hyväksyttäviä, vaikka sanatarkka korjausmenetelmä olikin merkinnyt ne "vääriksi".

Testeistä 1 ja 2 molemmilla korjausmenetelmillä suoritettu osioanalyysi antaa tarkemman kuvan erottelukyvyn parantumisesta kontekstuaalisessa korjauksessa. Testissä 1 oli sanatarkalla menetelmällä korjattuna vain 5 osiota, joiden osiokorrelaatio oli merkitsevä 5 prosentin riskitasolla (eli korkeampi kuin .22). Kontekstuaalisessakin korjauksessa tällaisten hyvin toimineiden osioiden lukumäärä nousi vain 8:aan, eli testin todellinen erottelukyky jäi näiden kahdeksan osion varaan. Testistä 1 tuli nyt suurelta osin liiankin helppo, sillä siinä oli 16 osiossa ratkaisuprosentti yli 95% ja 20 osiossa yli 90%. Sanatarkalla menetelmällä korjattuna tällaisten hyvin helppojen osioiden määrä oli vain 7 osiota. Testissä 2 ("Ford") erottelukyky oli osiokorrelaatioilla mitattuna molemmissa tapauksissa huomattavasti parempi. Sanatarkasti korjattuna 13 osiota korreloi 5 prosentin riskitasolla merkitsevästi testin kokonaispistemäärään, ja kontekstuaalisessa korjauksessa tällaisten osioiden lukumäärä oli jo 19, eli 45%. Ratkaisuprosentit osoittavat myös, että aukon täydentäminen on joissakin tapauksissa hyvin vaikeaa ja sanatarkasti mahdotontakin. Tällaisia erittäin vaikeita osioita (ratkaisuprosentit alle 5%) oli sanatarkalla menetelmällä korjatessa testissä 1 kahdeksan ja testissä 2 kuusi. Näiden kohdalla kontekstuaalinen korjaus parantaa osioiden antamaa informaatiota testattavien välisistä eroista: testissä 1 oli vain yhden osion ratkaisuprosentti 5%, ja testissä 2 alin ratkaisuprosentti oli jo 13.9% (vrt. Liite 2). Näiltäkin osin testi siis paranee kontekstuaalisessa korjauksessa.

2.3. Cloze-testien korrelaatiot englannin valintakokeeseen. Jyväskylän ryhmässä suoritettiin myös tutkimus cloze-testien korrelaatioista samojen koehenkilöiden kesällä 1975 suorittamaan englannin kielen valintakokeeseen. Nämä valintakoemenestystä koskevat tiedot olivat kuitenkin käytettävissä vain 63 koehenkilöstä. Tutkimuksen tämä osa suoritettiin HYLPS-ohjelmalla, joka on käytettävissä Univac 1108-tietokonejärjestelmässä.

Englannin kielen valintakoe oli monivalintatekniikkaan perustuva testistö, joka kesällä 1975 muodostui seuraavista osatesteistä: 1) äänneiden tunnistaminen (25 osiota), 2) lyhyiden puhejaksojen pääsisällön ymmärtäminen (20 osiota), 3) puhutun

ymmärtäminen (30 osiota), 4) englanninkielinen luento (15 minuutin pituinen, 20 osiota), 5) sanasto ja idiomit I (testattavat kohteet irrallisissa lauseissa, 30 osiota), 6) sanasto ja idiomit II (27 osiota yhtäjaksoisessa tekstissä), 7) tekstin ymmärtäminen (23 osiota), 8) kielelliset rakenteet (50 osiota), 9) kielellinen päättely (kyky havaita kielellisiä säännönmukaisuuksia oudossa kielimateriaalissa, 20 osiota), sekä 10) valintapistemäärä, jossa nämä testit olivat mukana 70 prosentin vaikutusosuudella sekä 5 kouluarvosanaa 30 prosentin osuudella. Testiä 9) lukuunottamatta nämä testit olivat periaatteessa samanlaisia kuin Kohosen (1974) toimittamassa julkaisussa esiteltyt. Niiden tärkeimmät tilastolliset ominaisuudet on esitetty liitteessä 3, jossa on myös laskettu tähän analyysiin osallistuneiden 63 koehenkilön keskiarvot ja hajonnat kaikissa testeissä.

Tämän tarkastelun kannalta on ongelmallista se, että koehenkilöt edustavat valintakokeen suorittaneista vain 20 prosentin parhaimmistoa, joka kiintiöiden puitteissa voitiin ottaa sisään. Tämä merkitsee sitä, että valintakokeen testien varianssi on heidän kohdallaan luonnollisesti pienempi kuin koko kokeen suorittaneessa perusjoukossa. Kuten liitteestä 3 havaitaan, valintakokeen testit olivat kuitenkin siksi vaikeita, että tämän parhaimmistonkin tuloksissa on vielä huomattavasti hajontaa. Tämän vuoksi korrelaatioiden laskeminen on vielä perusteltua. Tuloksia tarkastellessa on kuitenkin pidettävä mielessä, että tässä työssä analysoitu joukko on kielellisesti hyvin valikoitunutta. Koeasetelman kannalta olisi ollut edullisempaa, jos cloze-testit olisi voitu suorittaa koko valintakokeen perusjoukolle. Tällöin korrelaatioiden perusteella olisi myös voinut tehdä päätelmiä siitä, missä määrin cloze-testien mittaustulos olisi antanut samanlaisen valintatuloksen kuin kokeessa käyetyt monivalintatestit.<sup>1</sup>

1 Tämän kysymyksen selvittämiseksi suoritettiin kesällä 1975 valintakokeen Turussa suorittaneessa ryhmässä (N=300) alustava cloze-kokeilu, eri testein kuin tässä työssä käytetyt. Koska kokeilu suoritettiin valintakokeen jälkeen ja siihen osallistuminenkin oli vapaaehtoista, vain vajaat 50 pyrkijää halusi suorittaa kokeen. Tuloksia ei kuitenkaan analysoitu loppuun, sillä pyrkijät eivät olleet jaksaneet enää tosissaan paneutua cloze-testien suorittamiseen.

