

**Peruskoululaisten yhteiskunnallisen osaamisen
estimointi sekamallilla: sovellus ICCS 2009
-aineistoon**

Tilastotieteen pro gradu -tutkielma

Tuomo Huttu
Matematiikan ja tilastotieteen laitos
Jyväskylän yliopisto
7.6.2013

JYVÄSKYLÄN YLIOPISTO
Matematiikan ja tilastotieteen laitos

Huttu, Tuomo: Peruskoululaisten yhteiskunnallisen osaamisen estimointi sekamallilla: sovellus ICCS 2009 -aineistoon

Tilastotieteen pro gradu -tutkielma, 33 sivua, 9 liitettä
7.6.2013

TIIVISTELMÄ

Nuorten yhteiskunnallisia tietoja ja eri yhteiskunnallisten ilmiöiden ymmärtämistä, samoin kuin kansalaisuuteen ja yhteiskunnassa toimimiseen liittyviä asenteita tutkimalla voidaan selvittää kansalaisyhteiskunnan tulevaisuuden kannalta merkittäviä kysymyksiä. Nuorten halukkuuteen ja kiinnostukseen osallistua yhteiskunnalliseen toimintaan vaikuttavat nuorten yksilöllisen ja sosiaalisen taustan piirteet, kuten sukupuoli ja kielitausta. Kansainvälisessä ICCS-tutkimuksessa pyrittiin selvittämään piirteiden vaikutusta nuorten yhteiskunnallisiin tietoihin, osallistumiseen ja asenteisiin.

ICCS 2009 -aineiston pohjalta tutkittiin Suomen sisäistä vaihtelua nuorten yhteiskunnallisissa tiedoissa ja taidoissa sekä tähän vaihteluun vaikuttavia yksilöllisiä ja sosiaalisia piirteitä 8. luokkalaisilla nuorilla. Tutkimukseen osallistuneet oppilaat eivät vastanneet kaikkiin koekysymyksiin, vaan kysymykset jaettiin koevihkoihin, joista kukin oppilas sai ratkaistavakseen yhden. Tehtävien skaalauksessa käytettiin MRCML-mallia. Vastemuuttujina tutkimuksessa ovat osaamista kuvaavat Plausible Value -arvot, jotka on generoitu jokaiselle oppilaalle tämän taustatietojen ja koevastausten perusteella mallinnetusta posteriorijakaumasta. Taustamuuttujina ovat sukupuoli, kotona käytettävä kieli, viisi pääkomponenttia (kiinnostus politiikkaan, suhtautuminen tasa-arvoon, kodin sosioekonominen asema, demokraattisuus ja olosuhteet koulussa), osallistuminen koulun toimintaan, osallistuminen järjestötoimintaan, osallistuminen laittomaan protestointiin ja osallistuminen lailliseen protestointiin.

Mallinnus suoritettiin koulu- ja oppilasvaikutuksen sisältävällä varianssikomponenttimallilla, ja tuloksena saatiin 12 päävaikutusta ja seitsemän yhdysvaikutusta. Suomen koulujen välinen varianssikomponentti on pieni suhteessa kokonaisvarianssiin eli Suomen koulujen välinen vaihtelu on vähäistä. Yhdysvaikutuksia ei tule jättää pois mallinnuksesta niiden tuoman tilastollisesti merkitsevän informaation takia.

Avainsanat: Raschin malli, Plausible Value, sekamalli, ICCS 2009.

Sisältö

1	Johdanto	1
2	Tutkimusaineisto	3
2.1	Otanta	3
2.2	ICCS-tutkimuksen kysymykset	4
2.3	Muuttujat	5
3	Oppilaiden osaamista mittaavat pistemäärät	7
3.1	Raschin malleista	7
3.2	MRCML-malli	9
3.3	Populaatiomalli	11
3.4	Plausible Value -muuttujat	11
3.4.1	Monte Carlo -approksimointi	14
3.4.2	PV-arvojen generoiminen	16
4	Lineaarisisista malleista	17
4.1	Kiinteiden vaikutusten malli	17
4.1.1	Parametrien estimointi	17
4.2	Lineaarinen sekamalli	18
4.2.1	Parametrien estimointi	18
4.2.2	Sisäkorrelaatiosta	19
4.3	Pääkomponenttianalyysi	20
4.4	Uskottavuusosamäärätesti	20
5	Aineiston analysointi	21
5.1	Mallin valinta	21
5.2	Mallin estimaattien tulkinta	21
5.3	Diagnostiikkaa	28
6	Yhteenveto	30
	Viitteet	32
	Liitteet	34

1 Johdanto

ICCS 2009¹ on IEA-järjestön² organisoima kansainvälinen tutkimus, jossa selvitettiin nuorten yhteiskunnallista osaamista, osallistumista ja asenteita (Suoninen ym., 2010). Tutkimuksessa keskityttiin nuorten yhteiskunnallisiin tietoihin ja taitoihin, kansalaisuuteen asennoitumiseen ja haluun sitoutua aktiiviseen kansalaisuuteen. Vastauksien avulla haluttiin selvittää maiden välistä ja sisäistä vaihtelua nuorten tiedoissa, mahdollisia muutoksia CIVED 1999 -tutkimukseen³ verrattuna, nuorten kiinnostusta julkiseen ja poliittiseen toimintaan, nuorten kokemuksia kansalaisyhteiskuntaa uhkaavista tekijöistä, koulujen yhteyttä nuorten osaamiseen ja asenteisiin sekä nuorten taustaa kuvaavien piirteiden yhteyttä oppimistuloksiin ja asenteisiin.

ICCS 2009 -tutkimuksen aineisto saatiin koe- ja kyselylomakkeilla, jotka oppilaiden ja opettajien tuli täyttää. Koetehtävät on jaettu seitsemään koeviikkoon niin, että kukin oppilas vastasi niistä yhteen. Tutkimustietoa kerättiin 8. luokkalaisilta oppilailta ja opettajilta 38 eri maassa. Opettajille osoitetuissa kyselyissä kerättiin tietoa opettajien ammatillisesta taustasta ja heidän käsityksistään koulujen yhteiskunnallisesta opetuksesta ja kasvatuksesta. Opettajille suunnatuilla kysymyksillä pyrittiin saamaan tietoa siitä ympäristöstä, jossa nuoret oppivat yhteiskuntaan liittyviä asioita. Suomessa kyselyyn osallistui 186 koulusta 3307 oppilasta ja 2295 opettajaa. ICCS 2009 -tutkimuksessa saatujen vastausten perusteella suomalaisten nuorten yhteiskunnallinen tietämys on huippuluokkaa (Suoninen ym., 2010) ja maan sisäinen vaihtelu eri koulujen välillä on pientä.

Tässä tutkielmassa on ICCS 2009 -aineiston pohjalta analysoitu yksilöllisten ja sosiaalisten taustojen, kuten sukupuolen, kielitaustan ja sosioekonomisen taustan yhteyttä nuorten oppimistuloksiin ja asenteisiin sekä tarkasteltu Suomen sisäistä vaihtelua nuorten yhteiskunnallisissa tiedoissa ja taidoissa. ICCS 2009 -tutkimuksen päätuloksissa (Suoninen ym., 2010) eri taustamuuttujien vaikutuksia on tutkittu erikseen. Tässä työssä suurin osa aineiston muuttujista on yhdistetty pääkomponenteiksi ja niitä tarkastellaan samanaikaisesti valittujen sukupuoli- ja kielitaustamuuttujan kanssa. Samalla tutkitaan sukupuolen sekä kielimuuttujan ja muiden taustamuuttujien välisiä yhdysvai-

¹International Civic and Citizenship Education Study (suomeksi *Nuorten yhteiskunnallisen osaamisen, osallistumisen ja asenteiden tutkimus*)

²International Association for the Evaluation of Educational Achievement

³Civic Education Study

kutuksia. MRCML-mallin⁴ avulla saatiin määriteltyä oppilaan suorituskyyky tehtävien jokaiselle vastaustasolle. Oppilaiden osaamisen todennäköisyystiheyden ja taustamuuttujia hyödyntävän kykypriorin avulla saadaan laskettua oppilaiden kykyjen posteriorijakaumat, joista on generoitu oppilaan osaamista kuvaavat Plausible Value -muuttujat. Havaintojen mahdollisen korrelaation (sekä koulujen että oppilaiden sisällä) takia tässä työssä käytettiin mallinnusmenetelmänä varianssikomponenttimallia, joka sisältää sekä koulu- että oppilasvaikutuksen.

⁴The multidimensional random coefficients multinomial logit model

2 Tutkimusaineisto

ICCS-tutkimus selvittää nuorten asennoitumista kansalaisuuteen ja kansalaisyhteiskuntaan, sekä antaa tietoa nuorten valmiudesta ja halukkuudesta toimia itsenäisinä ja oma-aloitteisina jäseninä yhteiskunnassa. ICCS-tutkimukseen osallistui 38 maata tai aluetta ja se on osallistujamaiden yhteishanke. Tässä pro gradu -tutkielmassa analysoidaan vuoden 2009 Suomen aineistoa.

