Konfirmatorisesta faktorianalyysista ja sen käytöstä reliabiliteetti- ja validiteettitarkasteluissa yleisten kielitutkintojen syksyn 1998 englannin keskitason testissä

Raija Härkönen

Tilastotieteen pro gradu –tutkielma
7. joulukuuta 2000

Jyväskylän yliopisto
Tilastotieteen laitos

Tutkimusseloste

Raija Härkönen: *Konfirmatorisesta faktorianalyysista ja sen käytöstä reliabiliteetti- ja validiteettitarkasteluissa yleisten kielitutkintojen syksyn 1998 englannin keskitason testissä.*
Tilastotieteen pro gradu –tutkielma, Jyväskylän yliopisto, 7. joulukuuta 2000.
Sivuja 91.

Tutkielman aiheena oli selvittää yleisten kielitutkintojen englannin kielen testin reliabiliteettia ja validiteettia useammasta näkökulmasta. Tarkoituksena oli tutkia, mittasivatko tehtävät sitä kielitaidon aluetta mitä pitikin, olivatko arvioijat yhdenmukaisia arvioidessaan tuottamisen osakokeita (kirjoittaminen ja puhuminen) sekä oliko kirjoittamisen ja puhumisen arviointi yhtä helppoa. Aineistona oli yleisten kielitutkintojen syksyn 1998 englannin kielen keskitason testi. Aineistossa oli 251 testisuoritusta, joista 94:lle oli tehty kaksoisarviointi. Tutkimusmenetelmänä käytettiin konfirmatorista faktorianalyysia.

Tulosten mukaan englannin kielen testin tehtävät mittasivat pääsääntöisesti hyvin kielitaidon eri osa-alueita eli tekstin ymmärtämistä, kirjoittamista, rakenteita ja sanastoa, puheen ymmärtämistä sekä puhumista, ja nämä osa-alueet puolestaan mittasivat luotettavasti yleistä kielitaitoa. Arvioijien välillä oli eroja; ykkösarvioijat olivat arvioineet suorituksia kakkosarvioijia lievemmin. Kirjoittamisen arvioiminen oli vaikeampaa kuin puhumisen arviointi, mutta ykkös- ja kakkosarvioijat olivat yhtä luotettavia arvioimaan tuottamisen osakokeita, ja kirjoittamisen sekä puhumisen arvioinnin laatu oli yhtä hyvää sekä ykkös- että kakkosarvioijien keskuudessa.

Avainsanoja: konfirmatorinen faktorianalyysi, reliabiliteetti, validiteetti, yleinen kielitutkinto

Sisällysluettelo

1 Johdanto

Tutkimuskohteena tässä pro gradu –työssä on yleisten kielitutkintojen syksyn 1998 englannin kielen keskitason testin reliabiliteetti ja validiteetti. Aihetta lähestytään kolmesta eri näkökulmasta eli tutkimuskysymyksiä ovat: mittaavatko tehtävät sitä kielitaidon aluetta mitä niiden pitäisi mitata, ovatko arvioijat yhdenmukaisia arvioidessaan tuottamisen osakokeita (kirjoittaminen ja puhuminen) sekä onko kirjoittamisen ja puhumisen arviointi yhtä helppoa. Aineistolle on tehty muutamia eksploratiivisia analyyseja, mutta pääasiassa tutkimusmenetelmänä käytetään konfirmatorista faktorianalyysia. Lähestymistapa on tutkimushypoteeseja varmentava.

Luku 2 käsittää tutkielman teoriaosuuden, joka koostuu konfirmatorisesta faktorianalyysista sekä reliabiliteetista ja validiteetista. Teoreettisen viitekehyksen pohjana on käytetty teoksia Faktorianalyysi (Leskinen, 1987) ja Structural Equations with Latent Variables (Bollen, 1989). Luvussa 3 esitellään aineisto ja tutkimusongelmat. Aineisto koostuu 251:stä yleisten kielitutkintojen englannin kielen testisuorituksesta syksyllä 1998. Arvioinnin laadun varmistamiseksi 94 suoritusta on kaksoisarvioitu. Luku 4 sisältää englannin kielen testin reliabiliteetti- ja validiteettitarkasteluja ja luku 5 käsittelee arvioijien yhdenmukaisuutta kirjoittamisen ja puhumisen arvioinnissa.

2 Konfirmatorinen faktorianalyysi

Faktorianalyysin perusideana on pyrkiä kuvaamaan muuttujien kokonaisvaihtelua pienemmällä muuttujien määrällä. Faktorianalyysimallit on rakennettu vähintään välimatka-asteikon tasoisille muuttujille, mutta mallia käytetään usein järjestysasteikollisillekin muuttujille. Faktorianalyysi perustuu selvään malliin, jolla etsitään havaittujen muuttujien avulla taustalla olevia tekijöitä, ns. piilomuuttujia eli faktoreita (Heikkilä, 1999).

Faktorianalyysin käyttö voidaan jakaa tutkimusotteen perusteella kahteen ryhmään, eksploratiiviseen ja konfirmatoriseen faktorianalyysiin. Eksploratiivisessa faktorianalyysissa etsitään kokeilemalla ominaisarvotarkastelujen ja rotaatiomenetelmien avulla sopiva tulkittavissa oleva faktorirakenne. Konfirmatorisessa faktorianalyysissa pyritään taas tarkentamaan ja varmentamaan sisällöllisen tutkimusteorian pohjalta nousseita teoreettisia hypoteeseja faktorirakenteista (Leskinen, 1987).

2.1 Konfirmatorinen faktorimalli

Oletetaan seuraavassa, että havaittujen y-muuttujien lukumäärä on p ($y_1, y_2, \dots, y_p$) ja η-faktoreiden lukumäärä on m (η_1, η_2, ..., η_m), $m < p$. Otetaan käyttöön seuraavat vektorimerkinnät:

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, \quad \underset{\sim}{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix}, \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

Faktorimallin perusmalli voidaan silloin esittää matriisimuodossa

$$\underset{\sim}{y} = \Lambda \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \ , \tag{2.1}$$

jossa Λ on $p \times m$-latausmatriisi

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix} .$$

Oletetaan edelleen, että $\mathrm{E}\eta_i\varepsilon_j = 0$, $\forall i,j$, $i = 1,\dots, m$, $j = 1,\dots, p$, ja että $\mathrm{E}\varepsilon_i\varepsilon_j = 0$, $\forall i,j$, $i \neq j$, $i,j = 1,\dots, p$. Merkintöjen ja laskutoimitusten yksinkertaistamiseksi oletetaan, että muuttujien odotusarvot ovat nollia. Merkitään η-faktoreiden $m \times m$-kovarianssimatriisia Ω:lla:

$$\mathrm{cov}(\underset{\sim}{\eta}) = E \underset{\sim}{\eta} \underset{\sim}{\eta}^T = \Omega = \begin{bmatrix} \omega_{11} & & & \\ \omega_{21} & \omega_{22} & & \\ \vdots & \vdots & \ddots & \\ \omega_{m1} & \omega_{m2} & \cdots & \omega_{mm} \end{bmatrix}$$

ja ε-jäännösten kovarianssimatriisia Θ:lla:

$$\mathrm{cov}(\underset{\sim}{\varepsilon}) = E \underset{\sim}{\varepsilon} \underset{\sim}{\varepsilon}^T = \Theta = \begin{bmatrix} \theta_1 & & & (0) \\ & \theta_2 & & \\ & & \ddots & \\ (0) & & & \theta_p \end{bmatrix} = \mathrm{diag}(\theta_1 \theta_2 \ldots \theta_p).$$

Silloin havaittujen y-muuttujien, satunnaisvektorin $\underset{\sim}{y}$ teoreettinen kovarianssimatriisi Σ on muotoa

$$\begin{aligned}\Sigma = \operatorname{cov}(\underset{\sim}{y}) &= E\,\underset{\sim}{y}\,\underset{\sim}{y}^T = E(\Lambda\underset{\sim}{\eta}+\underset{\sim}{\varepsilon})(\Lambda\underset{\sim}{\eta}+\underset{\sim}{\varepsilon})^T \\ &= E(\Lambda\underset{\sim}{\eta}+\underset{\sim}{\varepsilon})\,(\underset{\sim}{\eta}^T\Lambda^T+\underset{\sim}{\varepsilon}^T) \\ &= E(\Lambda\underset{\sim}{\eta}\,\underset{\sim}{\eta}^T\Lambda^T+\Lambda\underset{\sim}{\eta}\,\underset{\sim}{\varepsilon}^T+\underset{\sim}{\varepsilon}\,\underset{\sim}{\eta}^T\Lambda^T+\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}^T) \\ &= \Lambda E(\underset{\sim}{\eta}\,\underset{\sim}{\eta}^T)\Lambda^T+\Lambda E(\underset{\sim}{\eta}\,\underset{\sim}{\varepsilon}^T)+E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\eta}^T)\Lambda^T+E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}^T) \\ &= \Lambda\operatorname{cov}(\underset{\sim}{\eta})\Lambda^T+\operatorname{cov}(\underset{\sim}{\varepsilon}) \\ &= \Lambda\,\Omega\,\Lambda^{\mathrm{T}}+\Theta,\end{aligned}$$

koska $E(\underset{\sim}{\eta}\,\underset{\sim}{\varepsilon}^T)=0$ ja $E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\eta}^T)=0$. Faktorimalli (2.1) tuottaa siten y-muuttujien teoreettiseksi kovarianssimatriisiesitykseksi tuloksen

$$\Sigma = \Lambda\,\Omega\,\Lambda^{\mathrm{T}}+\Theta. \tag{2.2}$$

Faktorimallien valinnan ratkaisee ensisijaisesti se, kuinka tarkkoja hypoteeseja tutkijalla on tutkimusongelmastaan faktorianalyyseja ajatellen. Mikäli tutkimustilanne on sellainen, ettei tutkija voi etukäteen kiinnittää faktoreiden lukumäärää tai tutkijalla ei ole ennakkotietoa rajoitteista latausmatriisissa Λ, on eksploratiivinen faktorimalli ainoa vaihtoehto. Faktorimallin konstruointi ja sen tulkinnat suoritetaan ensisijaisesti aineiston antamaan informaatioon perustuen, ja tutkimus on silloin luonteeltaan eksploratiivista aineiston analysointia.

Jos tutkijalla on ennakkotietoa faktoreiden lukumäärästä ja hän pystyy tutkimushypoteesiensa nojalla asettamaan rajoituksia faktoreiden latausmatriisissa Λ ja faktoreiden kovarianssimatriisissa Ω, hän voi valita konfirmatorisen faktorimallin ja testata tilastollisesti valitun mallin sopivuutta aineistoon. Tutkimusote on silloin tutkimushypoteeseja varmentavaa eli konfirmatorista aineiston käsittelyä. Tutkimushypoteesit voivat olla peräisin aikaisemmista eksploratiivisten faktorianalyysien tuloksista tai ne voivat olla puhtaasti teoreettisen ajattelun synnyttämiä.

Konfirmatorisessa faktorimallissa tehdään siis rajoituksia faktoreiden latausmatriisiin Λ. Osa latauksista λ_{ij} kiinnitetään mallissa vakioiksi. Faktoreiden kovarianssimatriisi Ω voi olla rajoittamaton tai faktoreiden varianssejа ja/tai faktoreiden kovariansseja (korrelaatioita) voidaan kiinnittää vakioiksi. Parametrien välille voidaan tehdä rajoituksia, joista parametrien yhtäsuuruusrajoitukset ovat yleisimpiä. Latausten kiinnitykset ovat yleensä nollia tai ykkösiä. Faktoreiden varianssien kiinnitykset ovat yleensä ykkösiä, jolloin Ω on faktoreiden korrelaatiomatriisi. Faktoreiden kovarianssien (korrelaatioiden) kiinnitykset ovat yleensä nollia. Konfirmatorisissa faktorimalleissa jäännösten kovarianssimatriisi Θ ei välttämättä ole diagonaalinen vaan osa jäännöksistä voi korreloida keskenään (Leskinen, 1987).

2.1.1 Moniulotteinen normaalijakauma

Oletetaan, että havaitut y-satunnaismuuttujat noudattavat p-ulotteista normaalijakaumaa, jonka odotusarvovektori on

$$E\,\underset{\sim}{y} = \underset{\sim}{\mu}$$

ja kovarianssimatriisi

$$\mathrm{cov}\,(\underset{\sim}{y}) = E(\underset{\sim}{y} - \mu)(\underset{\sim}{y} - \mu)^T = \Sigma\,.$$

Merkitään tätä

$$\underset{\sim}{y} \sim N_p(\underset{\sim}{\mu}, \Sigma).$$

Satunnaismuuttujien varianssit ovat muotoa

$$\sigma_{ii} = \mathrm{E}\,(y_i - \mu_i)^2, \qquad i = 1,\ldots,p$$

ja kovarianssit ovat muotoa

$$\sigma_{ij} = \mathrm{E}\,(y_i - \mu_i)\,(y_j - \mu_j), \qquad i \neq j;\ i,j = 1,\ldots,p.$$

Satunnaismuuttujien y eli satunnaisvektorin $\underset{\sim}{y}$ teoreettinen kovarianssimatriisi on silloin muotoa

$$\Sigma = \begin{bmatrix} \sigma_{11} & & & \\ \sigma_{21} & \sigma_{22} & & \\ \vdots & \vdots & \ddots & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Faktoriesitystä käyttäen moniulotteisen normaalijakauman teoreettinen kovarianssimatriisi parametrisoidaan silloin kuten edellä (2.2):

$$\Sigma = \Lambda\,\Omega\,\Lambda^{\mathrm{T}} + \Theta.$$

Usein teoreettisen kovarianssimatriisin sijasta kiinnostuksen kohteena on y-muuttujien teoreettinen korrelaatiomatriisi

$$P = \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix},$$

jossa

$$\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}\,, \qquad i > j;\ i, j = 1, \ldots, p.$$

Matriisiesitysmuodossa korrelaatiomatriisin P ja kovarianssimatriisin Σ välinen yhteys on

$$P = D^{-1}\,\Sigma\,D^{-1},$$

jossa

$$D = diag(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\cdots\sqrt{\sigma_{pp}}).$$

Valittaessa korrelaatiomatriisi P faktorimallitarkastelujen pohjaksi, saadaan

$$\begin{aligned} P &= D^{-1}\,\Sigma\,D^{-1} \\ &= D^{-1}\,\Lambda\;\Omega\;\Lambda^{T} D^{-1} + D^{-1}\Theta D^{-1} \\ &= \Lambda\,(D^{-1}\;\Omega\;D^{-1})\,\Lambda^{T} + D^{-1}\;\Theta D^{-1} \\ &= \Lambda^*\;\Omega^*\;\Lambda^{*T} + \Theta^*. \end{aligned}$$

Korrelaatiomatriisiin perustuvan faktorimalliesityksen ja 'alkuperäisen' kovarianssimatriisiin perustuvan esityksen välille saadaan siten seuraavia skaalayhteyksiä:

$$\begin{aligned} \Lambda^* &= D^{-1}\,\Lambda, \\ \Omega^* &= \Omega, \\ \Theta^* &= D^{-1}\;\Theta D^{-1} \end{aligned}$$

ja

$$\Lambda^* = \Lambda,$$

$$\Omega^* = D^{-1}\,\Omega D^{-1},$$

$$\Theta^* = D^{-1}\,\Theta D^{-1}.$$

2.1.2 Havaintoaineisto sekä otoskovarianssi- ja otoskorrelaatiomatriisit

Oletetaan, että p-ulotteisesta normaalijakaumasta, perusjoukosta, on poimittu N:n havainnon otos. Havaintoaineisto voidaan silloin esittää seuraavan $N{\times}p$-havaintomatriisin avulla:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{Np} \end{bmatrix},$$

jossa k:nnen havainnon arvot ovat k:nnella rivillä, $k = 1,\ldots, N$, ja i:nnen y-muuttujan arvot ovat i:nnellä sarakkeella, $i = 1,\ldots, p$. Odotusarvovektorin suurimman uskottavuuden estimaattori on silloin otoskeskiarvovektori

$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_p \end{bmatrix}, \qquad \bar{y}_i = (1/N)\sum_{k=1}^{N} y_{ki}$$

ja kovarianssimatriisin harhaton estimaattori on otoskovarianssimatriisi

$$S = \begin{bmatrix} s_{11} & & & \\ s_{21} & s_{22} & & \\ \vdots & \vdots & \ddots & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix},$$

jossa

$$s_{ii} = [1/(N-1)] \sum_{k=1}^{N} (y_{ki} - \bar{y}_i)^2 \qquad i = 1,\ldots,p$$

ovat otosvariansseja ja

$$s_{ij} = [1/(N-1)] \sum_{k=1}^{N} (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j) \qquad i \neq j;\ i,j = 1,\ldots,p$$

ovat otoskovariansseja. Voidaan osoittaa, että otoskeskiarvovektori $\underset{\sim}{\bar{y}}$ ja otoskovarianssimatriisi S ovat perusjoukon odotusarvovektorin $\underset{\sim}{\mu}$ ja kovarianssimatriisin Σ tyhjentäviä tunnuslukuja. Ne siis sisältävät otoksesta kaiken sen informaation, joka koskee perusjoukon odotusarvovektoria $\underset{\sim}{\mu}$ ja kovarianssimatriisia Σ. Moniulotteisen normaalijakauman tapauksessa ne myös sisältävät kaiken otosinformaation jakaumasta. Vastaavasti korrelaatiomatriisin P tyhjentävä estimaattori on otoskorrelaatiomatriisi

$$R = \begin{bmatrix} 1 & & & \\ r_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix},$$

jossa

$$r_{ij} = s_{ij} / \sqrt{s_{ii} s_{jj}} \qquad i \neq j;\ i,j = 1,\ldots,p.$$

Näitä korrelaatiokertoimia kutsutaan Pearsonin korrelaatiokertoimiksi tai tulomomenttikorrelaatiokertoimiksi (Leskinen, 1987).

2.2 Konfirmatorisen faktorimallin rakentaminen

Konfirmatorisen faktorianalyysin käyttö tutkimuksen teossa vaatii tutkimusotetta, jossa tutkimusongelmat ja niitä koskevat hypoteesit ovat riittävän selvästi rajattuja. Tutkimushypoteesit on pystyttävä muotoilemaan niin, että niistä voidaan spesifioida identifioituvia faktorimalleja. Näiden mallien mahdollistamien parametrien estimointitulosten perusteella voidaan tarkastella, vastaako mallin tuottama kuvaus ja selitys tutkittavista ilmiöistä asetettuja hypoteeseja. Malleihin kohdistettujen tilastollisten testien ja muiden riittävyystarkastelujen avulla tutkitaan niiden sopivuutta tutkimusaineistoon, jolloin mallien osoittautuessa riittäviksi saadaan tukea ja varmennusta asetetuille tutkimushypoteeseille. Konfirmatorisen faktorimallin rakentaminen jaetaan viiteen päävaiheeseen, jotka ovat

1) mallin spesifiointi
2) mallin identifioituvuustarkastelut
3) mallin parametrin estimointi
4) mallia koskevien hypoteesien testaus
5) mallin riittävyystarkastelut.

Konfirmatorisen faktorimallin rakentaminen etenee tilastollisen teorian kannalta vaiheittain edellä olevassa järjestyksessä (Nummenmaa ym., 1997).

2.2.1 Mallin spesifiointi

Mallin spesifioinnilla eli täsmentämisellä tarkoitetaan faktorimallin valintaa tutkimushypoteesin pohjalta. Jokaiselle faktorimallin parametrimatriisien Λ-, Ω- ja Θ-parametrille siis on valittava jokin seuraavista vaihtoehdoista:

- parametri estimoidaan vapaasti ilman rajoituksia
- parametri estimoidaan jonkin toisen parametrin kanssa yhtäsuurena

- parametri kiinnitetään vakioksi (yleensä nollaksi tai ykköseksi)
- parametrin estimointi suoritetaan jonkin lineaarisen rajoitteen mukaisesti.

Se, mikä vaihtoehto kullekin parametrille valitaan riippuu oletetuista havaittujen muuttujien ja faktoreiden välisistä yhteyksistä (Λ), faktoreista ja niiden välisistä yhteyksistä (Ω) sekä mittausvirheiden variansseista (Θ) (Leskinen, 1987).

2.2.2 Identifioituvuustarkastelut

Konfirmatoriset faktorimallit voidaan jakaa ei-identifioituviin ja identifioituviin malleihin. Mallin identifioituvuus eli yksilöityvyys on mallin teoreettinen ominaisuus, joka ei riipu tutkimusaineistosta, otannasta tai käytettävästä estimointimenetelmästä.

Mallin parametrirakenne tuottaa vain yhden kovarianssirakenteen Σ; kun jokaisella mallin parametrilla on tietty arvo, saadaan siis vain yksi kovarianssimatriisi Σ. Voi kuitenkin olla useita parametrirakenteita, jotka tuottavat saman kovarianssirakenteen (ekvivalentit rakenteet). Jos yksittäisellä parametrilla on sama arvo kaikissa ekvivalenteissa rakenteissa, on kyseinen parametri identifioituva. Jos mallin kaikki rakenteet ovat identifioituvia, on koko malli identifioituva.

Edellä olevaa määritelmää on vaikea käytännössä soveltaa mallin identifioituvuuden tutkimiseen. Sovellusten kannalta käyttökelpoisempi määritelmä on, että mallin yksittäinen parametri on identifioituva, jos se on ratkaistavissa kovarianssimatriisin Σ avulla. Jos mallin kaikki parametrit ovat ratkaistavissa matriisin Σ avulla, on malli identifioituva. (Jöreskog, 1981)

Jos parametri voidaan ratkaista kovarianssimatriisista Σ useammalla tavalla käyttämällä eri yhtälöitä kovarianssimatriisiesityksessä (2.2), sanotaan kyseisen parametrin olevan yli-identifioituva. Yli-identifioituvuuden käsite ei ole ristiriidassa identifioituvuuden käsitteen kanssa, sillä eri ratkaisut tuottavat saman numeerisen arvon yli-identifioituville parametreille.