Cloze-testien korrelaatiot valintakokeen monivalintatesteihin on esitetty liitteessä 4. Korrelaatiomatriisista voidaan tehdä seuraavia päätelmiä. 1. Kontekstuaalisella korjausmenetelmällä arvostellut cloze-testit korreloivat paremmin valintakokeen testeihin kuin sanatarkalla menetelmällä korjatut. Tämä tulos on samansuuntainen useiden aikaisempien tutkimusten kanssa (Oller ja Conrad 1971; Oller 1972; Oller ym 1974; Stubbs ja Tucker 1974). Testissä 3 (muuttujat 13 ja 18 liitteessä 4) tulokset ovat kuitenkin osittain päinvastaisia: sanatarkan menetelmän korrelaatio valintapistemäärään (muuttuja 10) on merkitsevä 5%:n riskitasolla, mutta kontekstuaalisen menetelmän vastaava korrelaatio ei ole tilastollisesti merkitsevä. 2. Cloze-testit näyttävät hieman yllättäen korreloivan jonkin verran paremmin valintakokeen kuuntelutesteihin kuin kirjallisiin tehtäviin (cloze-testit 19 ja 20 sekä muuttujat 1-4 ja 5-9). Kuuntelukokeissa korrelaatio on paras äänneiden tunnistamistestiin. Kaikki korrelaatiot ovat kuitenkin samalla 1%:n riskitasolla merkitseviä myös puhejaksojen tunnistamis- ja puhutun ymmärtämistesteihin; englanninkielinen luento sen sijaan saa alhaisempia korrelaatioita. Näiltäkin osin tulokset ovat samansuuntaisia aikaisempien tutkimusten kanssa, joskin niissä todetut korrelaatiokertoimet ovat olleet huomattavasti korkeammat (vrt. Darnell 1970: clozentropia/kuullun ymmärtäminen .74 ja Oller ym. 1974: cloze/kuullun ymmärtäminen .76).

3. Cloze-testeistä saa parhaita korrelaatioita valintakoe-testeihin Fordia käsitellyt testi 2. Tässä testissä myös sanatarkalla menetelmällä korjatut tulokset (muuttuja 12) saavat tilastollisesti merkitseviä korrelaatioita kuuntelutesteihin. Äidinkielen testin 1 korrelaatiot ovat epäjohdonmukaisen erottelukyvyn vuoksi heikkoja kautta linjan, mutta parantuvat jonkin verran kontekstuaalisella menetelmällä korjattuina. Molemmat testit korreloivat merkitsevästi valintakokeen loppupistemäärään (muuttuja 10), ja Ford-testin korrelaatio siihen on merkitsevä 1%:n riskitasolla. 4. Englanninkieliset cloze-testit korreloivat valintakokeen kirjallisista tehtävistä parhaiten sanaston tuntemukseen ja tekstin ymmärtämiseen (5%:n riskitasolla). Nämäkin tulokset ovat samansuuntaisia aikaisempien tutkimusten kanssa, mutta kertoimet ovat alhaisempia (vrt. Stubbs ja Tucker 1974: cloze, kon-



tekstuaalinen korjaus/sanasto .65 ja luetun ymmärtäminen .70; Oller ja Conrad 1971: cloze/luetun ymmärtäminen .80). Voidaan siis todeta, että englanninkielisten cloze-testien korrelaatiot valintakokeen testeihin ovat samansuuntaisia useiden aikaisempien tutkimusten kanssa: ne osoittavat positiivista ja tilastollisesti merkittävää riippuvuutta cloze-testien ja integratiivista taitoa mittaavien kuullun ja luetun ymmärtämistestien välillä.

2.4. Sanatarkan ja kontekstuaalisen korjausmenetelmän sekä eri arvostelijoiden väliset korrelaatiot. Testit 1 ja 2 korjasi toisistaan riippumatta kaksi suomalaista ja kaksi syntyperäistä englantilaista puhujaa, kaikki yliopiston opettajia. Liitteessä 4 esitetyssä korrelaatiomatriisissa on laskettu myös näiden väliset korrelaatiot (tulomomenttikerroin). Matriisista havaitaan, että eri korjausmenetelmillä ja arvostelijoilla saadut tulokset ovat samansuuntaisia ja kohtalaisen korkeaa yksimielisyyttä osoittavia: korrelaatiokertoimet ovat .70-.80 suuruusluokkaa. Näistä tuloksista laskettiin myös järjestyskorrelaatiot (rho), jotka olivat seuraavanlaisia:

testi 1, arvostelijat A/B	.80
testi 2, arvostelijat C/D	.81

Sanatarkan ja kontekstuaalisen korjausmenetelmän antaman paremmuusjärjestyksen vertailemiseksi laskettiin testeissä 1 ja 2 arvostelijoiden välinen keskiarvo, jota korreloitiin sanatarkan menetelmän tuloksiin; testin 3 korjasi kuitenkin vain yksi syntyperäinen englantilainen arvostelija. Arvostelumenetelmien väliset järjestyskorrelaatiot olivat seuraavat:

testi 1, sanatarkka/A+B:n keskiarvo	.73
testi 2, sanatarkka/C+D:n keskiarvo	.73
testi 3, sanatarkka/kontekstuaalinen	.77 (yksi arvostelija)

Arvostelijoiden välisiä korrelaatioita on pidettävä tyydyttävän korkeina. Sen sijaan sanatarkan ja kontekstuaalisen menetelmän väliset korrelaatiot ovat huomattavasti alhaisempia kuin eräissä tutkimuksissa saadut (vrt. Oller ym. 1974: .94 ja Stubbs ja Tucker 1974: .97, r-korrelaatioita). Näiden tulosten perusteella näyttää siltä, että kontekstuaalisen menetelmän korvaaminen sanatarkalla arvostelulla aiheuttaisi ainakin melko huomattavaa testattavien paremmuusjärjystä koskevan informaation menetystä.

Kun ottaa huomioon, että kontekstuaalinen menetelmä osoittautui tilastollisesti paremmaksi ja sai myös korkeampia korrelaatiota kriteerinä käytettyyn valintakokeeseen, on ainakin näiden tulosten valossa perustellumpaa käyttää kontekstuaalista kuin sanatarkkaa korjaustapaa.