2.1 Otanta

Perusjoukkona tutkimuksessa on vähintään 13,5-vuotiaat 8. luokkalaiset oppilaat. Perusjoukon ulkopuolelle rajattiin Ahvenanmaan koulut ja erityiskoulut. Otanta suoritettiin kaksivaiheisena (Suoninen ym., 2010). Ensimmäisessä vaiheessa valittiin koulut ja toisessa vaiheessa kustakin koulusta valittiin yksi 8. vuosiluokan opetusryhmä. Suomi on jaettu viiteen alueeseen (Etelä-, Länsi-, Itä-, Pohjois-Suomi ja ruotsinkielinen alue). Alueilta valittavien koulujen lukumäärät olivat suhteessa alueiden 8. luokan oppilasmääriin. Kustakin alueesta koulut poimittiin PPS-otannalla⁵. Kustakin valitusta koulusta poimittiin yksi 8. luokan opetusryhmä yksinkertaisella satunnaisotannalla. Lopullinen otoskoko oli 3496 oppilasta, jotka valittiin 186 eri koulusta. Kussakin maassa otokseen tuli kuulua vähintään 150 koulua ja 3000-4500 oppilasta. Suomessa pieni osa (noin 1 %) oppilaista jäi tutkimuksesta pois toiminnallisen rajoitteen, vakavan oppimisrajoitteen tai kielitaitorajoitteen takia. Tutkimus tehtiin vuoden 2009 keväällä. Kaikki otokseen valitut koulut ja oppilaat eivät kuitenkaan päässeet osallistumaan tutkimukseen, joten Suomessa ICCS-tutkimukseen osallistui 176 yläkoulua ja 3307 oppilasta (94,6 % otoksesta).

Aineiston havaintomäärä oli kaiken kaikkiaan 3307 oppilasta. Oppilaista 468:n (noin 14,2 %) muuttujien arvot olivat puutteellisia joidenkin valittujen muuttujien osalta, ja nämä oppilaat poistettiin kokonaan tästä tutkielmasta. Tutkielman aineistoon jäi siis 2839 oppilasta.

Aineiston laatu varmistettiin toteuttamalla kaikki tutkimuksen vaiheet kansainvälisten laatustandardien mukaan ja dokumentoimalla ne huolellisesti. Tutkimus järjestettiin jokaisessa koulussa tähän tehtävään koulutetun henkilön (esimerkiksi koulun opinto-ohjaajan) johdolla. Joka kymmenennessä koulussa oli myös kansallinen tarkkailija arvioimassa tutkimuksen toteutus-

⁵Probability Proportional to Size

ta. Kansainvälisen ohjeistuksen pohjalta koulutettu arviointiryhmä pisteytti avoimet tehtävät.

2.2 ICCS-tutkimuksen kysymykset

Tutkimusaineisto kerättiin koe- ja kyselylomakkeilla koulujen oppilailta, opettajilta ja rehtoreilta. Oppilaiden yhteiskunnallista tietämystä arvioitiin 80 koetehtävän avulla (esimerkkejä tehtävistä, ks. liitteet 1 ja 2, Suoninen ym., 2010). Tehtävistä suurin osa oli monivalintatehtäviä mutta mukana oli myös avoimia kysymyksiä. Kysymyksien suuresta määrästä johtuen oppilaat eivät saaneet kaikkia tehtäviä ratkaistavaksi, vaan tehtävät oli rotatoitu 7 koevihkoon (ks. taulukot 1 ja 2, Schulz ym., 2011, 29), ja jokainen oppilas sai vastattavaksi yhden koevihkon. Koetehtävät on jaettu 7 klusteriin (ks. taulukko 1) ja jokainen koevihko sisältää kolmen klusterin tehtävät (ks. taulukko 2). Oppilaiden yhteiskunnallisten asioiden ymmärtämisestä ja niihin liittyvistä asenteista, sekä taustoista ja ominaispiirteistä kerättiin tietoa kyselylomakkeiden avulla. (Esimerkkejä kysymyksistä, ks. liitteet 3 ja 4)

Taulukko 1: Koevihkojen klusterit

Klusteri	Tehtäviä
C1	10 (9 monivalinta- ja 1 avoin teht.)
C2	10 (9 monivalinta- ja 1 avoin teht.)
C3	10 (9 monivalinta- ja 1 avoin teht.)
C4	11 (10 monivalinta- ja 1 avoin teht.)
C5	11 (10 monivalinta- ja 1 avoin teht.)
C6	11 (10 monivalinta- ja 1 avoin teht.)
C7	17 monivalintatehtävää

Taulukko 2: Koevihkot ja niiden sisältämät klusterit

Koevihko	Sijainti		
	1	2	3
1	C1	C2	C4
2	C2	C3	C5
3	C3	C4	C6
4	C4	C5	C7
5	C5	C6	C1
6	C6	C7	C2
7	C7	C1	C3

Kansallisesti saadaan edustava kokoelma vastauksia kaikkiin 80 tehtävään,

kun kokonaisuutena oppilaat tekevät tasaisesti kaikkia eri tehtäviä. Tällainen koeviikkojen käyttö on tavallinen menettely kansainvälisissä koulututkimuksissa.

2.3 Muuttujat

Tässä pro gradu -tutkielmassa tutkitaan sekä ICCS-aineistossa valmiiksi olleiden että aineistosta luotujen (pääkomponenttien) oppilaiden taustoja ja näkemyksiä kuvaavien muuttujien vaikutusta nuorten yhteiskunnallisiin tietoihin ja taitoihin. Tutkittavina vastemuuttujina ovat oppilaan osaamista kuvaavat Plausible Value -arvot (PV1-PV5), joista lisää luvussa 3.4. Taustamuuttujiksi valittiin sukupuoli ja kotona käytettävä kieli, koska ne ovat kansainvälisesti mielenkiintoisia muuttujia ICCS:n kansainvälisen raportin (Schulz ym., 2010, 242) mukaan. Nämä muuttujat ovat osoittautuneet analyyseissä tilastollisesti merkitseviksi ja sisällöllisesti mielenkiintoisiksi ja merkittäviksi monissa maissa. Näiden taustamuuttujien lisäksi raportissa mainittiin mm. kodin sosioekonominen asema ja oppilaan näkemys luokkahuonekeskustelujen avoimuudesta. Nämä muuttujat sisältyvät seuraavaksi esiteltäviin pääkomponentteihin.

Kotona käytettävä kieli (riippuen siitä, onko se sama vai eri kuin testissä käytettävä kieli) kuvaa sitä, onko oppilas tehnyt testin äidinkielellään. Muuttuja poimii vahvasti maahanmuuttajataustaiset oppilaat mutta myös esimerkiksi suomenkieliset oppilaat, jotka käyvät ruotsinkielistä koulua ja toisinpäin. Muuttuja kuvastaa myös vahvasti sitä, saako oppilas opetusta äidinkielellään tai kielellä, jota hän osaa ja jota hänen kotonaan ensisijaisesti käytetään.

Tässä työssä tarkastelun kohteena on myös muuttujia, jotka kuvaavat oppilaan mielipiteitä, näkemyksiä ja taustoja. Näitä muuttujia on yhteensä 29 (ks. taulukko 3 ja liite 5). Näistä muuttujista 23 on pistemäärämuuttujia (WLE, *weighted likelihood estimate*, Warm, 1989). Jokainen näistä WLE-pistemääristä on laskettu sarjasta vastauksia, joiden kysymykset liittyvät toisiinsa teoreettisesti tai empiirisesti (Schulz, Ainley & Fraillon, 2011). Osa muuttujista perustuu diskreetteihin vastauksiin. WLE-pistemäärät on skaalattu kansallisesti siten, että keskiarvo on 50 ja keskihajonta 10. Kaikissa näissä 29 muuttujassa isompi luku tarkoittaa korkeampaa tasoa, suurempaa määrää tai sitä, että on enemmän samaa mieltä.

Ryhmittelemällä näitä mainittuja aineiston muuttujia ja suorittamalla ryh-

mille pääkomponenttianalyysi (luku 4.3) saadaan vähennettyä käytettävien taustamuuttujien määrää huomattavasti säilyttäen kuitenkin iso osa muuttujien alkuperäisestä vaihtelusta. Käytettävät pääkomponentit eivät ole aineistossa valmiina, joten ne on laskettu erikseen. Pääkomponentit (PK) sekä muuttujat ATTCNT, PARTSCHL, PARTCOM, ILLPROT ja LEGPROT ovat tässä työssä skaalattu siten, että jokaisen keskiarvo on 0 ja keskihajonta 1. Kaikki muut paitsi sukupuoli ja kielimuuttuja ovat jatkuvia muuttujia. Valitut taustamuuttujat ovat:

Taulukko 3: Käytettävät taustamuuttujat

GEND	Sukupuoli
0 = Poika (47,1 % aineistosta)	
1 = Tyttö (52,9 % aineistosta)	
LANG	Kotona käytettävä kieli
0 = Muu kieli (3,4 % aineistosta)	
1 = Testissä käytettävä kieli (96,6 % aineistosta)	
POLIT	Kiinnostus politiikkaan (PK)
RGHT	Suhtautuminen tasa-arvoon (PK)
SOSEC	Kodin sosioekonominen asema (PK)
DEM	Demokraattisuus (PK)
COND	Olosuhteet koulussa (PK)
ATTCNT	Isänmaallisuus (WLE)
PARTSCHL	Osallistuminen koulun toimintaan (WLE)
PARTCOM	Osallistuminen järjestötoimintaan (WLE)
ILLPROT	Osallistuminen laittomaan protestointiin (WLE)
LEGPROT	Osallistuminen lailliseen protestointiin (WLE)

3 Oppilaiden osaamista mittaavat pistemäärät

3.1 Raschin malleista

Osana ICCS-tutkimuksessa käytettävää IRT-menetelmää⁶ (*item response theory*) on mm. yksinkertainen logistinen malli (alkuperäinen Raschin malli, Rasch, 1960).

Dikotomisten tehtävien tapauksessa voidaan käyttää yksinkertaista logistista mallia:

$$P(x_{is}|\alpha_i, \theta_s) = \frac{\exp(x_{is}(\theta_s - \alpha_i))}{1 + \exp(\theta_s - \alpha_i)}, \quad (1)$$

missä $P(x_{is}|\theta_s)$ on oppilaan s todennäköisyys ratkaista tehtävä i oikein ($x_{is} = 1$) tai väärin ($x_{is} = 0$), θ_s on oppilaan s kyky ja α_i on tehtävän i vaikeus.

