Konfirmatorisista faktorimalleista ja rakenneyhtälömalleista ja niiden soveltamisesta 30 vuoden pitkittäistutkimusaineistoon

Sami Saarikivi

Pro gradu -tutkielma

16. marraskuuta 2000

Jyväskylän yliopisto

tilastotieteen laitos

Sisältö

1 Johdanto **3**

2 Konfirmatorisesta faktorianalyysistä **5**

2.1 Konfirmatorisen faktorimallin rakentamisen vaiheet 6

2.1.1 Mallin spesifiointi 7

2.1.2 Identifioituvuustarkastelut 7

2.1.3 Parametrien estimointi 10

2.1.4 Hypoteesien testaaminen 12

2.1.5 Mallin riittävyystarkastelut 14

2.2 Faktoripistemäärien estimointi 18

2.3 Toisen kertaluvun faktorimalleista 19

3 Rakenneyhtälömalleista **23**

3.1 Rekursiiviset ja simultaaniset rakenneyhtälömallit 23

3.2 Rekursiivisten ja simultaanisten rakenneyhtälömallin rakentaminen 25

3.2.1 Mallin spesifiointi ja identifioituvuustarkastelut 25

3.2.2 Parametrien estimointi ja hypoteesien testaus 26

3.2.3 Mallin riittävyystarkastelut 27

4 Suhtautuminen koulutukseen - pitkittäistutkimusaineisto **29**

4.1 Tutkimusaineisto 29

4.2 Tutkimusongelmat 31

5 Suhtautuminen koulutukseen -faktorimallin rakentaminen 32

5.1 Aineiston muuttujat ja oletusten esitarkastelu 32

5.2 Eksploratiivinen faktorianalyysi 34

5.3 Konfirmatorinen faktorianalyysi 36

5.4 Faktoripistemäärämuuttujien estimointi 39

6 Suhtautuminen koulutukseen -rakenneyhtälömallin rakentaminen 41

6.1 Muuttujat ja oletusten tarkastelu 41

6.2 Rekursiivisen rakenneyhtälömallin rakentaminen 43

7 Suhtautuminen koulutukseen -toisen kertaluvun faktorimalli ja rakenneyhtälömalli 47

7.1 Toisen kertaluvun faktorimallin sovellus 47

7.2 Faktoripistemäärämuuttujan estimointi ja rekursiivinen rakenneyhtälömalli 49

8 Yhteenveto 53

Lähteet 55

Liitteet 58

1 Johdanto

Tutkielman on tarkoitus olla esimerkkikuvaus kronologisesti etenevästä tilastotieteellisestä tutkimuksesta, joka suoritetaan konfirmatorisen faktorianalyysin ja rakenneyhtälömallien avulla. Lähtökohtatilanteessa käytössä on analysoitava tutkimusaineisto ja siihen liittyvät tutkimusongelmat, joista on tarkoitus saada tilastollisesti kelvollisia tutkimustuloksia.

Tutkielman alkuosa muodostuu teoriasta. Ensimmäisessä teoriakappaleessa käsitellään konfirmatorista faktorianalyysiä. Siitä esitellään yleisesti konfirmatorisen faktorimallin ominaisuuksia sekä keskitytään teoreettisesti tarkemmin konfirmatorisen faktorimallin rakentamisen vaiheisiin. Ensimmäisen teoriakappaleen loppuosa käsittelee faktoripistemäärien ja faktoripistemäärämuuttujien estimointia, joka liittää teoreettisesti kahden laajemman kokonaisuuden, konfirmatoristen faktorimallien ja rakenneyhtälömallien rakentamisen. Kappaleessa konfirmatorista faktorianalyysiä laajennetaan vielä toisen kertaluvun faktorimalleihin.

Teoriaosan toinen kappale käsittelee siis rakenneyhtälömalleja. Kappaleessa tutustutaan lähemmin rekursiivisten ja simultaanisten rakenneyhtälömallien pääperiaatteisiin sekä rakenneyhtälömallien rakennusvaiheisiin.

Tutkielman loppuosassa sovelletaan edellä mainittuja malleja empiiriseen tutkimusaineistoon. Havainnollistavana aineistona käytetään 30 vuoden pitkittäistutkimusaineistoa, jossa tutkitaan lahjakkuuden, kotitaustan ja koulumenestyksen yhteyksiä koulutus-, ammatti- ja työuraan ja muihin elämänrakenteen ja elämänhistorian merkittäviin tekijöihin lapsuudesta aikuisuuteen.

Empiirisen osan alussa valitaan aineistosta 11 muuttujaa, jotka kuvaavat suhtautumista nykyiseen tai viimeksi käytyyn kouluun aikuisiässä. Alustavan eksploratiivisen faktorianalyysin jälkeen muuttujajoukolle suoritetaan konfirmatorinen faktorianalyysi. Saadut tulokset talletetaan faktoripistemäärämuuttujiin, joiden avulla analyysiä jatketaan vielä rekursiivisiin rakenneyhtälömalleihin. Empiirisen osan lopussa esitetään vielä lyhyesti vaihtoehtoinen analysointimenetelmä, jossa sovelletaan toisen kertaluvun faktorimallia. Siinä muodostetaan kokonaisvaltaisempi faktoripistemäärämuuttuja, jota käytetään edelleen rekursiivisen rakenneyhtälömallin rakentamisessa.

Tutkielman teoriaosa perustuu pääpiirteittäin lähdeteoksiin Leskinen, E. (1987). *Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1/1987* sekä teokseen Nummenmaa, T., Konttinen, R., Kuusinen, J. & Leskinen, E. (1997) *Tutkimusaineiston analyysi. Porvoo: WSOY.*

2 Konfirmatorisesta faktorianalyysistä

Konfirmatorisen faktorianalyysin perusmalli(Jöreskog & Sörbom 1979) on matriisimuotoa

$$y = \Lambda\eta + \epsilon, \tag{2.1}$$

jossa vektori y sisältää *p x 1* kappaletta havaittuja muuttujia ja η-vektori *k x 1* kappaletta latentteja muuttujia eli faktoreita. Siten on oltava voimassa k<p. Täten latausmatriisiksi Λ muodostuu *p x k* -dimensioinen matriisi. Mallin jäännösvektori ϵ sisältää *p x 1* kappalettta satunnaismuuttujia, jotka muodostuvat muuttujien mittausvirheistä ja spesifisistä faktoreista(Harman 1967).

Merkitään jatkossa faktoreiden kovarianssimatriisia (*k x k*) Ω:lla:

$$cov(\eta) = E(\eta\eta^T) = \Omega. \tag{2.2}$$

Lisäksi merkitään mittausvirheiden kovarianssimatriisia (*p x p*) Θ:lla, joka faktorimallin perustilanteessa on aina diagonaalinen:

$$cov(\epsilon) = E(\epsilon\epsilon^T) = \Theta. \tag{2.3}$$

Konfirmatorinen faktorimalli on tilastollinen malli, jossa faktoreiden latausmatriisiin Λ tehdään rajoituksia. Toisin kuin eksploratiivisessa faktorimallissa, osa latauksista λ_{ij} kiinnitetään vakioiksi. Yleisimpiä kiinnityksiä ovat 0 tai 1, jotka asetetaan aikaisemmin, esimerkiksi eksploratiivisen faktorianalyysin tuottamien faktorirakenteiden perusteella. Latausten lisäksi konfirmatorisessa faktorinalyysissä rajoituksia voidaan tehdä myös faktoreiden kovarianssimatriisiin Ω ja parametrien välille , esimerkiksi asettamalla yhtäsuuruusrajoite. Asettamalla kovarianssimatriisin Ω varianssit ykkösiksi, muodostuu matsiisista Ω faktoreiden korrelaatiomatriisi. Konfirmatorisessa faktorimallissa jäännökset voivat tietyissä olosuhteissa korreloida, jolloin jäännösten kovarianssimatriisi ei ole välttämättä diagonaalinen.

Konfirmatorisen faktorianalyysin perusmallissa oletetaan, että mittausvirheiden ja latenttien muuttujien odotusarvot ovat nollia:

$$E(\eta) = E(\epsilon) = 0. \tag{2.4}$$

Lisäksi faktoreiden ja mittausvirheiden väliset kovarianssit oletaan korreloimattomiksi:

$$E(\eta\epsilon) = 0. \tag{2.5}$$

2.1 Konfirmatorisen faktorimallin rakentamisen vaiheet

Konfirmatorisen faktorimallin rakentaminen voidaan jakaa viiteen selkeästi erotettavaan tutkimusvaiheeseen. Vaiheiden suoritusjärjestyksellä on selvä merkitys. Jonkin tutkimusvaiheen epäonnistunut toteuttaminen palauttaa mallin rakentamisen edeltävälle tasolle. Konfirmatorisen faktorimallin rakentamisen vaiheet suoritusjärjestyksessä ovat

1) mallin spesifiointi

2) identitifioituvuustarkastelut

3) parametrien estimointi

4) hypoteesien testaus

5) riittävyystarkastelut

2.1.1 Mallin spesifiointi

Konfirmatorisen faktorimallin rakentaminen aloitetaan aina mallin spesifioinnilla. Malli pyritään tässä vaiheessa täsmentämään tutkimushypoteesin mukaisesti. Tämä tapahtuu antamalla matriisien Λ, Ω ja Θ parametreille jokin tarkoituksenmukainen tehtävä. Parametri voidaan esimerkiksi jättää estimoitavaksi ilman mitään rajoituksia, estimoida yhtäsuurena jonkin toisen parametrin kanssa tai kiinnittää vakioksi. Valintaan vaikuttavia tekijöitä ovat muuttujien ja faktoreiden väliset yhteydet sekä faktoreiden väliset yhteydet.

2.1.2 Identifioituvuustarkastelut

Konfirmatorisen faktorimallin rakentamisen onnistumisen kannalta identifioituvuustarkasteluilla on hyvinkin tärkeä asema. Identifioituvuus eli yksilöityvyys on faktorimallin teoreettinen ominaisuus, joka on riippumaton estimointimenetelmästä ja tutkittavasta aineistosta. Identifioituvuustarkastelut kohdistuvat mallin parametrisoinnin tuottamaan kovarianssirakenteeseen Σ. Perusmallin(2.1) perusteella havaittujen muuttujien teoreettinen kovarianssimatriisi on muotoa

$$\Sigma = \Lambda\Omega\Lambda^T + \Theta. \tag{2.6}$$

Identifioituvuuden tarkastelulle löytyy useita teoreettisia vaihtoehtoja, mutta käytännön sovelluksia varten on identifioituvuudelle käyttökelpoinen määritelmä (Jöreskog 1981): *Mallin yksittäinen parametri on identifioituva, jos se on ratkaistavissa kovarianssimatriisin* Σ *avulla. Jos mallin kaikki parametrit ovat ratkaistavissa* Σ*:n avulla, on malli identifioituva.*

Jotta parametrit matriiseissa Λ, Ω ja Θ voitaisiin ratkaista kovarianssimatriisin avulla, täytyy mallin täyttää identifioituvuudelle välttämättömät ehdot. Ensinnäkin ehdon (2.7) pitää olla voimassa:

$$[pk + (1/2)k(k+1) + p] - r_1 \leq (1/2)p(p+1), \tag{2.7}$$

jossa pk on parametrien maksimimäärä latausmatriisissa Λ, $(1/2)k(k+1)$ parametrien maksimimäärä faktoreiden kovarianssimatriisissa Ω sekä p parametrien määrä mittausvirheiden kovarianssimatriisissa Θ. $(1/2)p(p+1)$ on yhtälöiden lukumäärä teoreettisessa kovarianssimatriisissa Σ ja r_1 parametrien rajoitusten lukumäärä yhteensä.

Pelkkä ehdon (2.7) toteutuminen ei takaa mallin identifioituvuutta, vaan mallin täytyy täyttää myös rotaatioyksikäsitteisyysehto(2.8). Sen mukaan malliin on asetettava rajoituksia r_2 vähintään k^2 kappaletta matriiseihin Λ ja Ω:

$$r_2 \geq k^2 \tag{2.8}$$

Yhdistämällä vielä ehdot (2.7) ja (2.8) saadaan faktorimallin rajoitusten lukumäärälle r^* välttämätön ehto

$$r^* \geq max(r_1, r_2) \tag{2.9}$$

Määritelmän (Jöreskog 1981) käyttäminen identifioituvuuden tarkistamiseksi ei aina ole välttämättä tarpeen. Identifioituvuusongelman voi ohittaa mallin spesifiointivaiheessa määrittämällä yksinkertaisen latausrakenteen omaavan mittausmallin. Se perustuu Thurstonen (1935) esittämään yksinkertaiseen faktorimalliin, jossa kukin havaittu muuttuja mittaa vain yhtä latenttia muuttujaa. Kun mallissa asetetaan vielä faktoreiden varianssit ykkösiksi, on faktorimalli aina identifioituva. Vaihtoehtoisesti voidaan jokaiselta faktorilta kiinnittää yksi lataus ykköseksi ja jättää faktoreiden varianssit vapaasti estimoitaviksi.