### 3. Tulosten tarkastelua

Tässä tutkimuksessa todettiin cloze-testeissä tilastolliselta kannalta ongelmalliseksi se, että umpimähkäinen aukotuseriaate ("standard cloze") tuottaa paljon heikkoja osioita, jotka ovat joko liian helppoja tai vaikeita antaakseen mitään relevanttia informaatiota testattavien välisistä suorituseroista. Mittaus on tällaisten osioiden kohdalla luonnollisesti epä-taloudellista. Jos testi on lyhyt, sen todellinen erottelukyky jää tällöin vain harvojen hyvin toimineiden osioiden varaan. Tässä työssä oli kontekstuaalisessakin korjauksessa testissä 1 vain 19% ja testissä 2 45% sellaisia osioita, joiden korrelaatio saman testin kokonaispistemäärään oli tilastollisesti merkittävä. Myös Carroll (1959; ks. Rankin 1965) on todennut saman ongelman standarditekniikalla suorittamisissaan englannin vieraan kielen cloze-testauksissa. Testien alhaisen reliabiliteetin vuoksi Carroll suhtautuikin epäilevästi cloze-testien käyttökelpoisuuteen yksilöiden välisessä erottelussa. Koska reliabiliteetti riippuu suuresti testin pituudesta, osiomäärien tulisi ilmeisesti olla suurempia kuin tässä tutkimuksessa käytetyt, ehkä 50-70 osiota.

Tässä todettu erottelukyvyn epäjohtonmukaisuus saattaa tosin olla osittain laskennallistakin. OPSAM-tietokoneohjelmassa edellytetään nimittäin, että testien tulisi olla mittauskohteiltaan homogeenisiä. Ellei asianlaita ole näin, osiokorrelaation alhaisuus saattaa johtua (moniselitteisyyden ja liiallisen helppouden tai vaikeuden ohella) myös siitä, että osio mittaa jotakin testin osioiden pääjoukosta riippumatonta kohdetta. Pienissä henkilöjoukoissa, kuten tässäkin tutkimuksessa (N=79), satunnaistekijöistä johtuva vaihtelu saattaa myös olla alhaisen osiokorrelaation syynä. Jos osiot mittaavat useita toisistaan riippumattomia kohteita, ne saattavat olla sinänsä mielekkäitä, mutta niistä ei muodostu yhtä, homogeenistä testiä. Kaikki nämä

tekijät alentavat tällöin testin sisäistä johdonmukaisuutta osoittavaa reliabiliteettia (Mikkonen ja Mikkonen 1971). On mahdollista, että eri aukkojen täydentämisessä tarvittavat päättelyprosessit saattavat olla luonteeltaan erilaisia, riipuen erilaisista syntaktisista ja semanttisista rajoituksista. Toisaalta voidaan myös ajatella, että sidosteisen normaali-tekstin virittämä maailmankuva ja vastaamisessa käytettävä yleisstrategia ("ennakointikieliopin mobilisointi") muodostavat jossain mielessä yhtenäisen mittauskohteen, minkä tulisi näkyä suhteellisen korkeina reliabiliteetteina. Tähän viittaavat eräissä englannin syntyperäisillä puhujilla suoritetuissa kokeissa saadut korkeat reliabiliteetti-indeksit, joissa on samoin käytetty sisäistä johdonmukaisuutta osoittavaa laskemisperiaatetta (esim. Bormuth 1963: .84-.87 välillä olevia arvoja). Myös tässä työssä saadut englanninkielisten testien reliabiliteetit osoittivat tyydyttävää sisäistä johdonmukaisuutta.

Näiden tulosten valossa Ollerin (1973a ja 1973b) havainnot noin 10 prosentin suuruusluokkaa olevista heikosti erottelevista cloze-osioista tuntuvat joka tapauksessa liiankin optimistisilta. Niiden puutteena on se, että Oller ei ilmeisesti ole suorittanut paljoakaan testiensä osioanalyysyjä, vaan puhuu niiden erottelukyvyistä ratkaisuprosenttien perusteella. Ratkaisuprosentit osoittavat kuitenkin vain testien vaikeustasoa, joka ei vielä sano mitään niiden erottelukyvyn johdonmukaisuudesta. Tässä tutkimuksessa saatuja tuloksia on kuitenkin koehenkilöjoukon pienuuden ja testien lyhyiden vuoksi pidettävä vielä alustavina. Erottelukyvyn ongelmaa olisi tämän vuoksi meilläkin tutkittava lisää suuremmissa testausryhmissä ja sidostuneisuudeltaan mahdollisesti erilaisissa, pitemmissä testeissä.

Standarditeknikan erottelukyvyn ongelmallisuus nostaa myös esille kysymyksen standardi- ja modifioidun cloze-tekniikan välisestä suhteesta. Eräässä näitä vertaillaessa tutkimuksessa todettiin, että molemmat korreloivat yhtä hyvin kriteerinä käytettyyn lukemistestiin; modifioitu cloze-testi oli sitä paitsi myös reliaabelimpi ja erottelukykyisempi kuin standardi-cloze (Greene 1964; ks. Bennett ym.). Myös Nygårdin ja Nordströmin

käyttämät modifioidut cloze-testit (tässä volyymissa) osoittautuivat reliaabeleiksi ja samalla OPSAM-ohjelmalla korjattuina hyvän erottelukyvyn omaaviksi. Myös näiden kahden tekniikan välistä suhdetta olisi mielenkiintoista tutkia suuremmissa, vertailukelpoisissa koehenkilöjoukoissa.

Kontekstuaalinen korjausmenetelmä osoittautui tässäkin tutkimuksessa tilastolliseksi paremmaksi. Arvostelijoiden väliset .80 suuruusluokkaa olevat korrelaatiokertoimet osoittavat myös melko hyvää korjaaajien välistä reliabiliteettia. Tätä tulosta voidaan verrata Hirvosen (1974) tutkimustulokseen ohjattujen kirjoitustehtävien (kirje ja kannanotto) subjektiivisesta arvostelusta. Hän käytti nimittäin samoin kahta riippumatonta arvostelijaa ja sai näiden väliseksi korrelaatioksi .63. Tähän verrattuna cloze-testien arvostelu on huomattavasti objektiivisempää. Siinä päästäisiin ilmeisesti täydelliseen yksimielisyyteen, jos arvostelijat etukäteen laativat kontekstin perusteella hyväksyttävistä vastauksista luettelon, jota voidaan tarpeen mukaan täydentää yhdessä neuvotellen. Tällaisen luettelon laatiminen (esim. pienen esikokeilun perusteella) myös nopeuttaisi arvostelua, joka on muutoin noin 4-5 kertaa hitaampaa kuin sanatarkka korjaus. Darnellin (1970) esittämä clozentropia-korjaus tuntuu tämän vuoksi lupaavalta, jos se saadaan meilläkin ohjelmoiduksi tietokoneella suoritettavaksi.