Tällöin oppilaan s todennäköisyys ratkaista tehtävä i oikein on

$$P(x_{is} = 1|\alpha_i, \theta_s) = \frac{\exp(\theta_s - \alpha_i)}{1 + \exp(\theta_s - \alpha_i)}. \quad (2)$$

Vastaavasti oppilaan s todennäköisyys tehdä tehtävä i väärin on

$$P(x_{is} = 0|\alpha_i, \theta_s) = \frac{1}{1 + \exp(\theta_s - \alpha_i)}.$$

On helppo osoittaa, että $P(x_{is} = 1|\alpha_i, \theta_s) + P(x_{is} = 0|\alpha_i, \theta_s) = 1$.

Oletetaan tehtävien vaikeudet sisältävä vektori α tunnetuksi ja oppilaan s vastaukset $\mathbf{x}_s = (x_{1s}, x_{2s}, \dots, x_{ns})$ riippumattomiksi (tehtäviä n kpl). Nyt oppilaan s kyky θ_s voidaan ratkaista maksimoimalla uskottavuusfunktio

$$L(\theta_s) = \prod P(\mathbf{x}_s|\theta_s) = \prod_{i=1}^n \frac{\exp(x_{is}(\theta_s - \alpha_i))}{1 + \exp(\theta_s - \alpha_i)}. \quad (3)$$

Uskottavuusfunktion logaritmistä nähdään, että uskottavuusfunktio riippuu vain oikeiden vastausten määrästä $\sum_{i=1}^n x_{is}$, joka on tyhjentävä tunnusluku:

$$\log L(\theta_s) = \sum_{i=1}^n x_{is}\theta_s - \sum_{i=1}^n \left[x_{is}\alpha_i + \log(1 + \exp(\theta_s - \alpha_i)) \right].$$

⁶Mallinnuksessa, analysoinnissa ja testien sekä kyselyiden pisteytyksessä käytetty menetelmä, jolla voidaan mitata kykyjä, asennetta ja muita muuttujia.

Uskottavuusfunktion maksimoiva θ_s ratkaistaan numeerisesti logaritmissen uskottavuusfunktion derivaatan nollakohdasta:

$$\frac{d}{d\theta_s} \log L = \sum_{i=1}^n \left(x_{is} - \frac{\exp(\theta_s - \alpha_i)}{1 + \exp(\theta_s - \alpha_i)} \right) =: 0$$

Logaritmissen uskottavuusfunktion toinen derivaatta

$$\frac{d^2}{d\theta_s^2} \log L = - \sum_{i=1}^n \left(\frac{\exp(\theta_s - \alpha_i)}{(1 + \exp(\theta_s - \alpha_i))^2} \right) < 0$$

jokaisella θ_s :n arvolla, joten uskottavuusfunktion maksimoiva θ_s on globaali maksimi.

Esimerkki: Vastausarja on $\mathbf{x} = (1, 1, 0, 0)$ ja tehtävien vaikeus on $\boldsymbol{\alpha} = (-1, -0.5, 0.5, 1)$. Nyt

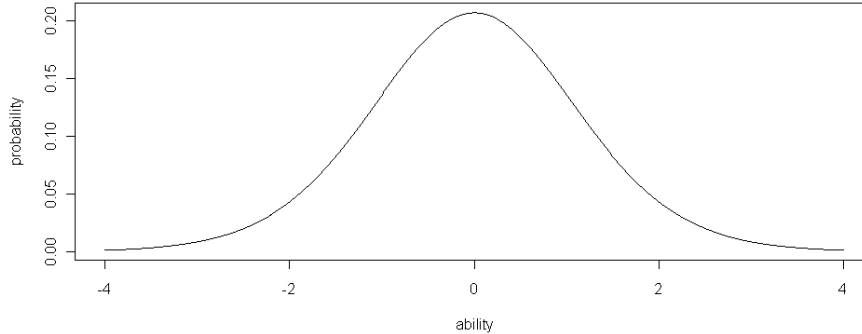
$$\begin{aligned} \frac{d}{d\theta_s} \log L|_{\theta_s=0} &= 2 - \frac{\exp(0 - (-1))}{1 + \exp(0 - (-1))} - \frac{\exp(0 - (-0.5))}{1 + \exp(0 - (-0.5))} \\ &\quad - \frac{\exp(0 - 0.5)}{1 + \exp(0 - 0.5)} - \frac{\exp(0 - 1)}{1 + \exp(0 - 1)} = 0. \end{aligned}$$

Kuvasta 1 nähdään, että uskottavuusfunktio maksimoituu kohdassa $\theta_s = 0$ ja se on tässä esimerkissä globaali maksimi. Taulukossa 4 on vastaava esimerkki vastausarjojen todennäköisyyksistä eri kykyparametrien arvoilla. Vastausarjan todennäköisyys saadaan sen sisältämien vastausten todennäköisyyksien tulona.

Taulukko 4: Esimerkki tehtävien vaikeuksista, vastausarjasta sekä vastauksien ja vastausarjojen todennäköisyyksistä eri kykyparametrien arvoilla

tehtävä i	vaikeus α_i	vastaus x_i	vastauksen tn		
			kyky $\theta_s = -1$	$\theta_s = 0$	$\theta_s = 1$
1	-1	1	0,50	0,73	0,88
2	-0,5	1	0,38	0,62	0,82
3	0,5	0	0,82	0,62	0,38
4	1	0	0,88	0,73	0,50
vastausarjan tn			0,14	0,21	0,14

Raschin malli on todennäköisyysjakauma, joka sisältää tehtävien vaikeuden ja oppilaan kyvyn. Oppilaan kyky ja tehtävien vaikeus on yhdistetty logistisella funktiolla. Tällä tavalla on mahdollista laskea oppilaan todennäköisyydet suoritua eri tehtävissä. Raschin mallin avulla voidaan luoda arvio jokaiselle oppilaalle siitä, kuinka tämä sijoittuu kunkin tehtävän osalta, vaikka



Kuva 1: Esimerkki kyvyn uskottavuudesta (kaava (3)), $\mathbf{x} = (1, 1, 0, 0)$, $\boldsymbol{\alpha} = (-1, -0.5, 0.5, 1)$

oppilas ei olisikaan tehnyt kuin osan tehtävistä. Tämä on mahdollista, koska oppilaiden antamien vastausten avulla voidaan ratkaista tehtävien vaikeudet ja oppilaiden kyvyt. Kun tunnetaan tehtävän vaikeus ja oppilaan kyky, niin kaavan (2) avulla voidaan laskea kyseisen oppilaan todennäköisyys ratkaista kyseinen tehtävä oikein. Raschin mallia käytetään laajalti oppilastutkimuksissa ja se on yksi suosituimmista sekä levinneimmistä IRT-malleista maailmalla (OECD, 2009).

3.2 MRCML-malli

ICCS-tutkimuksessa käytettiin sekä kansainvälisessä että kansallisessa tehtävien skaalauksessa IRT-menetelmänä MRCML-mallia (*the multidimensional random coefficients multinomial logit model*, Adams, Wilson & Wang, 1997). MRCML-malli on yleistetty Raschin malli, joka yhdistää monta olemassa olevaa Raschin mallia (mm. kappaleessa 3.1 esitellyt mallit).

Oletetaan, että tehtävät on indeksoitu $i = 1, \dots, n$ (tehtäviä n kpl) ja jokaiselle tehtävälle on $K_i + 1$ vastaustasoa ($0, 1, \dots, K_i$ pistettä). Käytetään vastaussarjoille vektoria $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iK_i})$, jossa $x_{ij} = 1$, jos tehtävän i vastaus kuuluu vastaustasoon j ja muulloin $x_{ij} = 0$.

Vektori $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \dots, \alpha_p)$ sisältää tehtävien vaikeudet (ts. jokaisen tehtävän jokaisen vastaustason vaikeudet). Vektorin $\boldsymbol{\alpha}$ lineaarikombinaatioita käytetään jokaisen tehtävän vastaustason empiiristen tunnuslukujen kuvailemi-

nessa. Lineaarikombinaatiot määritellään design-vektoreilla \mathbf{a}_{ik} ($i = 1, \dots, n$ ja $k = 1, \dots, K_i$), joiden jokaisen pituus on p . Design-vektoreista \mathbf{a}_{ik} saadaan design-matriisi $\mathbf{A}^T = (\mathbf{a}_{11}, \mathbf{a}_{12}, \dots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \dots, \mathbf{a}_{2K_2}, \dots, \mathbf{a}_{nK_n})$ (design-vektorit sisältävät arvoja 0 ja 1 siten, että kunkin tehtävän ensimmäisellä tasolla design-vektori on $(1, 0, \dots, 0)^T$, toisella tasolla $(1, 1, 0, \dots, 0)^T$ jne.).

MRCML-mallissa oletetaan, että D :n kyvyn sarja on oppilaan vastausten taustalla. Näitä yksilön kykyjä kuvaa vektori $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \dots, \theta_D)$. Mallissa on pisteytysominaisuus, joka mahdollistaa suorituskyvyn (*scoren*) määrittämisen jokaisen tehtävän jokaiselle vastaustasolle. Merkitään vastauksen suorituskyyä b_{ikd} (tehtävä i , vastaustaso k , kyky d) (0, jos vastaus ei riitä kyseiselle vastaustasolle ja 1, jos riittää) (määritelmän mukaan 0-vastaustasolla suorituskyy on 0). Kootaan suorituskyyt b_{ikd} vektoriksi $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \dots, b_{ikD})^T$ ja saadut vektorit \mathbf{b}_{ik} matriisiksi $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \dots, \mathbf{b}_{iK_i})^T$ ja vielä saadut matriisit \mathbf{B}_i pisteytysmatriisiksi $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n)^T$.