Tarkastellaan välttämättömiä ehtoja faktorimallin (2.1)

$$\underset{\sim}{y} = \Lambda \underset{\sim}{\eta} + \underset{\sim}{\varepsilon}$$

identifioituvuudelle eli milloin tarkasteltavat parametrit voidaan ratkaista kovarianssiesityksen (2.2)

$$\Sigma = \Lambda \Omega \Lambda^{\mathrm{T}} + \Theta$$

avulla. Symmetrisessä kovarianssimatriisissa Σ on eri kovariansseja ja variansseja eli yhtälöitä yhteensä (1/2) p (p + 1) kappaletta. Toisaalta kovarianssimatriisin parametrisoinnissa (2.2) on mahdollisia parametreja latausmatriisissa Λ pm kappaletta, faktoreiden kovarianssimatriisissa Ω (1/2) m (m + 1) kappaletta ja jäännösten kovarianssimatriisissa Θ p kappaletta eli yhteensä

$$\begin{aligned} s &= pm + (1/2)\, m\, (m+1) + p \\ &= (1/2)(2p+m)(m+1) \end{aligned}$$

kappaletta. Merkitään parametreille tehtävien rajoituksien lukumäärää r_1:llä. Näin saadaan välttämätön ehto parametrien Λ, Ω ja Θ ratkaisemiseksi yhtälöistä (2.2):

$$s - r_1 \leq (1/2)\, p\, (p+1). \tag{2.3}$$

Ehto (2.3) ei kuitenkaan vielä takaa mallin identifioituvuutta. Tarkastellaan edelleen esitystä (2.1), jossa ei ole rajoituksia parametrien suhteen. Olkoon M mielivaltainen epäsingulaarinen $m \times m$-matriisi, jolla suoritetaan rakenteella Λ, Ω muunnos, rotaatio

$$\Lambda^* = \Lambda M,$$
$$\Omega^* = M^{-1} \, \Omega \, (M^{-1})^{\mathrm{T}}.$$

Silloin ratkaisulle Λ^*, Ω^* on voimassa

$$\begin{aligned} \Lambda^* \Omega^* \Lambda^{*\,\mathrm{T}} &= (\Lambda M) \, M^{-1} \, \Omega \, (M^{-1})^{\mathrm{T}} \, (M^{\mathrm{T}} \Lambda^{\mathrm{T}}) \\ &= \Lambda \, \Omega \, \Lambda^{\mathrm{T}}, \end{aligned}$$

ts. on olemassa mielivaltainen määrä erilaisia ratkaisuja, jotka tuottavat saman kovarianssirakenteen Σ. Tarkasteltava faktorirakenne ei ole silloin identifioituva ilman rajoituksia latausmatriisissa Λ ja kovarianssimatriisissa Ω.

Koska ratkaisu Λ, Ω ei ole rotatoinnin suhteen yksikäsitteinen, on faktorimallin identifioituvuudelle asetettava välttämättömiä (lisä)ehtoja. Koska mielivaltainen, epäsingulaarinen matriisi M on muotoa $m \times m$, on parametreille matriiseissa Λ ja Ω asetettava vähintään m^2 rajoitusta yksikäsitteisyyden edellytykselle rotatoinnin suhteen. Edellä olevasta ehdosta (2.3) saadaan toinen välttämätön ehto rajoitusten lukumäärälle r_2:

$$r_2 \geq m^2. \tag{2.4}$$

Yhdistämällä ehdot (2.3) ja (2.4) saadaan seuraava välttämätön ehto rajoitusten lukumäärälle:

$$r^* \geq \max(r_1, r_2).$$

Yleisiä ja käytännössä toimivia riittävyysehtoja faktorimallin (2.1) identifioituvuudelle ei voida yleisessä tapauksessa johtaa (McDonald & Krane 1977, Jöreskog 1981). Kuten aiemmin määriteltiin, on faktorimalli identifioituva, jos sen parametrit ovat ratkaistavissa mallin teoreettisesta kovarianssimatriisiesityksestä. Yksittäisiä, pieniä ja suhteellisen yksinkertaisen rakenteen omaavia faktorimalleja tarkasteltaessa tämä on käyttökelpoinen menettely mallin identifioituvuuden selvittämiseksi. Identifioituvuuden määritelmään perustuvat identifioituvuustarkastelut ovat kuitenkin yleisessä tapauksessa työläitä tehdä. Riittävyysehdot liittyvät keskeisesti rajoitusten sijaintiin parametrimatriiseissa. Eräs tapa arvioida valitun faktorimallin identifioituvuutta on tarkastella riittävien ehtojen voimassa oloa faktorimallin rotaatio-yksikäsitteisyydelle. Mallin rotaatio-yksikäsitteisyys ei kuitenkaan ole vielä riittävä ehto mallin identifioituvuudelle, vaan myös sillä, ovatko parametrien kiinnitykset nollia vai nollasta eroavia, on vaikutusta mallin identifioituvuuteen (Leskinen, 1983a).

Kovarianssiesitys voi olla rajoittamaton tai rajoitettu. Tarkastellaan faktorimallia (2.1) vastaavaa kovarianssiesitystä (2.2). Jos rajoitukset matriiseissa Λ ja Ω on valittu siten, että matriisin

$$\Omega^* = \Lambda \, \Omega \Lambda^{\mathrm{T}}$$

alkioista w^*_{ij} yksikään ei ole kiinteä, on kyseessä rajoittamaton ratkaisu (*unrestricted solution*). Jos rajoitukset matriiseissa Λ, Ω tai Θ on valittu siten, että yksikin alkioista w^*_{ij} on kiinteä, on kyseessä rajoitettu ratkaisu (*restricted solution*). Rajoittamattomassa ratkaisussa yksikään θ_i ei ole kiinteä, koska se johtaisi rajoitettuun ratkaisuun. Rajoittamaton tai rajoitettu ratkaisu voivat olla identifioituvia tai identifioimattomia. Rajoitettu ratkaisu saattaa tuoda lisäongelmia identifioituvuustarkasteluihin. Esimerkiksi jos joku $w^*_{ij} = 0$, $i \neq j$, niin tästä seuraa kovarianssin σ_{ij} menetys, koska nyt yhtälö $\sigma_{ij} = w^*_{ij} = 0$ ei anna informaatiota valitun faktorirakenteen parametrien suhteen. Toisaalta jos kiinnitetään joku θ_i nollasta eroavaksi, saadaan varianssi $\sigma_{ii} = w^{*ii} = \theta_i$ käyttöön w^*_{ii} :n sisältämien parametrien identifioituvuustarkasteluissa (Leskinen, 1987).

2.2.3 Parametrien estimointi

Konfirmatorisen faktorimallin parametrien estimoinnin lähtökohtana on spesifioitu malli, jonka tuntemattomat parametrit latausmatriisissa Λ, faktoreiden kovarianssimatriisissa Ω ja jäännösten kovarianssimatriisissa Θ ovat identifioituvia. Estimoinnin tarkoituksena on löytää tuntemattomille parametreille sellaiset arvot, jotka mahdollisimman hyvin vastaavat faktorimallin parametrien arvoja tutkittavassa perusjoukossa. Parametrit estimoidaan siten, että niiden tuottama kovarianssimatriisin Σ sovite

$$\hat{\Sigma} = \hat{\Lambda}\,\hat{\Omega}\,\hat{\Lambda}^T + \hat{\Theta}$$

olisi mahdollisimman lähellä otoskovarianssimatriisia S, johon on kerätty tilastollisesti tyhjentävästi otoksen antama informaatio perusjoukon teoreettisesta kovarianssimatriisista Σ. Konfirmatorisen faktorimallin luonteesta johtuu, ettei parametrien estimaattoreita voida johtaa yleisessä tapauksessa analyyttisesti. Kun sopiva estimointikriteeri on valittu, suoritetaan parametrien estimointi sopivan tietokoneohjelman avulla iteratiivisesti (Leskinen, 1987).

Jos x-muuttujat noudattavat moniulotteista normaalijakaumaa, mittaukset on suoritettu vähintään välimatka-asteikolla ja havaintoja on riittävästi ($N > 100$), voidaan estimoinnissa käyttää tehokasta suurimman uskottavuuden menetelmää (ML, *maximum likelihood*). Jos nämä oletukset eivät ole voimassa, voidaan parametrien estimointiin käyttää esimerkiksi yleistettyä pienimmän neliösumman menetelmää (GLS, *generalized least-squares*). Käytettäessä järjestysasteikolla mitattujen muuttujien polykorisia korrelaatiokertoimia, voidaan estimointi suorittaa yleisesti painotetulla pienimmän neliösumman menetelmällä (WLS, *generally weighted least-squares*) (Nummenmaa ym., 1997).

Pienimmän neliösumman (pns) menetelmässä minimoidaan kohdefunktio

$$F_1(\Lambda, \Omega, \Theta) = (1/2)\,\mathrm{tr}\,[\Sigma\,(\Lambda, \Omega, \Theta) - S]^2$$

$$= (1/2)\sum_{i=1}^{p}\sum_{j=1}^{p}\left[\sigma_{ij}(\Lambda,\Omega,\Theta) - s_{ij}\right]^2$$

parametrimatriiseissa Λ, Ω ja Θ olevien estimoitavien parametrien suhteen.

Suurimman uskottavuuden (su) menetelmässä minimoidaan kohdefunktio

$$F(\Lambda,\Omega,\Theta) = \ln|\Sigma(\Lambda,\Omega,\Theta)| - \ln|S| + \mathrm{tr}[\Sigma^{-1}(\Lambda,\Omega,\Theta)\,S] - p \qquad (2.5)$$

parametrimatriiseissa Λ, Ω ja Θ olevien estimoitavien parametrien suhteen.

Pienimmän neliösumman estimointimenetelmässä ei siis käytetä jakaumaoletuksia havaituista muuttujista. Se soveltuu käytettäväksi silloin, kun havaitut y-muuttujat eivät noudata moniulotteista normaalijakaumaa. Suurimman uskottavuuden menetelmää voidaan käyttää silloin, kun havaitut muuttujat noudattavat moniulotteista normaalijakaumaa ja kun otoskoko on riittävän suuri. Pns-menetelmä on y-muuttujien skaalauksesta riippuvainen, mutta sen sijaan su-menetelmä on skaalainvariantti.

Suurimman uskottavuuden estimointimenetelmässä oletetaan, että havaitut y-muuttujat noudattavat p-ulotteista normaalijakaumaa

$$\underset{\sim}{y} \sim N_p(\underset{\sim}{\mu}, \Sigma),$$

ja että tästä jakaumasta on poimittu N:n alkion otos. Silloin otoksen $\underset{\sim}{y}_1, \underset{\sim}{y}_2, \dots, \underset{\sim}{y}_N$ uskottavuusfunktio on muotoa

$$L(\underset{\sim}{\mu}, \Sigma; \underset{\sim}{y}_1 \dots, \underset{\sim}{y}_N) = \prod_{j=1}^{N} f(\underset{\sim}{y}_j ; \underset{\sim}{\mu}, \Sigma)$$

$$= (2\pi)^{-pN/2} |\Sigma|^{-N/2} \exp\left\{ -(1/2) \sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\mu}) \right\}.$$

Ottamalla tästä luonnollinen logaritmi, saadaan

$$\ln L(\underset{\sim}{\mu}, \Sigma) = -(pN/2) \ln 2\pi - (N/2) \ln|\Sigma| - (1/2) \sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\mu})$$

Neliösummalauseke voidaan jakaa kahteen komponenttiin seuraavasti:

$$\sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\mu}) = \sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}} + \underset{\sim}{\bar{y}} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}} + \underset{\sim}{\bar{y}} - \underset{\sim}{\mu})$$

$$= \sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}}) + \sum_{j=1}^{N} (\underset{\sim}{\bar{y}}_j - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\bar{y}} - \underset{\sim}{\mu})$$

Edelleen saadaan matriisin jäljen ominaisuutta käyttäen

$$\sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}}) = \sum_{j=1}^{N} tr (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})^T \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})$$

$$= \sum_{j=1}^{N} tr\, \Sigma^{-1} (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})\, (\underset{\sim}{y}_j - \underset{\sim}{\bar{y}})^T$$

$$= (N-1)\, tr\, (\Sigma^{-1} S)$$

jossa S on otoskovarianssimatriisi

$$S = \left[1/(N-1)\right] \sum_{j=1}^{N} (\underset{\sim}{y}_j - \underset{\sim}{\overline{y}})(\underset{\sim}{y}_j - \underset{\sim}{\overline{y}})^T \ .$$

Neliösummalausekkeen jälkimmäinen komponentti sieventyy muotoon

$$\sum_{j=1}^{N} (\underset{\sim}{\overline{y}} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\overline{y}} - \underset{\sim}{\mu}) = N((\underset{\sim}{\overline{y}} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\overline{y}} - \underset{\sim}{\mu}).$$

Koska odotusarvoa $\underset{\sim}{\mu}$ ei parametrisoida faktorimallissa, saadaan sille suurimman uskottavuuden estimaattori

$$\underset{\sim}{\hat{\mu}} = \underset{\sim}{\overline{y}} \ ,$$

jonka estimointi on riippumatonta parametrien Λ, Ω ja Θ estimoinnista. Näin ollen jälkimmäinen komponentti neliösummalausekkeesta häviää. Kun eliminoidaan uskottavuusfunktion logaritmista vielä vakiokomponentti $-(pN / 2) \ln 2\pi$, saadaan

$$\ln L(\Sigma) = -[(N-1)/2] \ln |\Sigma| - [(N-1)/2] \operatorname{tr} (\Sigma^{-1} S),$$

jossa kovarianssimatriisi Σ on estimoitavien parametrien Λ, Ω ja Θ funktio. Funktio $\ln L[\Sigma(\Lambda,\Omega,\Theta)]$ saavuttaa estimoitavien parametrien suhteen maksiminsa samassa pisteessä kuin funktio F saavuttaa miniminsä, koska

$$\ln L[\Sigma(\Lambda, \Omega, \Theta)] = -[(N-1)/2]F + \ln |S| - p.$$

Estimointiteknisesti on kuitenkin helpompaa käyttää minimoitavana funktiona funktiota F,

$$F(\Lambda,\Omega,\Theta) = \ln |\Sigma(\Lambda,\Omega,\Theta)| - \ln |S| + \operatorname{tr}[\Sigma^{-1}(\Lambda,\Omega,\Theta)S] - p.$$

Voidaan osoittaa, että

$$F(\Lambda, \Omega, \Theta) \geq 0.$$

Kohdefunktiota F tarkastelemalla havaitaan, että otoskovarianssimatriisin S tulee olla positiivisesti definiitti, ts. $|S| > 0$. Vastaavasti myös kovarianssimatriisin $\Sigma(\Lambda, \Omega, \Theta)$ tulee olla positiivisesti definiitti.

Suurimman uskottavuuden estimointimenetelmällä on olemassa seuraavat hyvät estimointiteoreettiset ominaisuudet:

1) estimaattorit ovat tarkentuvia,
2) estimaattoreilla on minimivarianssiominaisuudet,
3) estimaattorit ovat asymptoottisesti normaalijakautuneita.

Nämä ominaisuudet perustuvat riittävän suureen otoskokoon: mitä suurempi otoskoko on, sitä tarkempia estimaatteja suurimman uskottavuuden estimointimenetelmällä saadaan. Koska faktorimallin parametrien estimointi perustuu otoskovarianssi- tai otoskorrelaatiomatriisin käyttöön, asettaa otoskovarianssi- tai otoskorrelaatiokertoimien estimointi omia alarajavaatimuksia riittävälle otoskoolle. Otoskoon alarajaksi on yleisesti määritelty 100-200 havaintoa.

Jos faktorimalli on identifioituva, on mahdollista konstruoida sen parametreille tarkentuvia estimaattoreita. On kuitenkin huomattava, että identifioituvuus ei takaa tarkentuvien estimaattoreiden olemassaoloa. Toisaalta, jos malli ei ole identifioituva, ei mikään estimointimenetelmä tuota yksikäsitteisiä estimaatteja identifioitumattomille parametreille. Identifioitumattoman mallin tapauksessa ei myöskään voida laskea keskivirhearvioita yhdellekään mallin parametrin estimaattorille, koska informaatiomatriisi on silloin singulaarinen. Identifioitumaton parametri saa silloin ratkaisun, joka on vain yksi ratkaisu mielivaltaisesta määrästä muita ratkaisuja. Tällainen ratkaisu riippuu voimakkaasti lähtöarvojen asettamisesta iteratiiviselle estimointialgoritmille. Identifioitumattoman mallin identifioituvat parametrit ovat sen sijaan estimoitavissa yksikäsitteisesti. Identifioitumaton malli voi kuitenkin olla käyttökelpoinen käytännön sovelluksissa.

Konfirmatorisen faktorimallin rakentamiseen liittyy myös ongelmatilanteita, jotka voivat johtaa estimoinnin epäonnistumiseen. Parametrien estimointialgoritmin toiminnassa voi esiintyä seuraavia ongelmia:

1) uskottavuusfunktiota ei pystytä laskemaan (su-menetelmä),
2) estimointiaika loppuu,
3) iteraatiokierrosten maksimilukumäärä 250 ylittyy,
4) saavutetaan lokaalinen, paikallinen minimi,
5) saadaan kelvottomat parametrien estimaatit.

Yleisin syy edelläoleville virhetilanteille on se, että on valittu väärä, aineistoon sopimaton faktoriesitys.

Jos suurimman uskottavuuden menetelmässä uskottavuusfunktiota (kohdefunktio F) ei pystytä laskemaan, ei otoskovarianssi- tai otoskorrelaatiomatriisi tai parametrien Λ, Ω ja Θ lähtöarvoilla laskettu kovarianssimatriisi Σ ole positiivisesti definiitti. Matriisien S tai R ei-positiivinen definiittisyys voi johtua lähinnä kolmesta syystä:

1) havaintoja on liian vähän suhteessa muuttujiin eli $N \leq p$,
2) puuttuvia havaintoja on paljon ja korrelaatiot tai kovarianssit on laskettu muuttujapareittain täyden informaation menetelmällä,
3) havaitut muuttujat ovat lineaarisesti riippuvia (esimerkiksi joku otoskorrelaatioista on yksi tai lähellä sitä).

Nämä syyt liittyvät huonosti suunniteltuun ja toteutettuun tutkimukseen, jolloin otosaineisto ei täytä konfirmatorisen faktorimallin tilastollisen rakentamisen oletuksia. Eräs mahdollisuus otosaineiston parantamiseksi tässä tilanteessa on karsia voimakkaasti muuttujia, mutta tämäkään ei välttämättä auta, mikäli otoskoko N on liian pieni. Jos parametrien lähtöarvoilla laskettu kovarianssimatriisi Σ ei ole positiivisesti definiitti, syy on huonosti valituissa lähtöarvoissa tai huonosti spesifioidussa faktorimallin rakenteessa. Tutkijan olisikin kyettävä etukäteen ennakoimaan faktorimallinsa parametrien mahdollisia arvoja mielekkäiden lähtöarvojen valitsemiseksi. Lähtöarvoja valittaessa on yleensä riittävää antaa estimoitaville latauksille nollasta eroavat lähtöarvot, estimoitaville faktoreiden korrelaatioille itseisarvoltaan pienehköt lähtöarvot ja jäännösvariansseille positiiviset, nollasta eroavat lähtöarvot.

Jos uskottavuusfunktiota ei pystytä laskemaan su-menetelmällä, voidaan parametrien estimoinnissa käyttää pns-menetelmää. Näin on tehtävä erityisesti, jos havaittujen muuttujien jakaumat poikkeavat voimakkaasti normaalijakaumasta ja/tai otoskoko N on pieni.

Jos algoritmin toiminnalle varattu estimointiaika loppuu tai jos iteraatiokierrosten maksimimäärä ylittyy, on kyseessä väärin spesifioitu ja/tai aineistoon sopimaton malli. Varsinkin iteraatiokierrosten maksimimäärän ylittyminen on vakava virhesignaali tällaisesta tilanteesta. Estimointiajan loppuminen voi johtua myös faktorimallin estimoitavien parametrien suurehkosta lukumäärästä, jolloin algoritmille on varattava enemmän estimointiaikaa.

Lokaalisen minimikohdan havaitseminen parametrien estimoinnissa on yleensä hankala tehtävä. Mikäli tällaista tilannetta epäilee, on estimointeja toistettava erilaisilla parametrien lähtöarvokokeiluilla. Estimointialgoritmi voi toimia myös laskentateknisesti oikein, mutta tuottaa tilastollisesti kelvottomia parametrien estimaatteja, kuten esimerkiksi negatiivisia virhevariansseja tai ei-positiivisesti definiittejä faktoreiden korrelaatiomatriiseja (Leskinen, 1987).

2.2.4 Hypoteesien testaaminen

Faktorimallia koskevien hypoteesien testaus edellyttää suurimman uskottavuuden estimointimenetelmän käyttöä faktorimallin parametrien estimoinnissa. Oletetaan siten edelleen, että havaitut y-muuttujat noudattavat p-ulotteista normaalijakaumaa ja että otoskoko on riittävän suuri. Faktorimallin estimoinnin tuloksena on saatu parametreille suurimman uskottavuuden estimaatit $\hat{\Lambda}$, $\hat{\Omega}$, $\hat{\Theta}$ minimoimalla kohdefunktio (2.5)

$$F(\Lambda,\Omega,\Theta) = \ln\ |\Sigma(\Lambda,\Omega,\Theta)|\ \text{-}\ln |S| + \mathrm{tr}\ [\Sigma^{-1}(\Lambda,\Omega,\Theta)\ S] - p$$

estimoitavien parametrien suhteen matriiseissa Λ, Ω ja Θ. Suurimman uskottavuuden estimoinnin tuottama sovite Σ:lle on silloin

$$\hat{\Sigma} = \hat{\Lambda}\ \hat{\Omega}\ \hat{\Lambda}^T + \hat{\Theta}\ .$$

Johdetaan seuraavassa uskottavuussuhteen testi, jolla voidaan testata valitun faktorimallin sopivuutta aineistoon. Kohdan (2.2.3) mukaisesti voidaan otoksen uskottavuusfunktion logaritmi esittää ilman vakiota muodossa

$$\ln L = -[(N\text{-}1)/2]\ [\ln |\Sigma| + \mathrm{tr}\ (\Sigma^{-1}S)].$$

Valitaan nyt nollahypoteesiksi

$$H_0\text{: } \Sigma = \Lambda\ \Omega\ \Lambda^{\mathrm{T}} + \Theta,$$

ts. valittu faktorimalli on voimassa ja olkoon vastahypoteesina

H_1: Σ:ssa ei ole rajoituksia.