Vaikka cloze-testien ilmeisvaliditeetti tuntuu perustelulta (vrt. Enkvist ja Kohonen tässä volyymissa), jotkut testattavat kokevat silti tekniikan epäoikeudenmukaisena. Varsinkin sanatarkkaa täydentämistä pidetään kohtuuttoman vaikeana. Tämä saattaa osittain johtua tekniikan outoudestakin (Hirvonen 1973). Tämän kysymyksen selvittelemiseksi testattavia pyydettiin tässä tutkimuksessa esittämään vastauslomakkeella mieliteensä käytettyjen cloze-testien vaikeudesta ja soveltuvuudesta kielitaidon mittausvälineeksi. Vastaukset pyydettiin merkitsemään asteikolle 1... 5<sup>1</sup>. Vastausprosentit olivat noin 74%.

1 Kysymykset on esitetty testin lopussa liitteessä 1. Tarkat vastausprosentit olivat, kysymyksittäin, 75.6%, 74.8% (kys. 1) ja 74.8% ja 73.2% (kys. 2, a ja b).

Kyselyn tuloksista ilmeni, että englanninkielisiä testejä pidettiin selvästi vaikeampina kuin äidinkielistä: "helppona" (luokat 1-2) piti testejä 2 ja 3 vain  $3/95=3.2\%$ , vastaavan luvun ollessa testin 1 kohdalla  $36/96=37.5\%$ . Testejä 2 ja 3 piti "vaikeina" (luokat 4-5)  $66/95=69.5\%$ , kun tätä vastaava arviointi testistä 1 oli  $18/96=18.8\%$ . Näiden mielipiteiden ristiintaulukointi sanatarkalla menetelmällä arvosteltuun menestykseen antoi tulokseksi, että äidinkielen testin vaikeustasoa koskevat mielipiteet olivat 1:n riskitasolla riippuvia menestyksestä tässä testissä, mutta englanninkielisten testien vaikutta koskeneet mielipiteet eivät riippuneet menestyksestä.

Mielipiteitä cloze-testien soveltuvuudesta kielitaitokoikeiksi (kysymys 2) voidaan pitää varauksellisen positiivisina. Rungas kolmannes vastanneista piti sekä äidinkielistä että englanninkielisiä testejä hyvinä tai erinomaisina (luokat 4-5; prosenttiluvut vastaavasti 34.8% ja (engl.testit) 35.5%). Noin neljännes piti niitä kuitenkin heikkoina kielitaidon mittareina (prosenttiluvut (suom.testit) 28.4% ja 24.7%). On mielenkiintoista todeta, että nämä testien onnistuneisuutta koskeneet mielipiteet eivät olleet riippuvia kummassakaan tapauksessa sanatarkalla korjauksella arvostellusta testimenestyksestä: ristiintaulukoinnista lasketut  $\chi^2$ -arvot eivät olleet tilastollisesti merkitseviä.

Asennoituminen käytettävää testiä kohtaan näyttää joka tapauksessa olevan yksi cloze-testien pulmakysymys, sillä negatiivinen asennoituminen tuo tuloksiin kielitaidosta riippumattonta virhevarianssia (vrt. myös Johanssonin 1973 tulokset melutestistä). Eräiden vastaajien vapaamuotoisista kommenteista voi päätellä, että testit koetaan liian "psykologisina". Kontekstuaalisen arvostelumenetelmän käyttö ilmeisesti lieventää tätä ongelmaa, jos arvosteluperusteet selvitetään testattaville ennen testiä. Tämäkin näkökohta, parempien tilastollisten ominaisuuksien ohella, puoltaa siis kontekstuaalisen korjaustavan tai vastaavanlaisen tietokonesovelluksen käyttöä.

## Lähdeluettelo

- Bennett, S. ym. 1967. "The Cloze procedure: A Review of the Literature and Possible Applications to the Study of Deviant Language Functions", Center for Research on Language and Language Behavior. (ilm.vuosi ei tiedossa)
- Bormuth, J. 1963. "Cloze as a Measure of Readability", International Reading Association Conference Proceedings. New York: Scholastic Magazines 8:131-34.
- Darnell, D.K. 1970. "Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students", Speech Monographs 37:36-46.
- Gradman, H. 1973. "Reduced Redundancy Testing: A Reconsideration", teoksessa M.C.O'Brien (toim.), Testing in Second Language Teaching: New Dimensions, Dublin, 41-48.
- Hirvonen, P. 1973. Englannin kielen taidon mittaaminen lukion päättyessä: 3, Lukukoe ja kirjoituskoe. Turku: AFinLA:n julkaisuja 7.
- Hirvonen, P. 1974. Englannin kielen taidon mittaaminen lukion päättyessä: 5, Kielitaitokoe. Turku: AFinLA:n julkaisuja 9.
- Johansson, S. 1973. "An Evaluation of the Noise Test", IRAL 11:107-31.
- Kohonen, V. (toim.). 1974. Inter-University Student Selection for English 1974. Jyväskylä: Kielikeskuksen julkaisuja 5.
- Mikkonen, V. ja J.Mikkonen. 1971. OPSAM: opintosaavutusten mitta. Helsinki: Tammi.
- Oller, J.W., Jr. ja C.A.Conrad. 1971. "The Cloze Technique and ESL Proficiency", Lang. Learning 21:183-95.
- Oller, J.W., Jr. ja N.Inal. 1971. "A Cloze Test of English Prepositions", TESOL Quarterly 5:315-26.
- Oller, J.W., Jr. 1972. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language", MLJ 56:151-58.
- Oller, J.W., Jr. 1973a. "Cloze Tests of Second Language Proficiency and What They Measure", Lang. Learning 23:105-18.

- Oller, J.W., Jr. 1973b. "Pragmatic Language Testing", Lang. Sciences 28:7-12.
- Oller, J.W., Jr. ja J.C.Richards (toim.). 1973. Focus on the Learner: Pragmatic Perspectives for the Language Teacher. Rowley, Mass.: Newbury House Publishers.
- Oller, J.W., Jr. ym. 1974. "Cloze, Dictation, and the Test of English as a Foreign Language", Lang. Learning 24, 245-52.
- Perren, G.E. ja J.L.M. Trim (toim.). 1971. Applications of Linguistics. Cambridge UP.
- Rankin, E.F., Jr. 1965. "The Cloze Procedure - A Survey of Research", Fourteenth Yearbook of the National Reading Conference, 133-50.
- Spolsky, B. 1971. "Reduced Redundancy as a Language Testing Tool", teoksessa Perren, G.E. ja J.L.M. Trim (toim.) 1971, 383-90.
- Spolsky, B. 1973. "What Does It Mean to Know a Language; or How Do You Get Someone to Perform His Competence?", teoksessa Oller, J.W., Jr. ja J.C.Richards (toim.) 1973, 164-76.
- Stubbs, J.B. ja G.R.Tucker. 1974. "The Cloze Test as a Measure of English Proficiency", MLJ 58:239-41.