MRCML-malli:

$$P(x_{ij} = 1 | \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \frac{\exp(\mathbf{b}_{ij}^T \boldsymbol{\theta} + \mathbf{a}_{ij}^T \boldsymbol{\alpha})}{\sum_{k=0}^{K_i} \exp(\mathbf{b}_{ik}^T \boldsymbol{\theta} + \mathbf{a}_{ik}^T \boldsymbol{\alpha})}. \quad (4)$$

Merkitään nyt vastaussarjan todennäköisyyttä:

$$f(\mathbf{x} | \boldsymbol{\alpha}, \boldsymbol{\theta}) = \Psi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \exp[\mathbf{x}^T (\mathbf{B}\boldsymbol{\theta} + \mathbf{A}\boldsymbol{\alpha})], \quad (5)$$

missä

$$\Psi(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \left\{ \sum_{\mathbf{z} \in \Omega} \exp[\mathbf{z}^T (\mathbf{B}\boldsymbol{\theta} + \mathbf{A}\boldsymbol{\alpha})] \right\}^{-1},$$

jossa Ω sisältää kaikki mahdolliset vastaussarjat.

Esimerkiksi neljän vastaustason tehtävän tapauksessa

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{ja} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Näiden matriisien ja yhtälön (4) avulla saadaan

$$\begin{aligned} P(x_{11} = 1 | \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\theta}) &= \exp(\theta_1 + \alpha_1)/c, \\ P(x_{12} = 1 | \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\theta}, x_{11} = 1) &= \exp(\theta_1 + \theta_2 + \alpha_1 + \alpha_2)/c, \\ P(x_{13} = 1 | \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\theta}, x_{12} = 1) &= \exp(\theta_1 + \theta_2 + \theta_3 + \alpha_1 + \alpha_2 + \alpha_3)/c, \end{aligned}$$

joissa

$$c = 1 + \exp(\theta_1 + \alpha_1) + \exp(\theta_1 + \theta_2 + \alpha_1 + \alpha_2) + \exp(\theta_1 + \theta_2 + \theta_3 + \alpha_1 + \alpha_2 + \alpha_3).$$

3.3 Populaatiomalli

Kappaleessa 3.2 esitelty MRCML-malli on ehdollinen latentilla muuttujalla θ . Jotta voitaisiin arvioida oppilaiden osaamista, täytyy määritellä θ :lle jakaumatiheys $g(\theta|\mu, \sigma^2)$. Oletetaan, että oppilaat ovat normaalijakautuneesta populaatiosta keskiarvolla μ ja varianssilla σ^2 . Saadaan jakaumatiheys

$$g(\theta|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2}\right]. \quad (6)$$

Laajentamalla jakaumatiheyttä (6) korvaamalla μ regressiomallilla $\mathbf{Y}^T \boldsymbol{\beta}$ saadaan

$$g(\theta|\mathbf{Y}, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(\theta - \mathbf{Y}^T \boldsymbol{\beta})^T(\theta - \mathbf{Y}^T \boldsymbol{\beta})\right]. \quad (7)$$

Yleistämällä vielä lisää yhtälöä (7) käyttämällä skalaarin θ sijaan vektoria $\boldsymbol{\theta}$ ($D \times 1$) saadaan oppilaalle s populaatiomalli (Adams, Wu & Macaskill, 1997)

$$g(\boldsymbol{\theta}_s|\mathbf{W}_s, \gamma, \Sigma) = (2\pi)^{-D/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)^T \Sigma^{-1} (\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)\right], \quad (8)$$

missä γ ($D \times u$ -matriisi) sisältää regressiokertoimet, Σ on $D \times D$ -kovarianssimatriisi ja \mathbf{W}_s ($u \times 1$ -vektori) sisältää taustamuuttujat.

Saatua populaatiomallia (8) käytetään posteriorijakaumassa priorijakaumana määrittettäessä oppilaiden kykyjä.

3.4 Plausible Value -muuttujat

Oppilaat eivät tehneet kaikkia tehtäviä, minkä vuoksi on generoitu osaamista kuvaavat Plausible Value -muuttujat (PV), jotka ovat moni-imputaatioita latentista saavutuksesta (Wu, 2005). PV:t on generoitu aineistoon ACER ConQuest -ohjelmalla (Wu, Adams, Wilson & Haldane, 2007). Jokaiselle oppilaalle on mallinnettu oma jakaumansa taustatietojen (kyselylomakkeiden)

ja koevastausten (tehtävälomakkeiden) perusteella. PV:t ovat generoituja arvoja näistä jakaumista ja jokaiselle oppilaalle on arvottu 5 PV:ta. Toisin sanoen nämä arvotut PV:t ovat realisaatioita kunkin oppilaan kyvystä eli θ_s :sta.

Sen sijaan, että laskettaisiin oppilaan osaamiselle vain estimaatti, PV:iden avulla saadaan estimoitua todennäköisyysjakauma, joka sisältää myös arvon epävarmuudelle. Käyttämällä vain yhtä PV:ta saadaan harhattomia estimaatteja tutkittaville parametreille, mutta testin luotettavuuteen vaikuttavaa oppilaan tekemättömien tehtävien aiheuttamaa imputaatiovirhettä ei saada estimoitua (OECD, 2009, 43). Käyttämällä useampaa kuin yhtä PV:ta saadaan paremmin otettua huomioon puuttuvan tiedon aiheuttama virhe. Tässä työssä käytetään kaikkia viittä aineistossa olevaa PV:ta. ICCS-tutkimuksessa PV:t skaalattiin kansainvälisesti siten, että keskiarvoksi tuli 500 ja keskihajonnaksi 100.

Kaikille oppilaskyselylomakkeen muuttujille on kansallisesti tehty pääkomponenttianalyysi (PCA, luku 4.3) (Adams, 2002). Pääkomponentteja on valittu niin monta, että ne sisältävät vähintään 90 % alkuperäisten muuttujien vaihtelusta. Näitä valittuja pääkomponentteja on käytetty priorijakaumassa oppilaiden taustamuuttujina. Näitä PV-arvojen laskemisessa käytettyjä pääkomponentteja ei ole aineistossa, joten ne eivät ole samat kuin kappaleessa 2.3 esitellyt ja sekamallissa käytetyt pääkomponentit, vaikka ne molemmat muodostetaankin pääosin samoista muuttujista.

Oppilaan s osaamista kuvaava jakauma saadaan posteriorijakaumana (Wu, 2005 ja Adams, Wu & Macaskill, 1997):

$$h(\theta_s | \mathbf{W}_s, \alpha, \gamma, \Sigma, \mathbf{x}_s) = \frac{f(\mathbf{x}_s | \alpha, \theta_s) g(\theta_s | \mathbf{W}_s, \gamma, \Sigma)}{\int f(\mathbf{x}_s | \alpha, \theta_s) g(\theta_s | \mathbf{W}_s, \gamma, \Sigma) d\theta_s}, \quad (9)$$

missä $f(\mathbf{x}_s | \alpha, \theta_s)$ on vastaussarjan todennäköisyysstiheys (yhtälö (5)) ja priorijakauma $g(\theta_s | \mathbf{W}_s, \gamma, \Sigma)$ on populaatiomalli (8).

Parametreja γ , Σ ja α ei voida analyttisesti ratkaista yhtäaikaaisesti, joten ne ratkaistaan seuraavasti (Adams, Wu & Macaskill, 1997 ja Adams, Wilson & Wu, 1997):

Otoksesta satunnaisesti valitun oppilaan vastaussarjan \mathbf{x} todennäköisyys on

$$f(\mathbf{x} | \alpha, \mathbf{W}_s, \gamma, \Sigma) = \int_{\theta} f(\mathbf{x} | \alpha, \theta) g(\theta | \mathbf{W}_s, \gamma, \Sigma) d\theta. \quad (10)$$

Tutkittavana on N oppilaan otos. Oletetaan oppilaat toisistaan riippumattomiksi, joten uskottavuusfunktio on

$$\Lambda(\boldsymbol{\alpha}, \gamma, \Sigma | \mathbf{x}_s) = \prod_{s=1}^N \int_{\boldsymbol{\theta}_s} f(\mathbf{x}_s | \boldsymbol{\alpha}, \boldsymbol{\theta}_s) g(\boldsymbol{\theta}_s | \mathbf{W}_s, \gamma, \Sigma) d\boldsymbol{\theta}_s \quad (11)$$

ja logaritminen uskottavuusfunktio on

$$\lambda(\boldsymbol{\alpha}, \gamma, \Sigma | \mathbf{x}_s) = \sum_{s=1}^N \log \int_{\boldsymbol{\theta}_s} f(\mathbf{x}_s | \boldsymbol{\alpha}, \boldsymbol{\theta}_s) g(\boldsymbol{\theta}_s | \mathbf{W}_s, \gamma, \Sigma) d\boldsymbol{\theta}_s. \quad (12)$$