Vastahypoteesin mukaan kovarianssimatriisi on yleistä muotoa

$$\Sigma = [\sigma_{ij}],$$

jossa ei ole olemassa faktorimallin mukaista parametrisointia. Hypoteesipari voidaan esittää myös muodossa

$$H_0\colon\ \underset{\sim}{y} \sim N_p(\underset{\sim}{0}, \Lambda\,\Omega\,\Lambda^T + \Theta)$$

$$H_1\colon\ \underset{\sim}{y} \sim N_p(\underset{\sim}{0}, \Sigma).$$

Uskottavuussuhteen testi tälle hypoteesiparille on silloin muotoa

$$\lambda = \frac{\max\limits_{H_0} L}{\max\limits_{H_1} L},$$

jossa osoittajassa on uskottavuusfunktion maksimi estimoitavien parametrien Λ, Ω ja Θ suhteen hypoteesin H_0 ollessa voimassa ja nimittäjässä uskottavuusfunktion maksimi estimoitavien parametrien σ_{ij} suhteen hypoteesin H_1 ollessa voimassa. Suurimman uskottavuuden periaatteen nojalla maksimi H_0:n ollessa voimassa saadaan, kun

$$\hat{\Sigma} = \hat{\Lambda}\,\hat{\Omega}\,\hat{\Lambda}^T + \hat{\Theta},$$

ja maksimi H_1:n ollessa voimassa saadaan, kun

$$\hat{\Sigma} = S.$$

Uskottavuussuhteen testistä tunnetaan edelleen, että H_0:n ollessa voimassa testisuure -2 ln λ noudattaa asymptoottisesti χ^2-jakaumaa vapausastein t_1 - t_0 , jossa t_1 on H_1-hypoteesin määrittämässä mallissa olevien estimoitavien parametrien lukumäärä ja t_0 on H_0-hypoteesin määrittämän mallin estimoitavien parametrien lukumäärä. Uskottavuussuhteen testifunktioksi saadaan näin ollen

$$
\begin{aligned}
-2\ln\lambda &= -2(\max_{H_0} L / \max_{H_1} L) \\
&= -2\ln(\max_{H_0} L) + 2\ln(\max_{H_1} L) \\
&= (N-1)\left[\ln\left|\hat{\Sigma}\right| + tr\,(\hat{\Sigma}^{-1}S)\right] - (N-1)\left[\ln|S| + p\right] \\
&= (N-1)\left[\ln\left|\hat{\Sigma}\right| + tr\,(\hat{\Sigma}^{-1}S) - \ln|S| - p\right],
\end{aligned}
$$

joka on toisaalta

$$= (N-1)\,F\,(\hat{\Lambda}, \hat{\Omega}, \hat{\Theta}).$$

Asymptoottisesti χ^2-testin testisuureen arvo saadaan siten suurimman uskottavuuden estimoinnin kohdefunktion F minimin avulla.

Vapausasteet χ^2-testille saadaan kaavasta

$$t_1 - t_0 = (1/2)\,p\,(p+1) - t_0,$$

jossa t_1 = (1/2) p (p +1) on hypoteesin H_1 mallin estimoitavien parametrien lukumäärä eli estimoitavien kovarianssien ja varianssien lukumäärä matriisissa Σ ja jossa t_0 on hypoteesin H_0 määrittämän faktorimallin estimoitavien parametrien lukumäärä. Tässä oletetaan, että teoreettinen kovarianssimatriisi on rajoittamaton.

Identifioituvan mallin vapausasteiden lukumäärä voidaan siis laskea helposti vähentämällä yhtälöiden lukumäärästä vapaasti estimoitavien parametrien lukumäärä. Jotta testi olisi vapausasteiden suhteen käyttökelpoinen, on oltava

$$t_0 < (1/2)\, p\, (p + 1).$$

Vapausasteiden lukumäärän voidaan tulkita ilmaisevan sitä, kuinka paljon faktorimalliesityksen avulla on pystytty tiivistämään informaatiota otoskovarianssimatriisin alkioiden lukumäärästä $(1/2)\ p\ (p + 1)$. Tilastollisessa testauksessa lasketaan χ^2-testisuureen arvo, vapausasteet ja estimoitua χ^2-arvoa vastaava p-arvo eli merkitsevyystaso. Se kertoo todennäköisyyden

$$p = P(\chi^2(df) \geq \hat{\chi}^2(df))$$

eli estimoidun mallin merkitsevyystason H_0:n ollessa voimassa. Pienet p-arvot ilmaisevat siten estimoidun mallin huonoa yhteensopivuutta ja suuret p-arvot puolestaan hyvää yhteensopivuutta. Konfirmatorisen faktorimallin rakentamisessa χ^2-testiä käytetään yleensä yhtenä yhteensopivuusmittana estimoidun mallin sopivuudesta aineistoon. Testiä ei tule silloin käyttää pelkästään tilastollisena merkitsevyystestinä eli estimoitua mallia ei tule hylätä tai hyväksyä pelkästään χ^2-testin tuloksen perusteella mallin rakentamisessa (Leskinen, 1987). Käyttökelpoinen sääntö χ^2-testin käytölle on: Jos χ^2-testin saama arvo on suuri suhteessa testin vapausasteisiin, on estimoitu malli liian yksinkertainen ja sitä tulee modifioida vapauttamalla lisää parametrejä estimoitavaksi. Toisaalta jos χ^2-testin saama arvo on pieni verrattuna vapausasteisiinsa, voi estimoitu malli olla liian yksityiskohtainen, jolloin sen antamien tulosten yleistettävyys on kyseenalaista (Jöreskog, 1969).

Paitsi että asymptoottisen χ^2-testin avulla testataan H_0-hypoteesin faktorimallia kokonaisuudessaan yleistä vastahypoteesia vastaan, voidaan χ^2-testiä käyttää myös faktorimallin osaa tai sen yksittäistä parametria koskevien hypoteesien testauksessa. Oletetaan lähtökohtana, että H_0-hypoteesin määrittämä faktorimalli on oikea ja merkitään tätä mallia M_1:llä. Usein on kiinnostavaa tarkentaa tutkimushypoteesia testaamalla jotakin M_1-mallin erikoistapausta. Merkitään tätä faktorimallia M_2:lla. Malli M_2 kuuluu siis mallin M_1 erikoistapauksena mallin M_1 määrittämien mallien joukkoon. Silloin voidaan testata uutta nollahypoteesia

$$H_0\text{: Malli } M_2 \text{ oikea: } \underset{\sim}{y} = \Lambda_2 \underset{\sim}{\eta_2} + \underset{\sim}{\varepsilon_2}$$

aikaisempaa nollahypoteesimallia M_1 vastaan

$$H_1\text{: Malli } M_1 \text{ oikea: } \underset{\sim}{y} = \Lambda_1 \underset{\sim}{\eta_1} + \underset{\sim}{\varepsilon_1}\,,$$

jossa $H_0 \subset H_1$. Näitä sisäkkäisiä hypoteeseja voidaan testata nyt peräkkäisten χ^2-testien periaatteella: jos uusi H_0 on voimassa, noudattaa testisuure

$$\chi_2^2 - \chi_1^2$$

asymptoottisesti χ^2-jakaumaa vapausastein df_2 - df_1. Peräkkäistestissä χ_2^2 on mallin M_2 nollahypoteesia vastaava testisuure vapausastein df_2, missä vastahypoteesina on yleinen vastahypoteesi H_1: $\Sigma = [\sigma_{ij}]$. Mallin M_1 nollahypoteesia vastaava testisuure on puolestaan χ_1^2 vapausastein df_1, mikä on saatu mallin M_1 testaamisesta yleistä H_1-vastahypoteesia vastaan. Toisaalta vapausasteiden lukumäärä saadaan myös seuraavasti:

$$\begin{aligned} df_2 - df_1 &= 1/2\, p(p+1) - t_2 - [1/2\, p(p+1) - t_1] \\ &= t_1 - t_2, \end{aligned}$$

jossa t_1 ja t_2 ovat mallien M_1 ja M_2 estimoitavien parametrien lukumäärä.

Myöskin otoskokoon N on kiinnitettävä huomiota käytettäessä χ^2-testiä, sillä χ^2-testisuureen arvo on suoraan verrannollinen otoskokoon, koska

$$\hat{\chi}^2 = (N-1)(\hat{\Lambda}, \hat{\Omega}, \hat{\Theta}) .$$

Asymptoottisena testinä χ^2-testin käyttö vaatii riittävän suuren otoskoon, mutta käytännössä konfirmatorista faktorimallia koskevien hypoteesien testaamisessa syntyy ongelmia, jos otoskoko N on hyvin suuri, esimerkiksi $N > 500$. Testi hylkää silloin hyvin herkästi estimoidun mallin aineistoon sopimattomana, vaikka poikkeamat kovarianssimatriisin sovitteen $\hat{\Sigma}$ ja otoskovarianssimatriisin S välillä olisivatkin käytännön kannalta merkitsemättömän pieniä. Hyvin suurten otoskokojen aiheuttamien ongelmien eliminoimiseen ovat Bentler & Bonett (1980) esittäneet χ^2-testisuureeseen perustuvaa normeerattua yhteensopivuusindeksiä (*Normed Fit Index*). Valitaan sen tarkasteluun mallimerkinnät M_0 ja M_1, jossa M_0 on yleinen hypoteettinen nollamalli ja M_1 valittu estimoitu malli. Olkoon näitä malleja vastaavat χ^2-testisuureet

$$\chi_0^2 = (N-1)\, F_0$$

ja

$$\chi_1^2 = (N-1)\, F_1 .$$

Normeerattu yhteensopivuusindeksi on silloin muotoa

$$\Delta_{01} = \frac{\chi_0^2 - \chi_1^2}{\chi_0^2}$$

$$= \frac{(N-1)F_0 - (N-1)F_1}{(N-1)F_0}$$

$$= \frac{F_0 - F_1}{F_0} ,$$

jossa otoskoon N vaikutus on poistettu. Konfirmatorisen faktorimallin tapauksessa yleiseksi nollahypoteesiksi voidaan valita malli

$$\underset{\sim}{y} = \underset{\sim}{\varepsilon} , \tag{2.6}$$

jolloin

$$\operatorname{cov}(\underset{\sim}{y}) = \Sigma = \Theta = diag(\sigma_{11}\sigma_{22}\ldots\sigma_{pp}) .$$

Nollamallissa (2.6) havaitut muuttujat oletetaan siten korreloimattomiksi. Tällainen malli sopii silloin χ^2-testin mielessä mahdollisimman huonosti sellaiseen aineistoon, jossa havaitut muuttujat korreloivat keskenään. Kun estimoitua mallia M_1 verrataan tällaiseen nollamalliin, saadaan normeeratun yhteensopivuusindeksin Δ_{01} avulla arvio siitä, kuinka paljon malli M_1 on selittänyt havaittujen muuttujien kovariansseja tai korrelaatioita suhteessa nollamalliin, joka ei selitä niitä lainkaan.

Normeeratun yhteensopivuusindeksin Δ_{01} arvoalue on $0 \leq \Delta_{01} \leq 1$. Mitä lähempänä ykköstä mallin M_1 arvo on, sitä paremmin malli sopii indeksin Δ_{01} mielessä aineistoon. Bentler ja Bonett esittävät alarajaksi .90. Jos estimoidulle mallille M_1 indeksi $\Delta_{01} \geq .90$, voidaan mallia normeeratun yhteensopivuusindeksin mielessä pitää riittävänä. Jos taas indeksin arvo on alle .90, ei estimoitu malli M_1 ole indeksin Δ_{01} mielessä riittävä.

Paitsi että suuret otoskoot aiheuttavat ongelmia χ^2-testiin, myöskin pienet otoskoot (esimerkiksi $N < 100$) ovat ongelmallisia. Kun N on pieni, eivät suurehkotkaan poikkeamat sovitteen Σ ja otoskovarianssimatriisin S välillä ole välttämättä tilastollisesti merkitseviä. Silloin χ^2-testin mielessä useat erilaiset mallivaihtoehdot voivat olla aineistoon sopivia. Testi ei tällöin ole kovin tehokas ja sen käyttöä mallin valinnan kriteerinä tulisikin tällöin välttää.

Bentlerin ja Bonettin nollamalliajattelua ja normeerattua yhteensopivuusindeksiä voidaan käyttää myös pienten otoskokojen tapauksessa arvioitaessa χ^2-testin tehokkuutta. Oletetaan tällöin, että estimoitu malli M_1 sopii χ^2-testin mielessä aineistoon, kun vastahypoteesina on yleinen H_1-hypoteesi. Tässäkin tapauksessa nollamallina voidaan käyttää korreloimattomien havaittujen muuttujien mallia (2.6). Tutkitaan ensin peräkkäistestin avulla, onko peräkkäistestin

$$\chi^2(df) = \chi_0^2(df_0) - \chi_1^2(df_1) \,, \qquad df = df_0 - df_1,$$

saama arvo tilastollisesti merkitsevä. Jos testisuureen χ^2 (df) p-arvo on suurempi kuin .05, aiheutuu mallin M_1 yhteensopivuus χ^2-testin mielessä ennemminkin χ^2-testin tehottomuudesta hylätä malli, kuin siitä, että malli M_1 olisi oleellisesti parannus nollamallivaihtoehdosta. Jos taas peräkkäistestin saama arvo on tilastollisesti merkitsevä, $p < .05$ ja normeeratun yhteensopivuusindeksin Δ_{01} saama arvo on suurempi kuin .90, voidaan estimoitua mallia M_1 pitää oleellisena parannuksena nollamallista. Mallin M_1 yhteensopivuus χ^2-testin mielessä ei silloin johdu pienen otoskoon aiheuttamasta tehottomuudesta χ^2-testin käytössä (Leskinen, 1987).

2.2.5 Mallin riittävyystarkastelut

Estimoidulle faktorimallille on mahdollista suorittaa riittävyystarkasteluja. Ajatuksena on valita ensin tutkimushypoteesin pohjalta mahdollisimman yksinkertainen faktorimalli ja laajentaa sitä lisäämällä malliin parametreja riittävyystarkastelujen perusteella. Konfirmatorisia faktorimalleja modifioidaan siten eksploratiivisesti aineiston antaman informaation perusteella. Riittävyystarkastelut voidaan jakaa neljään ryhmään:

1) koko mallia koskevat tarkastelut,
2) muuttujakohtaiset tarkastelut,
3) parametrikohtaiset tarkastelut ja
4) havaintokohtaiset tarkastelut.

Koko mallia koskevissa riittävyystarkasteluissa χ^2-testillä on hyvin keskeinen asema. Käytännössä mallin rakentamisessa χ^2-testiä ei useinkaan käytetä pelkästään hypoteesien testaamisen mielessä, vaan pikemminkin antamassa arviota mallin riittävyydestä aineiston kuvaajana. Siten χ^2-testisuureelle muodostuukin usein ennemminkin yhteensopivuusmitan kuin merkitsevyystestin luonne.

Toisena koko mallia koskevana yleisenä riittävyysmittana voidaan käyttää yhteensopivuusindeksiä GFI (*Goodness of fit index*), joka määritellään seuraavasti:

$$\mathrm{GFI} = 1 - tr\,(\hat{\Sigma}^{-1}S - I)^2 \,/\, tr\,(\hat{\Sigma}^{-1}S)^2 \qquad \text{(su-estimointi)}$$

ja

$$\mathrm{GFI} = 1 - tr\,(S - \hat{\Sigma})^2 \,/\, tr\,(S)^2 \qquad \text{(pns-estimointi)}$$

(Jöreskog & Sörbom, 1981). Yhteensopivuusindeksi GFI on riippumaton otoksen koosta N ja suhteellisen robusti havaittujen muuttujien normaalijakaumasta poikkeaville jakaumille. Indeksin arvoalue on välillä (0, 1]. Kun kovarianssimatriisin sovite $\hat{\Sigma}$ on lähellä otoskovarianssimatriisia S, ovat osoittajat yhteensopivuusindeksin GFI lausekkeiden toisessa termissä lähellä nollaa, jolloin GFI saa korkeita, lähellä ykköstä olevia arvoja. Korkeat indeksin arvot ilmaisevat siten hyvää yhteensopivuutta. Käytännössä yhteensopivuusindeksin GFI tulee saada suurempia arvoja kuin .90, jotta mallia voitaisiin tältä osin pitää riittävänä.

Kolmas koko mallia koskeva riittävyysmitta on keskimääräistä jäännöskovarianssia ja jäännösvarianssia mittaava indeksi RMR (*Root Mean Square Residual*), joka määritellään seuraavasti:

$$\mathrm{RMR} = \sqrt{2 \sum_{j=1}^{p} \sum_{i=1}^{j} (s_{ij} - \hat{\sigma}_{ij})^2 / [p(p+1)]}\ ,$$

jossa

$$s_{ij} = [S]_{ij} \quad \text{ja} \quad \hat{\sigma}_{ij} = [\hat{\Sigma}]_{ij}$$

(Jöreskog & Sörbom, 1981). Tätä indeksiä on tulkittava suhteessa havaittuihin kovariansseihin ja variansseihin. RMR-indeksistä on johdettu SRMR (*Standardized Root Mean Square Residual*), jonka pitäisi olla mahdollisimman lähellä nollaa (pienempi kuin .03), jotta malli olisi hyväksyttävissä.

Neljäs koko mallia koskeva riittävyysmitta on RMSEA (*Root Mean Square Error of Approximation*), jolla arvioidaan mallin 'yksinkertaistamisesta' johtuvaa approksimointivirhettä. Mallia voidaan pitää hyvänä, jos RMSEA:n arvo on alle .05.

Myöskin AIC (*Akaike's Information Criterion*) mittaa koko mallin sopivuutta aineistoon. Tätä indeksiä käytetään samalle aineistolle estimoitujen eri mallien vertailussa; mitä pienempi on mallin AIC, sitä parempi malli on kyseessä. AIC määritellään seuraavasti:

$$\begin{aligned}\mathrm{AIC} &= (N-1)\hat{F} + 2t \\ &= \chi^2 + 2t\ ,\end{aligned}$$

jossa t on estimoitujen parametrien lukumäärä.

Kriittisen otoskoon CN (Critical N) avulla voidaan arvioida otoskokoa, jolla malli on vielä riittävä χ^2-testin mielessä merkitsevyystasolla α (yleensä oletetaan, että α = 0). CN:n arvo saadaan kaavasta:

$$\text{CN} = (N-1)\frac{\chi^2_{1-\alpha}(df)}{\hat{\chi}^2(df)} + 1 \, .$$

Jos kriittisen otoskoon CN arvo on suurempi kuin todellinen otoskoko N, on estimoitu malli riittävä.

Näiden koko mallin yhteensopivuutta mittaavien riittävyysindeksien käyttöön liittyy kaksi ongelmaa:

1) mallien sopiessa edellä esitettyjen riittävyysmittojen mielessä hyvin aineistoon, voi mallissa silti olla riittämättömiä yksityiskohtia, ja
2) mallin sopiessa edellä mainittujen riittävyysmittojen mielessä huonosti aineistoon, mitat eivät ilmaise riittämättömyyden syitä.

Tämän takia mallin riittävyysarvioinneissa tarvitaan myös yksityiskohtaisempia tarkasteluja.

Muuttujakohtaisista tarkasteluista tärkein on jokaiselle y-muuttujalle määriteltävä riittävyysindeksi $\hat{R}_i^2$ (S*quared Multiple Correlation*):

$$\hat{R}_i^2 = 1 - \hat{\theta}_i / s_{ii} \, , \qquad i = 1, \ldots, p.$$

Indeksit $\hat{R}_i^2$ voidaan tulkita havaittujen muuttujien reliabiliteetti- tai kommunaliteettikertoimiksi ja ne kuvaavat, kuinka hyvin kukin y_i toimii faktoreiden mittaamisessa. Indeksien teoreettinen arvoalue on välillä [0, 1]. Suuri indeksin arvo ilmaisee vastaavan havaitun muuttujan hyvää mittauskykyä. Jos jonkin havaitun y-muuttujan riittävyysindeksi on lähellä nollaa, voi se olla merkki siitä, ettei kyseinen muuttuja toimi lainkaan indikaattorina, jolloin se tulee poistaa mallista. Toinen mahdollisuus on mallin väärä spesifiointi kyseisen muuttujan suhteen, jolloin mallia on modifioitava uudelleen.

Faktorimallin riittävyystarkasteluista keskeisimpiä ovat parametrikohtaiset tarkastelut, joita voidaan suorittaa sekä vapaasti estimoitujen, yhtä suurina estimoitujen että kiinnitettyjen parametrien osalta. Parametrien estimaattien arvojen tulee olla mallin sisällöllisen tulkinnan kannalta mielekkäitä. Jos estimaatteja ei voida tulkita sisällöllisesti, on mallin käyttö kyseenalaista. Estimaattien arvojen on myöskin oltava tilastotieteellisesti kelvollisia. Faktoreiden varianssien estimaattien $\hat{\omega}_{jj}$, $j = 1, \ldots, m$ ja jäännösvarianssien estimaattien $\hat{\theta}_i$, $i = 1, \ldots, p$ on oltava positiivisia. Esimerkiksi jäännösvarianssien estimaattien negatiivisuus saattaa olla merkki väärin spesifioidusta mallista, pienestä otoksesta tai ei-identifioituvan mallin estimoinnista. Faktoreiden välisten korrelaatiokertoimien tulee olla välillä (-1, 1) ja estimoitujen kovarianssimatriisien $\hat{\Omega}$ ja $\hat{\Theta}$ on oltava positiivisesti definiittejä. Mikäli mallissa on tilastollisesti kelvottomia parametrien estimaatteja, mallia ei saa käyttää.

Mikäli estimointimenetelmänä on käytetty suurimman uskottavuuden menetelmää ja estimoitavana on ollut identifioituva malli, voidaan tarkastella parametrien estimaattoreiden keskivirheitä ja estimaattoreiden välisiä korrelaatioita. Parametrien estimaattoreiden keskivirheiden käyttö on klassinen tapa arvioida parametrien estimointitarkkuutta ja nollasta eroavuutta eli tilastollista merkitsevyyttä. Jos parametrin estimaattorin keskivirhe on pieni, on kyseisen parametrin estimointi tilastollisesti tarkkaa ja parametrin luottamusvälit ovat kapeita. Esimerkiksi 95%:n luottamusväli voidaan laskea kaavan

$$\text{parametrin estimaatti} \pm 2 \times \text{keskivirhe}$$

avulla. Luottamusväli sisältää silloin 95%:n todennäköisyydellä tuntemattoman parametrin. Suuret keskivirheet ilmaisevat puolestaan estimoinnin epäluotettavuutta, joka useimmiten johtuu liian pienestä otoksesta. Vertaamalla saatua parametrin estimaattia vastaavaan keskivirheeseen, saadaan

$$t\text{-arvo} = \frac{\text{parametrin estimaatti}}{\text{keskivirhe}},$$

jonka avulla voidaan arvioida parametrin nollasta eroavuutta. Jos parametri tulkitaan nollaksi, kun $|\ t\text{-arvo}\ | < 2$, niin menettely vastaa t-testin käyttöä likimain 5%:n merkitsevyystasolla. Tätä sääntöä ei kuitenkaan tule käyttää mekaanisesti, vaan jos estimoitu malli on muuten riittävä ja yksinkertainen, ei sitä t-testin mielessä tule täsmentää loppuun asti. Lisäksi on huomioitava, että parametrien estimaattoreiden korreloidessa keskenään eivät yksittäisiin parametreihin kohdistetut tilastolliset merkitsevyysarvioinnit ole toisistaan riippumattomia. Parametrien estimaattoreiden korrelaatiotarkastelut liittyvät estimaattoreiden välisiin voimakkaisiin korrelaatioihin; jos näitä esiintyy, merkitsee se kyseisten parametrien vahvaa riippumista toisistaan ja malli voidaan tulkita silloin näiden parametrien suhteen lähes ei-identifioituvaksi. Tätä ongelmaa voidaan eliminoida vähentämällä mallista estimoitavia parametreja.