## TEKSTIN YMMÄRTÄMISKOKEILU

Oheisilla testeillä pyritään selvittämään, missä määrin pystyt ymmärtämään suomen- tai englanninkielistä tekstiä, josta on määrävällein umpimähkäisesti jätetty pois osa sanoista. Tehtäväsi on tekstin asiayhteyden perusteella täydentää mielestäsi alkuperäisessä tekstissä ollut sana tai sitä mahdollisimman tarkoin vastaava synonyymi. Kunkin viivan paikalle tulee YKSI SANA, joka voi olla mistä tahansa sanaluokasta ja myös yhdyssana, mutta silti vain yksi sana.

Pyydän Sinua ystävällisesti paneutumaan huolellisesti näiden tehtävien ratkaisemiseen, sillä ne ovat osa tekstin ymmärtämistä koskevaa tutkimusta, jolloin jokaisen suoritus vaikuttaa ratkaisevasti analyysituloksiin. Saat halutessasi pitää nämä tehtäväpaperit sekä kokeen jälkeen jaettavat oikeat vastaukset, joista voit tarkistaa pistemääräsi. Testit arvostellaan tietokoneella, ja voit tarkistaa sijoituksesi kokeiluun osallistuvassa joukossa laitoksen ilmoitustaululle noin kahden viikon kuluttua ilmestyvästä tulosluiskasta.

Suoritusohjeet: 1. Lue tekstiote kertaalleen läpi merkitysisällön hahmottamiseksi. 2. Täydennä tehtävämönisteesen tekstiyhteyden perusteella mielestäsi oikea, alkuperäisestä tekstistä poisjätetty sana. 3. Tarkista vastauksesi lukemalla tekstiote läpi vielä kerran. 4. Kopioi vastaukset (selvästi tekstaten!) vastauslomakkeelle. 5. Varmista, että kaikki vastaukset on siirretty vastauslomakkeelle, ja että myös nimesi on lomakkeen oikeassa yläreunassa. 6. Luovuta vastauslomake kokeen valvojalle.

Suoritus aika: 60 minuuttia, eli n. 20 minuuttia/ testi.

### Testi 1. Lähi-Idän tilanne

Egyptin ja Israelin äskettäin allekirjoittama sopimus osoittaa molempien osapuolien suostuneen huomattaviin myönnytyksiin, joilla rintamien (1)\_\_\_\_\_ vyöhykettä edelleen laajennetaan. Joukkojen (2)\_\_\_\_\_ ohella on nyt ilmeisesti (3)\_\_\_\_\_ myös sodan vaaraa useilla (4)\_\_\_\_\_. Myös periaatteellisella tasolla on (5)\_\_\_\_\_ sopimukseen eräistä kysymyksistä, jotka (6)\_\_\_\_\_ olleet Israelin olemassaolon tunnustamisen (7)\_\_\_\_\_: kolmen vuoden "välirauha", lupaus (8)\_\_\_\_\_



ryhtymättä sotaan tai sillä (9)\_\_\_\_\_, merkitsee ainakin väliaikaista Israelin (10)\_\_\_\_\_. Solmiessaan kahdenkeskisen sopimuksen Israelin (11)\_\_\_\_\_ Egypti on siten sanoutunut (12)\_\_\_\_\_ arabimaiden yhteisen rintaman säilyttämislupauksestaan. (13)\_\_\_\_\_ kyseessä ei olekaan lopullinen (14)\_\_\_\_\_, Kairon ratkaisu on selväkö (15)\_\_\_\_\_ myönnytys, jota arabinationalistit eivät (16)\_\_\_\_\_ suopein silmin. Sopimukseen päästäkseen (17)\_\_\_\_\_ myös israelilaiset joutuneet antamaan (18)\_\_\_\_\_ vaatimuksestaan Siinain hallussapitämisestä, (19)\_\_\_\_\_ he ovat nyt myös (20)\_\_\_\_\_ selvemmin sitoneet tulevaisuutensa Yhdysvaltojen (21)\_\_\_\_\_ varaan.

Ulkoministeri Kissingerille sopimus (22)\_\_\_\_\_ huomattavaa diplomaattista menestystä, joka (23)\_\_\_\_\_ keväisen sovittelukierroksen epäonnistumisen tuottamat (24)\_\_\_\_\_. Jonkin verran arveluttavalta näyttää (25)\_\_\_\_\_ sijaan se, että Yhdysvallat (26)\_\_\_\_\_ lähettämään siviilitekniikoita Siinaihin. Tämä (27)\_\_\_\_\_ tuoda mukanaan arvaamattomia riskejä (28)\_\_\_\_\_ räjähdysalttiilla alueella. Toisaalta on (29)\_\_\_\_\_, että Lähi-Idässä ei ilman (30)\_\_\_\_\_ päästäisikään eteenpäin; niitä ovat (31)\_\_\_\_\_ tapauksessa nyt ottaneet kaikki (32)\_\_\_\_\_. Sopimus on osoitus liennytyso politiikan (33)\_\_\_\_\_, mutta se tuo samalla (34)\_\_\_\_\_ tekijän Lähi-Idän tilanteeseen: yhden (35)\_\_\_\_\_ muodollisen sitoutumisen tilanteen pitämiseksi (36)\_\_\_\_\_. Jos kongressi hyväksyy sopimuksen (37)\_\_\_\_\_ siviilitekniikoiden lähettämistä Siinaihin, nämä (38)\_\_\_\_\_ mahdollisessa kriisitilanteessa vaarassa jäädä (39)\_\_\_\_\_. Jotkut amerikkalaiset ovat jo (40)\_\_\_\_\_ sopimuksessa vaaran Yhdysvaltojen kietoutumisesta (41)\_\_\_\_\_ ongelmavyöhyteen samalla tavoin kuin (42)\_\_\_\_\_ Vietnamin sotaan.