Logaritmisen uskottavuusfunktion (12) avulla saadaan välivaihein seuraavat osittaisderivaatat ja ne asetetaan nolliksi:

$$\begin{aligned} \frac{\partial \lambda}{\partial \boldsymbol{\alpha}} &= \frac{\partial}{\partial \boldsymbol{\alpha}} \left\{ \sum_{s=1}^N \log \int_{\boldsymbol{\theta}_s} f(\mathbf{x}_s | \boldsymbol{\alpha}, \boldsymbol{\theta}_s) g(\boldsymbol{\theta}_s | \mathbf{W}_s, \gamma, \Sigma) d\boldsymbol{\theta}_s \right\} \\ &\vdots \\ &= \mathbf{A}^T \sum_{s=1}^N \left[\mathbf{x}_s - \int_{\boldsymbol{\theta}_s} E_z(\mathbf{z} | \boldsymbol{\theta}_s) h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s \right] \\ &=: \mathbf{0}, \end{aligned} \quad (13)$$

missä $E_z(\mathbf{z} | \boldsymbol{\theta}_s) = \Psi(\boldsymbol{\theta}_s, \boldsymbol{\alpha}) \sum_{z \in \Omega} \mathbf{z} \exp \{ \mathbf{z}^T (\mathbf{B} \boldsymbol{\theta}_s + A \boldsymbol{\alpha}) \}$ (jossa $\Psi(\boldsymbol{\theta}_s, \boldsymbol{\alpha})$ kuten kaavassa (5));

$$\begin{aligned} \frac{\partial \lambda}{\partial \gamma} &= \sum_{s=1}^N \int_{\boldsymbol{\theta}_s} \frac{\partial \log g(\boldsymbol{\theta}_s | \gamma, \Sigma)}{\partial \gamma} h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s \\ &\vdots \\ &= \Sigma \left\{ \sum_{s=1}^N \bar{\boldsymbol{\theta}}_s \mathbf{W}_s^T - \sum_{s=1}^N \gamma^T \mathbf{W}_s \mathbf{W}_s^T \right\} \\ &=: \mathbf{0}, \end{aligned} \quad (14)$$

missä $\bar{\boldsymbol{\theta}}_s = \int_{\boldsymbol{\theta}_s} \boldsymbol{\theta}_s h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s$; ja

$$\begin{aligned}
\frac{\partial \boldsymbol{\lambda}}{\partial \Sigma} &= \sum_{s=1}^N \int_{\boldsymbol{\theta}_s} \frac{\partial \log g(\boldsymbol{\theta}_s | \gamma, \Sigma)}{\partial \Sigma} h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s \\
&\vdots \\
&= -\frac{N}{2\Sigma^2} \left[\Sigma - \frac{1}{N} \sum_{s=1}^N \int_{\boldsymbol{\theta}_s} (\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)(\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)^T h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s \right] \\
&=: \mathbf{0}.
\end{aligned} \tag{15}$$

Yhtälöistä (14) ja (15) saadaan ratkaistua

$$\hat{\gamma} = \left(\sum_{s=1}^N \bar{\boldsymbol{\theta}}_s \mathbf{W}_s^T \right) \left(\sum_{s=1}^N \mathbf{W}_s \mathbf{W}_s^T \right)^{-1} \tag{16}$$

ja

$$\hat{\Sigma} = \frac{1}{N} \sum_{s=1}^N \int_{\boldsymbol{\theta}_s} (\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)(\boldsymbol{\theta}_s - \gamma \mathbf{W}_s)^T h(\boldsymbol{\theta}_s | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) d\boldsymbol{\theta}_s. \tag{17}$$

3.4.1 Monte Carlo -approksimointi

Yhtälöiden (13), (16) ja (17) yhtäaikaisessa ratkaisemisessa käytetään EM-algoritmia⁷ (Dempster, Laird & Rubin, 1977) tavalla, joka on esitelty viitteessä Bock & Aitken (1981) ja Newtonin menetelmää⁸. Yhtälöiden (13), (16) ja (17) integraalit approksimoidaan numeerisesti Monte Carlo -menetelmällä (Volodin & Adams, 2002).

Generoidaan populaatiomallista (8) arvot $\Theta_1, \Theta_2, \dots, \Theta_P$, jotka ovat D -ulotteisia vektoreita. Vastaussarjan todennäköisyyttä (10) approksimoidaan käyttäen

$$f(\mathbf{x} | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma) = \frac{1}{P} \sum_{p=1}^P f(\mathbf{x} | \boldsymbol{\alpha}, \Theta_p) \tag{18}$$

⁷Expectation-maximization, iteratiivinen menetelmä uskottavuuden maksimin tai posteriorin parametriestimaattien maksimien löytämiseksi

⁸Numeerinen menetelmä funktion nollakohtien löytämiseksi

ja posteriorijakaumaa (9) approksimoidaan käyttäen

$$h(\Theta_q | \mathbf{W}_s, \boldsymbol{\alpha}, \gamma, \Sigma, \mathbf{x}_s) = \frac{f(\mathbf{x}_s | \boldsymbol{\alpha}, \Theta_q)}{\sum_{p=1}^P f(\mathbf{x}_s | \boldsymbol{\alpha}, \Theta_p)}, \quad (19)$$

missä $q = 1, \dots, P$.

Käytettävä EM-algoritmi etenee seuraavasti:

Vaihe 1:

Generoidaan arvot Θ_q estimaattien $\gamma^{(t)}$ ja $\Sigma^{(t)}$ (iteraatio t) perusteella standardoidusta multinormaalijakaumasta $N(\gamma^{(t)} \mathbf{W}_s, \Sigma^{(t)})$ jokaiselle iteraatiolle t .

Vaihe 2:

Lasketaan diskreetti approksimaatio posteriorijakauman θ_s :n tiheydelle iteraatiossa t käyttäen

$$h(\Theta_q | \mathbf{W}_s, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \Sigma^{(t)}, \mathbf{x}_s) = \frac{f(\mathbf{x}_s | \boldsymbol{\alpha}^{(t)}, \Theta_q)}{\sum_{p=1}^P f(\mathbf{x}_s | \boldsymbol{\alpha}^{(t)}, \Theta_p)},$$

missä $\boldsymbol{\alpha}^{(t)} = \hat{\boldsymbol{\alpha}}^{(t)}$, $\gamma^{(t)} = \hat{\gamma}^{(t)}$ ja $\Sigma^{(t)} = \hat{\Sigma}^{(t)}$.

Vaihe 3:

Ratkaistaan $\hat{\boldsymbol{\alpha}}^{(t+1)}$ yhtälöstä

$$\mathbf{A}^T \sum_{s=1}^N \left[\mathbf{x}_s - \sum_{r=1}^P E_z(\mathbf{z} | \Theta_r) h(\Theta_r | \mathbf{W}_s, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \Sigma^{(t)}, \mathbf{x}_s) \right] = \mathbf{0}$$

käyttäen Newtonin menetelmää.

Vaihe 4:

Estimoidaan $\gamma^{(t+1)}$ ja $\Sigma^{(t+1)}$ käyttäen

$$\gamma^{(t+1)} = \left(\sum_{s=1}^N \bar{\Theta}^s \mathbf{W}_s^T \right) \left(\sum_{s=1}^N \mathbf{W}_s \mathbf{W}_s^T \right)^{-1}$$

missä $\bar{\Theta}^s = \sum_{r=1}^P \Theta_r h(\Theta_r | \mathbf{W}_s, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \Sigma^{(t)}, \mathbf{x}_s)$, ja

$$\hat{\Sigma}^{(t+1)} = \frac{1}{N} \sum_{s=1}^N \sum_{r=1}^P (\Theta_r - \gamma^{(t+1)} \mathbf{W}_s) (\Theta_r - \gamma^{(t+1)} \mathbf{W}_s)^T h(\Theta_r | \mathbf{W}_s, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \Sigma^{(t)}, \mathbf{x}_s).$$

Vaihe 5:

Palataan vaiheeseen 1.

3.4.2 PV-arvojen generoiminen

Kaavassa (9) verrannollisuusvakiota ei voida analyttisesti integroida, mutta sitä voidaan arvioida Monte Carlo -integroinnilla⁹ (Adams, Wu & Macaskill, 1997 ja Gelman ym., 1995, 312):

$$\int f(\mathbf{x}_s|\boldsymbol{\alpha}, \boldsymbol{\theta}_s)g(\boldsymbol{\theta}_s|\mathbf{W}_s, \gamma, \Sigma)d\boldsymbol{\theta}_s \approx \frac{1}{M} \sum_{m=1}^M f(\mathbf{x}_s|\boldsymbol{\alpha}, \varphi_{ms}) \equiv \mathfrak{S},$$

missä $\{\varphi_{ms}\}_{m=1}^M$ arvotaan multinormaalijakaumasta $g(\boldsymbol{\theta}_s|\mathbf{W}_s, \gamma, \Sigma)$. Alaindeksi s viittaa oppilaaseen. Arvon M tulee olla suuri. ICCS-tutkimuksessa on käytetty $M = 2000$.

Lasketaan arvot

$$p_{ms} = f(\mathbf{x}_s|\boldsymbol{\alpha}, \varphi_{ms})g(\varphi_{ms}|\mathbf{W}_s, \gamma, \Sigma)$$

ja saadaan parit $(\varphi_{ms}, \frac{p_{ms}}{\mathfrak{S}})_{m=1}^M$, joita voidaan käyttää posteriorijakauman (9) tiheyden approksimaationa.

Nyt φ_{sj} :n arvontatodennäköisyys on

$$q_{sj} = \frac{p_{ms}}{\sum_{m=1}^M p_{ms}}.$$

PV-arvoiksi valitaan ne φ_{i_0} , jotka täyttävät ehdon

$$\sum_{c=1}^{i_0-1} q_{cs} < \eta_i \leq \sum_{c=1}^{i_0} q_{cs},$$

missä $\{\eta_i\}_{i=1}^L \sim \text{Tas}(0, 1)$.

Priorijakauma sisältää taustamuuttujia, kuten esimerkiksi sukupuolen ja sosio-ekonomisen aseman. Tästä syystä esimerkiksi kahdella samalla tavalla ko-keessa vastanneella oppilaalla, joilla on sama koetulos, on todennäköisesti erilaiset PV:t, koska todennäköisesti näiden oppilaiden posteriorijakaumat ovat erilaiset.

⁹Menetelmä, jolla voidaan mm. arvioida monimutkaisia integraaleja helposti simuloitavien jakaumien avulla.

4 Lineaarista malleista

4.1 Kiinteiden vaikutusten malli

Oletetaan, että \mathbf{y} on vastemuuttujan havaitut arvot ($n \times 1$ -vektori). Oletetaan myös $p - 1$ tuntematonta kiinteää vaikutusta.