Konfirmatorisissa faktorimalleissa voidaan parametreja kiinnittää vakioiksi tai niitä voidaan rajoittaa yhtä suurina estimoitaviksi. Tällaisten kiinnitysten sopivuutta voidaan arvioida parametrien modifikaatioindeksien (MI) avulla. Faktorimallin kiinnitetyille ja rajoitetusti estimoiduille parametreille laskettu modifikaatioindeksin arvo kertoo, kuinka paljon estimoidun mallin χ^2-testisuure vähintään laskee mikäli ko. yksittäinen parametri vapautetaan. Tällöin myöskin menetetään yksi vapausaste. Tarkastelemalla suurimpia modifikaatioindeksien arvoja, saadaan arvio mallin yhteensopivuuden parantamisen suunnasta ja sen vaikutuksesta χ^2-testisuureen arvoon. Vapaasti estimoiduille parametreille modifikaatioindeksin arvo on nolla. Jos sen sijaan kiinnitetyille tai rajoitetusti estimoiduille parametreille modifikaatioindeksi on nolla, on se yleensä merkki siitä, ettei kyseinen parametri ole identifioituva. Mikäli korkeita modifikaatioindeksin saaneita parametreja vapautetaan kaksi tai useampia yhtäaikaa estimoitaviksi, ei χ^2-testisuureen arvo yleensä vähene niin paljon kuin mitä on näiden vapautettujen parametrien modifikaatioindeksien summa. Lisäksi modifikaatioindeksien käytössä on huomioitava, että mallin uudelleen spesifiointi on suoritettava vain sellaisten parametrien suuntaan, jotka ovat tutkimusongelman kannalta mielekkäitä.

Estimoidun faktorimallin riittävyyttä voidaan tarkastella myös havaintojen eli otoskovarianssien ja –varianssien suhteen. Kun malli on estimoitu, lasketaan mallin tuottamat jäännökset kovariansseille ja variansseille:

$$S-\hat{\Sigma}=\left[s_{ij}-\hat{\sigma}_{ij}\right], \qquad i,j=1,\ldots,p.$$

Suuret jäännökset $s_{ij}-\hat{\sigma}_{ij}$ ilmaisevat silloin, mihin osaan aineistoa malli ei sovi. Suurten jäännösten arvioiminen voi kuitenkin olla hankalaa otoskovarianssimatriisin sisältäessä toisistaan suuruudeltaan poikkeavia varianssejä, minkä takia on suoritettava jäännösten normalisointi. Asymptoottisesti on voimassa, että

$$\mathrm{E}s_{ij}\approx\sigma_{ij}, \qquad i,j=1,\ldots,p$$

ja

$$\mathrm{var}(s_{ij})\approx(\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2)/N, \qquad i,j=1,\ldots,p$$

(Jöreskog & Sörbom, 1981). Jäännösten normalisointi suoritetaan silloin seuraavasti

$$NR_{ij}=\frac{s_{ij}-\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}+\hat{\sigma}_{ij}^2/N}}, \qquad i,j=1,\ldots,p.$$

Jos estimoitu malli on oikea, saadaan

$$\mathrm{E}\,(NR_{ij}) \approx 0, \qquad i,j = 1, \ldots, p$$

ja

$$\mathrm{var}\,(NR_{ij}) \approx 1, \qquad i,j = 1, \ldots, p.$$

Silloin normalisoidut jäännökset, joille on voimassa $|\, NR_{ij} \,| > 2$, ilmaisevat mallin riittämättömyyttä. Riittämättömyys paikantuu silloin tällaisia normalisoituja jäännöksiä vastaavien havaittujen muuttujien y_i ja y_j välisiin kovariansseihin tai korrelaatioihin. Normalisoitujen jäännösten avulla saadaan useimmiten myös vihje siitä, miten faktorimallia tulisi parametritasolla spesifioida uudelleen (Leskinen, 1987).

2.3 Toisen kertaluvun faktorimallit

Jos konfirmatorisen faktorimallin (2.1)

$$\underset{\sim}{y} = \Lambda \underset{\sim}{\eta} + \underset{\sim}{\varepsilon}$$

η-faktorit korreloivat, ajatellaan, että niiden taustalla on toisia faktoreita, jotka vaikuttavat η-faktoreihin ja aiheuttavat korrelaatiot. Silloin voidaan ajatella, että η-faktoreiden avulla mitataan näitä laajempia, latentteja yleiskäsitteitä. Näitä latentteja muuttujia kutsutaan toisen kertaluvun (*second-order*) faktoreiksi (Jöreskog & Sörbom, 1979; Leskinen, 1983b). Merkitään toisen kertaluvun faktorirakennetta seuraavasti:

$$\underset{\sim}{\eta} = \Gamma \underset{\sim}{\xi} + \underset{\sim}{\zeta}\,, \tag{2.7}$$

jossa $\underset{\sim}{\xi}$ on $n\times 1$-satunnaisvektori, joka sisältää toisen kertaluvun faktorit ξ_1, ξ_2, ..., ξ_n, Γ on $m\times n$-latausmatriisi ja $\underset{\sim}{\zeta}$ on $m\times 1$-jäännösvektori. Jäännökset voidaan tulkita myös spesififaktoreiksi. Olkoon ξ–faktoreiden kovarianssimatriisi

$$\mathrm{cov}(\underset{\sim}{\xi}) = \Phi$$

ja ζ–jäännösten kovarianssimatriisi

$$\mathrm{cov}(\underset{\sim}{\zeta}) = \Psi .$$

Oletetaan, että $\underset{\sim}{\xi}$ ja $\underset{\sim}{\zeta}$ ovat keskenään korreloimattomia. Toisen kertaluvun faktorirakenne (2.7) tuottaa silloin seuraavan kovarianssiesityksen ensimmäisen kertaluvun η-faktoreille:

$$\begin{aligned}\mathrm{cov}(\underset{\sim}{\eta}) = \Omega = E\underset{\sim}{\eta}\underset{\sim}{\eta}^T \\ &= E(\Gamma\underset{\sim}{\xi} + \underset{\sim}{\zeta})(\underset{\sim}{\xi}^T\Gamma^T + \underset{\sim}{\zeta}^T) \\ &= \Gamma E(\underset{\sim}{\xi}\underset{\sim}{\xi}^T)\Gamma^T + E\underset{\sim}{\zeta}\underset{\sim}{\zeta}^T \\ &= \Gamma\,\Phi\,\Gamma^{\mathrm{T}} + \Psi,\end{aligned} \tag{2.8}$$

joka on analoginen ensimmäisen kertaluvun faktoreiden tuottaman havaittujen y-muuttujien kovarianssimatriisiesityksen kanssa. Toisen kertaluvun faktorimalli on kokonaisuudessaan silloin muotoa

$$\begin{cases} \underset{\sim}{y} = \Lambda\underset{\sim}{\eta} + \underset{\sim}{\varepsilon}, \\ \underset{\sim}{\eta} = \Gamma\underset{\sim}{\xi} + \underset{\sim}{\zeta}, \end{cases} \tag{2.9}$$

jonka tuottama kovarianssimatriisiesitys havaituille y-muuttujille on

$$\operatorname{cov}(\underset{\sim}{y}) = \Sigma = \Lambda\,\Omega\,\Lambda^{T} + \Theta$$

$$= \Lambda\,[\,\Gamma\,\Phi\,\Gamma^{T} + \Psi\,]\,\Lambda^{T} + \Theta.$$

Toisen kertaluvun faktorimallin käsitettä on edelleen mahdollista laajentaa useamman kertaluvun faktorimalliksi.

2.3.1 Toisen kertaluvun faktorimallin identifioituvuus

Toisen kertaluvun faktorimallin (2.9) identifioituvuudelle pätevät samat säännöt kuin ensimmäisen kertaluvun faktorimalleillekin eli toisen kertaluvun faktorimalli on identifioituva, jos sen parametrit Λ, Ω, Θ ja Γ, Φ, Ψ ovat ratkaistavissa kovarianssimatriisin Σ avulla. Identifioituvuustarkastelut kannattaa kuitenkin suorittaa vaiheittain seuraavasti:

1) tutkitaan ensin ensimmäisen kertaluvun faktorimallin parametrien Λ, Ω ja Θ identifioituvuus kovarianssimatriisin Σ avulla,
2) tämän jälkeen tutkitaan toisen kertaluvun faktorirakenteen parametrien Γ, Φ ja Ψ identifioituvuus Ω:n avulla.

Ensimmäisen kertaluvun faktorimallin identifioituvuustarkastelut tehdään kuten aiemmin, ja kun faktorimallin parametrit Λ, Ω ja Θ on todettu identifioituviksi, voidaan jatkotarkastelut perustaa ensimmäisen kertaluvun faktoreiden kovarianssiesitykseen (2.8) $\Omega = \Gamma\ \Phi\ \Gamma^{T} + \Psi$. Jos toisen kertaluvun faktorirakenteen parametrit Γ, Φ ja Ψ voidaan ratkaista Ω:n avulla, on myös toisen kertaluvun faktorirakenne identifioituva. Tuntemattomien, estimoitavien parametrien lukumäärän maksimi Γ:ssa, Φ:ssä ja Ψ:ssä on $(1/2)\ m\ (m + 1)$, joka on Ω:n määrittämien yhtälöiden lukumäärä. Huomioitavaa on myöskin, että ensimmäisen kertaluvun faktoreita on oltava vähintään kolme, jotta toisen kertaluvun yhden faktorin malli olisi identifioituva.

2.3.2 Toisen kertaluvun faktorimallin rakentaminen

Toisen kertaluvun faktorimallin rakentaminen on vaiheiltaan yhdenmukaista konfirmatorisen faktorimallin rakentamisen kanssa. Myös siinä voidaan erottaa mallin spesifiointi, identifioituvuustarkastelut, parametrien estimointi, hypoteesien testaus ja riittävyystarkastelut. Käytännössä mallin rakentaminen kannattaa suorittaa rakentamalla ensin ensimmäisen kertaluvun faktorimalli. Jos se osoittautuu riittäväksi, voidaan sen jälkeen suorittaa toisen kertaluvun faktoriosan rakentaminen.

Toisen kertaluvun faktorimallin spesifioinnissa pyritään löytämään ensimmäisen kertaluvun faktoreiden kovarianssirakenteelle vähäparametrinen, selkeästi tulkittavissa oleva toisen kertaluvun faktorimalliesitys. Mallin spesifioinnin yhteydessä voidaan tutkia toisen kertaluvun faktoriosan identifioituvuus.

Kun toisen kertaluvun faktorirakenteelle on spesifioitu identifioituva malliesitys, suoritetaan toisen kertaluvun faktorimallin kaikkien parametrien Λ, Θ, Γ, Φ ja Ψ samanaikainen estimointi. Estimaattien $\hat{\Gamma}$, $\hat{\Phi}$ ja $\hat{\Psi}$ avulla saadaan η-faktoreiden kovarianssirakenteelle sovite

$$\hat{\Omega} = \hat{\Gamma}\,\hat{\Phi}\,\hat{\Gamma}^T + \hat{\Psi}\,.$$

Toisen kertaluvun faktoriosan ollessa riittävä, eivät ensimmäisen kertaluvun faktorimallin parametrien Λ ja Θ estimaatit yleensä oleellisesti muutu verrattuna ensimmäisen kertaluvun faktorimallin parametrien estimointituloksiin.

Hypoteesien testaus toisen kertaluvun faktorimalleissa tapahtuu samoin kuin ensimmäisenkin kertaluvun faktorimalleissa. Jos halutaan testata tilastollisesti toisen kertaluvun faktorirakenteen yhteensopivuutta ensimmäisen kertaluvun faktoreiden kovarianssirakenteen kuvaamisessa, on identifioituvia ja estimoitavia parametreja Γ:ssa, Φ:ssä ja Ψ:ssä oltava vähemmän kuin yhtälöitä Ω:ssa eli vähemmän kuin $(1/2)\ m\ (m + 1)$ kappaletta.

Toisen kertaluvun faktorirakenteen riittävyystarkastelut ovat täysin samat kuin konfirmatorisen faktorimallin riittävyystarkastelut.

2.4 Faktoreiden tasovertailumallit

Laajennettaessa konfirmatorisia faktorimalleja tasovertailumalleiksi, otetaan kovarianssimatriisin lisäksi mukaan havaittujen muuttujien odotusarvovektorit ja otoskeskiarvovektorit eri ryhmissä. Yhden ryhmän tapauksessa tutkitaan muutosta mittauskerrasta toiseen. Tarkastelun pohjana on tasoparametrinen faktorimalli

$$\underset{\sim}{y} = \underset{\sim}{\tau} + \Lambda \underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \ ,$$

jossa $E\underset{\sim}{\eta} = \underset{\sim}{\alpha}$ ja $E\underset{\sim}{\varepsilon} = \underset{\sim}{0}$, sekä $\mathrm{cov}(\underset{\sim}{\eta}) = \Omega$ ja $\mathrm{cov}(\underset{\sim}{\varepsilon}) = \Theta$. Silloin havaittujen muuttujien odotusarvovektori on muotoa

$$E\underset{\sim}{y} = \underset{\sim}{\tau} + \Lambda \underset{\sim}{\alpha} \tag{2.10}$$

ja havaittujen muuttujien kovarianssimatriisiesitys on edelleen muotoa

$$\begin{aligned}\Sigma &= E(\underset{\sim}{y} - \underset{\sim}{\mu})(\underset{\sim}{y} - \underset{\sim}{\mu})^T \\ &= \Lambda\, \Omega\, \Lambda^{\mathrm{T}} + \Theta.\end{aligned}$$

Odotusarvovektorin $\underset{\sim}{\mu}$ parametrisoinnissa (2.10) parametrit $\underset{\sim}{\tau}$ kuvaavat havaittujen muuttujien konstruoinnista aiheutunutta mittaamisen yleistasoa ja komponentti $\Lambda\underset{\sim}{\alpha}$ faktorirakenteen aiheuttamaa tasoa. Yleisessä muodossa (2.10) $p\times 1$-odotusarvovektori $\underset{\sim}{\mu}$ sisältää p yleistasoa τ_i, $i = 1, \ldots, p$ ja m faktoreiden odotusarvoa α_j, $j = 1, \ldots, m$, joten τ- ja α-parametreja ei voida ratkaista ilman lisäoletuksia odotusarvovektorin $\underset{\sim}{\mu}$ avulla. Parametrisointi (2.10) ei siten sellaisenaan ole identifioituva. Käytännössä yleisin tapa saada malli identifioituvaksi on esimerkiksi yhden faktorin tilanteessa olettaa, että

$$\tau_1 = \tau_2 \text{ ja } \alpha_1 = 0, \alpha_2 \text{ estimoidaan.}$$

2.4.1 Faktoreiden tasovertailumallien rakentaminen

Käytännössä suositeltavin tapa on rakentaa ensin konfirmatoriset faktorimallit ja vertailla niiden faktorirakenteita. Mielekkäin faktoreiden tasovertailutilanne syntyy tapauksissa, joissa faktoreiden latausrakenteet Λ ja mahdollisesti myös faktoreiden kovarianssimatriisit Ω voidaan olettaa yhtäsuuriksi eri ryhmissä. Faktorit ja niiden väliset suhteet voidaan tulkita silloin samanlaisiksi eri ryhmissä, mikä selkeyttää niiden tasovertailujen sisällöllistä tulkintaa.

Varsinaisessa tasoparametrisessa faktorimallin rakentamisessa on mukana samanaikaisesti havaittujen muuttujien kovarianssimatriisin faktorimalliesitykseen liittyvät Λ, Ω ja Θ –parametrit sekä odotusarvojen tasoparametrisointiin liittyvät $\underset{\sim}{\tau}$- ja $\underset{\sim}{\alpha}$-parametrit. Tällainen tasoparametrinen faktorimalli estimoidaan samanaikaisesti kahdessa tai useammassa ryhmässä ja suoritetaan estimoidun mallin tilastollinen testaus.

Parametrien estimoinnin lähtökohtana on silloin havaitun muuttujajoukon teoreettinen momenttimatriisi ja otosmomenttimatriisi kussakin ryhmässä. Havaittujen muuttujien teoreettinen momenttimatriisi $\mathrm{E}\,\underset{\sim}{y}\,\underset{\sim}{y}^T$ voidaan esittää muodossa

$$\begin{aligned}\mathrm{E}\,\underset{\sim}{y}\,\underset{\sim}{y}^T &= E\left[(\underset{\sim}{y}-\underset{\sim}{\mu})+\underset{\sim}{\mu}\right]\left[(\underset{\sim}{y}-\underset{\sim}{\mu})+\underset{\sim}{\mu}\right]^T \\ &= E\,(\underset{\sim}{y}-\underset{\sim}{\mu})\,(\underset{\sim}{y}-\underset{\sim}{\mu})^T+\underset{\sim}{\mu}\,\underset{\sim}{\mu}^T \\ &= \Sigma+\underset{\sim}{\mu}\,\underset{\sim}{\mu}^T\end{aligned}$$

ja sitä vastaava otosmomenttimatriisi muodossa

$$S+\underset{\sim}{\bar{y}}\,\underset{\sim}{\bar{y}}^T\ ,$$

jossa $\underset{\sim}{\bar{y}}$ on otoskeskiarvovektori.

Voidaan osoittaa, että suurimman uskottavuuden menetelmässä yhden perusjoukon tapauksessa kaikkien estimoitavien parametrien suhteen minimoitava kohdefunktio on muotoa

$$F^* = \ln|\Sigma^*| + tr\,(\Sigma^{*-1}M) - \ln|M| - (p+1)\ ,$$

jossa

$$\Sigma^* = \begin{bmatrix} \Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T & \underset{\sim}{\mu} \\ \underset{\sim}{\mu}^T & 1 \end{bmatrix}$$

on $p + 1$ –vektorin $(\underset{\sim}{y}^T, 1)^T$ teoreettinen momenttimatriisi ja

$$M = \begin{bmatrix} S + \underset{\sim}{\bar{y}}\underset{\sim}{\bar{y}}^T & \underset{\sim}{\bar{y}} \\ \underset{\sim}{\bar{y}}^T & 1 \end{bmatrix}$$

on vastaava otosmomenttimatriisi. Kohdefunktio F* minimoidaan nyt parametrivektoreissa $\underset{\sim}{\tau}$ ja $\underset{\sim}{\alpha}$ sekä parametrivektoreissa Λ, Ω ja Θ olevien estimoitavien parametrien suhteen.

Tasoparametrinen faktoreiden rakenneyhtälömalli on muotoa

$$\begin{cases} \underset{\sim}{y} = \underset{\sim}{\tau} + \Lambda\underset{\sim}{\eta} + \underset{\sim}{\varepsilon} \\ \underset{\sim}{\eta} = \underset{\sim}{\alpha} + B\underset{\sim}{\eta} + \underset{\sim}{\zeta}. \end{cases}$$

Malli tuottaa seuraavat havaittujen muuttujien ja faktoreiden odotusarvojen rakenneparametrisoinnit

$$E\underset{\sim}{y} = \underset{\sim}{\mu} = \underset{\sim}{\tau} + \Lambda E\underset{\sim}{\eta} = \underset{\sim}{\nu} + \Lambda\underset{\sim}{\alpha},$$

jossa

$$E\underset{\sim}{\eta} = \underset{\sim}{\kappa} = (I - B)^{-1}\underset{\sim}{\alpha} \,.$$

Tasoparametristen faktoreiden rakenneyhtälömallien käyttö liittyy yleisesti ryhmävertailutilanteisiin. Kuitenkin esimerkiksi seurantatutkimuksissa faktoreiden tasovertailumallien käyttö on mielekästä myös yhden ryhmän (perusjoukon) tapauksessa tutkittaessa faktoreiden tasomuutoksia mittauskertojen välillä (Leskinen, 1987).

2.5 Reliabiliteetti – ja validiteettianalyysit

Mittausmalli liittää mitattavan käsitteen yhteen tai useampaan latenttiin muuttujaan, jotka taas ovat yhteydessä havaittuihin muuttujiin. Koska latenttia muuttujaa ei voida suoraan havaita, tarvitaan sen mittaamiseen useita havaittuja muuttujia. Tällöin käytetään ns. mittausmallia, joka muodostuu havaituista muuttujista, latentista muuttujasta ja mittausvirheestä (Bollen, 1989).

Konfirmatorinen mittausmalli on muotoa

$$y_i = a_i f + e_i, \qquad i = 1,\ldots, p, \tag{2.11}$$

jossa y_i :t ovat havaittuja muuttujia, f on latentti muuttuja eli faktori, a_i :t ovat faktorilatauksia ja e_i:t ovat havaitun muuttujan mittausvirheitä. Oletuksena on, etteivät latentti muuttuja ja mittausvirhe korreloi ($\rho(f,\ e_i) = 0$), että mittausvirheiden odotusarvo on nolla (E(e_i = 0)) ja etteivät mittausvirheet korreloi keskenään ($\rho(e_i, e_j) = 0$) (Bollen, 1989).

2.5.1 Reliabiliteetti

Reliabiliteetti kuvaa sitä, miten luotettavasti mittari mittaa latenttia muuttujaa. Jos mittarilla on hyvä reliabiliteetti, samalla mittarilla tulisi saada samankaltaisia tuloksia eri mittauksissa. Tutkimuksen sisäinen reliabiliteetti voidaan todeta mittaamalla sama tilastoyksikkö useampaan kertaan. Jos mittaustulokset ovat samat, mittaus on reliaabeli. Tutkimuksen ulkoinen reliabiliteetti tarkoittaa sitä, että mittaukset ovat toistettavissa myös muissa tutkimuksissa ja tilanteissa. Puutteellinen reliabiliteetti johtuu yleensä satunnaisvirheestä.

Reliabiliteetti määritellään havaitun ja latentin muuttujan korrelaation neliönä

$$\mathrm{rel}(y_i) = \rho^2(y_i, f), \qquad i = 1, \ldots, p.$$

Reliabiliteetti voidaan määritellä myös seuraavasti:

$$\mathrm{rel}(y_i) = 1 - \frac{\mathrm{var}(e_i)}{\mathrm{var}(y_i)}, \qquad i = 1, \ldots, p$$

(Liukkonen & Leskinen, 1999).