Havainnollistetaan esimerkin avulla, kuinka määritelmän (Jöreskog 1981) käyttö identifioituvuuden määrittämiseksi tapahtuu käytännössä. Valitaan tarkasteltavaksi esimerkiksi neljä y-muuttujaa ja kaksi η-faktoria(ks. taulukko 5.3, muuttujat TARPEEN, JORIIT, RASKVAI ja HELPPOA, faktorit TARPEELL ja HELPPOUS), jossa kiinnitykset on tehty seuraavasti:

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \end{bmatrix}, \Omega = \begin{bmatrix} 1 & \\ w_{21} & 1 \end{bmatrix}, \Theta = diag(\theta_1 \theta_2 \theta_3 \theta_4).$$

Siten saadaan teoreettiselle kovarianssimatriisille Σ esitys

$$\Sigma = \begin{bmatrix} \sigma_{11} & & & \\ \sigma_{21} & \sigma_{22} & & \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} = \begin{bmatrix} \lambda_{11}^2 + \theta_1 & & & \\ \lambda_{11}\lambda_{21} & \lambda_{21}^2 + \theta_2 & & \\ w_{21}\lambda_{11}\lambda_{32} & w_{21}\lambda_{21}\lambda_{32} & \lambda_{32}^2 + \theta_3 & \\ w_{21}\lambda_{11}\lambda_{42} & w_{21}\lambda_{21}\lambda_{42} & \lambda_{32}\lambda_{42} & \lambda_{42}^2 + \theta_4 \end{bmatrix}$$

Välttämättömät ehdot identifioituvuudelle (2.7 ja 2.8) ovat tässä esimerkkimallissa voimassa. Ratkaistaan ensimmäiseksi korrelaatio w_{21} kovarianssimatriisista Σ kovarianssien σ_{31} , σ_{42} , σ_{21} ja σ_{43} avulla:

$$\frac{\sigma_{31}\sigma_{42}}{\sigma_{21}\sigma_{43}} = \frac{w_{21}\lambda_{11}\lambda_{32}w_{21}\lambda_{21}\lambda_{42}}{\lambda_{11}\lambda_{21}\lambda_{32}\lambda_{42}} = w_{21}^2,$$

josta saadaan

$$w_{21} = \pm\sqrt{\frac{\sigma_{31}\sigma_{42}}{\sigma_{21}\sigma_{43}}}.$$

Siten faktoreiden välinen korrelaatio w_{21} on saatu ratkaistua teoreettisen kovarianssimatriisin Σ avulla ja on täten identifioituva. Seuraavaksi ratkaistaan lataukset λ_{11} ja λ_{21} käyttämällä apuna kovariansseja σ_{31} , σ_{32} ja σ_{21} :

$$\frac{\sigma_{31}}{\sigma_{32}} = \frac{w_{21}\lambda_{11}\lambda_{32}}{w_{21}\lambda_{21}\lambda_{32}} = \frac{\lambda_{11}}{\lambda_{21}},$$

$$\sigma_{21} = \lambda_{11}\lambda_{21},$$

joista saadaan ratkaistua

$$\lambda_{11} = \frac{\lambda_{21}\sigma_{31}}{\sigma_{32}} = (\frac{\sigma_{21}}{\lambda_{11}})\frac{\sigma_{31}}{\sigma_{32}},$$

joten

$$\lambda_{11}^2 = \frac{\sigma_{21}\sigma_{31}}{\sigma_{32}},$$

$$\Leftrightarrow \lambda_{11} = \pm\sqrt{\frac{\sigma_{21}\sigma_{31}}{\sigma_{32}}}$$

ja

$$\lambda_{21} = \frac{\sigma_{21}}{\lambda_{11}} = \pm\sqrt{\frac{\sigma_{21}\sigma_{32}}{\sigma_{31}}}.$$

Siten myös lataukset λ_{11} ja λ_{21} ovat identifioituvia. Vastaavasti saadaan ratkaistua lataukset λ_{32} ja λ_{42} käyttämällä apuna kovariansseja σ_{32} , σ_{42} ja σ_{43}. Näin ollen mallin kaikki lataukset ovat identifioituvia. Tämän ominaisuuden ansiosta saadaan ratkaistua puolestaan jäännösvarianssit θ_1 , θ_2 , θ_3 ja θ_4 varianssiesityksistä σ_{ii}, kun $i = 1, ..., 4$. Koko malli on siten identifioituva.

2.1.3 Parametrien estimointi

Konfirmatorisen faktorimallin parametrien estimoinnissa oletetaan, että otos on poimittu perusjoukosta ja siten saatu otoskovarianssimatriisi S. Estimoinnin tarkoituksena on etsiä spesifioidun mallin tuntemattomille identifioituville parametreille matriiseissa Λ, Ω ja Θ arvot siten, että kovarianssimatriisin Σ sovite

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Omega}\hat{\Lambda}^T + \hat{\Theta}. \tag{2.10}$$

olisi mahdollisimman lähellä otoskovarianssimatriisia S.

Faktorimallin estimointimenetelmän valitsemiseen vaikuttavia tekijöitä on useita, esimerkiksi havaittujen muuttujien jakaumien muoto, mitta-asteikot sekä otoskoko. Mallin parametrien estimointi tapahtuu iteratiivisesti ja laskennallisesti intensiivisesti. Tämä johtaa väistämättä tietokoneavusteisten laskentaohjelmistojen käyttöön, esimerkiksi LISREL 8(Jöreskog & Sörbom 1999).

Kaikki konfirmatorisessa faktorianalyysissä käytettävät estimointimenetelmät perustuvat kohdefunktion F minimoimiseen. Suurimman uskottavuuden (SU-)estimointimenetelmässä minimoidaan kohdefunktiota

$$F(\Lambda, \Omega, \Theta) = ln|\sum(\Lambda, \Omega, \Theta)| - ln|S| + tr[\sum^{-1}(\Lambda, \Omega, \Theta)S] - p \tag{2.11}$$

matriisien Λ, Ω ja Θ estimoitavien parametrien suhteen. Suurimman uskottavuuden estimointimenetelmää käytetään, kun havaitut muuttujat noudattavat moniulotteista normaalijakaumaa. Lisäksi otoskoon täytyy olla riittävän suuri. Otoskoon alarajaksi on esitetty useita eri vaihtoehtoja. Esimerkiksi Boomsma (1982) ja Grönroos (1984) ovat esittäneet riittäväksi alarajaksi 100-200 havaintoa. Lawley & Maxwell (1971) ovat ehdottaneet alarajaksi lievemmän ehdon $N - 1 - p > 50$.

Oletusten ollessa voimassa SU-estimointimenetelmä on estimointimenetelmistä tehokkain, sillä menetelmällä on hyviä estimointiin liittyviä teoreettisia ominaisuuksia. Estimaattorit ovat tarkentuvia ja asymptoottisesti normaalijakautuneita ja niillä on minimivarianssiominaisuudet.

Jos SU-estimoinnin vaatima multinormaalisuusoletus ei ole täysin voimassa, voidaan käyttää yleistettyä pienimmän neliösumman (GLS-) estimointimenetelmää, jossa minimoidaan kohdefunktiota (Jöreskog & Sörbom 1996)

$$F(\Lambda, \Omega, \Theta) = (1/2)tr[(I - S^{-1}\Sigma)^2]. \tag{2.12}$$

Jos havaitut muuttujat ei ole edes likimain multinormaalisti jakautuneita, mallin parametrit voidaan estimoida käyttämällä yleistettyä painotettua pienimmän neliösumman (WLS-)estimointimenetelmää (Browne 1984). Jotta menetelmä olisi täysin käyttökelpoinen, havaittujen muuttujien jakauma pitää olla symmetrinen ja otoskoon täytyy olla riittävän suuri. Menetelmässä minimoidaan kohdefunktiota (Jöreskog & Sörbom 1996)

$$\begin{aligned} F(\Lambda, \Omega, \Theta) &= (s - \sigma)^T W^{-1} (s - \sigma) \\ &= \sum_{g=1}^{k} \sum_{h=1}^{g} \sum_{i=1}^{k} \sum_{j=1}^{i} w^{gh,ij} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij}), \end{aligned} \tag{2.13}$$

missä s on havaittujen muuttujien kovarianssimatriisin S alakolmiomatriisista muodostettu vektori. Vastaavasti σ on teoreettisen kovarianssimatriisin Σ alakolmiomatriisista muodostettu vektori. Funktiossa W^{-1} on painomatriisi, jossa painot $w_{gh,ij}$ saadaan kaavalla

$$w_{gh,ij} = m_{ghij} - s_{gh}s_{ij}, \tag{2.14}$$

missä m_{ghij}:t ovat havaittujen muuttujien neljänsiä keskusmomentteja, jotka saadaan kaavasta

$$m_{ghij} = (1/N)\sum_{a=1}^{N}(y_{ag} - \bar{y}_g)(y_{ah} - \bar{y}_h)(y_{ai} - \bar{y}_i)(y_{aj} - \bar{y}_j). \tag{2.15}$$

Jos aiemmin mainittuja estimointimenetelmiä ei voida oletusten rikkoutumisien takia käyttämään, käytetään parametrien estimoinnissa pienimmän neliösumman menetelmää(ULS), jossa minimoidaan kohdefunktiota (Jöreskog & Sörbom 1996)

$$F(\Lambda, \Omega, \Theta) = (1/2)tr[(S - \Sigma)^2] \tag{2.16}$$

matriisien Λ, Ω ja Θ estimoitavien parametrien suhteen.

2.1.4 Hypoteesien testaaminen

Kun konfirmatorisen faktorimallin parametrit on estimoitu, testataan mallin yhteensopivuutta havaintoaineiston kanssa. Tarkasteltava nollahypoteesi H_0 ja vastahypoteesi H_1 on

$H_0 : \Sigma = \Lambda\Omega\Lambda^T + \Theta$

H_1 : Kovarianssimatriisissa Σ ei ole rajoituksia.

Nollahypoteesin paikkaansapitävyyden eli onko faktorimalli voimassa, testaaminen johtaa asymptoottisen χ^2-testin käyttöön. Testisuure saadaan estimoinnin tuottaman kohdefunktion minimin F_{min} avulla:

$$\chi^2 = (N - 1)F_{min}(\hat{\Lambda}, \hat{\Omega}, \hat{\Theta}). \tag{2.17}$$

Testin vapausasteet df saadaan kaavasta

$$df = (1/2)p(p+1) - s, \tag{2.18}$$

jossa (1/2)p(p+1) on vastahypoteesin määrittämän rajoittamattoman faktorimallin estimoitavien parametrien lukumäärä eli estimoitujen kovarianssien ja varianssien lukumäärä teoreettisessa kovarianssimatriisissa Σ . Kaavassa s on puolestaan nollahypoteesin määrittämän faktorimallin vapaasti estimoitavien parametrien lukumäärä. Vapausasteluku on siis määrä kuinka paljon faktoriesityksen avulla on pystytty tiivistämään informaatiota otoskovarianssimatriisin alkioiden lukumäärästä (1/2)p(p+1). Jotta testi olisi käyttökelpoinen, pitää estimoitavien parametrien määrä nollahypoteesin mukaisessa mallissa olla pienempi kuin vastahypoteesin määrittämässä mallissa, eli

$$s < (1/2)p(p+1). \tag{2.19}$$

Päätös nollahypoteesin hyväksymisestä/hylkäämisestä voidaan tehdä estimoitua χ^2-arvoa vastaavan p-arvon eli merkitsevyystason avulla. Laskettu p-arvo ilmaisee todennäköisyyden

$$p = P(\chi^2(df) \geq \hat{\chi}^2(df)) \tag{2.20}$$

eli estimoidun faktorimallin merkitsevyystason nollahypoteesin ollessa tosi. Suuret p-arvot ilmaisevat hyvää estimoidun mallin sopivuutta aineiston kuvaajaksi ja pienet p-arvot puolestaan huonoa yhteensopivuutta valitulla merkitsevyystasolla. Esimerkiksi 5%:n merkitsevyystasolla saadaan ehdot: jos p-arvo ≥ 0.05, mallia voidaan pitää riittävänä ja jos p-arvo < 0.05, malli ei ole riittävä.

Lopullista päätöstä nollahypoteesin hyväksymisestä/hylkäämisestä ei aina kannata tehdä pelkän χ^2-testin perusteella. Jöreskog (1969) onkin ehdottanut nyrkkisääntöä χ^2-testin käytölle:

$$\hat{\chi}^2 < 2df. \tag{2.21}$$

Jos ehto pitää paikkansa, on faktorimalli riittävä. Toisaalta liian yksityiskohtaisen mallin, jolloin χ^2-testisuureen arvo suhteessa vapausasteisiin on liian pieni, estimoidut tulokset eivät välttämättä ole yleistettävissä.

χ^2-testin käytössä ilmenee ongelmia varsinkin suurilla otoiskoilla (N > 500). χ^2-testi hylkää silloin hyvin herkästi mallin sopimattomana, vaikka estimoidun kovarianssimatriisin ja otoskovarianssimatriisin välinen poikkeama olisikin häviävän pieni. Sen takia päätöksenteossa kannattaa käyttää χ^2-testin lisäksi muita yhteensopivuusindeksejä, kuten Bentlerin ja Bonettin (1980) kehittämää normeerausindeksiä NFI (normed fit index). Sen avulla arvioidaan otoskoon vaikutusta χ^2-testisuureeseen sekä myös mallin sopivuutta havaintoaineistoon. NFI vertaa estimoidun faktorimallin ns. nollamallin χ^2-testisuurearvoja keskenään. Nollamallissa havaitut muuttujat oletetaan korreloimattomiksi. Indeksi on riippumaton otoskoosta:

$$NFI = \frac{\chi_0^2 - \chi_1^2}{\chi_0^2} = \frac{(N-1)F_0 - (N-1)F_1}{(N-1)F_0} = \frac{F_0 - F_1}{F_0} \tag{2.22}$$

NFI-indeksin saamat arvot vaihtelevat 0:n ja 1:n välillä. Korkeat NFI-indeksin arvot ilmaisevat estimoidun mallin hyvyyttä. Yleisenä nyrkkisääntönä voidaan pitää ehtoa: jos NFI > 0.90, malli on riittävä.

2.1.5 Mallin riittävyystarkastelut

Konfirmatorisen faktorimallin rakentamisen viimeisessä vaiheessa suoritetaan vielä riittävyystarkasteluja. Tässä vaiheessa mallia voidaan vielä modifioida eksploratiivisesti aineiston antaman informaation perusteella. Yksinkertaisen faktorirakenteen omaavaan malliin voidaan esimerkiksi mahdollisesti lisätä estimoitavia parametreja tai päinvastoin yliestimoidusta mallista poistaa ylimääräisiä parametreja . Faktorimallin riittävyystarkastelut voidaan jakaa neljään eri ryhmään:

1) koko mallia koskevat tarkastelut

2) muuttujakohtaiset tarkastelut

3) parametrikohtaiset tarkastelut

4) havaintokohtaiset tarkastelut

Tarkastelujen suoritusjärjestyksellä ei ole merkitystä, mutta kaikkien tarkasteluvaiheiden läpikäyminen on ehdotonta onnistuneen faktorimallin rakentamisessa. Esimerkiksi koko mallia koskevissa tarkasteluissa ei voida paikantaa mahdollisen riittämättömyyden syitä ja toiseksi vaikka koko mallia koskevat riittävyystunnusluvut puoltaisivat hyvää estimoitua mallia, saattaa malli olla yksityiskohtaisesti riittämätön.