Kiinteiden vaikutusten malli matriisimuodossa (McCulloch & Searle, 2001):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (20)$$

missä \mathbf{X} on kiinteiden selittävien muuttujien arvot ($n \times p$ -matriisi) ja $\boldsymbol{\beta}$ on kiinteät vaikutukset ($p \times 1$ -vektori, joka sisältää vakioparametrin ja regressio-kertoimet). Oletetaan, että jäännökset ovat riippumattomia sekä toisistaan että selittäjistä. Oletetaan myös $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$. Tästä saadaan

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}).$$

4.1.1 Parametrien estimointi

Parametrit $\boldsymbol{\beta}$ ja σ^2 voidaan estimoida suurimman uskottavuuden menetelmällä (ML, *maximum likelihood*). Uskottavuusfunktio on

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right). \quad (21)$$

Uskottavuusfunktion logaritmi on

$$\log L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (22)$$

Asettamalla uskottavuusfunktion logaritmin osittaisderivaatat nolliksi,

$$\frac{\partial \log L(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \frac{\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}}{\sigma^2} =: 0$$

ja

$$\frac{\partial \log L(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2(\sigma^2)^2} + \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} =: 0,$$

ja ratkaisemalla niistä $\boldsymbol{\beta}$ ja σ^2 saadaan ML-estimaattorit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad \text{ja} \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$

olettaen, että $(\mathbf{X}^T \mathbf{X})^{-1}$ on olemassa.

4.2 Lineaarinen sekamalli

Havainnot mahdollisesti korreloivat sekä koulujen että oppilaiden sisällä ja siksi on järkevää käyttää mallia, joka sisältää sekä koulu- että oppilasvaikutuksen.

Lineaarinen sekamalli matriisimuodossa (McCulloch & Searle, 2001):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \boldsymbol{\epsilon}, \quad (23)$$

missä \mathbf{y} on vastemuuttuja ($n \times 1$ -vektori, n havaittua arvoa, merkitään $n = 5s$, s oppilasta ja 5 PV:ta), \mathbf{X} on kiinteän osan design-matriisi ($n \times p$, p parametria), $\boldsymbol{\beta}$ on kiinteät vaikutukset ($p \times 1$ -vektori), \mathbf{Z}_1 on kouluun liittyvän satunnaisosan design-matriisi ($n \times q$, q koulua), \mathbf{u}_1 on kouluun liittyvät satunnaisvaikutukset ($q \times 1$ -vektori), \mathbf{Z}_2 on oppilaaseen liittyvän satunnaisosan design-matriisi ($n \times s$), \mathbf{u}_2 on oppilaaseen liittyvät satunnaisvaikutukset ($s \times 1$ -vektori) ja $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I})$ on satunnaisvirhe.

Sekamalli on kuten kiinteiden vaikutusten malli (kappale 4.1) mutta siihen on lisätty satunnaisosat $\mathbf{Z}_1\mathbf{u}_1$ ja $\mathbf{Z}_2\mathbf{u}_2$. Sekamallissa oletetaan, että $\mathbf{u}_1 \sim N(0, \sigma_{u_1}^2 \mathbf{I})$, $\mathbf{u}_2 \sim N(0, \sigma_{u_2}^2 \mathbf{I})$, $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I})$ ja $\text{cov}(\mathbf{u}, \boldsymbol{\epsilon}) = 0$. Näistä saadaan

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_{u_1}^2 \mathbf{Z}_1 \mathbf{Z}_1^T + \sigma_{u_2}^2 \mathbf{Z}_2 \mathbf{Z}_2^T + \sigma_\epsilon^2 \mathbf{I}).$$

Tässä pro gradu -tutkielmassa käytetään mallinnusmenetelmänä varianssi-komponenttimallia, joka on lineaarisen sekamallin (23) erikoistapaus.

Varianssikomponenttimalli skalaarimuodossa:

$$y_{kir} = \mathbf{x}_{ki}^T \boldsymbol{\beta} + \gamma_k + \nu_{ki} + \epsilon_{kir}, \quad (24)$$

missä nyt k on koulu, i on oppilas, r on mittaus (5 mittaus/oppilas), \mathbf{x}_{ki}^T on selittävien muuttujien vektori, $\boldsymbol{\beta}$ on kiinteän osan parametrivektori (kiinteät vaikutukset), $\gamma_k \sim N(0, \sigma_\gamma^2)$ on kouluun liittyvä satunnaisefekti, $\nu_{ki} \sim N(0, \sigma_\nu^2)$ on oppilaaseen liittyvä satunnaisefekti ja $\epsilon_{kir} \sim N(0, \sigma_\epsilon^2)$ on oppilaaseen liittyvä satunnaisvirhe.

4.2.1 Parametrien estimointi

Tässä työssä mallin parametrien estimointi tehdään ML-menetelmällä. Merkitään $\text{cov}(\mathbf{y}) = \mathbf{V}$. Uskottavuusfunktio on

$$L(\boldsymbol{\beta}, \mathbf{V}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right). \quad (25)$$

Uskottavuusfunktion logaritmi on

$$\log L(\boldsymbol{\beta}, \mathbf{V}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (26)$$

Edellyttäen, että $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ on olemassa, kiinteiden vaikutusten estimaattoriksi saadaan

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

ja estimaattorin kovarianssimatriisiksi saadaan

$$\text{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

Jos matriisia \mathbf{V} ei tunneta, niin myös se voidaan estimoida ML-menetelmällä. Yhtälöön (25) sijoitetaan $\boldsymbol{\beta}$ paikalle $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$ ja tämän jälkeen uskottavuusfunktio maksimoidaan matriisin \mathbf{V} alkioiden suhteen. Matriisi \mathbf{V} ratkaistaan siis komponenteittain ja merkitään $\mathbf{V} = \mathbf{V}(\sigma_\gamma^2, \sigma_\nu^2, \sigma_\epsilon^2)$. Maksimointi suoritetaan yleensä Newtonin menetelmällä tai EM-algoritmillä (Pinheiro & Bates, 2000, 79).

4.2.2 Sisäkorrelaatiosta

Satunnaisvaikutuksilla voidaan ottaa huomioon aineiston kovarianssirakenne, koska satunnaisvaikutuksen luokat ovat sisäisesti korreloituneita. Tämän sisäisen korreloituneisuuden voimakkuuden mittaamisessa keskeinen käsite on sisäkorrelaatio (ICC, *intraclass correlation*). Havaintojen tilastollista riippuvuutta voidaan arvioida sisäkorrelaation avulla. Sisäkorrelaatiokerroin saadaan jakamalla koulujen välinen varianssi muuttujan kokonaisvarianssilla (Goldstein, 2011).

Sisäkorrelaation kaava on nyt varianssikomponenttimallin (24) merkinnöin

$$\rho = \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\nu^2 + \sigma_\epsilon^2}. \quad (27)$$

Sisäkorrelaatio mittaa koulujen oppilaiden homogeenisuutta koko havaintoaineistossa. Mikäli koulujen välinen varianssikomponentti σ_γ^2 on nolla, myös koulujen sisäkorrelaatio on nolla. Tällöin havaintoyksiköiden välillä ei ole keskinäistä riippuvuutta koulujen sisällä. Kun muuttujan sisäkorrelaatio on nollaa suurempi mutta alle yhden, saman koulun oppilaat ovat keskenään homogeenisempia kuin oppilaat koko havaintoaineistossa. Tällöin oppilaiden

välillä on riippuvuutta, joka on huomioitava aineiston tilastollisessa analyysissä. Mitä suurempi sisäkorrelaatio on, sitä voimakkaampaa on havaintojen keskinäinen riippuvuus ja sitä enemmän tulokset poikkeavat oikeista, jos tätä ominaisuutta ei oteta tilastollisissa analyyseissä huomioon.

4.3 Pääkomponenttianalyysi

Aineistossa on suuri määrä muuttujia, joten muuttujamäärän vähentämiseksi käytetään pääkomponenttianalyysia (Jolliffe, 2002). Näin saadaan säilytettyä riittävän suuri osa aineiston vaihtelusta, mutta käsiteltävänä on huomattavasti pienempi määrä muuttujia.

Aineiston käytettävistä muuttujista 24 muuttujaa ryhmitellään viiteen ryhmään (ks. liite 5) siten, että samankaltaiset muuttujat ovat omissa ryhmissään. Jokaiselle ryhmälle suoritetaan erikseen pääkomponenttianalyysi (luku 4.3) ja saadut ensimmäiset pääkomponentit otetaan selittäjiksi. Kiinnostus politiikkaan selittää 44,0 % sisältämiensä muuttujien alkuperäisestä vaihtelusta. Vastaavasti suhtautuminen tasa-arvoon selittää 56,4 %, kodin sosio-ekonominen asema 49,5 %, demokraattisuus 49,1 % ja olosuhteet koulussa 50,0 %.

4.4 Uskottavuusosamäärätesti

Uskottavuusosamäärätestillä voidaan verrata kahta samaa aineistoa käsittelevää mallia, jotka ovat sisäkkäiset. Sisäkkäisillä malleilla tarkoitetaan sellaisia malleja, joissa monimutkaisempi malli sisältää kaikki yksinkertaisemman mallin parametrit. Vähemmän parametreja sisältävässä mallissa ei siis ole yhtään parametria, jota ei olisi laajemmassa mallissa. Laajemmassa mallissa vastaavasti on yksi tai useampi parametri enemmän kuin yksinkertaisemmassa mallissa, koska muuten mallit olisivat identtiset.

Malleista lasketaan uskottavuusfunktioiden arvot, joiden avulla saadaan laskettua testisuure. Testisuure noudattaa χ^2 -jakaumaa, ja nyt saadaan laskettua p-arvo. Suuren p-arvon tapauksessa ($p > 0,05$) voidaan pitää yksinkertaisempaa mallia riittävänä verrattuna monimutkaisempaan malliin. Vastaavasti p-arvon ollessa pieni yksinkertaisempaa mallia voidaan pitää liian rajoittavana siitä puuttuvien parametrien vuoksi. Menetelmää käytetään tässä työssä mallin riittävyyden tarkastelussa valittaessa mallia.