Reliabiliteettikerroin vaihtelee välillä [0, 1] ja suuret kertoimen arvot kertovat korkeasta reliabiliteetista eli jos havaittu muuttuja mittaa hyvin latenttia muuttujaa, on reliabiliteetti lähellä ykköstä. Suurilla mittausvirheen varriansseilla reliabiliteetti on lähellä nollaa ja jos mittausvirheen varianssi on nolla, on reliabiliteetti tasan 1. Siitä, miten suuri kertoimen pitäisi olla, ei ole olemassa yksiselitteistä rajaa, mutta luku saisi mielellään olla yli .70 (Heikkilä, 1999).

2.5.2 Skaalan reliabiliteetti

Skaalalla S tarkoitetaan osioiden $y_1, \ldots, y_p$ painotettua summaa

$$S = c_1y_1 + c_2y_2 + \ldots + c_py_p, \tag{2.12}$$

jossa c_i:t ovat painokertoimia. Skaalan S reliabiliteettikerroin voidaan määritellä

$$\text{rel}\,(S) = \frac{(\sum_{i=1}^{p} c_i a_i)^2 \operatorname{var}(f)}{(\sum_{i=1}^{p} c_i a_i)^2 \operatorname{var}(f) + \sum_{i=1}^{p} c_i^2 \operatorname{var}(e_i)}$$

(Liukkonen & Leskinen, 1999). Jos kaikki mittausmallin (2.11) lataukset a_i ovat yhtä suuria eli τ-ekvivalentteja ja jos skaalan (2.12) painokertoimet c_i on valittu ykkösiksi, on reliabiliteettikerroin sama kuin Cronbachin α-kerroin:

$$\alpha = \left(\frac{p}{p-1}\right)\left(1 - \frac{\sum_{i=1}^{p} \operatorname{var}(y_i)}{\operatorname{var}(S)}\right)$$

(Bollen, 1989). Jos lataukset a_i eivät ole yhtäsuuria, niin Cronbachin α-kerroin aliestimoi skaalan reliabiliteettia. Cronbachin α-kerroin voidaankin tulkita reliabiliteetin alarajaksi.

Skaala ja sen reliabiliteetti voidaan määritellä konfirmatorisen faktorianalyysin avulla. Lataukset a_i ovat faktorilatauksia ja skaalan painokertoimet c_i ovat faktoripistemäärien painokertoimia. Faktoripistemäärien painokertoimet saadaan kaavasta

$$\hat{c} = \hat{\Omega}\hat{\Lambda}^T\hat{\Sigma}^{-1},$$

jossa $\hat{\Omega}$ on faktoreiden kovarianssimatriisi, $\hat{\Lambda}$ on latausmatriisi ja $\hat{\Sigma}$ on kovarianssimatriisin sovite

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Omega}\hat{\Lambda}^T + \hat{\Theta}$$

(Liukkonen & Leskinen, 1999).

2.5.3 Validiteetti

Validiteettikertoimella tarkoitetaan klassisessa testiteoriassa mittarin ja kriteerimuuttujan välistä korrelaatiokerrointa. Kriteerimuuttujalla tarkoitetaan sellaista muuttujaa, jonka jo ennestään tiedetään mittaavan kyseistä piirrettä hyvin. Kriteerinä voidaan periaatteessa pitää mitä tahansa muuttujaa, jonka valinnalle tutkija voi esittää perusteluja. Validiteetti on siis määritelty niin, että miltei yhtä hyvin voitaisiin puhua yksinkertaisemmin muuttujan korrelaatiokertoimesta ulkopuoliseen muuttujaan (Nummenmaa ym., 1997).

Standardien mukaan mittarin ja mittausten validiteetilla tarkoitetaan niiden päätelmien sopivuutta, mielekkyyttä ja käyttökelpoisuutta, joita mittaustuloksista tehdään (Standards for Educational and Psychological Testing, 1985). Bollen (1989) määrittelee mittarin validiksi, jos se mittaa sitä, mitä sen on oletettukin mittaavan. Perinteisesti validiteetti on jaettu neljään lajiin: sisältövaliditeettiin (*content validity*), kriteerivaliditeettiin (*criterion validity*), rakennevaliditeettiin (*construct validity*) ja konvergentti- ja diskriminanttivaliditeettiin (*convergent and discrimination validity*). Näistä sisältövaliditeetti perustuu testin teoreettiseeen käsitteeseen, jota on vaikea empiirisesti todentaa. Sensijaan muita validiteetin lajeja voidaan analysoida empiirisin menetelmin (Bollen, 1989).

Sisältövaliditeetti kuvaa laadullista validiteettia, jossa tutkija päättää teoreettisen määritelmän pohjalta sen, onko mittarin sisältö tarkoitetun kaltainen. Kriteerivaliditeetti puolestaan kuvaa mittauksen ja kriteerimuuttujan vastaavuutta, jota mitataan yleensä näiden välisellä korrelaatiokertoimella eli validiteettikertoimella:

$$\mathrm{val}(y) = \rho(y, c).$$

Yhden havaitun muuttujan ja yhden latentin muuttujan tapauksessa korrelaatio saadaan kaavasta

$$\rho(y, c) = \frac{a_{11}a_{21}\operatorname{var}(f)}{\sqrt{\operatorname{var}(y_1)\operatorname{var}(c_1)}},$$

jossa y_1 on havaittu muuttuja ja c_1 on kriteerimuuttuja, johon mittaria verrataan; a_{11} ja a_{21} ovat muuttujien latauksia. Validiteettiin ei vaikuta siis pelkästään se, kuinka hyvin havaittu muuttuja mittaa latenttia muuttujaa, vaan myös muuttujien väliset lataukset. Kriteerivaliditeetin määrittelyä vaikeuttaa se, että kriteerimuuttujan valinnalla voidaan vaikuttaa validiteettiin vaikkei mittaria muutettaisikaan, ja kriteerimuuttujan määrittäminen on usein myös hankalaa.

Rakennevaliditeetti osoittaa, kuvaako mittari todella tavoiteltua teoreettista konstruktiota ja sitä käytetään yleensä silloin, kun sisältövaliditeettia ei tunneta tai kun ei olemassa sopivaa kriteerimuuttujaa.

Konvergentti- ja diskriminanttivaliditeetteja käytetään monimutkaisempiin rakenneyhtälömalleihin, joissa esim. useita piirrefaktoreita rakennetaan kahden tai useamman menetelmäfaktorin avulla.

Validiteetin määrittämiseen voidaan käyttää myös menetelmiä, jotka pohjautuvat korrelaatioiden sijasta rakenneyhtälöihin. Standardoimaton validiteettikerroin on havaitun muuttujan ja latentin muuttujan välinen lataus a_i. Standardoitu validiteettikerroin määritellään seuraavasti:

$$a_i^s = a_i\sqrt{\frac{\operatorname{var}(f)}{\operatorname{var}(y_i)}},$$

jossa siis latentin muuttujan ja havaitun muuttujan keskihajonnan suhde kerrotaan niiden välisellä latauksella (Bollen, 1989).

3 Aineiston esittely ja tutkimusongelmat

Tämän tutkielman aineistona on yleisten kielitutkintojen syksyn 1998 englannin kielen keskitason testi. Testissä arvioidaan osallistujan kielitaidon eri osa-alueita ja näiden perusteella kielitaidon yleistasoa.

3.1 Yleiset kielitutkinnot

Yleiset kielitutkinnot ovat aikuisille suunnattuja kielitutkintoja, joihin voi osallistua riippumatta siitä, miten ja missä on kielitaitonsa hankkinut. Ensimmäiset tutkinnot järjestettiin vuonna 1994 ja tähän mennessä osallistujia on ollut yli 11 000 (31.12.1999). Tutkinnon kehittämisestä vastaavat Opetushallitus ja Jyväskylän yliopisto. Tutkinto on kriteeriviitteinen eli osallistujia ei verrata toisiinsa vaan yleisten kielitutkintojen kriteereihin. Käytössä on eurooppalaisen mallin mukaan kehitetty kielitaidon tasoasteikko, jonka mukaan kielitaito jaetaan taitotasoihin lähtien alkeista lähes täydelliseen kielen hallintaan saakka. Tutkinnon käytössä asteikosta ovat taitotasot 1-8.

Testikieliä on yhdeksän: englanti, espanja, italia, ranska, ruotsi, saame, saksa, suomi ja venäjä. Yleisenglannin lisäksi tarjolla on myös kaupallisen ja teknisen englannin erikoistestit. Useimmissa kielissä voi suorittaa joko perus- , keski- tai ylimmän tason testin. Perustason testi on suunnattu taitotasoasteikon tasoille 1-3 eli kielen perustaidot omaaville. Keskitason testi on puolestaan tarkoitettu taitotasoille 3-5 (aiemmin taitotasoille 1-5) eli henkilöille, joiden kielentuntemus ja kielitaito ovat jo melko monipuolisia. Ylimmän tason testi käsittää taitotasot 5-8 ja se sopii henkilöille, jotka käyttävät paljon kieltä, osaavat sopeuttaa kieltään tilanteiden mukaisesti ja ymmärtävät kieltä melko vaivattomasti.

Testit koostuvat viidestä osakokeesta: tekstin ymmärtämisestä, kirjoittamisesta, rakenteista ja sanastosta, puheen ymmärtämisestä ja puhumisesta. Tekstin ymmärtämisen osakokeessa on 3-4 eri aihetta käsittelevää tekstiä, joihin liittyy joko monivalinta-, oikein/väärin- tai avokysymyksiä. Kirjoittamisen kokeessa tehtävänä on kirjoittaa 2-3 erityyppistä tekstiä ohjeiden mukaan. Rakenne ja sanasto –osiossa on 4-6 tehtävää, joissa keskitytään testin vaatimustason mukaisten rakenteiden ja sanaston hallintaan. Tehtävätyyppeinä käytetään mm. monivalintaa ja lyhyitä täydennystehtäviä. Puheen ymmärtämisen osakoe suoritetaan yleensä studiossa. Nauhalta kuullaan lyhyitä keskusteluja, haastatteluja ja katkelmia, joihin liittyy erityyppisiä tehtäviä. Puhumisen osakoe järjestetään yleensä myöskin studiossa, ylimmän tason ja suomen kielen testeissä on lisäksi kasvokkain suoritettava haastattelu. Suoritukset äänitetään ja haastatteluosa videoidaan. Mahdollisia puhumisen tehtävätyyppejä ovat simuloitu keskustelu (toisen puhujan puheenvuorot tulevat nauhalta), kysymyksiin vastaaminen, lauseiden täydentäminen, tilanteissa reagointi ja tapahtumasarjan kertominen tai omien mielipiteiden esittely. Perus- ja keskitason testit kestävät noin kolme tuntia ja ylimmän tason testi noin viisi tuntia.

Suorituksia arvioivat arvostelijoiksi koulutetut kieltenopettajat. Arvioinnin tasapuolisuuden takaamiseksi arvioija ei saa olla testiin osallistujan oma opettaja tai tuttu. Kirjoittamisen ja puhumisen tehtävät arvioidaan suoraan taitotasoille ja muiden osakokeiden vastaukset pisteitetään ja summista määritellään tasoarviot. Kaikkien viiden osakokeen tasoarvioiden perusteella määritellään yleistasoarvio, joka on pyöristetty keskiarvo tasoarvioista (Yleiset kielitutkinnot –esite, 1998). Osalle testattavista tehdään kaksois- tai jopa kolmoisarviointi arvioinnin luotettavuuden takaamiseksi.

3.2 Englannin kielen keskitason syksyn 1998 testi

Englannin keskitason testi on perinteisesti yleisten kielitutkintojen suosituin testi, ja siten osallistujamäärät ovat siinä suurimmat. Syksyllä 1998 englannin keskitason testiin osallistui 251 henkilöä ja heistä 94:lle tehtiin kaksoisarviointi. Keskitason tasoarvioasteikko oli syksyllä 1998 vielä <1, 1-5, mutta koska sen jälkeen siirryttiin keskitasolla asteikkoon <3, 3-5, myöskin syksyn –98 aineiston osalta on tasoarviot <1, 1 ja 2 koodattu tasoarvioksi <3 (analyyseissa tasoarvio <3 on 2).

Tekstin ymmärtämisen osakokeessa oli kolme tehtävää. Taulukossa 3.1 on esitetty tehtävien nimet, tehtävätyypit, osioiden lukumäärät, vastausvaihtoehdot, pisteitys ja tehtävistä saatujen pisteiden minimit, maksimit ja keskiarvot.

Taulukko 3.1 Tekstin ymmärtämisen tehtävien kuvaus

	TEHT. NIMI	TEHT. TYYPPI	OSIO LKM	VAST. MAHD.	PISTEI-TYS	MIN	MAX	K.A
TY_T1	Mainosteksti	Monivalinta	8	A, B, C	0/2	0/16	16/16	13.64
TY_T2	Lehtiartikkeli	Avokysymys	6	-	2*0/1 4*0/1/2	2/10	10/10	8.46
TY_T3	Lehtiartikkeli	O/V -väittämä	10	A, B	0/1	2/10	10/10	8.56

Tekstin ymmärtämisen osakokeen maksimipistemäärä oli 36 pistettä. Alin saatu pistemäärä oli 14 pistettä ja korkein 36 pistettä. Kokonaispisteiden keskiarvo oli 30.55. Tasoarviot olivat määräytyneet kokonaispisteistä seuraavasti:

- tasoarvio 2 = 0 –17 pistettä
- tasoarvio 3 = 18 – 28 pistettä
- tasoarvio 4 = 29 – 34 pistettä
- tasoarvio 5 = 35 – 36 pistettä.

Kirjoittaminen osakokeessa oli kolme kirjoittamistehtävää, joista saatujen tasoarvioiden lähimpään kokonaislukuun pyöristetty keskiarvo oli kirjoittamisen kokonaistaitotasoarvio. Kirjoittamisen tehtävien aiheet olivat postikortti, vastine yleisönosastonkirjoitukselle ja valituskirje. Kirjoittamisen arvioinnissa käytettyjä kriteereitä ovat yleiskuvaus (esim. tuttavallinen/virallinen teksti), viestivyys sekä sanasto ja rakenteet. Kirjoittamisen tasoarvioiden keskiarvo oli 4.04.

Rakenne ja sanasto -osakokeessa oli viisi tehtävää. Taulukossa 3.2 on esitetty tehtävien kuvaus.

Taulukko 3.2 Rakenne ja sanasto -osakokeen tehtävien kuvaus

	TEHT. NIMI	TEHT. TYYPPI	OSIO LKM	VAST. MAHD.	PISTEI-TYS	MIN	MAX	K.A
RS_T7	Sanastotehtävä	Monivalinta	40	A, B, C, D	0/1	13/40	39/40	28.76
RS_T8	Rakennetehtävä	Monivalinta	11	A, B,C, D	0/1	0/11	11/11	8.81
RS_T9	Sanastotehtävä	Aukkotäydennys	9	-	0/1	1/9	9/9	7.35
RS_T10	Rakennetehtävä	Avovastaus	5	-	0/1/2	1/10	10/10	8.33
RS_T11	Rakennetehtävä	Aukkotäydennys	14	-	0/1	0/14	14/14	8.37

Rakenne ja sanasto -osakokeen kokonaispistemäärän maksimi oli 84. Alin saatu pistemäärä oli 24 pistettä ja korkein 83 pistettä. Kokonaispisteiden keskiarvo oli 61.45. Tasoarviot olivat määräytyneet kokonaispisteistä seuraavasti:

- tasoarvio 2 = 0 – 41 pistettä
- tasoarvio 3 = 42 – 57 pistettä
- tasoarvio 4 = 58 – 70 pistettä
- tasoarvio 5 = 71 – 84 pistettä.

Puheen ymmärtämisen osakoe muodostui neljästä tehtävästä. Taulukossa 3.3 on esitetty tehtävien kuvaukset.

Taulukko 3.3 Puheen ymmärtämisen tehtävien kuvaus

	TEHT. NIMI	TEHT. TYYPPI	OSIO LKM	VAST. MAHD.	PISTEI-TYS	MIN	MAX	K.A
PY_T1	Katkelma	Aukkotäydennys	20	-	0/1	1/20	20/20	13.72
PY_T2	Keskustelu-tilanteita	Avovastaus	8	-	4*0/1 4*0/1/2	0/12	12/12	8.75
PY_T3	Kertomus	O/V- väittämä	6	A, B	0/1	0/6	6/6	4.38
PY_T4	Tiedotteita	Monivalinta	5	A, B, C, D	0/1	1/5	5/5	4.48

Puheen ymmärtämisen osakokeen kokonaispistemäärän maksimi oli 43 pistettä. Alin saatu pistemäärä oli 12 pistettä ja korkein 43 pistettä. Kokonaispisteiden keskiarvo oli 31.20. Tasoarviot olivat määräytyneet kokonaispisteistä seuraavasti:

- tasoarvio 2 = 0 – 21 pistettä
- tasoarvio 3 = 22 – 29 pistettä
- tasoarvio 4 = 30 – 36 pistettä
- tasoarvio 5 = 37 – 43 pistettä.

Puhumisen osakokeeseen kuului viisi puhumistilannetta, joista ensimmäistä ei arvioitu lainkaan. Puhumisen tehtävät syksyn 1998 keskitason testissä olivat kysymyksiin vastaaminen, viestin jättäminen, repliikkejä tilanteissa ja lyhyt puheenvuoro. Jokaisessa puhumistilanteessa osallistujan suoritus arvioitiin viiden kriteerin mukaisesti (viestin välittyminen, ääntäminen, sujuvuus, sanasto ja rakenteet). Näiden kriteerien perusteella annettiin kustakin tilanteesta tasoarviot, joiden lähimpään kokonaislukuun pyöristetty keskiarvo oli lopullinen puhumisen kokonaistaitotasoarvio. Puhumisen tasoarvioiden keskiarvo oli 4.22.

3.3 Tutkimusongelmat

Tutkielman ensimmäisenä tutkimuskohteena on se, mittaavatko tehtävät sitä asiaa, mitä niiden pitäisi mitata eli esimerkiksi mittaavatko tekstin ymmärtämisen tehtävät tekstin ymmärtämistä ja onko eri osakokeitten tehtävien välillä spesififaktoreita. Tämän jälkeen tutkitaan, mittaavatko viisi eri osakoetta yleistä kielitaitoa. Tavoitteena on siis selvittää testin rakennevaliditeettia.

Toisena tutkimusongelmana on arvioijien yhdenmukaisuus kirjoittamisen ja puhumisen osakokeissa. Tuottavia taitoja on huomattavasti hankalampi arvioida kuin esimerkiksi monivalintatehtäviä, joissa vastaus on aina joko oikein tai väärin. Tämän vuoksi kirjoittamisen ja puhumisen arvioinneissa yleensä onkin huomattavia eroja ykkös- ja kakkosarvioijien välillä, jolloin joudutaan suorittamaan kolmoisarviointi. Ykkös- ja kakkosarvioijien antamista tasoarvioista tutkitaan, ovatko ykkösarvioijat tiukempia vai löysempiä arvioijia kuin kakkosarvioijat, ja ovatko erot suurempia kirjoittamisessa vai puhumisessa. Lisäksi selvitetään sitä, kumman taidon arviointi on vaikeampaa, kirjoittamisen vai puhumisen.

4 Englannin kielen keskitason testin reliabiliteetti- ja validiteettitarkastelut

Tutkimusaineistona on yleisten kielitutkintojen englannin keskitason testi syksyltä 1998. Osallistujat ja arvioijat ovat täyttäneet lomakkeen, joka on luettu optisesti. Aineisto on analysoitu SPSS 9.0 ohjelmalla (SPSS Base 9.0 Applications Guide, 1999) ja konfirmatorinen faktorianalyysi on toteutettu LISREL8.30 –ohjelmalla (Lisrel 8: New Statistical Features, 1999). Koska mallien rakenteesta on selkeät oletukset, ei faktoreiden lukumäärän arviointia ole tehty eksploratiivisella faktorianalyysilla, vaan mallien toimivuutta on testattu suoraan konfirmatorisella faktorianalyysilla. Estimointimenetelmänä on käytetty suurimman uskottavuuden menetelmää (ML), joka vaatii muuttujien normaalijakautuneisuutta. Kirjoittamisen ja puhumisen tasoarviot eivät täyttäneet oletusta, mutta uudelleenluokittelu (tasoarviot <1, 1 ja 2 koodattiin tasoarvioksi <3) paransi tilannetta. Tekstin ymmärtämisessä, rakenteissa ja sanastossa sekä puheen ymmärtämisessä käytettiin suoria summamuuttujia, jotka eivät myöskään olleet normaalisti jakautuneita. Yleistetty pienimmän neliösumman estimointimenetelmä (GLS) ei vaadi muuttujien normaalisuutta, mutta estimointimenetelmänä käytettiin kuitenkin ML-menetelmää, koska sillä saadut estimaatit poikkesivat erittäin vähän GLS-estimaateista.

Mallien sopivuutta on tarkasteltu χ^2-testin ja riittävyysindeksien avulla. Testin oletukset ovat muuttujien normaalisuus ja riittävä otoskoko, ja sen *p*-arvon tulee olla yli .05, jotta estimoitu malli olisi riittävä aineistoon. Riittävyysindekseistä RMSEA:n (*Root Mean Square Error of Approximation*) arvon täytyisi olla alle .08 (malli kohtalainen) tai alle .05 (malli hyvä). NFI:n (*Normal Fit Index*) ja GFI:n (*Goodness of Fit Index*) arvojen tulisi yli .90, jotta malli olisi riittävä. SRMR:n arvo on noin nolla, kun malli on kunnossa. Mallien vertailuun on käytetty AIC-indeksiä (*Akaiken Information Criterion*). Otoskoon arvon, jolla malli on vielä riittävä, ilmaisee CN:n arvo (*Critical N*).

4.1 Yhden faktorin mallit osakokeille

Tekstin ymmärtämisen, rakenteiden ja sanaston sekä puheen ymmärtämisen osakokeiden reliabiliteettia ja validiteettia tutkittiin tekemällä yhden faktorin mallit kullekin osakokeelle. Tekstin ymmärtämisen muuttujat ovat TY-alkuisia, rakenteiden ja sanaston RS-alkuisia ja puheen ymmärtämisen PY-alkuisia. Osakokeen nimen lisäksi muuttujan nimessä on tehtävänumero, esimerkiksi TY_T1 on tekstin ymmärtämisen ensimmäinen tehtävä.