## Testi 2. President Ford.

Not even his opponents can deny that Ford has done far better than anyone had expected. (1)\_\_\_\_\_ no intellectual, Ford has gained (2)\_\_\_\_\_ by his diligence and common (3)\_\_\_\_\_; few people any more crack (4)\_\_\_\_\_ about his inability to chew (5)\_\_\_\_\_ and walk at the same (6)\_\_\_\_\_. Ford's current position is based (7)\_\_\_\_\_ his recent successes: a series (8)\_\_\_\_\_ well-thought vetoes of some bills (9)\_\_\_\_\_ by Congress, the dramatic rescue (10)\_\_\_\_\_ an American merchant vessel captured (11)\_\_\_\_\_ the Cambodians, and his forcefulness (12)\_\_\_\_\_ dealing with foreign leaders. His (13)\_\_\_\_\_ rests, above all, on the (14)\_\_\_\_\_ of tone that he has (15)\_\_\_\_\_ to the White House. What (16)\_\_\_\_\_ important is that he obviously

(17)\_\_\_\_\_ political life, and feels secure (18)\_\_\_\_\_ to take criticism and attack (19)\_\_\_\_\_ necessary. By his open behaviour (20)\_\_\_\_\_ has blotted out the shame (21)\_\_\_\_\_ the Nixon years. Even his (22)\_\_\_\_\_ are perceived as pluses: he (23)\_\_\_\_\_ trusted, in part, because he (24)\_\_\_\_\_ not seem to aim very (25)\_\_\_\_\_. He has, in effect, restored (26)\_\_\_\_\_ presidency to the American people, (27)\_\_\_\_\_ the response has been one (28)\_\_\_\_\_ relief and gratitude.

In the (29)\_\_\_\_\_ of a free and open (30)\_\_\_\_\_ of ideas, Ford has assembled (31)\_\_\_\_\_ of the most impressive Cabinets (32)\_\_\_\_\_ recent history. All of its (33)\_\_\_\_\_ have had distinguished careers in (34)\_\_\_\_\_ fields. Ford does not, however, (35)\_\_\_\_\_ to be the least embarrassed (36)\_\_\_\_\_ the high-powered brains around (37)\_\_\_\_\_. In fact, he is giving (38)\_\_\_\_\_ considerable independence and also expects (39)\_\_\_\_\_ to disagree with him. For (40)\_\_\_\_\_ moment, Ford is politically secure, (41)\_\_\_\_\_ stands a good chance of (42)\_\_\_\_\_ the 1976 presidential election.

Testi 3. On propaganda. (A)

The term 'propaganda' can be defined as a deliberate attempt to exercise an influence on the emotional attitudes of others. The word takes its (1)\_\_\_\_\_ from the Latin verb propagare, (2)\_\_\_\_\_ meant the gardeners' practice of (3)\_\_\_\_\_ the fresh shoot of a (4)\_\_\_\_\_ into the earth to produce (5)\_\_\_\_\_ plants. Attempts to change the (6)\_\_\_\_\_ of others are older than (7)\_\_\_\_\_ history and have probably originated (8)\_\_\_\_\_ the development of speech. It (9)\_\_\_\_\_ speech that gives one the (10)\_\_\_\_\_ to manipulate others without necessarily (11)\_\_\_\_\_ to violence. Physical force or (12)\_\_\_\_\_ of it may produce submission (13)\_\_\_\_\_ the will of another individual (14)\_\_\_\_\_ group. Since thoughts are primarily (15)\_\_\_\_\_ through the spoken or written (16)\_\_\_\_\_, it is obvious that even (17)\_\_\_\_\_ these cases the weapons are (18)\_\_\_\_\_ verbal or symbolic at any (19)\_\_\_\_\_, and the results aimed at (20)\_\_\_\_\_.

Within the present century, the (21)\_\_\_\_\_ of the term has undergone (22)\_\_\_\_\_ changes, and the word has (23)\_\_\_\_\_ a derogatory sense, being frequently (24)\_\_\_\_\_ as a synonym for 'brainwashing'. (25)\_\_\_\_\_ authorities even go so far (26)\_\_\_\_\_ to take the view that (27)\_\_\_\_\_ are at the mercy of (28)\_\_\_\_\_ media, and suggest that brainwashing (29)\_\_\_\_\_ other similar techniques are irresist-

ible (30)\_\_\_\_\_ lead to permanent changes in (31)\_\_\_\_\_ political or religious outlook. One (32)\_\_\_\_\_ to admit that there are (33)\_\_\_\_\_ which it would be highly (34)\_\_\_\_\_ to change, such as attitudes (35)\_\_\_\_\_ mental and public health, those (36)\_\_\_\_\_ by many towards racial matters, (37)\_\_\_\_\_ sexual freedom. It is certainly (38)\_\_\_\_\_ considering how campaigns promoting better (39)\_\_\_\_\_ in such areas might best (40)\_\_\_\_\_ run.<sup>1</sup>

-----

Kiitoksia avustasi! Pyydän Sinua vielä vastaamaan vastauslomakkeella esitettyihin kysymyksiin, joissa tiedustellaan mielipiteitäsi tämän-tyyppisestä testauksesta; muut mahdolliset kommentit vastauslomakkeen kääntöpuolelle!

Testien laatija:

Viljo Kohonen  
Turku

-----

Vastauslomakkeella olleet kysymykset:

1. Testien vaikeustaso? a) suom. 1 2 3 4 5  
b) engl. 1 2 3 4 5

(vastausohje: rengasta sopiva vaihtoehto: 1 = helppo .....  
5 = vaikea)

2. Miten hyvin tällaiset testit mittaavat mielestäsi kieli-  
taitoa? a) suom. 1 2 3 4 5  
b) engl. 1 2 3 4 5

(vastausohje: 1 = heikosti ..... 5 = erinomaisesti)

---

1 Tästä testistä poistettiin tulosten käsittelyvaiheessa osio 18 (oikea vastaus essentially), koska testin Turussa ja Jo-suussa suoritettussa versiossa oli tässä kohtaa toimitusvirheestä johtuva eri sana. Osioiden numerointi oikeissa vastauksissa jatkuu siten, että tällä lomakkeella oleva osio n:o 19 = n:o 18 seuraavalla sivulla, jne. Testin molemmilla versioissa olivat testattavat sanat tämän korjauksen jälkeen täsmälleen samat, mutta lauseiden järjestys erilainen.

Tekstin ymmärtämiskokeilu: vastauksetTesti 1.