Koko mallia koskevissa tarkasteluissa χ^2-testillä on hyvin keskeinen asema. χ^2-testi soveltuu hypoteesien testaamisten lisäksi hyvin myös arvioitaessa faktorimallin sopivuutta tutkimusaineiston kuvaajana (ks. kappale 2.1.4). Suurien otoskoiden tapauksessa kannattaa kuitenkin käyttää otoskoista riippumattomia mallin riittävyysindeksejä, kuten normeerausindeksi NFI (ks. kappale 2.1.4) ja GFI-indeksi (Goodness of fit index)(Tanaka & Huba 1984, Jöreskog & Sörbom 1996). GFI-indeksi on kuitenkin riippuvainen estimointimenetelmästä, joten esimerkiksi yleistetyn painotetun PNS-estimaattori indeksille saadaan kaavalla

$$GFI = 1 - \frac{(s-\hat{\sigma})^T W^{-1}(s-\hat{\sigma})}{s^T W^{-1} s}, \tag{2.23}$$

jossa osoittaja on kohdefunktion minimi sen jälkeen, kun malli on sovitettu. Nimittäjä on puolestaan kohdefunktion minimi ennen, kun malli on sovitettu eli painotettu otoskovarianssimatriisi. SU-estimointimenetelmällä GFI saadaan kaavalla

$$GFI = 1 - \frac{tr(\hat{\Sigma}^{-1}S - I)^2}{tr(\hat{\Sigma}^{-1}S)^2} \tag{2.24}$$

ja PNS-estimoinnilla

$$GFI = 1 - \frac{tr(S-\hat{\Sigma})^2}{tr(S^2)}. \tag{2.25}$$

GFI-indeksi voi saada arvoja 0:n ja 1:n väliltä. Jos indeksin arvo ylittää 0.90, voidaan koko mallia pitää tältä osin riittävänä.

Koko mallin riittävyyttä voidaan tarkastella myös keskimääräistä jäännöskovarianssia ja jäännösvarianssia mittaavalla RMR-indeksillä (root mean square residual) (Jöreskog & Sörbom 1981), joka saadaan kaavalla

$$RMR = \sqrt{\frac{2\sum_{j=1}^{p}\sum_{i=1}^{j}(s_{ij} - \hat{\sigma}_{ij})^2}{p(p+1)}}, \tag{2.26}$$

jossa s_{ij}:t ovat otoskovarianssimatriisin S ja $\hat{\sigma}_{ij}$:t kovarianssimatriisin $\hat{\Sigma}$ alkioita. Mitä pienempi on RMR-indeksin saama arvo, sitä parempi on estimoitu malli.

Riittävyysindeksiä AIC(Akaike 1974,1987) käytetään, kun halutaan verrata samalle aineistoille tehtyjä faktorimalleja. Indeksin arvo saadaan kaavasta

$$AIC = \hat{\chi}^2 + 2s, \tag{2.27}$$

jossa s on riippumattomien estimoitujen parametrien lukumäärä mallissa. Vertailussa parempi malli saa pienemmän indeksin arvon.

Mallin yksinkertaistamisesta johtuvaa approksimointivirhettä voidaan puolestaan mitata RMSEA-indeksillä (Steiger 1990, Jöreskog & Sörbom 1993), joka saadaan kaavasta

$$\epsilon = \sqrt{\frac{\hat{F}_0}{df}}, \tag{2.28}$$

jossa $\hat{F}_0$ saadaan kaavasta

$$\hat{F}_0 = max(\hat{F} - (\frac{df}{n}), 0), \tag{2.29}$$

jossa $\hat{F}$ on estimoitavan kohdefunktion minimi ja df on vapausasteet. Jos RMSEA:n saama arvo $\epsilon < 0.05$, on malli riittävä.

Muuttujakohtaiset tarkastelut voidaan tehdä riittävyysindeksin $\hat{R}_i^2$ (squared multiple correlation) avulla:

$$\hat{R}_i^2 = 1 - \frac{\hat{\theta}_{ii}}{\hat{\sigma}_i^2}, \tag{2.30}$$

jossa $\hat{\theta}_{ii}$ on estimoitu jäännösvarianssi ja $\hat{\sigma}_i^2$, mallin tuottama sovite havaittujen muuttujien varianssille ja i=1,...,p. Jokaiselle havaitulle muuttujalle laskettu $\hat{R}_i^2$ voidaan tulkita niiden reliabiliteetti- tai kommunaliteettikertoimiksi. Kerroin kertoo, kuinka hyvin kukin havaittu muuttuja mittaa faktoreita. Indeksi voi teoriassa saada arvoja 0:sta 1:seen. Mitä suurempi indeksin arvo, sitä paremmin muuttuja mittaa faktoreita.

Parametrikohtaisia tarkasteluja voidaan tehdä kaikille mallin spesifiointivaiheessa määritetyille parametrityypeille eli vapaasti estimoitaville, yhtäsuurina estimoitaville sekä kiinnitetyille parametreille. Parametrien estimaattien pitää olla tilastollisesti kelvollisia ja sisällöllisesti tulkittavissa. Siten esimerkiksi faktoreiden varianssien ja jäännösvarianssien estimaattien pitää olla positiivisia, korrelaatiokertoimien estimaattien itseisarvoltaan alle ykkösiä ja estimoitujen kovarianssimatriisien $\hat{\Omega}$ ja $\hat{\Theta}$ tulee olla positiivisesti definiittejä. Lisäksi kaikki malliin valittujen parametrien pitää olla nollasta eroavia eli tilastollisesti merkitseviä. Tämä tapahtuu vertaamalla parametrin estimaattia sitä vastaavaan keskivirheeseen. Kun sovelletaan t-testin menettelytapaa 5%:n merkitsevyystasolla, saadaan t-arvolle ehto

$$|\text{t-arvo}| = |\frac{\text{parametrin estimaatti}}{\text{keskivirhe}}| > 1.96, \tag{2.31}$$

jolloin parametri on tilastollisesti nollasta eroava ja siten kelvollinen malliin.

Yhtäsuurina estimoitaville ja kiinnitetyille parametreille tehdyt riittävyystarkastelut perustuvat parametrien modifikaatio-indekseihin (MI). Ne liittyvät estimointimenetelmän kohdefunktion ensimmäisiin ja toisiin derivaattoihin rajoitettujen tai kiinnitettyjen parametrien suhteen,(Jöreskog & Sörbom 1981 ja Sörbom 1986). Jos modifikaatioindeksin saama arvo on suuri, yhtäsuuruusrajoite tai kiinnitys parametrille ei ole faktorimallin kannalta sopiva. Modifikaatio-indeksin arvo ilmoittaa myös, paljonko estimoidun mallin χ^2-testisuuren arvo laskee, jos kyseinen parametri vapautetaan estimoitavaksi. Jos vapautus on sisällöllisesti mielekäs ja estimointi suoritetaan, vapausasteluku vähenee yhdellä.

Havaintokohtaiset tarkastelut suoritetaan laskemalla estimoidun mallin tuottamat jäännökset kovariansseille ja variansseille:

$$S - \hat{\Sigma} = [s_{ij} - \hat{\sigma}_{ij}], \tag{2.32}$$

jossa s_{ij} $(i, j = 1, ..., p)$ on otoskovarianssimatriisin S alkioita, $\hat{\sigma}_{ij}$:t vastaavia estimoidun mallin sovitteita. Suuret jäännökset $s_{ij} - \hat{\sigma}_{ij}$ kertovat, mihin osaan aineistoa malli ei sovi. Tilanteissa, joissa otoskovarianssimatriisi sisältää suuruudeltaan toisistaan poikkevia variansseja, on suurten jäännösten arvioiminen hankalaa. Ongelman ratkaisemiseksi jäännökset voidaan normalisoida kaavallla

$$NR_{ij} = \frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{(\hat{\sigma}_{ii}\hat{\sigma}_{jj} + \hat{\sigma}_{ij}^2)/N}}, \tag{2.33}$$

koska asymptoottisesti on voimassa (Jöreskog & Sörbom 1981):

$$E(s_{ij}) \approx \sigma_{ij} \text{ ja } var(s_{ij}) \approx (\sigma_{ii}\sigma_{jj} + \sigma_{ij}^2)/N, \tag{2.34}$$

kun i,j=1,...,p. Jos estimoitu faktorimalli on oikea, saadaan normalisoitujen jäännösten NR_{ij} odotusarvot likimain nolliksi ja varianssit ykkösiksi. Siten saadaan mallin riittävyyttä mittaava ehto normalisoiduille jäännöksille:

$$|NR_{ij}| < 2. \tag{2.35}$$

Jos ehto on voimassa, malli on selittänyt tarpeeksi kyseistä havaintoa eli normalisoitua jäännöstä vastaavaa havaittujen muuttujien y_i ja y_j kovarianssia(korrelaatiota). Jos malli sisältää suuria, positiivisia normalisoituja jäännöksiä, malli aliestimoi kyseistä kovarianssia ja malliin pitää siten lisätä parametreja. Jos malli sisältää puolestaan pieniä, negatiivisia normalisoituja jäännöksiä, malli yliestimoi kovarianssia ja mallista pitää poistaa estimoituja parametreja.

2.2 Faktoripistemäärien estimointi

Onnistuneen konfirmatorisen faktorimallin rakentamisen jälkeen faktorirakenteen sisältämä informaatio voidaan tallentaa jatkoanalyysejä varten faktoripistemäärämuuttujiin. Silloin lasketaan faktorimallin avulla faktoreiden arvot kullekin havainnolle otoksessa. Estimointi poikkeaa tavallisesta kiinteän parametrin tilastollisesta estimoinnista, sillä nyt on estimoitavana satunnaismuuttujien arvoja. Estimointia hankaloittaa myös faktorimallin parametrien dimensiot (ks. perusmalli 2.1), jolloin faktoripistemäärien laskeminen on mahdotonta. Siten jatkotarkasteluissa käytetään estimoituja faktoripistemäärämuuttujia $\tilde{\eta}_1, ..., \tilde{\eta}_k$. Faktoripistemäärät voidaan ratkaista esimerkiksi regressiomenetelmän avulla kaavasta (Lawley & Maxwell 1971)

$$\tilde{\eta} = \Omega\Lambda^T\Sigma^{-1}y. \tag{2.36}$$

Faktoripistemäärien estimaatit saadaan silloin kaavasta

$$\tilde{\tilde{\eta}} = \tilde{\Omega}\tilde{\Lambda}^T S^{-1}y, \tag{2.37}$$

kun faktorit korreloivat ja kaavasta

$$\tilde{\tilde{\eta}} = \tilde{\Lambda}^T S^{-1}y, \tag{2.38}$$

kun faktorit eivät korreloi. Siten saadaan kullekin otoshavainnolle $j = 1, ..., N$ faktoripistemäärämuuttuja-arvo $\tilde{\tilde{\eta}}_i$, kun $i = 1, ..., k$. Faktoripistemäärien teoreettisten ominaisuuksien, esimerkiksi faktoripistemäärämuuttujien harhaisuuden takia estimointiin liittyy kuitenkin aina epätarkkuutta. Ongelman välttämiseksi voidaan faktoripistemäärien sijasta käyttää vaihtoehtoisesti esimerkiksi kovarianssirakennemallien LISREL-sovellusta, jossa voidaan rakentaa faktoreille regressiomalleja estimoimatta faktoripistemääriä (Alanen, Kinnunen & Leskinen 1984). Tässä työssä kuitenkin käytetään empiirisessä sovellusosassa muuttujien havaintoarvoilla ja faktoripainokertoimien avulla muodostettuja faktoripistemäärämuuttujia.

2.3 Toisen kertaluvun faktorimalleista

Tarkastellaan seuraavaksi konfirmatorisen faktorimallin laajentamista toisen kertaluvun faktorimalleihin. Lähtökohtana on perusmalli 2.1, jossa η-faktorit korreloivat keskenään. Tällöin voidaan ajatella, että η-faktoreiden taustalla on vielä toisia laaja-alaisempia, toisen kertaluvun(Jöreskog 1979, Jöreskog & Sörbom 1979 ja Leskinen 1983a) faktoreita, jotka vaikuttavat η-faktoreihin ja siten selittävät η-faktoreiden väliset korrelaatiot. Toisen kertaluvun faktorirakenne on muotoa

$$\eta = \Gamma\xi + \zeta, \tag{2.39}$$

jossa Γ on $(m\ x\ n)$ -latausmatriisi ja ξ $(n\ x\ l)$ -satunnaisvektori, joka sisältää faktorit $\xi_1, \xi_2, ..., \xi_n$ ja ζ on $(m\ x\ l)$ -jäännösvektori. Faktorimallin jäännökset voidaan tulkita myös erityisfaktoreiksi. Olkoon ξ-faktoreiden kovarianssimatriisi

$$cov(\xi) = \Phi \tag{2.40}$$

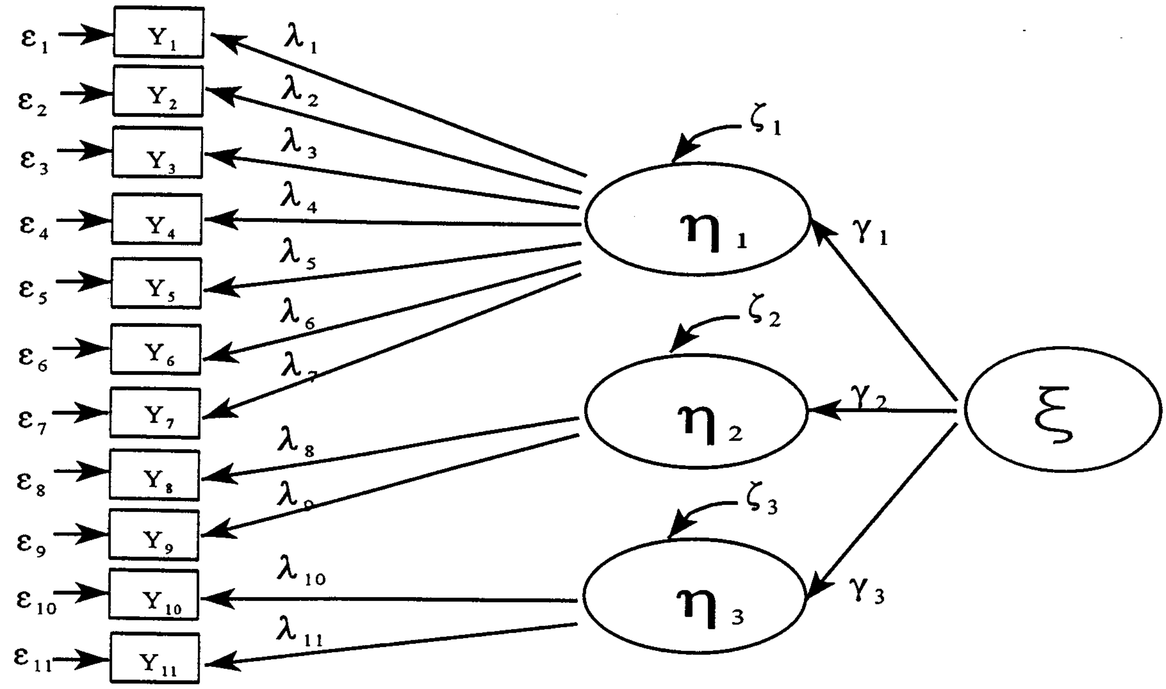
ja ζ-jäännösten kovarianssimatriisi

$$cov(\zeta) = \Psi. \tag{2.41}$$

Kun satunnaisvektori ξ ja jäännösvektori ζ oletetaan keskenään korreloimattomiksi, tuottaa toisen kertaluvun faktorirakenne ensimmäisen kertaluvun η-faktoreille kovarianssimatriisiesityksen, joka on analoginen ensimmäisen kertaluvun faktoreiden tuottaman havaittujen y-muuttujien kovarianssimatriisiesityksen kanssa. Täten toisen kertaluvun faktorimalli on kokonaisuudessaan muotoa

$$\begin{cases} y = \Lambda\eta + \epsilon, \\ \eta = \Gamma\xi + \zeta. \end{cases} \tag{2.42}$$

Toisen kertaluvun faktorimallin idea graafisesti esitettynä on kuviossa 2.1.