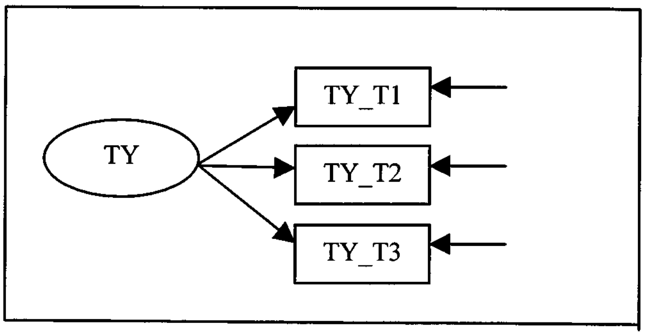
Taulukossa 4.1 on esitetty tehtävien väliset Pearsonin korrelaatiot. Kaikki korrelaatiot ovat positiivisia, mikä onkin luonnollista, sillä kaikki tehtävät mittaavat samaa asiaa, kielitaitoa, eri 'näkökulmista'. Korrelaatiot vaihtelevat välillä .158 - .679 ja ne ovat tilastollisesti vähintään melkein merkitseviä.

Taulukko 4.1 Tehtävien väliset Pearsonin korrelaatiot (*N*=251)

	TY T1	TY T2	TY T3	KI	RS T7	RS T8	RS T9	RS T10	RS T11	PY T1	PY T2	PY T3	PY T4	PU
TY_T1	1													
TY_T2	.419	1												
TY_T3	.578	.522	1											
KI	.327	.519	.473	1										
RS_T7	.416	.546	.554	.566	1									
RS_T8	.280	.424	.470	.563	.556	1								
RS_T9	.427	.531	.540	.450	.669	.463	1							
RS_T10	.190	.391	.288	.629	.484	.519	.416	1						
RS_T11	.338	.559	.491	.670	.679	.580	.538	.584	1					
PY_T1	.395	.396	.487	.545	.579	.536	.513	.504	.567	1				
PY_T2	.357	.412	.463	.439	.415	.386	.440	.325	.373	.589	1			
PY_T3	.289	.230	.341	.208	.262	.212	.331	.158	.269	.331	.354	1		
PY_T4	.327	.294	.411	.259	.281	.227	.306	.164	.235	.351	.351	.292	1	
PU	.339	.501	.447	.669	.622	.546	.552	.628	.646	.626	.490	.235	.304	1

4.1.1 Tekstin ymmärtäminen

Koska tekstin ymmärtämisen osakokeessa on vain kolme tehtävää, yhden faktorin malli on saturoitu. Mallin rakenne on kuviossa 4.1 ja taulukossa 4.2 on mallin *M1a* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja faktoripistemäärien painokertoimet. Kaikki estimoidut parametrit mallissa ovat tilastollisesti merkitseviä.



Kuvio 4.1 Tekstin ymmärtämisen yhden faktorin malli *M1a*

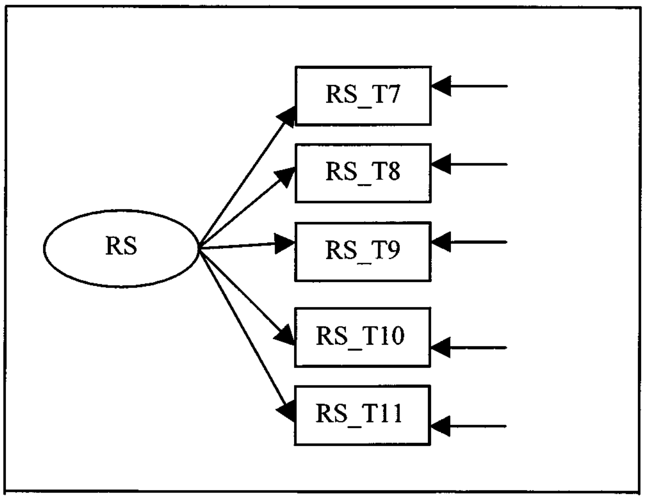
Taulukko 4.2 Tekstin ymmärtämisen mallin *M1a* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja faktoripistemäärien painokertoimet (N = 251).

	λ_y (s.e)	θ_ε (s.e)	R^2	Painokerroin
TY_T1	.68 (.07)	.54 (.07)	.46	.25
TY_T2	.62 (.07)	.62 (.07)	.38	.20
TY_T3	.85 (.07)	.28 (.08)	.72	.60

Validiteetti kuvaa, missä määrin on onnistuttu mittaamaan juuri sitä mitä pitikin mitata (Heikkilä, 1999) ja λ_y:n arvojen mukaan paras validiteetti on viimeisellä tehtävällä. Mittauksen reliabiliteetti kuvaa kykyä tuottaa ei-sattumanvaraisia tuloksia eli toistettaessa testi samoilla henkilöillä ja samoissa olosuhteissa tulosten tulisi olla samoja (Crocker & Algina, 1986). Selkeästi paras reliabiliteetti (.72) on myöskin viimeisellä tehtävällä. Faktoripistemäärien painokertoimista nähdään, että suurin paino on tehtävällä 3 ja pienin tehtävällä 2. Tehtävä kolme on siis selkeästi parempi kuin ensimmäinen ja toinen tehtävä, jotka ovat suunnilleen samantasoisia. Jos lataukset asetetaan yhtä suuriksi, malli *M1b* ei ole riittävä (χ^2 (2) = 6.57, p = .037); tehtävien validiteetit ja reliabiliteetit eivät siis ole yhtä hyviä. Koska malli on saturoitu, tämä tulos on myöskin suoraan mallien *M1b* ja *M1a* peräkkäistestin tulos.

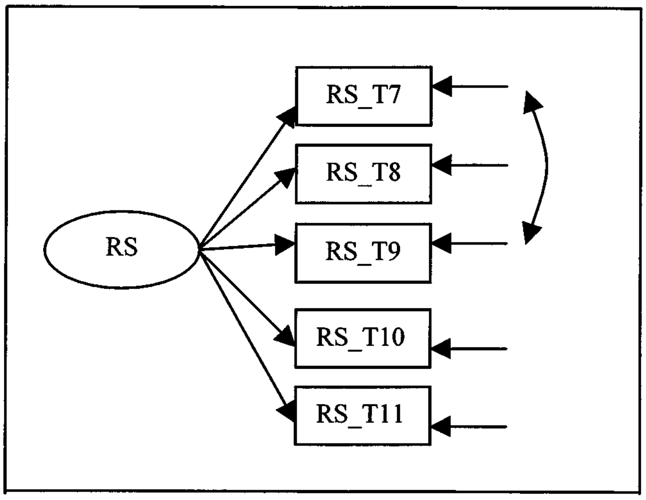
4.1.2 Rakenteet ja sanasto

Rakenne ja sanasto -osakokeessa on viisi tehtävää. Ensimmäinen malli *M2a*, jonka rakenne on esitetty kuviossa 4.2, ei ole riittävä (χ^2 (5) = 28.57, p = .00, RMSEA = .14, NFI = .95, SRMR = .039, GFI = .96, AIC = 48.57, CN = 143.07).



Kuvio 4.2 Hypoteesi rakenteiden ja sanaston yhden faktorin mallista (*M2a*)

Tehtävien 7 ja 9 välinen suuri modifikaatioindeksi (MI(θ_{31}) = 23.19) kertoo näiden jäännösten korreloivan. Tämä on itseasiassa loogista, koska molemmat tehtävät testaavat sanastoa ja kolme muuta tehtävää puolestaan rakenteita. Kun muuttujien RS_T7 ja RS_T9 välinen θ_{31} on estimoitu, on malli *M2b* kaikkien kriteereiden mukaan riittävä. Mallin rakenne on esitetty kuviossa 4.3 ja tulokset on esitetty taulukossa 4.3. Kaikki estimoidut parametrit mallissa ovat tilastollisesti merkitseviä.



Kuvio 4.3 Lopullinen rakenteiden ja sanaston yhden faktorin malli *M2b*

Taulukko 4.3 Rakenne ja sanasto -osakokeen mallin *M2b* estimoidut parametrit (keskivirheet suluissa), selitysasteet, faktoripistemäärien painokertoimet ja riittävyystarkastelut ($N = 251$).

	λ_y **(s.e)**	θ_ε **(s.e)**	R^2	**Painokerroin**
RS_T7	.78 (.06)	.39 (.05)	.61	.24
RS_T8	.71 (.06)	.50 (.05)	.50	.20
RS_T9	.63 (.06)	.60 (.06)	.40	.08
RS_T10	.68 (.06)	.54 (.06)	.46	.17
RS_T11	.85 (.06)	.27 (.05)	.73	.44

$\theta_{31} = .18$ (.04)

χ^2 (4) = 5.58 (p = .23), RMSEA = .040, NFI = .99, SRMR = .018, GFI = .99, AIC = 27.58, CN = 574.56

Paras validiteetti on λ_y:n arvojen mukaan tehtävällä 11. Tällä tehtävällä on myöskin paras reliabiliteetti ja suurin faktoripistemäärän painokerroin. Huonoin validiteetti, reliabiliteetti ja painokerroin puolestaan on tehtävällä 9.

Silmämääräisesti tarkasteltuna tehtävien 11 ja 7 validiteetit vaikuttavat olevan muita tehtäviä parempia. Jos lataukset asetetaan yhtä suuriksi (malli *M2c*), malli ei ole riittävä (χ^2 (8) = 20.32 (p = .009); tehtävien validiteetit eivät siis ole yhtä hyviä. Tämä tulos "tarkistetaan" vielä mallien peräkkäistestillä eli testataan, onko viimeinen malli *M2c* edellisen mallin *M2b* erikoistapaus.

Mallien χ^2-peräkkäistesti:

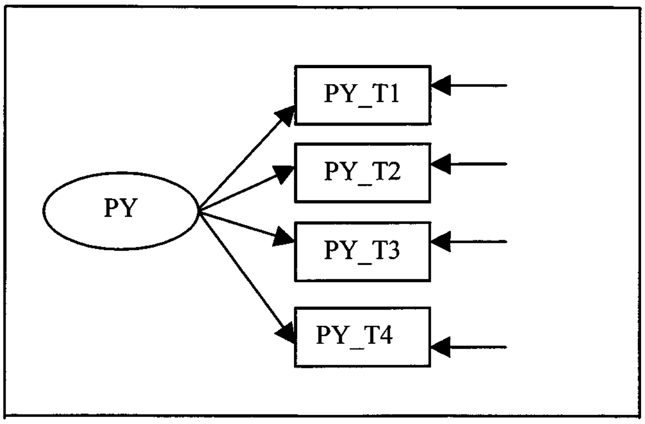
$$D = \chi^2_{M2c} - \chi^2_{M2b} \sim \chi^2\,(df) = 20.32 - 5.58 = 14.74$$

$$df = df_{M2c} - df_{M2b} = 8 - 4 = 4$$

Peräkkäistestin tulokseksi saadaan χ^2 (4) = 14.74, jonka p-arvo on .005. Malli *M2c* ei siis ole mallin *M2b* erikoistapaus, ja tehtävien validiteetit ja reliabiliteetit ovat todellakin erilaiset.

4.1.3 Puheen ymmärtäminen

Puheen ymmärtämisen osakokeeseen kuuluu neljä tehtävää. Kuviossa 4.4 on esitetty yhden faktorin mallin *M3a* rakenne, ja sitä koskevat estimoidut parametrit, selitysasteet, faktoripistemäärien painokertoimet ja riittävyysindeksit esitellään taulukossa 4.4. Kaikki estimoidut parametrit mallissa ovat tilastollisesti merkitseviä.



Kuvio 4.4 Puheen ymmärtämisen yhden faktorin malli *M3a*

Taulukko 4.4 Puheen ymmärtämisen osakokeen mallin *M3a* estimoidut parametrit (keskivirheet suluissa), selitysasteet, faktoripistemäärien painokertoimet ja riittävyystarkastelut ($N = 251$).

	λ_y **(s.e)**	θ_ε **(s.e)**	R^2	**Painokerroin**
PY_T1	.75 (.07)	.43 (.07)	.57	.40
PY_T2	.77 (.07)	.40 (.08)	.60	.44
PY_T3	.47 (.07)	.78 (.08)	.22	.14
PY_T4	.48 (.07)	.77 (.08)	.23	.14

χ^2 (2) = 2.68 (p = .26), RMSEA = .037, NFI = .99, SRMR = .024, GFI = .99, AIC = 18.68, CN = 864.58

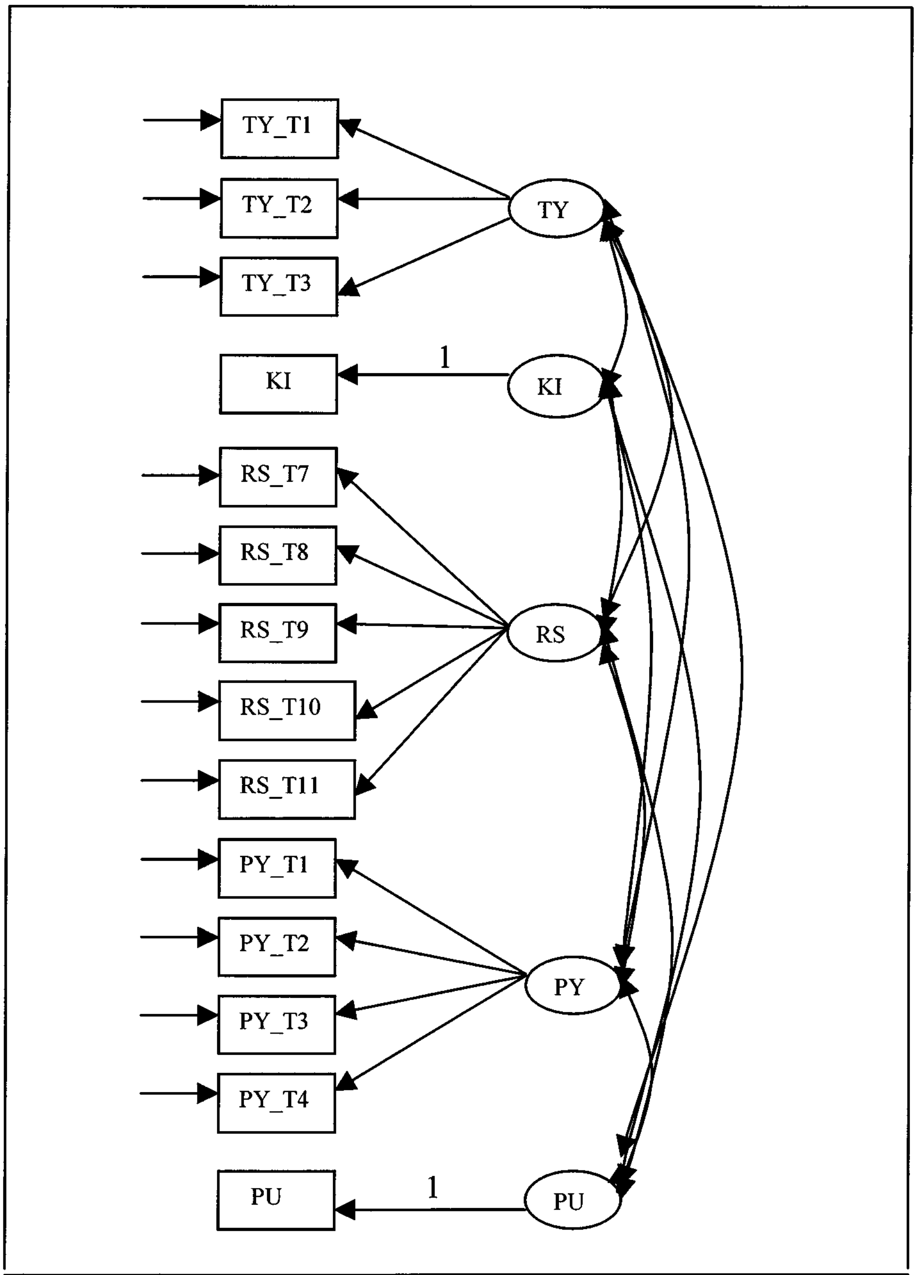
Parhaimmat validiteetit ovat λ_y:n arvojen mukaan tehtävillä 1 ja 2. Näillä tehtävillä ovat myöskin parhaat reliabiliteetit ja faktoripistemäärien painokertoimet. Tehtävät 3 ja 4 ovat puolestaan samalla tasolla ja huomattavasti huonompia kuin tehtävät 1 ja 2. Kuten tehtävien validiteettien suurista eroista voikin päätellä, asetettaessa lataukset yhtä suuriksi (malli *M3b*), malli ei ole riittävä (χ^2 (5) = 24.21, p = .000). Puheen ymmärtämisen tehtävien validiteetit eivät siis ole yhtä hyviä. Tämä tulos todetaan vielä mallien χ^2-peräkkäistestillä:

$$D = \chi^2_{M3b} - \chi^2_{M3b} \sim \chi^2\,(df) = 24.21 - 2.68 = 21.53$$
$$df = df_{M3b} - df_{M3b} = 5 - 2 = 3$$

Peräkkäistestin tulokseksi saadaan χ^2 (3) = 21.53, jonka p-arvo on .000. Malli *M3b* ei siis ole mallin *M3a* erikoistapaus, ja tehtävien validiteetit ja reliabiliteetit eivät ole yhtä hyviä.

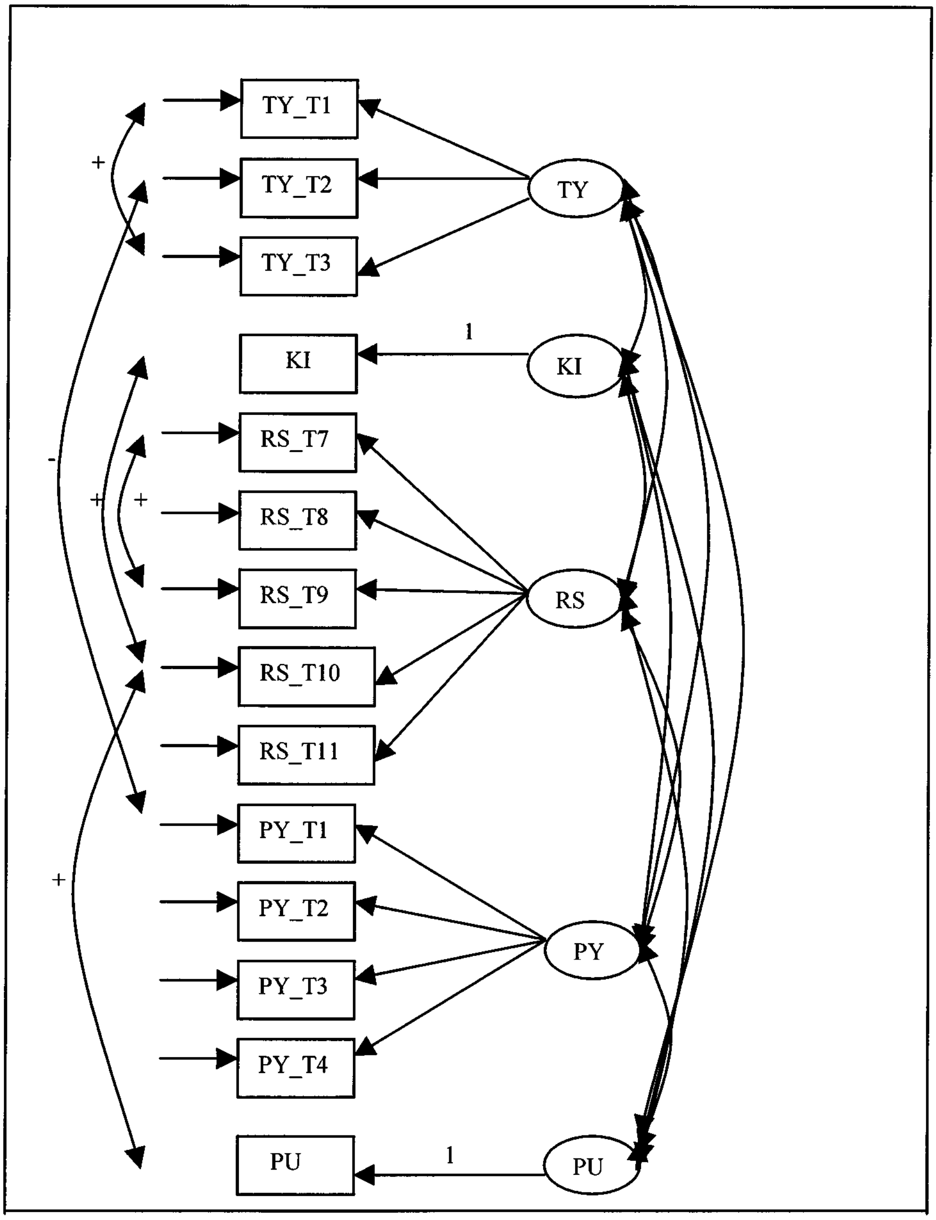
4.2 Viiden faktorin malli englannin kielen testin aineistolle

Viiden faktorin mallin avulla tutkitaan, mittaavatko esimerkiksi tekstin ymmärtämisen tehtävät todellakin vain tekstin ymmärtämistä, vai onko eri osakokeitten tehtävien välillä spesififaktoreita. Kuviossa 4.5 on oletus mallin rakenteesta. Mallin *M4a* mukaan tehtävät mittaavat vain sitä taitoa, jota niiden kuuluukin mitata. Riittävyysindeksien ja χ^2-testin mukaan malli ei kuitenkaan sovi aineistoon: χ^2 (69) = 207.17 (p = .000), RMSEA = .090, NFI = .90, SRMR = .056, GFI = .89, AIC = 279.17, CN = 136.32 (N= 251).



Kuvio 4.5 Hypoteesi viiden faktorin mallista (*M4a*)

Koska ainoastaan NFI:n arvo tukee mallia ja jäännösten modifikaatioindekseistä löytyy suuria arvoja (max MI = MI(θ_{75}) = 18.78), tehtävien välisiä jäännösten kovariansseja on estimoitava. Lopullisen mallin *M4b* rakenne on esitetty kuviossa 4.6. Jäännösten kovariansseja on estimoitu yhteensä viisi.



Kuva 4.6 Lopullinen viiden faktorin malli *M4b*.

Taulukossa 4.5 on esitetty mallia *M4b* koskevat estimoidut parametrit, selitysasteet ja riittävyysindeksit. Kaikki estimoidut parametrit mallissa ovat tilastollisesti merkitseviä.

Taulukko 4.5 Viiden faktorin mallin *M4b* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut ($N = 251$).