5 Aineiston analysointi

5.1 Mallin valinta

Mallin valinta aloitettiin tutkimalla valittujen taustamuuttujien vaikutusta vasteeseen, eli Plausible Value -arvoihin (oppilaiden yhteiskunnallisen ymmärtämisen aste). Mallien tutkiminen aloitettiin mallista, jossa oli kaikki päävaikutukset ja ne parittaiset interaktiot, jotka sisälsivät sukupuoli- tai kielimuuttujan. Mallista poistettiin p -arvojen perusteella vähiten tilastollisesti merkitseviä interaktioita yksitellen, kunnes kaikkien jäljelle jääneiden p -arvot olivat pienempiä kuin 0,05. Jokaisen interaktion poiston yhteydessä yksinkertaisemman mallin riittävyys varmistettiin uskottavuusosamäärätestillä (luku 4.4). Mallien sovitukset tehtiin R-ohjelmistolla (versio 2.7.2, R Development Core Team, 2008) käyttäen lme-funktiota (Pinheiro & Bates, 2000).

Päädytään malliin (symbolinen esitysmuoto, : tarkoittaa yhdysvaikutusta, McCullagh & Nelder, 1989):

$$PV \sim GEND + LANG + POLIT + RGHT + SOSEC + DEM + COND + ATTCNT + PARTSCHL + PARTCOM + ILLPROT + LEGPROT + GEND:SOSEC + GEND:ATTCNT + GEND:LEGPROT + LANG:POLIT + LANG:RGHT + LANG:ILLPROT + LANG:LEGPROT.$$

Taulukossa 5 on valitun mallin parametriestimaatit, keskivirheet sekä t - ja p -arvot. Taulukossa GEND=1 tarkoittaa tyttöä ja LANG=1 tarkoittaa, että testin kieli on sama kuin kotona käytettävä kieli. Käytettävien muuttujien väliset korrelaatiot on esitetty liitteessä 10.

5.2 Mallin estimaattien tulkinta

Vasteena mallissa on yhteiskunnallisia tietoja ja taitoja kuvaava Plausible Value -muuttuja (PV, luku 3.4). Pääkomponentit ovat jatkuvia muuttujia ja ne on skaalattu siten, että jokaisen pääkomponentin keskiarvo on 0 ja keskihajonta 1. Taulukosta 5 nähdään valitun mallin parametriestimaatit ja niiden keskivirheet, t -arvot ja p -arvot.

Demokraattisuus (DEM, 16.26) ja osallistuminen koulun toimintaan (PARTSCHL, 4.54) vaikuttavat positiivisesti vasteeseen. Osallistuminen järjestötoimintaan (PARTCOM, -7.04) ja olosuhteet koulussa (COND, -10.18) vaikuttavat ne-

Taulukko 5: Valitun mallin parametriestimaatit, keskivirheet, t -arvot ja p -arvot

	Est.	SE	t-arvo	p-arvo
(Intercept)	549.87	7.13	77.07	<0.001
GEND	0.51	2.54	0.20	0.841
LANG	31.45	6.97	4.51	<0.001
POLIT	-7.56	6.33	-1.20	0.232
RGHT	-4.29	6.77	-0.63	0.526
SOSEC	24.05	1.64	14.65	<0.001
DEM	16.26	1.39	11.72	<0.001
COND	-10.18	1.21	-8.43	<0.001
ATTCNT	-6.75	1.54	-4.39	<0.001
PARTSCHL	4.54	1.21	3.75	<0.001
PARTCOM	-7.04	1.14	-6.20	<0.001
ILLPROT	-40.93	7.01	-5.84	<0.001
LEGPROT	29.78	7.20	4.13	<0.001
GEND:SOSEC	-6.28	2.17	-2.89	<0.001
GEND:ATTCNT	-4.39	2.23	-1.96	0.050
GEND:LEGPROT	8.52	2.21	3.85	<0.001
LANG:POLIT	13.97	6.43	2.17	0.030
LANG:RGHT	17.96	6.82	2.63	0.008
LANG:ILLPROT	28.17	7.10	3.97	<0.001
LANG:LEGPROT	-19.25	7.25	-2.66	0.008

gatiivisesti vasteeseen (liitteissä 7–9 on kysymykset, joista pääkomponentin sisältämät muuttujat on muodostettu).

Valitussa mallissa on kolme yhdysvaikutustermiä sukupuoli-muuttujan (GEND) kanssa. Kodin sosioekonomisella asemalla (SOSEC) on positiivinen vaikutus vasteeseen pojilla (24.05) ja tytöillä vähemmän positiivinen (17.77) (ks. kuva 2). Isänmaallisuus (ATTCNT) vaikuttaa negatiivisesti vasteeseen pojilla (−6.75) ja tytöillä negatiivisemmin (−11.43) (ks. kuva 3). Lailliseen protestointiin osallistumisella (LEGPROT) on positiivinen vaikutus vasteeseen niillä pojilla, joiden kotona käytettävä kieli (LANG) on muu kuin testissä käytettävä kieli (29.78). Tytöillä tämä on suurempi (8.52 verran) ja oppilailta, joiden kotona käytettävä kieli on testissä käytettävä kieli vähemmän (−19.25 verran) (ks. kuvat 4 ja 8).

Valitussa mallissa on lisäksi kolme yhdysvaikutustermiä kielitaustan kanssa. Kiinnostus politiikkaan (POLIT) vaikuttaa vasteeseen negatiivisesti (−7.56) niillä, joiden kotona käytetään ensisijaisesti muuta kuin testissä käytettävää kieltä ja positiivisesti (6.41) niillä, joiden kotona käytettävä kieli on testissä

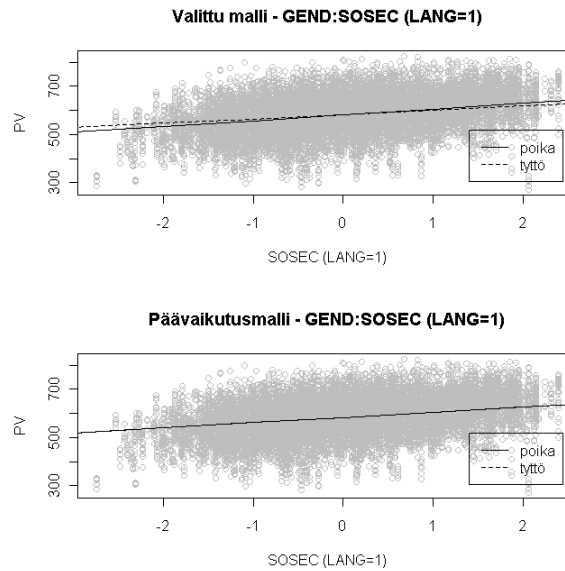
käytettävä kieli (ks. kuva 5). Suhtautumisella tasa-arvoon (RGHT) on negatiivinen vaikutus vasteeseen oppilailla, joiden kotona käytetään ensisijaisesti muuta kuin testissä käytettävää kieltä (-4.29) ja positiivinen niillä, joiden kotona käytettävä kieli on testissä käytettävä kieli (13.67) (ks. kuva 6). Laittomaan protestointiin osallistumisella (ILLPROT) on negatiivinen vaikutus vasteeseen niillä, joiden kotona käytetään ensisijaisesti muuta kuin testissä käytettävää kieltä (-40.93) ja vähemmän negatiivinen oppilailla, joiden kotona käytettävä kieli on testissä käytettävä kieli (-12.76) (ks. kuva 7).

Yhdysvaikutusten tärkeyttä mallissa voidaan arvioida vertaamalla valittua mallia päävaikutusmalliin (taulukko 6), joka sisältää vain valitun mallin sisältämät päävaikutukset. Huomataan, että yhdysvaikutusermien sisältämien pääkomponenttien vaikutukset voivat muuttua oleellisesti, jos yhdysvaikutukset jätetään pois (ks. kuvat 2–8). Esimerkiksi kielitaustan ja kiinnostuksen politiikkaan välisen yhdysvaikutuksen tapauksessa (ks. kuva 5) päävaikutusmallin mukaan kiinnostus politiikkaan vaikuttaa positiivisesti vasteeseen, mutta valitun mallin mukaan huomataan, että positiivinen vaikutus koskee niitä, jotka käyttävät kotona ensisijaisesti testissä käytettävää kieltä.

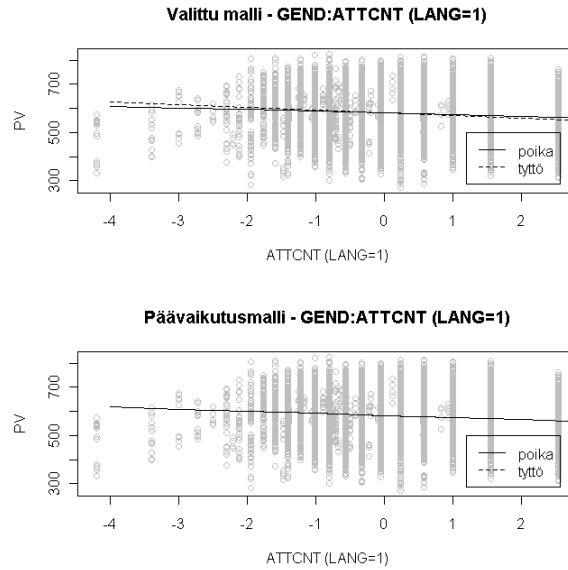
Valitussa mallissa varianssikomponentit ovat: $\sigma_\gamma^2 = 430.51$, $\sigma_\nu^2 = 2844.47$ ja $\sigma_\epsilon^2 = 941.29$ (ks. liite 6). Valitun mallin sisäkorrelaatio on 10.2 %, joka on kansainvälisesti pieni ja tarkoittaa, että Suomen koulujen välinen varianssikomponentti on pieni suhteessa kokonaisvarianssiin. Suomen koulujen välinen vaihtelu on vähäistä.