Kuvio 2.1 Toisen kertaluvun faktorimallin graafinen esitys

Toisen kertaluvun faktorimallin rakentaminen tapahtuu analogisesti konfirmatorisen faktorimallin rakentamisen mukaisesti. Käytännössä malli rakennetaan yleensä tavalla, jossa ensin rakennetaan riittävä ensimmäisen kertaluvun faktorimalli ja vasta sitten siirrytään toisen kertaluvun faktorimallin rakentamiseen. Toisen kertaluvun faktorimallia on mahdollista laajentaa myös vielä useamman kertaluvun faktorimalliksi, esimerkiksi kolmannen kertaluvun faktorimalliksi(Leskinen 1983b).

Toisen kertaluvun mallin spesifioinnin lähtökohtana on tarkoitus löytää ensimmäisen kertaluvun faktoreiden kovarianssirakenteelle selkeä toisen kertaluvun faktoriesitys. Tämä suoritetaan spesifioimalla mahdollisimman vähäparametrinen malli.

Mallin spesifionnin jälkeen suoritetaan identifioituvuustarkastelut. Tarkastelut voidaan tehdä jälleen määritelmän Jöreskog (1981) mukaisesti: toisen kertaluvun faktorimalli on identifioituva, jos sen parametrit Λ, Θ ja Γ, Φ, Ψ ovat ratkaistavissa teoreettisen kovarianssimatriisin Σ avulla. Identifioituvuustarkastelut kannattaa jakaa osiin kertaluvuttain. Eli ensin ratkaistaan ensimmäisen kertaluvun faktorimallin parametrit Λ, Ω ja Θ ja vasta tämän jälkeen ratkaistaan toisen kertaluvun faktorimallin parametrit Γ, Φ ja Ψ kovarianssimatriisin Σ avulla.

Toisen kertaluvun faktorimallin parametrien estimointi suoritetaan estimoimalla parametrit Λ, Θ ja Γ, Φ, Ψ samanaikaisesti. Estimaattien $\hat{\Gamma}, \hat{\Phi}$ ja $\hat{\Psi}$ avulla saadaan η-faktoreiden kovarianssirakenteelle sovite

$$\hat{\Omega} = \hat{\Gamma}\hat{\Phi}\hat{\Gamma}^T + \hat{\Psi}. \tag{2.43}$$

Ensimmäisen kertaluvun faktorimallin parametrien Λ ja Θ estimaatit eivät poikkea oleellisesti verratuna ensimmäisen kertaluvun faktorimallin parametrien estimointituloksista, jos toisen kertaluvun faktorimalli on riittävä.

Hypoteesien testauksessa toisen kertaluvun faktorirakenteen yhteensopivuutta ensimmäisen kertaluvun faktoreiden kovarianssirakenteen kuvaamisessa pitää ottaa huomioon rajoittava ehto. Sen mukaan identifioituvia ja estimoitavia parametreja matriiseissa Γ, Φ ja Ψ on oltava vähemmän kuin yhtälöitä matriisissa Ω eli vähemmän kuin (1/2)m(m+1) kappaletta. Riittävyystarkastelut suoritetaan samalla tavalla kuin konfirmatorisen faktorimallin tapauksessa(ks. kappale 2.1.5).

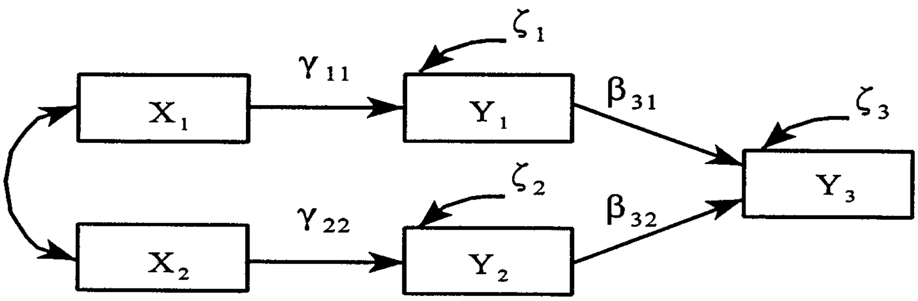
Vaihtoehtoisesti edellä tarkastellun, kokonaisvaltaisen toisen kertaluvun faktorimallin rakentamisen sijasta voidaan käyttää myös muita rakennustapoja. Esimerkiksi voidaan estimoida faktoripistemääräkertoimien ja otoshavaintojen avulla η-faktoripistemäärät, ja laskea η-faktoreiden otoskovarianssimatriisi, jonka avulla puolestaan suoritetaan uusi konfirmatorinen faktorianalyysi. Tätä menettelytapaa ollaan sovellettu kappaleessa 7.

3 Rakenneyhtälömalleista

Tässä kappaleessa keskitytään LISREL-mallien erikoistapauksiin, joissa rakenneyhtälömalli(structural equation models,SEM) muodostetaan havaittujen muuttujien välille. Polkumallityyppisistä malleista keskitytään rekursiivisiin ja simultaanisiin rakenneyhtälömalleihin. Tämän kappaleen pohjana on käytetty pääsääntöisesti lähteitä Nummenmaa, Konttinen, Kuusinen ja Leskinen (1997) sekä Leskinen (1987).

3.1 Rekursiiviset ja simultaaniset rakenneyhtälömallit

Rakenneyhtälömallien peruslähtökohtaa voidaan tarkastella havainnollistavan nuolikaavion avulla. Kuviossa 3.1 x-muuttujat ovat selittäviä, eksogeenisiä muuttujia ja y-muuttujat selitettäviä, endogeenisiä muuttujia. Muuttujien x_1 ja x_2 välinen, kaksisuuntainen nuoli tarkoittaa suoraa assosiaatiota, jonka syy-seuraus -suhdetta ei voida määrittää. Muiden kuviossa esiintyvien suhteet ovat ilmeisiä. Esimerkiksi x_1 on syy-muuttuja ja y_1 on seuraus-muuttuja. Näiden muuttujien välinen vaikutus on suora. Epäsuora vaikutus esiintyy esimerkiksi muuttujien x_1 ja y_3 välillä. Siinä x_1:n vaikutus y_3:een välittyy y_1:n kautta.



Kuvio 3.1 Esimerkki rakenneyhtälömallista

Mallia, joka sisältää vain endogeenisten muuttujien yhdensuuntaisia vaikutuksia, sanotaan rekursiiviseksi. Simultaanisessa rakenneyhtälömallissa endogeenisten muuttujien välillä on vastavuoroista, samanaikaista vaikutusta toisiinsa.

Matriisimuodossa esitetty yleinen rakenneyhtälömalli havaittujen muuttujien välille on

$$y = By + \Gamma x + \zeta, \tag{3.1}$$

jossa y-vektori sisältää endogeeniset muuttujat $y_1, ..., y_q$, x-vektori eksogeeniset muuttujat $x_1, ..., x_p$ ja ζ-vektori jäännöstermit $\zeta_1, ..., \zeta_q$. Parametreista q x q matriisi B sisältää endogeenisten muuttujien väliset yhteydet ja q x p matriisi Γ endogeenisten ja eksogeenisten muuttujien väliset yhteydet. Merkitään ζ-jäännösten kovarianssimatriisia Ψ:llä. Koska rekursiivisessa rakenneyhtälömallissa ei sallita endogeenisten muuttujien välisiä vastavuoroisia yhteyksiä, voidaan parametrimatriisi B määrittää joko ylä- tai alakolmiomatriisina. Simultaanisessa rakenneyhtälömallissa ylä- tai alakolmiomatriisiesitys ei puolestaan ole mahdollinen. Rakenneyhtälömalleissa ollaan tässä tapauksessa kiinnostuneita vain korrelatiivista rakenteista, siksi esimerkiksi malli (3.1) ei sisällä mitään keskiarvoihin liittyvää vakioparametria β_0 eikä muitakaan tasoparametreja.

3.2 Rekursiivisten ja simultaanisten rakenneyhtälömallin rakentaminen

Rekursiivisten ja simultaanisten rakenneyhtälömallien rakentaminen toteutetaan pääpiirteittäin hyvin samoin periaattein kuin konfirmatorisen faktorimallin rakentamisessa (ks. kappale 2.1). Mallin rakentaminen voidaan edelleen jakaa viiteen selkeästi erotettavaan tutkimusvaiheeseen.

3.2.1 Mallin spesifiointi ja identifioituvuustarkastelut

Rakenneyhtälömallin rakentaminen alkaa konfirmatorisen faktorimallin rakentamisen tapaan mallin spesifionnilla. Lähtökohdaksi kannattaa valita yleensä hyvin yksinkertainen malli, joka perustuu sisällöllisiin tutkimushypoteeseihin. Muuttujien ja relaatioiden valintaan vaikuttaa jo mallin spesifiointivaiheessa erilaiset identifioituvuusvaatimukset. Yksinkertaisimmassa tapauksissa rekursiivinen rakenneyhtälömalli on aina identifioituva, jos jäännösten kovarianssimatriisi Ψ on diagonaalinen eli

$$\Psi = cov(\zeta) = diag(\psi_{11}, ..., \psi_{qq}). \tag{3.2}$$

Jos jäännösten kovarianssimatriisiΨ ei ole diagonaalinen eli jäännökset korreloivat, pitää ehdon (3.3) olla voimassa eli

$$s \leq (1/2)(p+q)(p+q+1), \tag{3.3}$$

jossa s on estimoitavien $\beta-, \gamma-$ ja $\psi-$ parametrien lukumäärä ja $(1/2)(p+q)$ $(p+q+1)$ yhtälöiden lukumäärä kovarianssimatriiseissa Σ_{yy}(q x q),Σ_{xy}(p x q) ja Σ_{xx}(p x p). Simultaanisissa rakenneyhtälömalleissa pitää ottaa edellisten ehtojen lisäksi huomioon selittävien muuttujien lukumäärää rajoittava välttämätön määrällinen ehto, ns. order-ehto, jos jäännöskovarianssimatriisi Ψ on rajoittamaton.

Sen mukaan yhdessäkään rakenneyhtälössä ei saa olla selittäviä x- ja y-muuttujia enempää kuin on x-muuttujia (p kappaletta) koko estimoitavassa mallissa.

3.2.2 Parametrien estimointi ja hypoteesien testaus

Rekursiivisten ja simultaanisten rakenneyhtälömallien parametrien estimointi voidaan suorittaa kahdella eri tavalla. Rakenneyhtälöt voidaan estimoida joko yhtälö kerrallaan tai kaikki yhtälöt samanaikaisesti. Jos rakenneyhtälömallista halutaan saada kokonaisuudessaan ja mallin sopivuudesta havaintoaineistoon parempi kuva, pitää mallin parametrit estimoida kaikki yhtälöt samanaikaisesti. Tämä ns. systeemiestimointi suoritetaan yleensä SU-estimointimenetelmän avulla. Silloin endogeenisten muuttujien pitää noudattaa moniulotteista normaalijakaumaa ja olla vähintään välimatka-asteikolla mitattuja.

Jos mallien yhtälöt estimoidaan yksitellen, käytetään estimointimenetelmänä PNS-estimointia, kaksivaiheista PNS-estimointia tai SU-estimointimenetelmää (ks. kappale 2.1.3). Mainitut estimointimenetelmät tuottavat rekursiivisissa rakenneyhtälömalleissa tehokkaita estimaatteja, jos jäännökset eivät korreloi ja normaalijakaumaoletus pitää endogeenisten muuttujien osalta paikkansa. Simultaanisten mallien parametrien estimoinnissa yhtälö kerrallaan käytetään yleensä kaksivaiheista PNS-estimointimenetelmää. Käyttö on perusteltua, koska se tuottaa tarkentuvia estimaattoreita simultaanisissa malleissa. Myös epäsuoran pienimmän neliäsumman estimointimenetelmän käyttö on mahdollista, jos mallin yhtälöt ovat kaikki täsmälleen identifioituvia.

Mallin yhteensopivuustestinä voidaan käyttää konfirmatorisen faktorimallin rakentamisen tapaan χ^2-testiä. Testattavat hypoteesit ovat:

H_0: Kovarianssimatriisi Σ on rajoitettu ja

H_1: Kovarianssimatriisi Σ on rajoittamaton.

χ^2-testin vaatimat vapausasteet df saadaan kaavasta

$$df = (1/2)(p+q)(p+q+1) - s, \tag{3.4}$$

jossa s on estimoitavien parametrien lukumäärä mallissa. Estimoitu rakenneyhtälömalli on riittävä, jos χ^2-testiä vastaava p-arvo on suurempi kuin 0.05. Kuten konfirmatorisessakin faktorimallissa, lopullisten johtopäätösten tekemisessä, kannattaa käyttää apuna muita, otoskoosta riippumattomia yhteensopivuusindeksejä, kuten NFI- ja GFI-indeksejä.