			Λ(s.e)				
	TY	**KI**	**RS**	**PY**	**PU**	θ_{ε}**(s.e)**	R^2
TY_T1	.53 (.06)	0*	0*	0*	0*	.72 (.07)	28
TY_T2	.73 (.06)	0*	0*	0*	0*	.47 (.06)	.53
TY_T3	.70 (.06)	0*	0*	0*	0*	.51 (.06)	.49
KI	0*	1	0*	0*	0*	-	1
RS_T7	0*	0*	.81 (.05)	0*	0*	.35 (.04)	.65
RS_T8	0*	0*	.71 (.06)	0*	0*	.50 (.05)	.50
RS_T9	0*	0*	.69 (.06)	0*	0*	.52 (.05)	.48
RS_T10	0*	0*	.65 (.06)	0*	0*	.58 (.06)	.42
RS_T11	0*	0*	.83 (.05)	0*	0*	.31 (.04)	.69
PY_T1	0*	0*	0*	.84 (.06)	0*	.30 (.05)	.70
PY_T2	0*	0*	0*	.69 (.06)	0*	.52 (.05)	.48
PY_T3	0*	0*	0*	.42 (.06)	0*	.83 (.08)	.17
PY_T4	0*	0*	0*	.46 (.06)	0*	.79 (.07)	.21
PU	0*	0*	0*	0*	1	-	1

$\theta_{31} = .21$ (.05), $\theta_{75} = .11$ (.03), $\theta_{84} = .15$ (.04), $\theta_{10,2} = -.14$ (.04), $\theta_{14,8} = .12$ (.04).

* = kiinnitetty

χ^2 (64) = 126.81 (p = .000), RMSEA = .063, NFI = .94, SRMR = .047, GFI = .93, AIC = 208.81, CN = 205.14

AIC-indeksin arvo on pienentynyt verrattuna edelliseen malliin *M4a*, mikä kertoo mallin paranemisesta. Vaikka malli ei χ^2-testin mukaan olekaan vielä riittävä, jäännösten kovariansseja ei estimoitu enempää, koska mallin tulkinta olisi tullut silloin huomattavasti hankalammaksi. Osakokeiden tehtävien validiteetit ja reliabiliteetit ovat hyvin samansuuntaisia kuin osakokeiden yhden faktorin malleissakin.

Kirjoittamisen osakokeen ja rakenne- ja sanastotehtävän 10 jäännösten korrelointi (parametri θ_{84}) voidaan selittää sillä, että molemmat ovat tavallaan tuottamistehtäviä (RS_T10:ssä muodostetaan vastauksia kysymyksiin kahdesta valmiiksi annetusta sanasta). Tehtävän 10 jäännökset korreloivat myöskin toisen tuottamisosakokeen eli puhumisen kanssa (parametri $\theta_{14,8}$). Rakenne- ja sanastotehtävien 7 ja 9 välinen yhteys (parametri θ_{75}) oli havaittavissa jo rakenne- ja sanasto –osion yhden faktorin mallissa, jossa siis sanastotehtävien jäännökset korreloivat. Tekstin ymmärtämisen tehtävien 1 ja 3 jäännösten korreloimiselle ei löydy ilmiselvää selitystä (parametri θ_{31}). Myöskään tekstin ymmärtämisen tehtävän 2 ja puheen ymmärtämisen tehtävän 1 jäännösten negatiivista korreloimista ei voi helposti selittää (parametri $\theta_{10,2}$).

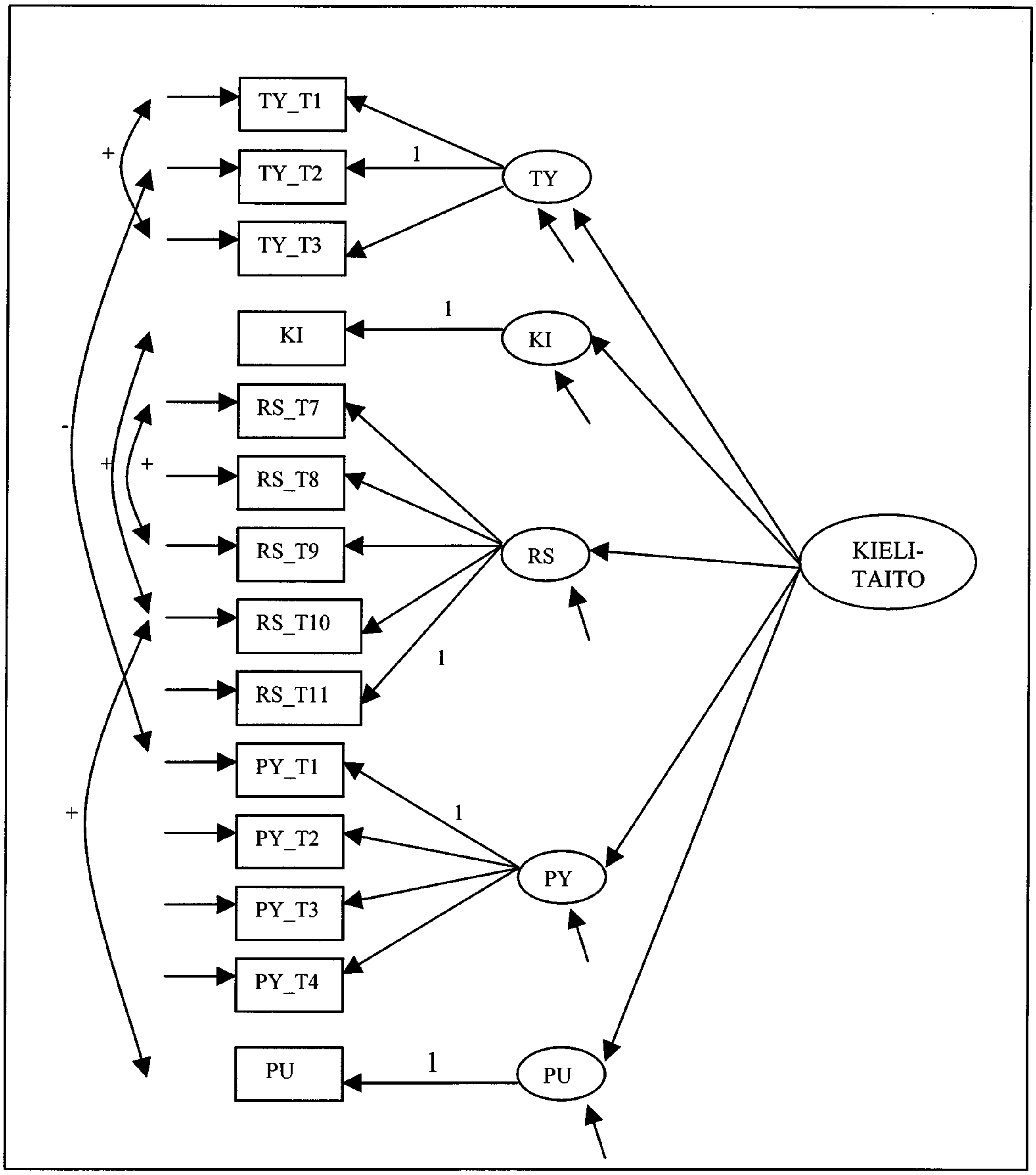
Taulukossa 4.6 on esitetty faktoreiden väliset korrelaatiot, jotka kaikki ovat suuria. Etenkin tekstin ymmärtämisen faktori korreloi voimakkaasti rakenteiden ja sanaston sekä puheen ymmärtämisen faktoreiden kanssa. Osakokeiden suurista korrelaatioista voi päätellä, että osakokeiden taustalla voisi olla yleisen kielitaidon faktori, joten seuraavaksi mallia laajennetaankin toisen kertaluvun faktorimalliksi.

Taulukko 4.6 Faktoreiden väliset korrelaatiot

	TY	KI	RS	PY	PU
TY	1				
KI	.72	1			
RS	.91	.75	1		
PY	.89	.63	.82	1	
PU	.69	.67	.78	.72	1

4.3 Toisen kertaluvun faktorimalli aineistolle

Toisen kertaluvun faktorimallin avulla tutkitaan, mittaavatko eri osakokeet ns. yleistä kielitaitoa vai onko osakokeitten välillä jokin/joitakin muu/muitakin faktori/faktoreita kuin yleinen kielitaito –faktori. Mallin *M5* oletus on esitetty kuviossa 4.7 (mallissa on automaattisesti mukana viiden faktorin mallin jäännösten korrelaatioiden estimaatit). Tekstin ymmärtämisen, rakenteiden ja sanaston sekä puheen ymmärtämisen tehtävien suurin λ_y-lataus on kiinnitetty ykköseksi, jotta malli olisi identifioituva. Malli ei χ^2-testin mukaan ole riittävä, mutta koska muut riittävyysindeksit tukevat mallin hyväksymistä, ei parametreja estimoitu enempää. Estimoidut parametrit keskivirheineen, selitysasteet ja riittävyysindeksit esitetään taulukossa 4.7. Kaikki estimoidut parametrit mallissa ovat tilastollisesti merkitseviä.



Kuvio 4.7 Toisen kertaluvun faktorimallin *M5* rakenne

Taulukko 4.7 Toisen kertaluvun faktorimallin *M5* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 251).

	Λ(s.e)					θ_ε(s.e)	R^2
	TY	**KI**	**RS**	**PY**	**PU**		
TY_T1	.72 (.09)	0*	0*	0*	0*	.71 (.07)	.29
TY_T2	1*	0*	0*	0*	0*	.45 (.06)	.55
TY_T3	.95 (.09)	0*	0*	0*	0*	.50 (.06)	.50
KI	0*	1	0*	0*	0*	-	1
RS_T7	0*	0*	.97 (.07)	0*	0*	.36 (.04)	.64
RS_T8	0*	0*	.85 (.07)	0*	0*	.50 (.05)	.50
RS_T9	0*	0*	.83 (.07)	0*	0*	.52 (.05)	.48
RS_T10	0*	0*	.79 (.07)	0*	0*	.56 (.05)	.43
RS_T11	0*	0*	1*	0*	0*	.31 (.04)	.69
PY_T1	0*	0*	0*	1*	0*	.28 (.05)	.72
PY_T2	0*	0*	0*	.82 (.07)	0*	.52 (.06)	.48
PY_T3	0*	0*	0*	.49 (.08)	0*	.83 (.08)	.17
PY_T4	0*	0*	0*	.53 (.08)	0*	.80 (.07)	.20
PU	0*	0*	0*	0*	1	-	1

θ_{31} = .20 (.05), θ_{75} = .11 (.03), θ_{84} = .13 (.03), $\theta_{10,2}$ = -.11 (.03) ja $\theta_{14,8}$ = .11 (.03).

* = kiinnitetty

	γ(s.e)	ψ(s.e)	R^2
TY	.68 (.06)	.09 (.04)	.84
KI	.78 (.07)	.39 (.05)	.61
RS	.80 (.05)	.05 (.02)	.93
PY	.73 (.06)	.19 (.05)	.73
PU	.81 (.07)	.35 (.06)	.65

χ^2 (69) = 143.93 (p = .000), RMSEA = .066, NFI = .93, SRMR = .051, GFI = .92, AIC = 215.93, CN = 191.11

Eri osakokeet mittaavat siis yleistä kielitaitoa suhteellisen hyvin. Voimakkaimmin yleinen kielitaito –faktorille latautuvat kirjoittamisen, rakenteiden ja sanaston sekä puheen ymmärtämisen faktorit. Tekstin ymmärtämisen ja puhumisen välinen faktoreiden jäännösten kovarianssin modifikaatioindeksi on suurehko (MI(ψ_{51}) = 7.89), ja sen olisi voinut estimoida. Tälle estimoinnille ei kuitenkaan ollut perusteellista syytä (oletusta, joka voisi aiheuttaa jäännösten korreloinnin), joten se jätettiin tekemättä. Testin parhaita tehtäviä ovat olleet sanastotehtävä 7, rakennetehtävä 11 ja puheen ymmärtämisen tehtävä 1. Sen sijaan puheen ymmärtämisen tehtävissä 3 ja 4 on korjaamista. Testin tehtävät mittaavat kuitenkin pääsääntöisesti hyvin kielitaidon eri osa-alueita ja nämä alueet puolestaan mittaavat luotettavasti kielitaitoa. Englannin keskitason syksyn 1998 testin rakenne on siis kunnossa ja testi toimii kuten sen on tarkoituskin.

5 Arvioijien yhdenmukaisuus kirjoittamisen ja puhumisen arvioinnissa

Testiin osallistuneista 251 henkilöstä kaksoisarviointi (A- ja B-arviointi) on tehty 94 osallistujalle. Kaksoisarviointeja tehneitä A-arvioijia on ollut yksitoista ja B-arvioijia viisi ja kaksi A-arvioijaa on myöskin osallistunut B-arviointiin. Yhteensä vertailtavia arvioijia on siis 14 kappaletta. Taulukossa 5.1 on esitetty eri arvioijakombinaatiot; esimerkiksi A-arvioija 1 on arvioinut yhteensä 13 suoritusta, joista B-arvioija 8 on kaksoisarvioinut kuusi ja B-arvioija 12 seitsemän suoritusta.

Taulukko 5.1 A- ja B-arvioijien ristiintaulukko

	B-arvioijan tunnus					
	8	**9**	**12**	**13**	**14**	**Yht.**
A-arvioijan tunnus						
1	6		7			**13**
2				8		**8**
3	11					**11**
4		2				**2**
5	6					**6**
6			2			**2**
7			5			**5**
8				14		**14**
9					15	**15**
10		11				**11**
11		7				**7**
Yht.	**23**	**20**	**14**	**22**	**15**	**94**

Jokainen arvioija on siis arvioinut sekä puhumisen että kirjoittamisen osakokeet. Taulukosta 5.2 nähdään, että A- ja B-arvioitsijoiden antamien kirjoittamisen tasoarvioiden reunajakaumat eivät vaikuta symmetrisiltä: B-arvioijat ovat olleet tiukempia kuin A-arvioijat. Etenkin A:n tasoarvioiden luokassa 4 ja B:n tasoarvioiden luokassa 3 on isoja eroja arvioitsijoiden välillä.

Taulukko 5.2 A- ja B-arvioijan kirjoittamisesta antamien tasoarvioiden ristiintaulukko

	B:n kirjoittamisen tasoarviot				
	<3	3	4	5	**Yht.**
A:n kirjoittamisen tasoarviot					
<3	4	3			**7**
3	1	12	2	6	**15**
4	1	10	25	19	**42**
5		3	8	25	**30**
Yht.	**6**	**28**	**35**	**25**	**94**

Puhumisen jakauma vaikuttaa huomattavastikin symmetrisemmältä kuin kirjoittamisen jakauma (Taulukko 5.3). Ainoastaan B-arvioitsijoiden tasoarvioluokassa 3 hajonta on suurempaa eikä diagonaalilla ole suurinta frekvenssiä. Reunajakaumat ovat lähes identtiset.

Taulukko 5.3 A- ja B-arvioijan puhumisesta antamien tasoarvioiden ristiintaulukko

	B:n puhumisen tasoarviot				
	<3	3	4	5	**Yht.**
A:n puhumisen tasoarviot					
<3	1	1			**2**
3	2	8	5		**15**
4		8	27	7	**42**
5		1	9	25	**35**
Yht.	**3**	**18**	**41**	**32**	**94**

Taulukossa 5.4 on esitelty parittaisten vertailujen tunnuslukuja. Muuttuja A_KI_TASO on A-arvioijan antama kirjoittamisen tasoarvio ja muuttuja B_KI_TASO on vastaavasti B-arvioijan antama kirjoittamisen tasoarvio. Nämä muuttujat ovat kokonaislukuja, jotka arvioijat ovat antaneet kolmen kirjoittamistehtävän perusteella. Muuttuja A_KI on kolmesta A-arvioijan arvostelemasta kirjoittamisen tehtävästä laskettu keskiarvo ja B_KI samalla

tavalla laskettu keskiarvo B-arvioijalta. Muuttuja A_PU_TASO on A-arvioijan antama kokonaistasoarvio puhumisesta ja A_PU puhumisen neljän tehtävän kriteerien perusteella laskettu keskiarvo (yhteensä 4×5 kriteeriä). B_PU_TASO ja B_PU ovat vastaavat muuttujat B-arvioijilta.

Taulukko 5.4 Parittaisten vertailujen tunnuslukuja

		Keskiarvo	**Keskihajonta**
Pari 1	A_KI_TASO	4.01	.89
	B_KI_TASO	3.84	.90
Pari 2	A_KI	4.0408	.6959
	B_KI	3.8422	.7746
Pari 3	A_PU_TASO	4.17	.77
	B_PU_TASO	4.09	.81
Pari 4	A_PU	4.1981	.6188
	B_PU	4.0420	.7104

Sekä puhumisessa että kirjoittamisessa A-arvioijat vaikuttavat hieman löysemmiltä kuin B-arvioijat, sillä heidän antamiensa tasoarvioiden keskiarvot ovat suuremmat kuin B-arvioijilla.

A- ja B-arvioijien antamien kirjoittamisen ja puhumisen tasoarvioiden korrelaatiot ovat tilastollisesti erittäin merkitseviä: kirjoittamisen korrelaatio on .664 ja puhumisen korrelaatio .761.

Parittaisten keskiarvojen vertailun *t*-testin tulos osoittaa, että kirjoittamisen keskiarvot eivät ole samalla tasolla A- ja B-arvioitsijoiden ryhmissä käytettiinpä sitten kokonaistasoarvioita tai tehtävistä laskettua keskiarvoa. Kokonaistasoarvion tapauksessa ero on tilastollisesti melkein merkitsevä (p = .018) ja keskiarvotasoarvion tapauksessa jo merkitsevä (p = .002).

Puhumisessa eroja löytyy käytettäessä kriteerien keskiarvoa, mutta arvioitsijoiden itsensä antamien kokonaistasoarvioiden keskiarvot ovat samalla tasolla (p = .184).

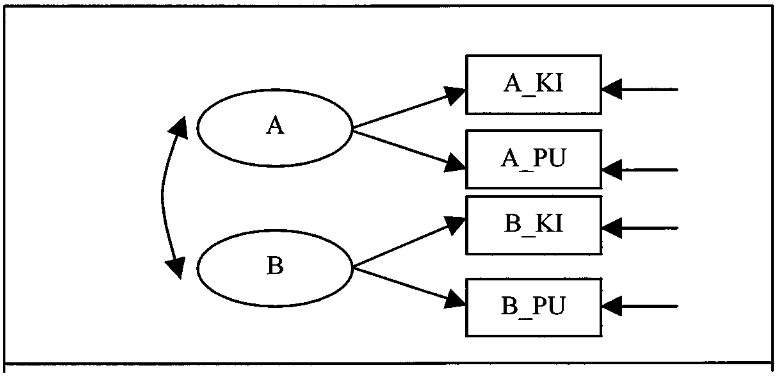
Muuttujien A_PU ja B_PU keskiarvot eroavat toisistaan tilastollisesti merkitsevästi ($p = .002$).

Luokittelijoiden yhdenmukaisuutta mittaava Kappa-kerroin K on kirjoittamiselle .481 ($p = .000$) ja puhumiselle .458 ($p = .000$). Kertoimen arvo vaihtelee välillä 0-1; 1 kertoo luokittelijoiden täydellisestä yksimielisyydestä ja 0 täydellisestä erimielisyydestä. Vaikka arvot ovat suhteellisen huonot (vain noin .5), kertoimen mukaan arvioijat ovat yksimielisiä tilastollisesti erittäin merkitsevästi. Eroja kuitenkin vaikuttaa olevan.

Jatkossa analyyseissa käytetään keskiarvomuuttujia A_KI, A_PU, B_KI ja B_PU, koska erot arvioitsijoiden välillä ovat suuremmat käytettäessä keskiarvoja kuin arvioijien itsensä antamia lopullisia tasoarvioita.

5.1 Arvioijien arviointi

Toisena tutkimusongelmana on selvittää, ovatko A- ja B-arvioijat samantasoisia vai arvioivatko A-arvioijat suorituksia tiukemmin tai lievemmin kuin B-arvioijat. Tutkimusongelmaa on lähdetty ratkaisemaan erilaisilla vaihtoehtoisilla malleilla. Mittausmalli *M6a* on esitetty kuviossa 5.1.



Kuvio 5.1 Arvioijien samantasoisuuden mittausmalli *M6a*

Ensimmäisen vaiheen mallia *M6a* koskevat estimaatit keskivirheineen, selitysasteet ja riittävyystarkastelut ovat taulukossa 5.5. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.5 Mallin *M6a* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut ($N = 94$).

	Λ(s.e)			
	A	**B**	θ_ε(s.e)	R^2
A_KI	.76 (.09)	0*	.43 (.08)	.57
A_PU	.88 (.09)	0*	.23 (.07)	.77
B_KI	0*	.81 (.09)	.35 (.07)	.65
B_PU	0*	.88 (.09)	.23 (.06)	.77

ψ_{21} = .96 (.04) * = kiinnitetty

χ^2 (1) = 8.54 (p = .003), RMSEA = .28, NFI = .96, SRMR = .032,

GFI = .96, AIC = 26.42, CN = 69.92

Malli *M6a* ei kriteereiden mukaan sovi aineistoon. Arvioijien väliset θ_ε :n modifikaatioindeksit ovat suurehkot (8.54), ja siten puhumisen arviointien (muuttujat A_PU ja B_PU) välinen kovarianssi θ_{42} on estimoitu (malli M6b). Yhtä hyvin olisi tosin voitu estimoida kirjoittamisen arviointien (muuttujat A_KI ja B_KI) välinen kovarianssi. Estimaatit keskivirheineen ja selitysasteet on esitetty taulukossa 5.6.

Taulukko 5.6 Mallin *M6b* estimoidut parametrit (keskivirheet suluissa) ja selitysasteet ($N = 94$).

	Λ(s.e)			
	A	**B**	θ_ε**(s.e)**	R^2
A_KI	.83 (.09)	0*	.31 (.08)	.69
A_PU	.80 (.09)	0*	.36 (.08)	.64
B_KI	0*	.89 (.09)	.20 (.08)	.80
B_PU	0*	.79 (.09)	.37 (.08)	.63

θ_{42} = .19 (.06), ψ_{21} = .90 (.05)

* = kiinnitetty

Malli *M6b* on saturoitu eli se sopii aineistoon täydellisesti. Puhumisen validiteettikertoimet ovat laskeneet, mutta se on seurausta puhumisen arvioinnin jäännösten kovarianssin estimoinnista. A- ja B-arvioijan välinen korrelaatio on voimakas (.90), mikä kertoo arvioijien yksimielisyydestä. Tämä korrelaatiokerroin ψ_{21} voidaan tulkita A- ja B-arvioijan väliseksi reliabiliteettikertoimeksi.

Seuraavaksi on kirjoittamisen ja puhumisen lataukset asetettu yhtäsuuriksi, ja siten on testattu, ovatko A- ja B-arvioijien validiteetit samalla tasolla. Tulokset, joissa A_PU ja B_PU muuttujien välistä kovarianssia ei ole estimoitu, ovat taulukossa 5.7 (malli *M6c*). Tulokset, joissa kovarianssin estimointi on suoritettu, on esitetty taulukossa 5.8 (malli *M6d*).