- |                   |                 |                       |
|-------------------|-----------------|-----------------------|
| 1. välistä        | 15. poliittinen | 29. ilmeistä          |
| 2. loitontamisen  | 16. katsele     | 30. riskejä           |
| 3. loittonnettu   | 17. ovat        | 31. joka              |
| 4. vuosilla       | 18. periksi     | 32. osapuolet         |
| 5. päästy         | 19. ja          | 33. jatkamisesta      |
| 6. ovat           | 20. entistä     | 34. uuden             |
| 7. esteenä        | 21. avun        | 35. suurvallan        |
| 8. olla           | 22. merkitsee   | 36. rauhallisena      |
| 9. uhkaamiseen    | 23. korvaa      | 37. kohdan            |
| 10. tunnustamista | 24. pettymykset | 38. ovat              |
| 11. kanssa        | 25. sen         | 39. sotatoimialueella |
| 12. irti          | 26. sitoutuvat  | 40. näkevinään        |
| 13. Vaikka        | 27. saattaa     | 41. Lähi-Idän         |
| 14. rauhansopimus | 28. edelleenkin | 42. aikoinaan         |
- 

Testi 2.

- |            |                 |              |             |
|------------|-----------------|--------------|-------------|
| 1. Though  | 12. in          | 23. is       | 33. members |
| 2. respect | 13. popularity  | 24. does     | 34. their   |
| 3. sense   | 14. change      | 25. high     | 35. appear  |
| 4. jokes   | 15. brought     | 26. the      | 36. by      |
| 5. gum     | 16. is          | 27. and      | 37. him     |
| 6. time    | 17. enjoys      | 28. of       | 38. them    |
| 7. on      | 18. enough      | 29. interest | 39. them    |
| 8. of      | 19. when        | 30. exchange | 40. the     |
| 9. passed  | 20. he          | 31. one      | 41. and     |
| 10. of     | 21. of          | 32. in       | 42. winning |
| 11. by     | 22. limitations |              |             |
- 

Testi 3.

- |             |                   |              |               |
|-------------|-------------------|--------------|---------------|
| 1. origin   | 11. resorting     | 21. radical  | 31. has       |
| 2. which    | 12. threat        | 22. acquired | 32. attitudes |
| 3. pinning  | 13. to            | 23. used     | 33. desirable |
| 4. plant    | 14. or            | 24. Some     | 34. towards   |
| 5. new      | 15. created       | 25. as       | 35. held      |
| 6. opinions | 16. word          | 26. we       | 36. or        |
| 7. recorded | 17. in            | 27. mass     | 37. worth     |
| 8. with     | 18. rate          | 28. and      | 38. attitudes |
| 9. is       | 19. psychological | 29. and      | 39. be        |
| 10. power   | 20. image         | 30. our      |               |
-

Liite 2. Testin 1 ("Lähi-Itä") osioanalyysin tulokset sanatarkalla ja kontekstuaalisella menetelmällä korjattuna Jyväskylän ryhmässä (N=79).

Sanatarkka			Kontekstuaalinen		
	r	%		r	%
1	.04	68.35	1	.16	73.42
2	.25	43.04	2	.20	62.03
3	.10	2.53	3	.15	77.22
4	.07	1.27	4	.29	5.06
5	-.24	86.08	5	.06	91.14
6	.00	100.00	6	.00	100.00
7	.28	18.99	7	.16	29.11
8	-.03	98.73	8	.00	100.00
9	.11	22.78	9	.14	74.68
10	.11	63.29	10	.20	65.82
11	.00	100.00	11	.00	100.00
12	.00	100.00	12	.00	100.00
13	.29	62.03	13	.23	64.56
14	.15	34.18	14	.23	81.01
15	-.05	20.25	15	.21	31.65
16	.03	96.37	16	-.12	97.47
17	.00	100.00	17	.00	100.00
18	-.12	83.54	18	.04	98.73
19	.09	20.25	19	-.06	44.30
20	-.01	43.04	20	.04	97.47
21	-.16	56.96	21	.15	98.73
22	.26	65.82	22	.22	97.47
23	.20	60.76	23	.09	93.67
24	.06	29.11	24	-.01	82.28
25	-.31	98.73	25	-.21	98.73
26	.02	2.53	26	.29	44.30
27	.06	64.56	27	.00	100.00
28	-.05	1.27	28	.11	81.01
29	.06	6.33	29	.06	96.20
30	.20	41.77	30	.11	44.30
31	-.21	72.15	31	.10	96.20
32	.33	58.23	32	.27	79.75
33	.13	2.53	33	-.08	82.28
34	.13	68.35	34	-.00	75.95
35	-.06	5.06	35	.28	50.63
36	-.04	12.66	36	.09	97.47
37	-.07	7.59	37	.21	93.67
38	-.02	83.54	38	.10	89.87
39	-.13	2.53	39	.30	53.16
40	.00	.00	40	.17	97.47
41	.07	50.63	41	.15	81.01
42	.01	67.09	42	.07	93.67
$\bar{x}$	.04	48.19	$\bar{x}$	.11	79.08
s	.14	34.71	s	.12	23.44

Osiokorrelaation merkitsevyysraja 5%:n riskitasolla: .22

Liite 2. Testin 2 ("Ford") osioanalyysin tulokset sanatar-  
kalla ja kontekstuaalisella menetelmällä korjat-  
tuna Jyväskylän ryhmässä (N=79).

Sanatarokka			Kontekstuaalinen		
	r	%		r	%
1	.18	16.46	1	.33	65.82
2	.11	2.53	2	.10	68.35
3	-.04	94.94	3	-.05	94.94
4	.38	10.13	4	.09	21.52
5	.18	26.58	5	.16	49.37
6	.10	79.75	6	.12	78.48
7	-.20	94.94	7	-.12	98.73
8	.13	96.20	8	.16	96.20
9	.11	7.59	9	.19	15.19
10	.19	86.08	10	.20	86.08
11	.02	93.67	11	.08	93.67
12	.16	63.29	12	.06	84.81
13	-.12	8.86	13	.28	44.30
14	.00	5.08	14	.25	18.99
15	.10	32.91	15	.28	45.57
16	-.14	72.15	16	-.12	77.22
17	.19	16.45	17	.38	75.95
18	.18	37.97	18	.10	43.04
19	-.01	10.13	19	.26	93.67
20	.13	91.14	20	.12	97.47
21	.05	79.75	21	.11	88.61
22	.00	.00	22	.34	44.30
23	.23	94.94	23	.20	94.94
24	.03	92.41	24	.07	93.67
25	.22	56.96	25	.20	56.96
26	.17	62.03	26	.25	62.03
27	.34	63.29	27	.33	67.09
28	.26	72.15	28	.25	72.15
29	.00	.00	29	.02	16.46
30	.25	1.27	30	.14	13.92
31	.37	68.35	31	.31	69.62
32	.08	69.62	32	.14	86.08
33	.34	75.95	33	.27	75.95
34	.25	11.39	34	.30	82.28
35	-.16	1.27	35	.16	78.48
36	.12	13.92	36	.34	51.90
37	.37	78.48	37	.38	78.48
38	.39	54.43	38	.23	67.09
39	.32	58.23	39	.18	79.75
40	.32	50.63	40	.33	50.63
41	.20	27.85	41	.25	58.23
42	.30	69.62	42	.35	69.62
$\bar{x}$	.14	48.79	$\bar{x}$	.19	66.85
s	.15	33.81	s	.12	24.20