3.2.3 Mallin riittävyystarkastelut

Rekursiivisissa ja simultaanisissa rakenneyhtälömalleissa myös mallin riittävyystarkastelut suoritetaan samaan tapaan kuin konfirmatorisen faktorimallin tilanteessa (ks. kohta 2.1.5). Suurimmat erot ovat kuitenkin muuttujakohtaisissa riittävyystarkasteluissa. Selittävien muuttujien selityskykyä voidaan mitata kussakin yhtälössä selitysasteilla $R^2_{y(i)}$, jotka ovat muotoa

$$R^2_{y(i)} = 1 - \frac{\psi_i}{var(y_i)}, \tag{3.5}$$

jossa ψ_i on i:nnen rakenneyhtälön jäännöksen ζ_i varianssi. Jos mallissa on endogeenisiä selittäviä muuttujia, kannattaa varsinkin simultaanisissa rakenneyhtälömalleissa käyttää selitysasteen laskemisessa apuna redusoitua muotoa:

$$(I - B)y = \Gamma x + \zeta, \tag{3.6}$$

josta saadaan

$$y = (I - B)^{-1}\Gamma x + \zeta^*, \tag{3.7}$$

jossa $\zeta^* = (I - B)^{-1}\zeta$.
Siten saadaan uusi selitysaste

$$R_i^{*2} = 1 - \frac{var(\zeta_i^*)}{var(y_i)}, \tag{3.8}$$

jonka avulla saadaan huomioitua kaikkien selittävien eksogeenisten muuttujien vaikutus y_i:n vaihtelusta (Jöreskog 1999).

Parametrikohtaisissa tarkasteluissa johtopäätökset tilastollisista merkitsevyyksistä tehdään taas keskivirheiden ja t-arvojen avulla (ks. kohta 2.1.5), mutta nyt tarkastelun kohteena ovat estimoidut $\beta-, \gamma-$ ja $\psi-$ parametrit. Nollaksi kiinnitetyille parametreille tehdyt johtopäätökset tehdään modifikaatioindeksien MI avulla samaan tapaan kuin konfirmatorisessa faktorianalyysissä.

4 Suhtautuminen koulutukseen - pitkittäistutkimusaineisto

Seuraavaksi havainnollistetaan konfirmatorisen faktorimallin ja rekursiivisen rakenneyhtälömallin rakentamista käytännön tutkimusaineiston avulla. Aineiston ja tutkimusongelmien esittelyn jälkeen esitetään alustavien analyysien tuloksia, joiden pohjalta rakennetaan konfirmatorinen faktorimalli. Analyysin tuottama informaatio tallennetaan faktoripistemäärämuuttujiin, joiden avulla tutkimusta laajennetaan rekursiivisiin rakenneyhtälömalleihin. Lopuksi esitetään vielä vaihtoehtoinen tarkastelutapa rakenneyhtälömalleihin, jossa sovelletaan toisen kertaluvun faktorimalleja.

4.1 Tutkimusaineisto

Tutkielmassa käytetty aineisto on osa professori Jorma Kuusisen johtamaa 30:n vuoden seuruututkimusta, jossa tutkitaan lahjakkuuden, kotitaustan ja koulumenestyksen yhteyksiä koulutus-, ammatti- ja työuraan ja muihin elämänrakenteen ja elämänhistorian merkittäviin tekijöihin lapsuudesta aikuisuuteen.

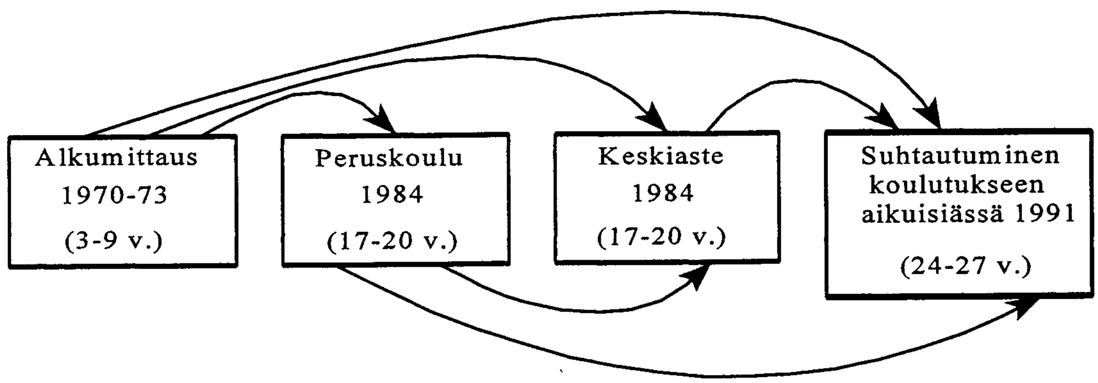
30:n vuoden seuruututkimus on jakaantunut neljään eri mittausajankohtaan. Tutkimus aloitettiin vuosina 1970-73 *alkumittauksella*, jolloin kerättiin taustatietoja 700:lta 3-9-vuotiailta lapsilta. Lapsilta saatiin taustatiedot vanhempien koulutuksesta, ammattiasemasta ja tuloista, sisarusten määrästä, syntymäjärjestyksestä, päiväkodista ja esikoulusta. Lapset testattiin ITPA-testillä, joka on kehitetty mittaamaan lasten verbaalista lahjakkuutta (Blåfield & Kuusinen 1974 ja Kuusinen & Blåfield 1974). Tutkimus jatkui vuonna 1984, jolloin lapset olivat suorittaneet *peruskoulun*. Tiedot menestyksestä kaikissa lukuaineissa kaikilta luokkatasoilta kerättiin rekisteritietoina 559:ltä 17-20 vuotiaalta oppilaalta. Kolmas mittauskerta tehtiin samana vuonna, jolloin nuoret olivat tehneet *keskiasteen* kouluvalinnan ja enemmistö oli jo jossakin keskiasteen oppilaitoksessa. Kouluvalinta- ja pääsytiedot saatiin 628 nuorelta, joista 259 ylioppilaaksi kirjoittaneelta on lisäksi lukion koulumenestys ja yo-kirjoitusten tulokset. Tähän mennessä viimeinen mittaus suoritettiin keväällä 1991, jolloin tutkittavat olivat saavuttaneet *aikuisiän*. 23-27 vuotiaille nuorille aikuisille lähetettiin kyselylomake, jonka avulla saatiin 437 henkilöltä tiedot koulutuksesta, työurasta, asumisesta, tuloista, ystävyyssuhteista, onnellisuudesta ja hyvinvoinnista. Tutkimusprojektin on tarkoitus jatkua vuonna 2000 samankaltaisella mittauksella, kuin oli vuoden 1991 aikuismittaus. Tämän hetkinen tutkimusasetelma on tiivistettynä taulukossa 4.1.

ALKUMITTAUS 1970 - 73	**PERUSKOULU** -1984	**KESKIASTE** - 1984	**AIKUISIKÄ** - 1991
Vuosina 1970-73 koottiin ITPAn normi-aineisto. Siinä oli 100 lasta kussakin 3-9-vuotiaiden kohortissa. Lisäksi koottiin tausta-tietoja: - vanhempien koulutus, ammattiasema ja tulot - sisarusten määrä - syntymäjärjestys - päiväkoti, esikoulu N = 700	Kaikki lapset olivat v. 1984 käyneet peruskoulun, jonka osalta tiedot menestyksestä kaikissa lukuaineissa kaikilla luokkatasoilla. N = 559	Kaikki nuoret olivat v. 1984 tehneet keski-asteen kouluvalinnan ja suurin osa oli jossakin keskiasteen oppilaitoksessa. Koulu-valinta- ja pääsytiedot saatiin 628 nuorelta. N = 628 Koulumenestys ja yo-kirjoitustiedot on käytettävissä 259 ylioppilaaksi kirjoittaneelta nuorelta. N = 259	Keväällä 1991 lähetettiin kyselylomake edellisen vaiheen niille nuorille aikuisille (23 - 27 v.), joiden osoite-tiedot oli käytettävissä (N = 543). Palautuspro-sentti oli 80%. - koulutus - työura - asuminen - tulot - ystävyyssuhteet - onnellisuus ja hyvinvointi N = 437

Taulukko 4.1 Tutkimuksen asetelma vuoteen 1991

4.2 Tutkimusongelmat

Tutkielman ensimmäinen sovellusosa rakentuu vuonna 1991 tehdyn survey-tutkimuksen aineistoon, joka kuvaa suhtautumista nykyiseen tai viimeksi käytyyn kouluun aikuisiässä. Ensimmäinen tutkimushypoteesi on, että millaisilla faktoreilla voidaan mitata suhtautumista kouluun aikuisiässä? Analysointimenetelminä kyseisen tutkimusongelman ratkaisemiseksi käytetään sekä eksploratiivista että konfirmatorista faktorianalyysiä. Toinen tutkimusongelma rakentuu konfirmatorisen faktorianalyysin tuottamien tulosten ympärille. Tarkoituksena on selvittää, mitkä tekijät vaikuttavat koulutukseen suhtautumiseen aina lapsuudesta aikuisuuteen. Tässä osassa sovelletaan rekursiivisia rakenneyhtälömalleja.



Kuvio 4.1 Tutkimusongelman rakenneyhtälömalligraafi.

5 Suhtautuminen koulutukseen -faktorimallin rakentaminen

5.1 Aineiston muuttujat ja oletusten esitarkastelu

Eksploratiivisessa ja konfirmatorisessa faktorianalyysissä tarkasteltavat muuttujat koostuvat 11 muuttujasta, jotka kuvaavat suhtautumista nykyiseen tai viimeksi käytyyn kouluun aikuisiässä. Muuttujat ovat väittämä-tyyppisiä kysymyksiä, joihin vastausvaihtoehdot ovat olleet:

1 = täysin eri mieltä
2 = jokseenkin eri mieltä
3 = jokseenkin samaa mieltä
4 = täysin samaa mieltä.

Kuten vastausvaihtoehdoista huomataan, mitta-asteikko tarkasteltavilla muuttujilla on järjestysasteikollinen. Täten muuttujien välillä on käytetty polykorisia korrelaatiokertoimia(Liite 1).

Analyysissä tarkasteltavat muuttujat ovat seuraavat:

PARHAAT	=	Elämäni parhaat hetket olen viettänyt koulussa
PITKÄST	=	Koulutyö on/oli minusta enimmäkseen pitkästyttävää
HAASTAV	=	Koulu tarjoaa/tarjosi minulle haastavia tavoitteita
VALMIUD	=	Koulu antaa/antoi valmiudet elämää varten

VIERAST	=	Koulu on/oli vieras todelliselle elämälle
TURHAA	=	Koulunkäynti on/oli minusta ihan turhaa
KUNNOL	=	Pyrin aina tekemään koulutyöni kunnolla
TARPEEN	=	Koulutuksen hankkiminen on elämäni eri vaiheissa vielä tarpeen
JORIITT	=	Minulle koulunkäynti saa jo riittää
RASKVAI	=	Koulunkäynti on/oli minulle liian raskasta ja vaikeaa
HELPPOA	=	Koulunkäynti on tavallisesti ollut minulle vaivatonta ja helppoa

Tutkittavien koehenkilöiden (N=437) vastauksissa esiintyi puuttuvia tietoja. Analyysin toteuttamista varten puuttuvien tietojen käsittelyn osalta on käytetty listwise-menetelmää. Sen mukaan analyysiin tulivat mukaan vain koehenkilöt, joilta ei puuttunut tietoja yhdestäkään muuttujasta. Lopulliseksi otoskooksi muodostui 416.

Koska estimointimenetelmät riippuvat jakaumista ja mitta-asteikoista, on muuttujille suoritettu Kolmogorov-Smirnovin normaalisuustestaus. Kuten taulukon 5.1 tuloksista nähdään, testiä vastaava nollahypoteesi hylätään jokaisen muuttujan osalta, eli mikään 11:sta muuttujasta ei ole normaalisti jakaantunut, joten muuttujajoukkokaan ei ole multinormaalinen. Suurimman uskottavuuden estimointimenetelmän oletus ei siis ole voimassa, joten pääanalyysissä käytettiin yleistetyn pienimmän neliösumman estimointimenetelmää.

Taulukko 5.1 Muuttujien normaalisuustestaus. Testisuure ja p-arvo

Muuttuja	Kolmogorov-Smirnov Z	p-arvo
PARHAAT	5.283	0.000
PITKÄST	5.069	0.000
HAASTAV	5.371	0.000
VALMIUD	5.102	0.000
VIERAST	4.497	0.000
TURHAA	9.225	0.000
KUNNOL	5.780	0.000
TARPEEN	6.680	0.000
JORIIT	6.337	0.000
RASKVAI	6.726	0.000
HELPPOA	5.977	0.000

5.2 Eksploratiivinen faktorianalyysi

Alustavana analyysinä normaalisuustarkastelujen lisäksi on suoritettu eksploratiivinen faktorianalyysi. Analyysin tarkoituksena on antaa taustatietoa konfirmatorisen faktorimallin rakentamiselle ja täten helpottaa faktorimalllin spesifiointia. Faktoreiden lukumäärän arviointi on suoritettu käyttämällä ominaisarvokriteeriä, jonka mukaan faktoreiden lukumäärä on yli ykkösten olevien ominaisarvojen lukumäärä. Faktorilataukset on estimoitu suurimman uskottavuuden estimointimenetelmän avulla ja faktoreiden rotatointi on suoritettu käyttämällä promax-rotatointia, jolloin faktorit voivat korreloida. Eksploratiivisen faktorianalyysin tuottamat latausmatriisi, muuttujien kommunaliteetit h_i^2 ja faktoreiden korrelaatiomatriisi ovat taulukossa 5.2.

Taulukko 5.2 Eksploratiivisen faktorianalyysin tuottamat latausmatriisi, kommunaliteetit ja faktoreiden korrelaatiomatriisi.