Taulukko 5.7 Mallin *M6c* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)			
	A	**B**	θ_ε**(s.e)**	R^2
A_KI	.78 (.08) $^{1)}$	0*	.42 (.08)	.59
A_PU	.88 (.08) $^{2)}$	0*	.23 (.06)	.77
B_KI	0*	.78 (.08) $^{1)}$	.36 (.07)	.63
B_PU	0*	.88 (.08) $^{2)}$	.23 (.06)	.77

ψ_{21} = .96 (.04)

* = kiinnitetty, 1) ja 2) = asetettu yhtäsuuriksi

χ^2 (3) = 8.71 (p = .033), RMSEA = .14, NFI = .96, SRMR = .038,

GFI = .96, AIC = 22.71, CN = 115.80

Malli *M6c* sopii suhteellisen hyvin aineistoon eli ykkös- ja kaksoisarvioijien kirjoittamisen ja puhumisen arvioinnin validiteetit ovat samalla tasolla. Kirjoittamisen arvioinnin reliabiliteetit sen sijaan ovat puhumisen arvioinnin reliabiliteetteja alhaisemmat.

Taulukko 5.8 Mallin *M6d* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)			
	A	**B**	θ_ε**(s.e)**	R^2
A_KI	.86 (.08) $^{1)}$	0*	.29 (.08)	.72
A_PU	.80 (.08) $^{2)}$	0*	.36 (.08)	.64
B_KI	0*	.86 (.08) $^{1)}$	.23 (.07)	.77
B_PU	0*	.80 (.08) $^{2)}$	.36 (.08)	.64

θ_{42} = .19 (.06), ψ_{21} = .90 (.04)

* = kiinnitetty, 1) ja 2) = asetettu yhtäsuuriksi

χ^2 (2) = 0.39 (p = .82), RMSEA = .00, NFI =1.0, SRMR = .021,

GFI = 1.0, AIC = 16.39, CN = 2200.10

Myös tämä malli *M6d* sopii hyvin aineistoon ja AIC-indeksinkin mukaan malli on parempi kuin aiemmat mallit. Kirjoittamisen arvioinnin validiteetit ovat kasvaneet ja puhumisen arvioinnin puolestaan laskeneet verrattuna malliin *M6c*, mikä johtuu puhumisen arvioinnin jäännösten kovarianssin estimoinnista. Samasta syystä reliabiliteeteissa on tapahtunut suuria muutoksia.

Kahdesta viimeisestä mallista *M6e* ja *M6f* on esitetty vain vaihtoehdot, joissa puhumisen arvioinnin jäännösten välinen kovarianssi on estimoitu. Estimoinnin vaikutus validiteetteihin ja reliabiliteetteihin ei ole enää yhtä voimakas kuin aiemmin, ja siten saadaan lähes samat tulokset ilman jäännösten kovarianssin estimointia kuin estimoinnin kanssakin.

Seuraavaksi kaikki lataukset ja jäännösten varianssit on asetettu yhtä suuriksi (malli *M6e*). Mallia koskevat estimaatit keskivirheineen, selitysasteet ja riittävyystarkastelut ovat taulukossa 5.9. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.9 Mallin *M6e* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut ($N = 94$).

	Λ(s.e)			
	A	**B**	θ_ε**(s.e)**	R^2
A_KI	.83 (.07) $^{1)}$	0*	.31 (.03) $^{(2}$	.69
A_PU	.83 (.07) $^{1)}$	0*	.31 (.03) $^{(2}$	.69
B_KI	0*	.83 (.07) $^{1)}$	.31 (.03) $^{(2}$	.69
B_PU	0*	.83 (.07) $^{1)}$	.31 (.03) $^{(2}$	.69

$\theta_{42} = .14\ (.04)$, $\psi_{21} = .90\ (.05)$

* = kiinnitetty, (1 ja (2 = asetettu yhtäsuuriksi

$\chi^2\,(6) = 1.87\ (p = .93)$, RMSEA = .00, NFI =.99, SRMR = .020,

GFI = .99, AIC = 9.87, CN = 801.91

Malli sopii aineistoon hyvin eli arvioijien validiteetit ja reliabiliteetit ovat hyviä sekä kirjoittamisen että puhumisen osakokeissa. Malleille *M6d* ja *M6e* tehdään vielä χ^2-peräkkäistesti:

$$D = \chi^2_{M6e} - \chi^2_{M6d} \sim \chi^2\,(df) = 1.87 - .39 = 1.48$$
$$df = df_{M6e} - df_{M6d} = 6 - 2 = 4$$

Peräkkäistestin tulokseksi saadaan χ^2 (4) = 1.48, jonka *p*-arvo on .83. Malli *M6e* on siis mallin *M6d* erikoistapaus ja arvioijien validiteetit ja reliabiliteetit ovat siis yhtä hyviä sekä kirjoittamisessa ja puhumisessa.

Viimeisessä mallissa *M6f* on analyysiin otettu mukaan myös keskiarvotarkastelut ja tehty tasovertailu A- ja B-arvioijien välille. A-arvioijan kirjoittamisen arvioinnin keskiarvo on 4.04 ja B-arvioijan 3.84. Puhumisen arvioinnin keskiarvo A-arvioijalla on 4.20 ja B-arvioijalla 4.04. B-arvioija vaikuttaisi siis tiukemmalta sekä kirjoittamisessa että puhumisessa kuin A-arvioija. Mallissa *M6f* on kaikki arvioijien validiteetit ja jäännösten varianssit edelleen asetettu yhtä suuriksi, ja lisäksi kirjoittamisen ja puhumisen keskiarvot on asetettu yhtä suuriksi arvioijien välille. Mallia koskevat estimaatit keskivirheineen, selitysasteet ja riittävyystarkastelut ovat taulukossa 5.10. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.10 Mallin *M6f* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)				
	A	**B**	θ_ε**(s.e)**	τ_y	R^2
A_KI	.59 (.05) 1)	0*	.15 (.02) (2	3.86 (.07)	.69
A_PU	.59 (.05) 1)	0*	.15 (.02) (2	4.04 (.07)	.69
B_KI	0*	.59 (.05) 1)	.15 (.02) (2	3.86 (.07)	.69
B_PU	0*	.59 (.05) 1)	.15 (.02) (2	4.04 (.07)	.69

θ_{42} = .08 (.02), ψ_{21} = .90 (.04), α_1 = .29 (.08)

* = kiinnitetty, (1 ja (2 = asetettu yhtäsuuriksi

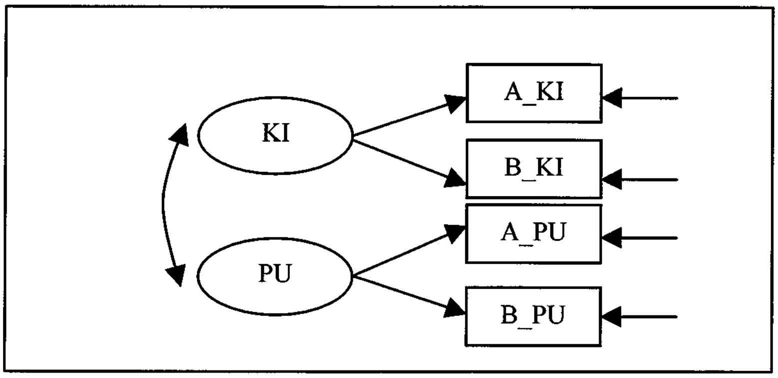
χ^2 (7) = 8.64 (p = .27), RMSEA = .050, NFI =.96, SRMR = .12,

GFI = .96, AIC = 22.64, CN = 185.27

Malli *M6f* on lähes kaikkien kriteereiden mukaan riittävä, ainoastaan SRMR:n ja χ^2-testin p-arvo ovat suurehkot. Estimoitujen parametrien erot verrattuna aiempiin malleihin johtuvat siitä, että aikaisemmin on analysoitu korrelaatiorakennetta, mutta nyt keskiarvotarkasteluissa pohjana on kovarianssimatriisi. Koska parametrin α_1 estimaatti on tilastollisesti merkitsevä, niin A-arvioijien keskiarvot ovat B-arvioijia korkeammat. Kakkosarvioijat ovat siis arvioineet osallistujia tiukemmin kuin ykkösarvioijat.

5.2 Kirjoittamisen ja puhumisen arvioinnin erot

Kolmantena tutkimusongelmana oli selvittää erilaisten vaihtoehtoisten mallien avulla, ovatko puhumisen ja kirjoittamisen mittaamisen validiteetit ja reliabiliteetit yhtä hyviä. Mittausmalli *M7a* on esitetty kuviossa 5.2.



Kuvio 5.2 Kirjoittamisen ja puhumisen arvioinnin malli *M7a*

Mallia *M7a* koskevat estimaatit keskivirheineen, selitysasteet ja riittävyystarkastelut on esitetty taulukossa 5.11. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.11 Mallin *M7a* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)			
	KI	**PU**	**θ_ε(s.e)**	**R^2**
A_KI	.78 (.09)	0*	.39 (.08)	.61
B_KI	.85 (.09)	0*	.28 (.07)	.72
A_PU	0*	.87 (.09)	.25 (.06)	.75
B_PU	0*	.88 (.09)	.23 (.06)	.77

ψ_{21} = .91 (.05)

* = kiinnitetty

χ^2 (1) = 5.74 (p = .017), RMSEA = .23, NFI = .97, SRMR = .023,

GFI = .97, AIC = 229.42, CN = 105.09

Ainoastaan RMSEA:n mukaan malli ei ole sopiva; muut indeksit sen sijaan tukevat mallin *M7a* hyväksymistä. Kirjoittamisen ja puhumisen välinen korrelaatio on voimakas (.91). Paras reliabiliteetti on B-arvioijien arvioimalla puhumisella. Puhumisen validiteetti näyttää hieman paremmalta kuin kirjoittamisen. Puhumisen validiteetit ovat likimain yhtä suuret A- ja B-arvioijien välillä, mutta kirjoittamisen arvioinnissa arvioijien välillä on enemmän eroa. Jäännösten θ_ε-modifikaatioindeksien (MI($\theta_{3,1}$) = MI($\theta_{3,2}$) = MI($\theta_{4,1}$) = MI($\theta_{4,2}$) = 5.74) perusteella joko A- tai B-arvioijien arviointien jäännösten kovarianssit voisi estimoida, mutta koska arvostelijoiden paremmuudesta ei haluta tehdä oletuksia, estimointi jätetään tekemättä.

Seuraavaksi asetetaan kirjoittamisen A- ja B-arvioijien λ_y -lataukset yhtä suuriksi, samoin tehdään puhumiselle eli tutkitaan, ovatko arvioijat yhtä valideja arvioinneissaan (malli *M7b*). Estimoinnin tulokset on esitetty taulukossa 5.12. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.12 Mallin *M7b* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)			
	KI	**PU**	θ_ε**(s.e)**	R^2
A_KI	.82 (.08) $^{1)}$	0*	.37 (.07)	.64
B_KI	.82 (.08) $^{1)}$	0*	.30 (.07)	.69
A_PU	0*	.87 (.07) $^{2)}$	.24 (.06)	.76
B_PU	0*	.87 (.07) $^{2)}$	.24 (.06)	.76

ψ_{21} = .92 (.05)

* = kiinnitetty, 1) ja 2) = asetettu yhtäsuuriksi

χ^2 (3) = 6.12 (p = .11), RMSEA = .11, NFI = .97, SRMR = .031,

GFI = .97, AIC = 20.12, CN = 167.90

Malli on AIC-indeksin mukaan parantunut huomattavasti mallista *M7a*. Tulosten mukaan A- ja B-arvioijat ovat yhtä hyviä arvioimaan kirjoittamista ja puhumista. Puhumisen arviointi vaikuttaa olevan validimpaa kuin kirjoittamisen. Puhumisen reliabiliteetti on arvioijilla samalla tasolla, mutta kirjoittamisessa B-arvioijilla on hieman A-arvioijia parempi reliabiliteetti.

Reliabiliteettien eroja selvitetään asettamalla jäännösten varianssit yhtä suuriksi (malli *M7c*). Estimoinnin tulokset on esitetty taulukossa 5.13. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.13 Mallin *M7c* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut ($N = 94$).

	Λ(s.e)			
	KI	**PU**	**θ_ε(s.e)**	**R^2**
A_KI	.81 (.08) 1)	0*	.34 (.05) 3)	.66
B_KI	.81 (.08) 1)	0*	.34 (.05) 3)	.66
A_PU	0*	.87 (.07) 2)	.24 (.04) 4)	.76
B_PU	0*	.87 (.07) 2)	.24 (.04) 4)	.76

$\psi_{21} = .91$ (.05)

* = kiinnitetty, 1), 2), 3) ja 4) = asetettu yhtäsuuriksi

χ^2 (5) = 6.61 ($p = .25$), RMSEA = .059, NFI = .97, SRMR = .027,

GFI = .97, AIC = 16.61, CN = 206.97

Malli *M7c* sopii kaikkien kriteereiden mukaan aineistoon. Kirjoittamisen ja puhumisen reliabiliteeteissa ei siis ole eroa A- ja B-arvioijien välillä. Puhumisen arvioinnin reliabiliteetti on parempi kuin kirjoittamisen arvioinnin; kirjoittamisen arvioiminen näyttäisi siis olevan vaikeampaa kuin puhumisen arviointi.

Lopuksi asetetaan kaikki lataukset ja jäännösten varianssit yhtä suuriksi (malli *M7d*). Estimoinnin tulokset on esitetty taulukossa 5.14. Kaikki estimoidut parametrit ovat tilastollisesti merkitseviä.

Taulukko 5.14 Mallin *M7d* estimoidut parametrit (keskivirheet suluissa), selitysasteet ja riittävyystarkastelut (N = 94).

	Λ(s.e)			
	KI	**PU**	θ_ε**(s.e)**	R^2
A_KI	.84 (.07) $^{1)}$	0*	.29 (.03) $^{2)}$	.71
B_KI	.84 (.07) $^{1)}$	0*	.29 (.03) $^{2)}$	.71
A_PU	0*	.84 (.07) $^{1)}$	.29 (.03) $^{2)}$	.71
B_PU	0*	.84 (.07) $^{1)}$	.29 (.03) $^{2)}$	.71

ψ_{21} = .91 (.05)

* = kiinnitetty, 1) ja 2) = asetettu yhtäsuuriksi

χ^2 (3) = 9.43 (p = .22), RMSEA = .061, NFI = .96, SRMR = .035,

GFI = .95, AIC = 15.43, CN = 178.61

Estimoitu malli *M7d* sopii aineistoon ja on AIC-kriteerin perusteella paras malli. Tämän mallin mukaan kirjoittamisen ja puhumisen arvioinnin laatu on siis yhtä hyvää sekä A- että B-arvioijien keskuudessa. Malleille *M7c* ja *M7d* tehdään vielä χ^2-peräkkäistesti:

$$D = \chi^2_{M7d} - \chi^2_{M7c} \sim \chi^2\,(df) = 9.43 - 6.61 = 2.82$$
$$df = df_{M7d} - df_{M7c} = 5 - 3 = 2$$

Peräkkäistestin tulokseksi saadaan χ^2 (4) = 2.82, jonka p-arvo on .24. Malli *M7d* on siis mallin *M7c* erikoistapaus.

6 Yhteenveto

Tutkielman tavoitteena oli tutkia yleisten kielitutkintojen englannin kielen keskitason testin reliabiliteettia ja validiteettia konfirmatorisen faktorianalyysin avulla. Aihetta lähestyttiin kolmesta eri näkökulmasta eli tutkittiin mittasivatko tehtävät sitä kielitaidon aluetta mitä niiden piti mitata, olivatko arvioijat yhdenmukaisia arvioidessaan tuottamisen osakokeita (kirjoittamista ja puhumista) sekä oliko kirjoittamisen ja puhumisen arviointi yhtä helppoa.

Tekstin ymmärtämisen, rakenteiden ja sanaston sekä puheen ymmärtämisen osakokeiden reliabiliteettia ja validiteettia tutkittiin tekemällä yhden faktorin mallit kullekin osakokeelle. Tekstin ymmärtämisen kolmannen tehtävän reliabiliteetti ja validiteetti olivat selkeästi kahta muuta tehtävää paremmat. Rakenteet ja sanasto –osiossa parhaiten toimivat tehtävät 7 (sanastotehtävä) ja 11 (rakennetehtävä) ja huonoimmin toimi puolestaan tehtävä 9 (sanastotehtävä). Osakokeen sanastotehtävien (tehtävät 7 ja 9) jäännökset korreloivat odotetusti, joten näiden tehtävien välillä oli spesififaktori. Puheen ymmärtämisen tehtävistä parhaat reliabiliteetit ja validiteetit olivat tehtävillä 1 ja 2. Tehtävät 3 ja 4 olivat puolestaan huomattavasti huonompia.

Viiden faktorin mallin avulla tutkittiin, mittasivatko esimerkiksi tekstin ymmärtämisen tehtävät todellakin vain tekstin ymmärtämistä, vai oliko eri osakokeitten tehtävien välillä spesififaktoreita. Jäännösten kovariansseja estimoitiin yhteensä viisi. Kirjoittamisen osakokeen ja rakenne- ja sanastotehtävän 10 jäännösten korrelointi voidaan selittää sillä, että molemmat ovat tavallaan tuottamistehtäviä (RS_T10:ssä muodostettiin vastauksia kysymyksiin kahdesta valmiiksi annetusta sanasta). Tehtävän 10 jäännökset korreloivat myöskin toisen tuottamisosakokeen eli puhumisen kanssa. Rakenne- ja sanastotehtävien 7 ja 9 välinen yhteys oli havaittavissa jo rakenne- ja sanasto –osion yhden faktorin mallissa, jossa siis sanastotehtävien jäännökset korreloivat. Tekstin ymmärtämisen tehtävien 1 ja 3 jäännösten korreloimiselle ei löytynyt selitystä. Myöskään tekstin ymmärtämisen tehtävän 2 ja puheen ymmärtämisen tehtävän 1 jäännösten negatiivista korreloimista ei voitu selittää.

Koska osakokeiden faktoreiden väliset korrelaatiot olivat suuria (.63 - .91), voitiin päätellä, että osakokeiden taustalla voisi olla yleisen kielitaidon faktori. Mallia laajennettiin toisen kertaluvun faktorimalliksi, jonka avulla tutkittiin, mittasivatko eri osakokeet yleistä kielitaitoa vai oliko osakokeiden välillä jokin/joitakin muu/muitakin faktori/faktoreita kuin yleinen kielitaito –faktori. Eri osakokeet mittasivat yleistä kielitaitoa suhteellisen hyvin. Voimakkaimmin yleinen kielitaito –faktorille latautuivat kirjoittamisen, rakenteiden ja sanaston sekä puheen ymmärtämisen faktorit. Testin parhaita tehtäviä olivat olleet sanastotehtävä 7, rakennetehtävä 11 ja puheen ymmärtämisen tehtävä 1. Sen sijaan puheen ymmärtämisen tehtävät 3 ja 4 olivat heikkoja. Testin tehtävät mittasivat kuitenkin pääsääntöisesti hyvin kielitaidon eri osa-alueita ja nämä alueet puolestaan mittasivat luotettavasti kielitaitoa.

Toisena tutkimusongelmana oli selvittää, olivatko A- ja B-arvioijat samantasoisia vai arvioivatko A-arvioijat kirjoittamisen ja puhumisen suorituksia tiukemmin tai lievemmin kuin B-arvioijat. Jo eksploratiiviset analyysit antoivat vihjeitä siitä, että A-arvioijat olisivat arvioineet suorituksia lievemmin. Konfirmatorinen faktorianalyysi vahvisti tämän; A-arvioijien antamien tasoarvioiden keskiarvot olivat todellakin B-arvioijien keskiarvoja korkeammat.

Kolmantena tutkimusongelmana oli selvittää, olivatko puhumisen ja kirjoittamisen mittaamisen validiteetit ja reliabiliteetit yhtä hyviä. Kirjoittamisen ja puhumisen reliabiliteeteissä ei ollut eroa A- ja B-arvioijien välillä. Puhumisen arvioinnin reliabiliteetti oli kuitenkin parempi kuin kirjoittamisen arvioinnin eli kirjoittamisen arvioiminen siis oli vaikeampaa kuin puhumisen arviointi. Puhumisen arviointi vaikutti olevan myös validimpaa kuin kirjoittamisen.

Tulosten mukaan A- ja B-arvioijat olivat siis yhtä hyviä arvioimaan tuottamisia ja kirjoittamisen sekä puhumisen arvioinnin laatu oli yhtä hyvää sekä A- että B-arvioijien keskuudessa.

Lähteet

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 88, 588-606.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Florida: Holt, Rinehart and Winston. Inc.

Heikkilä, T. (1999). *Tilastollinen tutkimus.* Helsinki: Edita

Jöreskog, K.G. (1969). A general approach to confirmatory factor analysis. *Psychometrica 34,* 183-202.

Jöreskog, K.G. (1981). Basic issues in the application of LISREL. *Data1,* 1-6.

Jöreskog, K.G. & Sörbom. D. (1979). *Advances in Factor Analysis and Structural Equation Models.* Cambridge, Mass: Abt Books.

Jöreskog, K.G. & Sörbom. D. (1981). *LISREL V. Analysis of linear relationships by maximum likelihood and least squares methods.* University of Uppsala, Department of Statistics. Research Report 81-8.

Jöreskog, K.G. & Sörbom. D. & du Toit, S. & du Toit M. (1999). *Lisrel 8: New Statistical Features.* Chicago: Scientific Software International Inc.

Leskinen, E. (1983a). *Identifioituvuustarkasteluja LISREL –faktorimallien rakentamisessa.* Jyväskylän yliopisto, tilastotieteen laitos. Julkaisuja 8/1983.

Leskinen, E. (1983b). *On explaining covariance structures of factors: examples of second-order and structural factor analysis models*. University of Jyväsylä, Department of statistics. Reports on Statistics 8/1983.

Leskinen, E. (1987). *Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen*. Jyväskylän yliopisto, tilastotieteen laitos. Julkaisuja 1/1987.

Liukkonen, J. & Leskinen, E. (1999). The reliability and validity of scores from the children's version of the perception of success questinnaire. *Educational and Psychological Measurement, 59(4)*. 4, 651-664.

McDonald, R.P. & Krane, W.R. (1977). A note on local identifiability and degrees of freedom in the asymtotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology 30*, 198-203.

Nummenmaa, T. & Konttinen, R. & Kuusinen, J. & Leskinen, E. (1997). *Tutkimusaineiston analyysi*. Porvoo: WSOY.

SPSS Base 9.0 Applications Guide. (1999). USA. SPSS Inc.

Standards for Educational and Psychological Testing. (1985). Washington, DC: American Psychological Association.

Yleiset kielitutkinnot –esite, 1998. Opetushallitus. Syrinx Oy / Libris Oy 1998.