Osiokorrelaation merkitsevyysraja 5%:n riskitasolla: .22

Liite 3. Vuoden 1975 englannin yhteisen valintakokeen testien tärkeimmät tilastolliset ominaisuudet (perusjoukon N=999) ja samojen testien keskiarvot ja hajonnat Jyväskylän cloze-testausryhmässä, sekä cloze-testien keskiarvot ja hajonnat. (N=63).

Valintakoe populaatio (N=999)						Cloze-ryhmä (N=63)	
testi	hyväks. osioita	ratkaisuprosentti	reliabiliteetti	$\bar{x}$	s	$\bar{x}$	s
1	25	78.6	.62	19.6	2.9	20.9	2.3
2	20	77.7	.48	15.5	2.3	17.1	1.8
3	30	47.8	.68	14.3	4.3	17.1	4.0
4	20	52.7	.56	10.5	3.0	12.3	2.7
5	30	47.0	.80	14.1	5.2	18.1	4.3
6	24	46.6	.60	11.7	3.2	13.5	2.9
7	22	65.2	.76	14.0	4.1	17.3	2.6
8	50	62.2	.82	31.1	6.9	35.8	3.8
9	20	49.9	.82	10.0	4.5	13.0	3.2
koko testi	247		.90	169.9	30.3	201.0	13.1

Valintakokeen testit: 1 Äänneiden tunnistaminen, 2 Puhejaksojen tunnistaminen, 3 Puhutun ymmärtäminen, 4 Englanninkielinen luento, 5 Sanasto ja idiomit I, 6 Sanasto ja idiomit II, 7 Tekstin ymmärtäminen, 8 Kielelliset rakenteet, 9 Kielellinen päättely, 10 Valintapistemäärä (=nämä valintakokeen testit (vaikutusosuus 70%) sekä 5 koulutodistusarvosanaa (pisteitys 5 x maksimi 9 = 45 pistettä korkein mahdollinen), vaikutusosuus 30%).

Cloze-testien keskiarvot ja hajonnat Jyväskylän ryhmässä (N=63):

	$\bar{x}$	s
11 "Lähi-Itä", sanatarkka	20.3	2.7
12 "Ford", sanatarkka	21.1	3.6
13 "Propaganda", sanatarkka	10.3	4.0
14 "Lähi-Itä", kontekstuaalinen A	33.4	3.0
15 "Lähi-Itä", " B	36.3	3.4
16 "Ford", kontekstuaalinen C	28.6	4.4
17 "Ford", " D	31.8	4.4
18 "Propaganda", " (D)	24.0	6.7
19 "Lähi-Itä", keskiarvo A+B	34.9	2.9
20 "Ford", keskiarvo C+D	30.2	4.2

Liite 4. Cloze-pistemäärien korrelaatiot englannin valintakokeen 1975 testeihin.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.00																			
2	.45	1.00																		
3	.38	.49	1.00																	
4	.36	.18	.22	1.00																
5	.19	.22	.29	.14	1.00															
6	.02	-.01	-.02	.02	.21	1.00														
7	.24	.07	-.03	.12	.16	.12	1.00													
8	.23	.14	-.01	.30	.36	-.03	-.06	1.00												
9	.01	.07	.13	.16	-.02	-.06	.23	-.13	1.00											
10	.58	.51	.45	.59	.54	.30	.29	.44	.26	1.00										
11	.31	.23	.29	.10	-.07	.24	.21	-.24	.11	.19	1.00									
12	.48	.28	.43	.36	.30	.36	.31	.05	-.02	.55	.42	1.00								
13	.20	.16	.30	.12	.30	.27	.08	.16	-.17	.30	.08	.36	1.00							
14	.46	.40	.46	.27	.38	.34	.33	.11	-.03	.58	.32	.83	.33	1.00						
15	.43	.30	.43	.02	.04	.24	.24	-.19	.10	.31	.75	.35	.14	.37	1.00					
16	.47	.30	.39	.11	.10	.12	.24	-.14	-.00	.34	.53	.31	.28	.41	.73	1.00				
17	.53	.35	.34	.11	.17	.31	.32	.01	-.13	.39	.35	.67	.31	.81	.46	.58	1.00			
18	.28	.15	.15	.02	.26	.18	.08	.08	-.19	.15	.24	.39	.77	.37	.16	.28	.47	1.00		
19	.48	.32	.42	.06	.07	.19	.26	-.16	.02	.34	.67	.34	.22	.41	.91	.93	.57	.25	1.00	
20	.50	.40	.45	.24	.33	.29	.30	.09	-.06	.52	.31	.78	.33	.95	.39	.48	.91	.45	.47	1.00

Testit: 1 äänteiden tunnistaminen, 2 puhejaksoiden tunnistaminen, 3 puhutun ymmärtäminen, 4 englanninkielinen luento, 5 sanasto ja idiomit I, 6 sanasto ja idiomit II, 7 tekstin ymmärtäminen, 8 kielelliset rakenteet, 9 kielellinen päättely, 10 valintapistemäärä, 11 "Lähi-Itä", sanatarikka, 12 "Ford", sanatarikka, 13 "propaganda", sanatarikka, 14 "Ford", kontekstuaalinen C, 15 "Lähi-Itä", kontekstuaalinen A, 16 "Lähi-Itä", kontekstuaalinen B, 17 "Ford", kontekstuaalinen D, 18 "propaganda, kontekstuaalinen, 19 "Lähi-Itä", keskiarvo A+B, 20 "Ford", keskiarvo C+D.