Muuttuja	1. Faktori	2. Faktori	3. Faktori	h_i^2
PARHAAT	**0.352**	-0.094	0.014	0.154
PITKÄST	**-0.705**	0.060	-0.077	0.564
HAASTAV	**0.661**	-0.031	-0.062	0.427
VALMIUD	**0.517**	0.229	-0.040	0.257
VIERAST	**-0.468**	-0.104	-0.020	0.210
TURHAA	**-0.524**	0.097	0.016	0.304
KUNNOL	**0.427**	-0.040	0.036	0.204
TARPEEN	-0.020	**-0.770**	-0.061	0.550
JORIIT	-0.007	**0.776**	-0.052	0.641
RASKVAI	-0.131	0.005	**-0.523**	0.333
HELPPOA	-0.081	0.022	**0.828**	0.640

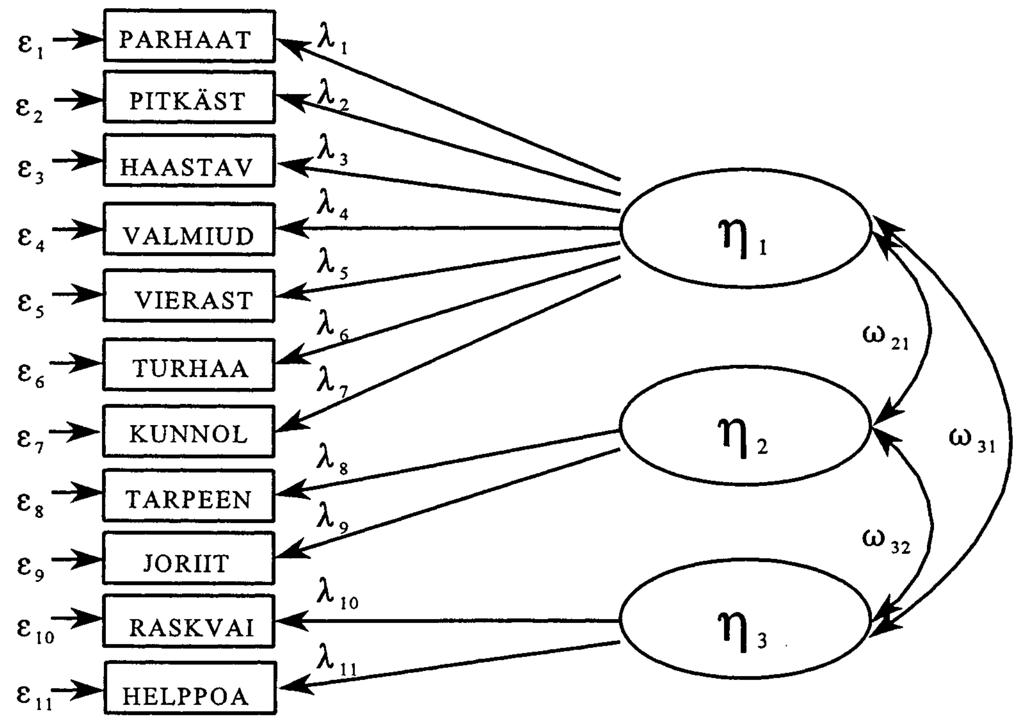
	1.Faktori	2.Faktori	3. Faktori
1.Faktori	1.000		
2.Faktori	0.253	1.000	
3.Faktori	0.290	0.417	1.000

Eksploratiivisen faktorianalyysin tuottamasta latausmatriisista huomataan, että ollaan saatu kolme selkeästi eroavaa faktoria. Ensimmäinen faktori sisältää suurimman osan latauksista, yhteensä seitsemän latausta ja kaksi muuta faktoria molemmat kaksi latausta. Latausten etumerkkejä ja muuttujien kuvauksia vertailemalla saadaan faktoreille alustavat nimet. Ensimmäinen faktori kuvaa selkeästi koulunkäynnin myönteisyyttä, toinen tarpeellisuutta ja kolmas sen helppoutta.

Myös positiivinen korrelaatiomatriisi tukee faktoreiden onnistunutta nimeämistä. Muuttujien kommunaliteetit h_i^2 ovat suuruudeltaan keskimäärin kohtalaisia.

5.3 Konfirmatorinen faktorianalyysi

Eksploratiivinen faktorianalyysi antoi hyvät lähtökohdat konfirmatorisen faktorimallin rakentamiselle. Saatujen tuloksien perusteella lähdettiin etsimään tilastollisesti kelvollista kolmen faktorin mallia. Ensimmäisenä kokeiltiin eksploratiivisen faktorianalyysin tuottamaa mallia, jonka mukaan jokainen muuttuja selittää vain yhtä faktoria ja jäännökset ovat korreloimattomia. Muut lataukset ovat asetettu nolliksi. Mallin spesifiointi graafisesti on kuviossa 5.1.



Kuvio 5.1 Konfirmatorisen faktorimallin spesifiointi lähtökohtatilanteessa.

Mallin riittävyystarkastelut ($\chi^2(41) = 369.95$, p-arvo = .00, GFI = .92, NFI = .63, RMSEA = .014, RMR = .098) osoittavat mallin olevan vielä hyvin keskeneräinen. Mallia parannettiin estimoimalla jäännöskovariansseja modifikaatioindeksien perusteella. Indeksin saama suuri arvo on tässä tapauksessa merkkinä muuttujien jäännösten korreloituneisuudesta. Parhaaseen mahdolliseen malliin päästiin estimoimalla yhteensä kahdeksan jäännösten kovarianssia, joiden estimaatit keskivirheineen ovat taulukossa 5.3. Suurimmaksi nollaksi kiinnitetyn parametrin MI-indeksiksi jäi muuttujien TURHAA ja VIERAST välinen jäännöskovarianssi eli $MI_{6,5}$=7.07. Myös uusien latausten estimointi olisi ollut modifikaatioindeksien perusteella aiheellista, mutta mitkään mahdollisista estimaateista ei ollut sisällöllisesti tulkittavissa, joten eksploratiivisen faktorianalyysin tuottama latausrakenne pysyi konfirmatorisessa faktorianalyysissä samankaltaisena.

Konfirmatoriseksi faktorimalliksi muodostui siten mittausmalli-tyyppinen latausratkaisu, jossa kukin muuttuja mittaa vain yhtä faktoria . Kun faktoreiden varianssit on lisäksi kiinnitetty ykkösiksi, on malli aina identifioituva. Siten faktorimallin teoreettiset, esimerkiksi määritelmään (Jöreskog 1981) perustuvat identifioituvuusominaisuudet ovat voimassa.

Lopullisen mallin riittävyysindeksit näyttävät jo huomattavasti paremmalta (taulukko 5.3), vaikka aivan ihanteellisiin tuloksiin ei päästykään, sillä χ^2-testiä vastaava p-arvo on vieläkin 0.00. Testisuureen saama suuri arvo selittyy kuitenkin suurella otoskoolla, jolle χ^2-testi on hyvin kriittinen. Kuitenkin vaihtoehtoisien, otoskoosta riippumattomien riittävyysindeksien NFI ja GFI saamista arvoista huomataan, että koko mallia voidaan pitää riittävänä kuvaamaan havaintoaineistoa.

Muuttujakohtaiset riittävyystarkastelut on suoritettu tutkimalla muuttujien kommunaliteettikertoimia $\hat{R}_i^2$. Muuttujat ovat selittäneet kohtalaisesti mitattuja faktoreita. Parhaiten faktoreita kuvaavat muuttujat ovat PITKAST, JORIIT ja RASKVAI. Pienimmät kertoimet saavat muuttujat VIERAST ja PARHAAT.

Myös parametrikohtaisten riittävyystarkastelujen osalta mallia voidaan pitää riittävänä. Faktoreiden varianssien ja jäännösvarianssien estimaatit ovat kaikki positiivisia ja korrelaatiokertoimet välillä (-1,1). Lisäksi taulukon 5.3 tuloksista nähdään, että mallin kaikki parametrien estimaatit ovat nollasta eroavia eli tilastollisesti merkitseviä.

Taulukko 5.3 Estimoidut lataukset Λ, jäännösvarianssit(suluissa keskivirheet), reliabiliteetit, koko mallia koskevat riittävyysindeksit, estimoidut jäännöskovarianssit ja faktoreiden korrelaatiomatriisi $\hat{\Omega}$ (suluissa keskivirheet)

X_i		$\hat{\Lambda}$(s.e.)		$\hat{\theta}$(s.e.)	$\hat{R}^2_{x(i)}$
	MYONTEIS	TARPEELL	HELPPOUS		
PARHAAT	.443(.049)	0*	0*	.768(.057)	.204
PITKÄST	-.881(.043)	0*	0*	.217(.037)	.781
HAASTAV	.662(.045)	0*	0*	.527(.042)	.454
VALMIUD	.510(.053)	0*	0*	.675(.057)	.279
VIERAST	-.432(.050)	0*	0*	.769(.057)	.196
TURHAA	-.684(.045)	0*	0*	.482(.040)	.493
KUNNOL	.486(.048)	0*	0*	.702(.053)	.251
TARPEEN	0*	.780(.060)	0*	.344(.072)	.639
JORIIT	0*	-.832(.059)	0*	.239(.076)	.744
RASKVAI	0*	0*	-.836(.075)	.264(.109)	.726
HELPPOA	0*	0*	.606(.064)	.581(.070)	.387

*) kiinnitetty

Koko mallia koskevat riittävyysindeksit:

$\chi^2(33) = 74.089$, p-arvo=.000, GFI=.975, NFI=.89, RMSEA=.0548, RMR=.0456

(Taulukko jatkuu seuraavalla sivulla)

Estimoidut jäännöskovarianssit:

$\hat{\theta}_{4,2}$=.183(.031), $\hat{\theta}_{5,4}$=-.141(.042), $\hat{\theta}_{8,2}$=.106(.024), $\hat{\theta}_{8,4}$=-.197(.041), $\hat{\theta}_{8,5}$=.146(.032), $\hat{\theta}_{9,4}$=.228(.039), $\hat{\theta}_{10,6}$=.088(.030), $\hat{\theta}_{11,9}$=-.094(.029)

$\hat{\Omega}(s.e.)$	MYONTEIS	TARPEELL	HELPPOUS
MYONTEIS	1.000		
TARPEELL	.392(.054)	1.000	
HELPPOUS	.370(.057)	.363(.058)	1.000

Konfirmatorisen faktorianalyysin tuottamat lataukset ja faktoreiden korrelaatiokertoimet pysyivät eksploratiiviseen faktorianalyysiin verrattuna samanmerkkisinä ja mitään uusia latauksia ei estimoitu. Täten latausrakenteiden tulkinnoista muodostui faktoreille lopullisesti selkokieliset nimet ja ominaisuudet:

1. Myönteisyys : on viettänyt parhaat hetket koulussa, koulutyö ei ollut pitkästyttävää, koulutyö tarjosi haastavia tavoitteita, antoi valmiudet elämää varten, koulutyö ei ollut vieras todelliselle elämälle, koulunkäynti ei ollut turhaa ja pyrkii tekemään koulutyön aina kunnolla.

2. Tarpeellisuus : koulutuksen hankkiminen on elämän eri vaiheissa vielä tarpeen ja koulunkäynti ei saa vielä riittää .

3. Helppous : koulunkäynti ei ollut raskasta ja vaikeaa ja koulunkäynti oli vaivatonta ja helppoa.

5.4 Faktoripistemäärämuuttujien estimointi

Jatkoanalyysejä varten on muodostettu kolme uutta faktoripistemäärämuuttujaa:

F1 = Myönteisyys-faktoripistemäärä

F2 = Tarpeellisuus-faktoripistemäärä

F3 = Helppous-faktoripistemäärä

Faktoripistemäärämuuttujien estimointi on suoritettu regressiomenetelmän avulla. Taulukossa 5.4 on faktoripistemäärämuuttujien painokertoimien lisäksi muuttujien keskiarvot $\bar{x}$ ja keskihajonnat s_x.

Taulukko 5.4 Estimoidut faktoripainokertoimet, keskiarvot ja keskihajonnat.

	η_i				
x_i	MYONTEIS	TARPEELL	HELPPOUS	$\bar{x}$	s_x
PARHAAT	**0.044**	-0.044	-0.004	1.477	1.098
PITKÄST	**-0.620**	-0.182	0.120	1.387	1.085
HAASTAV	**0.095**	-0.097	-0.008	1.846	1.257
VALMIUD	**0.267**	0.340	-0.059	1.658	1.227
VIERAST	**-0.028**	0.025	0.010	1.496	1.166
TURHAA	**-0.110**	0.121	-0.122	0.867	0.733
KUNNOL	**0.052**	-0.053	-0.005	1.929	1.246
TARPEEN	0.181	**0.425**	-0.090	2.235	1.378
JORIIT	0.017	**-0.648**	-0.010	1.143	1.080
RASKVAI	0.016	-0.061	**0.720**	1.049	0.890
HELPPOA	0.010	-0.098	**-0.225**	1.823	1.222

Esimerkiksi faktoripistemäärämuuttuja F1 on muodostettu seuraavasti:

$$\begin{aligned} \mathrm{F1} = \; & 0.044 * (\tfrac{PARHAAT-1.477}{1.098}) + (-0.620) * (\tfrac{PITKAST-1.387}{1.085}) \\ & +0.095 * (\tfrac{HAASTAV-1.846}{1.257}) + \ldots + 0.010 * (\tfrac{HELPPOA-1.823}{1.222}). \end{aligned}$$

6 Suhtautuminen koulutukseen -rakenneyhtälömallin rakentaminen

6.1 Muuttujat ja oletusten tarkastelu

Tarkoituksena on seuraavaksi tutkia rakenneyhtälömallien avulla, mitkä tekijät vaikuttuvat koulutukseen suhtautumiseen aikuisiässä eli saatuihin faktoripistemäärämuuttujiin F1, F2 ja F3. Kiinnostavia muuttujia on valittu 30 vuoden seuruututkimuksen eri vaiheista. Muuttujiksi kultakin mittauskerralta on pyritty valitsemaan keskeiset, elämän rakentumisen kannalta tärkeimmät muuttujat. Alkumittaus-hetkellä, jolloin tutkimushenkilöt olivat 3 - 9 -vuotiaita, on valittu muuttujat:

SP	=	Sukupuoli (1=poika ja 2=tyttö)
ITPA	=	ITPA-testin kokonaispistemäärä (0-99)
STATUS	=	STATUS-luokka (1=alin, 2=keski ja 3=ylin)
IAYRI72	=	Isän äyrit vuodelta 1972

Peruskouluajalta valittiin vain muuttuja:

KA = Kaikkien aineiden keskiarvo 9. luokalla (4.5-9.9)

Keskiaste-ajalta otettiin mukaan niinikään vain yksi muuttuja:

$$\text{YO} = \text{Ylioppilas (1=ei ja 2=kyllä)}$$

Aikuisikä-ajalta, jolloin tutkittavat olivat 23-27 -vuotiaita, mitattuja muuttujia ei otettu faktoripistemäärien lisäksi malliin mukaan ollenkaan, sillä samasta mittausajankohdasta johtuen kausaalisuussuhteiden määrittäminen koettiin ongelmalliseksi.

Valitut muuttujat ovat sekä diskreettejä että jatkuvia. Täten korrelaatiokertoimina on käytetty polykorisia, polyserialisia sekä Pearsonin korrelaatiokertoimia (Liite 2). Korrelaatiokertoimien laskemisessa ja puuttuvien havaintojen käsittelyssä on käytetty pairwise-menetelmää, jonka mukaan otoskoot vaihtelivat välillä $384 \leq N \leq 629$. Keskimääräiseksi otoskooksi on valittu 416, koska faktoripistemäärämuuttujat ovat muodostuneet kyseisellä otoskoolla.

Ajallisista eroavaisuuksista johtuen muuttujat on valittu analysoitavaksi siten, että muuttujat varhaislapsuudesta (1970-73) eli SP, ITPA, STATUS ja IAYRI72 ovat selittäviä, eksogeenisiä x-muuttujia ja KA, YO, F1 ja F3 ovat selitettäviä, endogeenisiä y-muuttujia. Muuttuja F2 on poistettu lopullisesta analyysistä, koska siihen kohdistuneet yhteydet muista muuttujista eivät olleet tilastollisesti merkitseviä.

Estimointimenetelmän valitsemiseksi on suoritettu jälleen muuttujien normaalisuustestaus. Rakenneyhtälömalleissa ollaan nyt kiinnostuneita vain selitettävien muuttujien jakaumista ja mitta-asteikoista. Taulukon 6.1 testisuureista ja niitä vastaavista p-arvoista huomataan, että SU-estimoinnin multinormaalisuusoletus ei ole voimassa, vaikka jotkut selitettävistä muuttujista ovatkin normaalisti jakaantuneita. Estimointimenetelmänä tullaan siten käyttämään yleistettyä PNS-estimointia.

Taulukko 6.1 Normaalisuuden testaus, testisuure ja p-arvo

Y_i	Kolmogorov-Smirnov Z	p-arvo
F1	0.961	0.314
F2	1.972	0.001
F3	3.311	0.000
KA	1.214	0.105

6.2 Rekursiivisen rakenneyhtälömallin rakentaminen

Rekursiivisen rakenneyhtälömallin rakentaminen noudattaa samoja tutkimusvaiheita kuin konfirmatorisen faktorimallinkin rakentaminen. Mallin spesifiointi lähtökohtatilanteessa on suoritettu jälleen valitsemalla mahdollisimman yksinkertaisen rakenteen omaava rakenneyhtälömalli. Lopullisessa mallissa on tilastollisesti merkitsevien rakenneparametrien lisäksi estimoitu rakenneyhtälöiden (F1 ja F3) välinen jäännösten kovarianssi. Vapautus on sallittua, sillä kyseisiä muuttujia ei pyritä selittämään keskenään. Estimoidut tulokset ovat taulukossa 6.2 sekä graafisesti esitettynä kuviossa 6.1. Lisäksi muuttujien väliset epäsuorat vaikutukset ovat taulukossa 6.3.

Koska muuttujien jäännöskovarianssimatriisi Ψ ei ole yhden vapautuksen takia diagonaalinen, täytyy mallin täyttää ehto (3.3) identifioituvuudelle. Koska estimoitavien parametrien lukumäärä 14 matriiseissa $\beta-, \gamma-$ ja $\psi-$ on pienempi kuin yhtälöiden lukumäärä 45 kovarianssimatriiseissa Σ_{yy},Σ_{xy} ja Σ_{xx}, on estimoitava rekursiivinen rakenneyhtälömalli identifioituva.

Rekursiivisen rakenneyhtälömallin riittävyystarkastelut osoittavat mallin riittäväksi. Koko mallia mittaavat riittävyysindeksit χ^2-testi, NFI, GFI, RMSEA ja RMR täyttävät kaikki mallin riittävyyteen vaadittavat ehdot (ks. kappale 2.1.5). Yhtälökohtaiset selitysasteet $R^2_{y(i)}$ on esitetty taulukossa 6.2. Pienet selityskertoimet johtuvat todennäköisesti suuresta mittausajankohtavälistä. Noin 20 vuotta tuntuu olevan liian suuri mittausväli tutkittaessa koulutukseen suhtautumista aina lapsuudesta aikuisuuteen. Myös parametri- ja havaintokohtaiset riittävyystarkastelujen mukaan mallia voidaan pitää riittävänä.

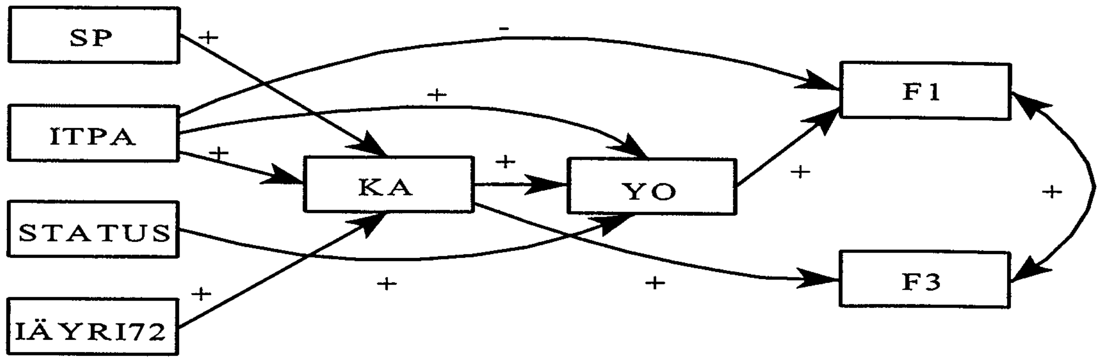
Taulukko 6.2 Rakenneparametrien estimaatit(suluissa keskivirheet), riittävyystarkastelut ja estimoitu jäännöskovarianssi

	KA	YO	F1	F3
SP	.353(.040)	0*	0*	0*
ITPA	.394(.041)	.116(.027)	-.200(.052)	0*
STATUS	0*	.095(.024)	0*	0*
IAYRI72	.160(.041)	0*	0*	0*
KA	0*	.792(.026)	0*	.244(.047)
YO	0*	0*	.308(.055)	0*
F1	0*	0*	0*	0*
F3	0*	0*	0*	0*
$\hat{\psi}(s.e.)$	.646(.046)	.217(.015)	.904(.064)	.921(.065)
$\hat{R}^2_{y(i)}$	.349	.780	.074	.060
*) kiinnitetty				

Riittävyystarkastelut:

$\chi^2(12) = 17.249$, p-arvo=.14, GFI=.99, NFI=.91, RMSEA=.0326, RMR=.028

Estimoitu yhtälöiden välinen jäännösten kovarianssi: $\hat{\psi}_{43}$=.320(.049)



Kuvio 6.1 Lopullisen rakenneyhtälömallin graafinen esitys

Taulukko 6.3 Epäsuorat vaikutukset.

	YO	F1	F3
SP	.280(.033)	.086(.019)	.086(.019)
ITPA	.312(.034)	.132(.027)	.096(.021)
STATUS		.029(.009)	
IAYRI72	.127(.033)	.039(.012)	.039(.013)
KA		.244(.044)	

Taulukon 6.2 tuloksista ja kuviosta 6.1 saadaan sisällölliset tutkimuksen lopputulokset. Alkumittauksen(1970-73) ja peruskoulun(1984) välinen rakenne muodostui seuraavaksi: mitä paremmat ovat verbaalisuutta mittavaan lahjakkuustestin(ITPA) tulokset ja isän tulot, sitä parempi on kaikkien aineiden keskiarvo 9. luokalla. Lisäksi saatiin selville, että sukupuolella on vaikutusta 9. luokan keskiarvoon: tytöt saavat parempia keskiarvoja kuin pojat. Alkumittauksen ja keskiasteen(1984) välisistä yhteyksistä saatiin selville, että korkeat lahjakkuustestin pistemäärät auttavat ylioppilaaksi pääsemistä. Toiseksi huomataan, että mi-

tä korkeampi on status-luokka, sitä paremmat mahdollisuudet on päästä ylioppilaaksi. Alkumittauksen ja aikuismittauksen(1991) väliltä saatiin mielenkiintoinen tutkimustulos. Sen mukaan lahjakkuustestissä menestyneet kokevat koulutukseen suhtautumisen kielteisenä. Peruskoulun ja keskiasteen välinen yhteys on ilmeinen. Mitä parempi on keskiarvo, sitä paremmat mahdollisuudet päästä ylioppilaaksi. Myös peruskoulun ja aikuisiän riippuvuus on looginen. Mitä parempi keskiarvo 9. luokalla, sitä helpompaa on koulunkäynti tulevaisuudessa. Lisäksi huomattiin, että ylioppilaat kokevat koulutuksen myönteisempänä kuin ei-ylioppilaat.

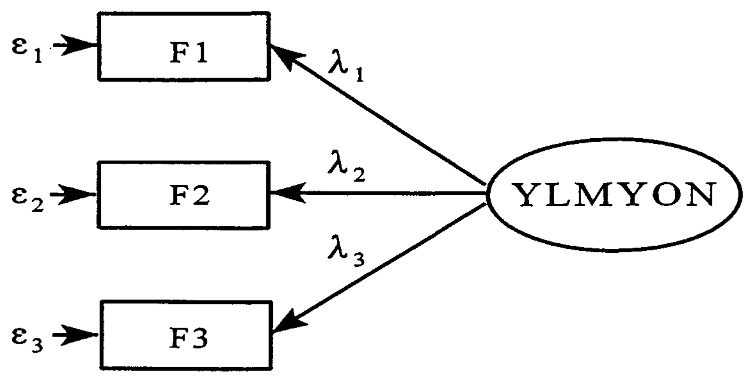
7 Suhtautuminen koulutukseen -toisen kertaluvun faktorimalli ja rakenneyhtälömalli

Kappaleessa 6.2 suoritetussa rekursiivisessa rakenneyhtälömallissa ei siis huomioitu faktoripistemäärämuuttujan F2 vaikutusta koulutukseen suhtautumisen kuvaajana, sillä vain muuttujiin F1 ja F3 kohdistui tilastollisesti merkitseviä vaikutuksia. Ottamalla kuitenkin muuttujan F2 vaikutus mukaan rekursiiviseen rakenneyhtälömalliin, tehdään muuttujille F1, F2 ja F3 uusi faktorianalyysi ja muodostetaan uusi faktoripistemäärämuuttuja. Näin saadaan koulutukseen suhtautumisesta yleisempi kokonaisuus, jonka vaikutusta rakenneyhtälömallissa tarkastellaan kappaleen loppuosassa.

7.1 Toisen kertaluvun faktorimallin sovellus

Kuten huomattiin jo taulukon 5.3 faktoreiden korrelaatiomatriisista, faktorit korreloivat positiivisesti ja olivat suuruudeltaan kohtalaisia. Tämä antaa syyn olettaa, että faktoreilla MYONTEIS, TARPEELL ja HELPPOUS olisi vielä laajempi, kokonaisvaltaisempi faktori, joka kuvaa koulutukseen suhtautumista aina lapsuudesta aikuisuuteen. Tutkimuksen ajattelun lähtökohtana voidaan pitää kuvion

7.1 toisen kertaluvun faktorimallia, jossa kolmen latentin muuttujan taustalla on vielä yksi yhdistävä latentti muuttuja.



Kuvio 7.1 Sovellettu ensimmäisen kertaluvun faktorimalli

Toisen kertaluvun faktorimallia on siis sovellettu käyttämällä ensimmäisen kertaluvun faktoreiden sijasta kappaleessa 5.4 estimoituja faktoripistemäärämuuttujia F1, F2 ja F3, joille on puolestaan suoritettu oma konfirmatorinen faktorianalyysi. Toisen kertaluvun faktorimallin rakentaminen on vaiheiltaan analogista ensimmäisen kertaluvun konfirmatorisen faktorimallin rakentamisen kanssa(ks. kappale 2.1).

Kuten taulukon 6.1 normaalisuustestauksen tuloksista nähtiin, faktoripistemäärämuuttujat eivät noudata moniulotteista normaalijakaumaa. Täten taulukon 7.1 lataukset on estimoitu käyttäen yleistettyä pienimmän neliösumman estimointimenetelmää. Taulukon 7.1 latauksista ja reliabiliteettikertoimista nähdään, että faktoripistemäärämuuttujat F1 ja F2 mittaavat parhaiten yleistä myönteisyyttä (YLMYON). Tässä tapauksessa malli on saturoitu ja faktorimallin sovite on siten täydellinen.

Taulukko 7.1 Estimoidut lataukset(suluissa keskivirheet), jäännösvarianssit(suluissa keskivirheet) ja muuttujien reliabiliteetit.

	YLMYON	$\hat{\theta}$(s.e.)	$\hat{R}^2_{x(i)}$
F1	.672(.063)	.548(.073)	.452
F2	.669(.063)	.553(.073)	.447
F3	.525(.058)	.724(.063)	.276

7.2 Faktoripistemäärämuuttujan estimointi ja rekursiivinen rakenneyhtälömalli

Rekursiivisen rakenneyhtälömallin rakentamista varten on sovelletusta toisen kertaluvun faktorimallista estimoitu uusi faktoripistemäärämuuttuja:

FYLMYON = Yleinen myönteisyys -faktoripistemäärä,

joka on estimoitu kertomalla taulukon 7.2 faktoripainokertoimet faktoripistemäärämuuttujien F1, F2 ja F3 havaintoarvoilla. Faktoripainokertoimet on estimoitu regressiomenetelmän avulla.

Taulukko 7.2 Toisen kertaluvun faktoripainokertoimet

	YLMYON
F1	.407
F2	.402
F3	.240

Rekursiivisen faktorimallin rakentamista varten muuttujat on asetettu analysoitavaksi periaatteeltaan lähes kappaleen 6.1 mukaisesti; eksogeeniset muuttujat (SP, ITPA, STATUS ja IAYRI72) ovat edelleen samoja, mutta endogeenisinä muuttujina on nyt KA:n ja YO:n lisäksi vain uusi faktoripistemäärämuuttuja FYLMYON. Muuttujien välinen korrelaatiomatriisi on liitteessä 2.

Faktoripistemäärämuuttujan FYLMYON jakauma ei ole normaalinen; Kolmogorov-Smirnovin testisuureen arvo 1.437 antaa p-arvoksi 0.032. Rakenneyhtälömallin endogeenisten muuttujien multinormaalisuusoletus ei ole siis voimassa. Siten rakenneparametrien estimoinnissa käytetään yleistettyä PNS-estimointimenetelmää.

Mallin spesifioinnin lähtökohtana oli kappaleen 6.2 tuottama rekursiivinen rakenneyhtälömalli. Tosin aikuisikäajalta oli tarkasteltavana vain muuttuja FYLMYON. Estimoitu malli on tässä tapauksessa identifioituva, sillä sen jäännöskovarianssimatriisi Ψ on diagonaalinen.

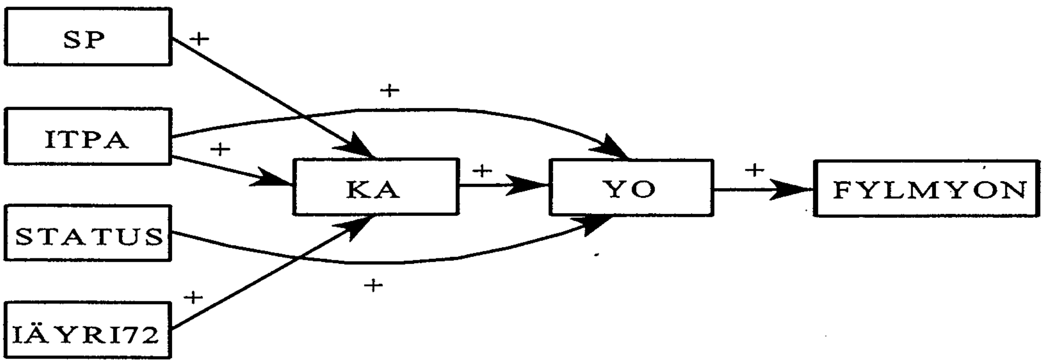
Koko mallia koskevat riittävyysindeksit, kuten $\chi^2(8) = 8.641$, p-arvo=.373, GFI=.994 ja NFI=.971 osoittavat mallin hyvää yhteensopivuutta havaintoaineiston kanssa. Muuttujakohtaisten riittävyystarkastelujen selityskertoimet $R^2_{y(i)}$ osoittavat, että erot ensimmäisen kertaluvun rakenneyhtälömalliin ovat hyvin pieniä. Esimerkiksi muuttujan FYLMYUON selityskerroin 0.033 osoittaa edelleen todennäköisesti liian pitkää aikaväliä mitattaessa koulutukseen suhtautumista aina lapsuudesta aikuisuuteen. Parametrien estimaatit, jotka ovat taulukossa 7.3, ovat kaikki tilastollisesti merkitseviä. Graafisesti esitettynä estimoitu malli on kuviossa 7.2.

Taulukko 7.3 Rakenneparametrien estimaatit(suluissa keskivirheet) ja riittävyystarkastelut.

	KA	YO	FYLMYON
SP	.340(.040)	0*	0*
ITPA	.389(.041)	.113(.027)	0*
STATUS	0*	.091(.024)	0*
IAYRI72	.164(.041)	0*	0*
KA	0*	.794(.026)	0*
YO	0*	0*	.182(.049)
F1	0*	0*	0*
F3	0*	0*	0*
$\hat{\psi}(s.e.)$	.656(.046)	.219(.015)	.945(.067)
$\hat{R}^2_{y(i)}$	.337	.778	.033
$*$) kiinnitetty			

Riittävyystarkastelut:

$\chi^2(8) = 8.641$, p-arvo=.373, GFI=.994, NFI=.971, RMSEA=.014, RMR=.025



Kuvio 7.2 Toisen kertaluvun rakenneyhtälömallin graafi

Kuten taulukon 7.3 tuloksista nähdään, malli ei eroa paljoakaan eksogeenisten ja endogeenisten muuttujien välisten parametrien estimaattien osalta verrattuna aikaisempaan malliin (taulukko 6.2). Uusia rakenneparametreja ei ole estimoitu, mutta ITPA-muuttujan ja aikuisiän faktoripistemäärämuuttujan välinen suora vaikutus on jouduttu kiinnittämään nollaksi. Kaikki estimoitujen rakenneparametrien estimaattien etumerkit ovat samanmerkkisiä mallien vertailussa. Endogeenisten muuttujien välillä suorista yhteyksistä ei ollut nollasta eroavia kuin jo aiemmin todettu muuttujien KA ja YO välinen sekä uutena muuttujien YO ja FYLMYON yhteys. Sen mukaan ylioppilaat suhtautuvat koulutukseen yleisesti myönteisemmin kuin "ei-ylioppilaat". Epäsuorien vaikutusten osalta ei myöskään tullut merkittäviä eroja suhteessa edelliseen malliin. Epäsuorien vaikutusten parametrien estimaatit keskivirheineen ovat taulukossa 7.4.

Taulukko 7.4 Epäsuorien vaikutusten estimaatit(suluissa keskivirheet).

	YO	FYLMYON
SP	.270(.033)	.049(.015)
ITPA	.309(.034)	.077(.022)
STATUS		.017(.006)
IAYRI72	.130(.033)	.024(.009)
KA		.145(.039)

8 Yhteenveto

Tutkielman teoriosassa käsiteltiin konfirmatorisia faktorimalleja ja niiden rakentamisen vaiheita. Konfirmatorista faktorianalyysiä laajennettiin vielä faktoripistemäärämuuttujien estimointiin sekä toisen kertaluvun faktorimalleihin. Lisäksi esitettiin teoreettisesti lähemmin rekursiivisia ja simultaanisia rakenneyhtälömalleja.

Tutkielman empiirisen osan aineistona käytettiin 30 vuoden pitkittäistutkimusaineistoa, jossa keskityttiin vain pieneen osaan suurta aineistoa. Tarkoituksena oli tutkia millä faktoreilla voidaan kuvata koulutukseen suhtautumista aikuisiässä, jolloin tutkittavat olivat 24-30 vuotiaita, ja toiseksi, mitkä tekijät vaikuttavat koulutukseen suhtautumiseen aina lapsuudesta aikuisuuteen. Tarkoituksena oli etsiä vain elämän kulun kannalta tärkeitä tekijöitä.

Alustavana analyysinä suoritettiin 11 muuttujalle eksploratiivinen faktorianalyysi, josta saatujen tulosten perusteella suoritettiin konfirmatorinen faktorianalyysi. Analyysi jakoi koulutukseen suhtautumista kuvaavat muuttujat selkeästi kolmeen eri faktoriin: myönteisyys-, tarpeellisuus- sekä helppous-faktoriin.

Seuraavaksi konfirmatorisen faktorianalyysin tuottamia tuloksia käytettiin faktoripistemäärämuuttujien avulla rekursiivisten rakenneyhtälömallien rakentamiseen. Rakenneyhtälömalleja tehtiin kahdenlaisia. Ensimmäisenä etsittiin koulutukseen suhtautumiseen muodostettuihin myöteisyys-, tarpeellisuus,- ja helppous-faktoripistemäärämuuttujiin vaikuttavia tekijöitä. Varhaislapsuudesta, kun lapset olivat 3-9 vuotiaita, koulutukseen suhtautumiseen vaikuttavia tekijöitä löydettiin neljä: sukupuoli, ITPA -älykkyystestin tulokset, status-luokka sekä isän tulot. Peruskouluajalta ja keskikouluajalta, jolloin koehenkilöt olivat 17-23 vuotiaita, merkitykselliseksi tekijöiksi todettiin 9. luokan keskiarvo ja tieto ylioppilaaksi pääsystä.

Toinen rakenneyhtälömalli rakennettiin soveltaen toisen kertaluvun faktorimalliajattelua. Kolmesta korreloivasta faktoripistemäärämuuttujasta muodostettiin uusi faktoripistemäärämuuttuja, joka tulkittiin toisen kertaluvun faktoriksi. Tämä faktori nimettiin yleiseksi myönteisyydeksi koulutukseen suhtautumista kohtaan. Erot ensimmäiseen rekursiiviseen rakenneyhtälömalliin olivat pieniä. Molemmissa malleissa rakenne pysyi pääpiirteiltään samankaltaisena. Ensimmäinen rakenneyhtälömalli havaittiin kuitenkin toista informatiivisemmaksi. Molemmista malleista kävi kuitenkin ilmi, että noin 20 vuotta koulutukseen suhtautumisen tutkimiseen on liian suuri aikaväli. Selitettävien faktopistemäärämuuttujien vaihtelusta pystyttiin selittämään vain hyvin pieni osa.

Lähteet:

Akaike, H. (1974): A new look at statistical model identification. *IEEE transactions on Automatic Control, 19*, 716-723.

Akaike, H. (1987): Factor analysis and AIC. *Psychometrica 52(3)*, 317-332.

Alanen, E., Kinnunen, V. ja Leskinen, E. (1984): *Eri-ikäisten miesten psyykkisen hyvinvoinnin rakenteen kuvailusta ja selittämisestä LISREL-mallien avulla.* Jyväskylän yliopisto. Psykologian laitoksen julkaisuja 265.

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

Blåfield, L. & Kuusinen, J. (1974): *Suomalaisen ITPAn psykometriset ominaisuudet.* Jyväskylä: Kasvatustieteiden tutkimuslaitoksen julkaisuja 241.

Boomsma, A. (1982): The robustness of LISREL against small sample sizes in factor analysis models. In Jöreskog, K.G. ja Wold, H. (ed.): *System Under Indiaect Observation, Causality-Structure-Prediction, Part I*, 149-174.

Browne, M.W. (1984): Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psycholog, 37*, 62-83.

Grönroos, M. (1984): *Confirmatory factor analysis in small samples: a Monte Carlo study of LISREL solutions.* University of Turku. Department of Statistics. Publications 1984-1.

Harman, H.H. (1967): *Modern Factor Analysis.* Second Edition. Chigago: University of Chigago Press

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

Jöreskog, K.G. (1979): Author's Addendum February, 1979. In Jöreskog, K.G. & Sörbom, D. (1979): *Advances in Factor Analysis and Structural Equation Models*. Cambridge, Mass: Abt Books, 40-43.

Jöreskog, K.G. (1981): Basic Issues in the application of LISREL. *Data 1*, 1-6.

Jöreskog, K.G. (1999): Karl's Corner: *Interpretation of R-square*. http://www.ssicentral.com/lisrel/column3.htm. (4.4.2000)

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. Reseach Report 81-8.

Jöreskog, K.G & Sörbom, D. (1993): *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chigago: Scientefic Software International, Inc.

Jöreskog, K.G & Sörbom, D. (1996): *LISREL 8: User's reference guide*. Chigago: Scientefic Software International, Inc.

Jöreskog, K.G, Sörbom, D., du Toit, M., du Toit, S. (1999): *LISREL 8: New Statistical Features*. Chigago: Scientefic Software International, Inc.

Kuusinen, J. & Blåfield, L. (1974): *Suomalaisen ITPAn faktorirakenne*. Jyväskylä: Kasvatustieteiden tutkimuslaitoksen julkaisuja 225.

Lawley, D.N. & Maxwell, A.E. (1971): *Factor Analysis as a Statistical Method*. London: Butterworths.

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

Leskinen, E. (1983b): *On third-order factor analysis models*. University of Jyväskylä. Department of Statistics. Reports on Statistics 9/1983.

Leskinen, E. (1987): Faktorianalyysi. Konfirmatoristen faktorimallien teoria ja rakentaminen. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja 1/1987.

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

Steiger, J.M. (1990): Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Sörbom, D. (1986): *Model modification*. University of Uppsala. Department of Statistics. Reseaerch Report 86-3.

Tanaka J.S, & Huba, G.J. (1984): *Confirmatory hierarchical factor analysis of psychological distress measures*. Journal of Personality and Social Psychology, 46, 621-635.

Thurstone, L.L. (1935): *The vectors of mind*. Chigago: University of Chigago Press.

Liite 1. Konfirmatorisen faktorianalyysin korrelaatiomatriisi (N=416)

	PARHAAT	PITKAST	HAASTAV	VALMIUD	RASKVAI	VIERAST
PARHAAT	1.000					
PITKAST	-0.408	1.000				
HAASTAV	0.239	-0.576	1.000			
VALMIUD	0.201	-0.248	0.384	1.000		
RASKVAI	-0.127	0.312	-0.206	-0.068	1.000	
VIERAST	-0.176	0.357	-0.271	-0.411	0.154	1.000
TURHAA	-0.267	0.575	-0.471	-0.393	0.306	0.389
KUNNOL	0.140	-0.462	0.359	0.175	-0.206	-0.183
TARPEEN	0.165	-0.173	0.177	-0.080	-0.242	0.049
JORIITT	-0.184	0.273	-0.153	0.108	0.299	0.092
HELPPOA	0.118	-0.214	0.071	-0.012	-0.538	-0.061

	TURHAA	KUNNOL	TARPEEN	JORIITT	HELPPOA
TURHAA	1.000				
KUNNOL	-0.291	1.000			
TARPEEN	-0.227	0.180	1.000		
JORIITT	0.282	-0.127	-0.703	1.000	
HELPPOA	-0.123	0.123	0.226	-0.344	1.000

Liite 2. Rakenneyhtälömallien korrelaatiomatriisi (N= 416)

	KA	YO	F1	F2	F3	FYLMYON
KA	1.000					
YO	0.871	1.000				
F1	0.152	0.194	1.000			
F2	0.066	0.061	0.450	1.000		
F3	0.237	0.187	0.353	0.351	1.000	
FYLMYON	0.175	0.179	0.812	0.826	0.636	1.000
SP	0.355	0.278	0.179	0.055	0.121	0.149
ITPA	0.445	0.497	-0.030	0.067	0.156	0.064
STATUS	0.287	0.358	0.040	0.000	0.120	0.054
IAYRI72	0.248	0.271	-0.006	0.010	0.094	0.029

	SP	ITPA	STATUS	IAYRI72
SP	1.000			
ITPA	0.051	1.000		
STATUS	0.038	0.335	1.000	
IAYRI72	-0.023	0.238	0.526	1.000