Taulukko 6: Pelkästään päävaikutukset sisältävän mallin parametriestimaatit, keskirhoeet, t -arvot ja p -arvot

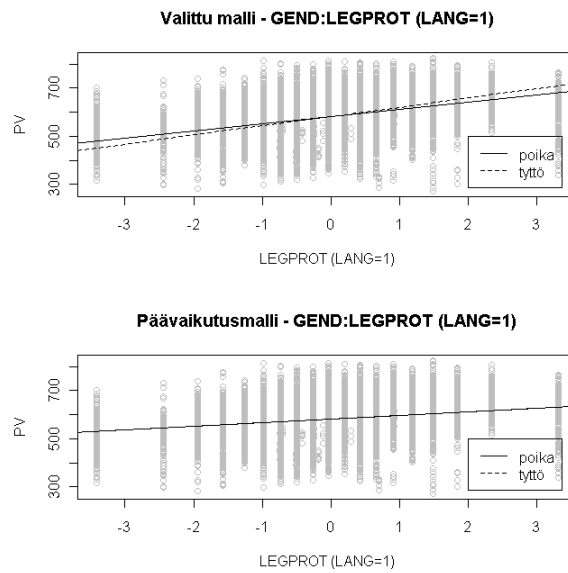
	Est.	SE	t-arvo	p-arvo
(Intercept)	537.21	6.42	83.67	<0.001
GEND	0.19	2.56	0.07	0.940
LANG	45.55	6.24	7.30	<0.001
POLIT	5.46	1.43	3.81	<0.001
RGHT	13.54	1.49	9.06	<0.001
SOSEC	21.15	1.20	17.66	<0.001
DEM	16.28	1.40	11.66	<0.001
COND	-10.51	1.21	-8.67	<0.001
ATTCNT	-8.69	1.23	-7.05	<0.001
PARTSCHL	4.90	1.22	4.02	<0.001
PARTCOM	-6.75	1.14	-5.92	<0.001
ILLPROT	-13.37	1.23	-10.90	<0.001
LEGPROT	14.90	1.43	10.44	<0.001



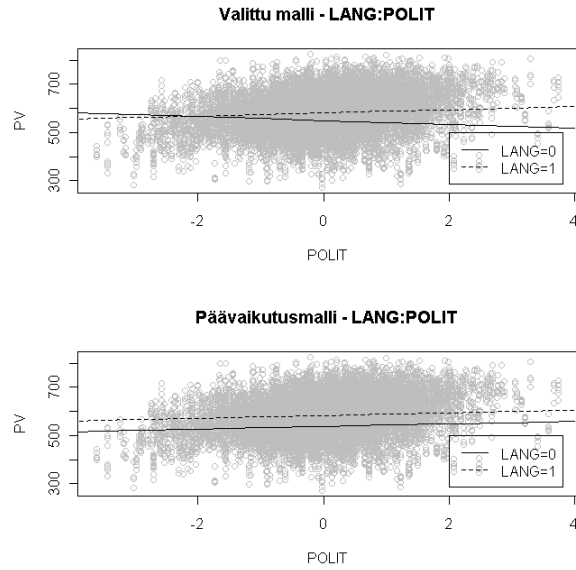
Kuva 2: Sukupuolen ja kodin sosioekonomisen aseman yhteinen vaikutus (GEND, SOSEC)



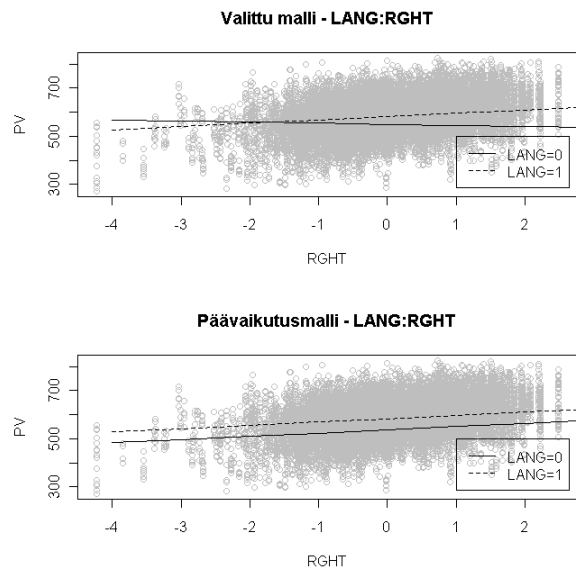
Kuva 3: Sukupuolen ja isänmaallisuuden yhteinen vaikutus (GEND, ATT CNT)



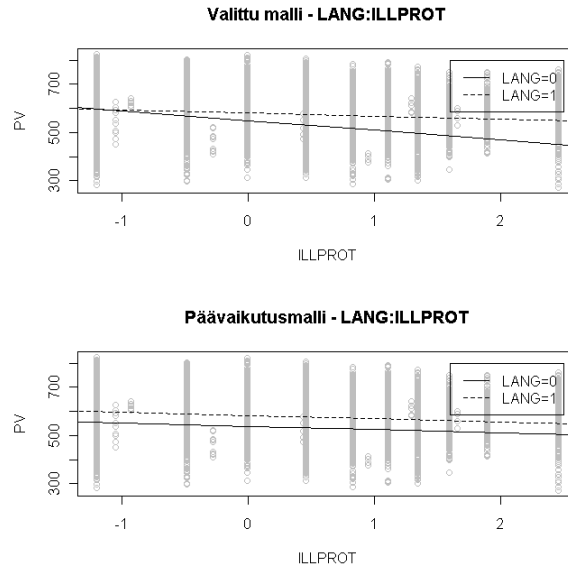
Kuva 4: Sukupuolen ja lailliseen protestointiin osallistumisen yhteinen vaikutus (GEND, LEG PROT)



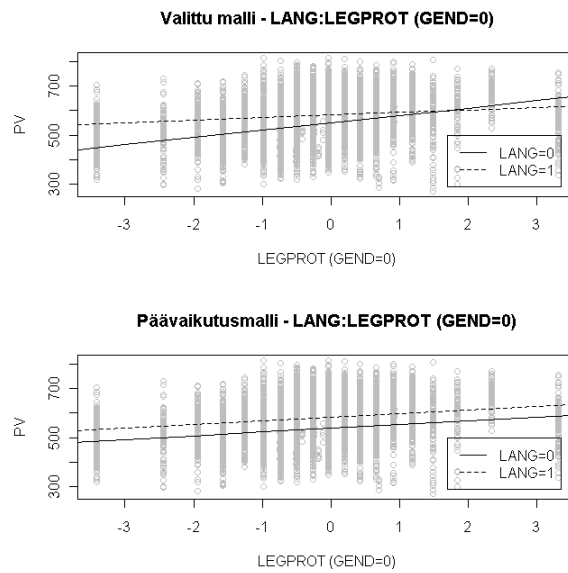
Kuva 5: Kielitaustan ja kiinnostuksen politiikkaan yhteinen vaikutus (LANG, POLIT)



Kuva 6: Kielitaustan ja tasa-arvoon suhtautumisen yhteinen vaikutus (LANG, RGHT)



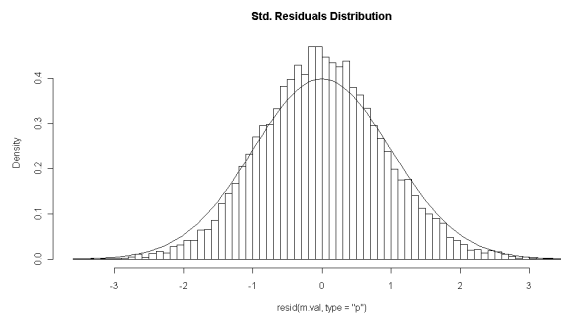
Kuva 7: Kielitaustan ja laittomaan protestointiin osallistumisen yhteinen vaikutus (LANG, ILLPROT)



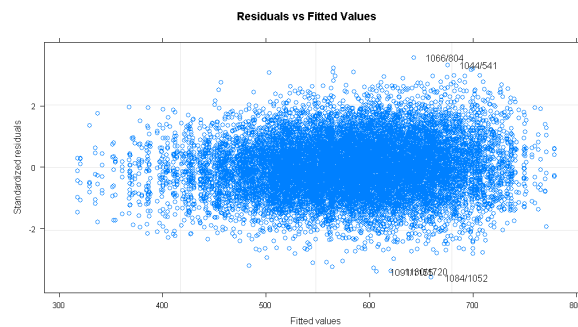
Kuva 8: Kielitaustan ja lailliseen protestointiin osallistumisen yhteinen vaikutus (LANG, LEGPROT)

5.3 Diagnostiikkaa

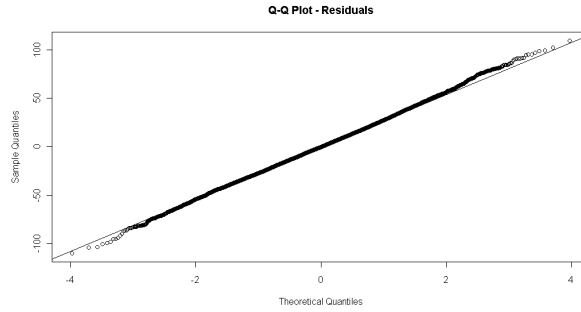
Kuvassa 9 on valitun mallin standardoitu residuaalijakauma, joka näyttäisi noudattavan kohtalaisesti normaalijakaumaa. Kuvassa 10 on valitun mallin standardoidut jäännökset pystyakselilla ja sovitteet vaaka-akselilla. Kuviossa ei näy lineaarista riippuvuutta, joten jäännökset eivät korreloi sovitteiden eivätkä selittäjien kanssa. Jäännösten tasainen vaihtelu kuvastaa virhevarianssin homoskedastisuutta. Kuvan 11 kvantiilikuviosta ei löydy erityisemmin poikkeavia havaintoja ja havaintoja voidaan pitää melko normaalisenä. Kuvan 12 perusteella myös kouluun liittyvä satunnaisvaikutus on melko normaalin joitakin kouluja lukuunottamatta. Sovitteet näyttäisivät noudattavan alkuperäisiä arvoja melko hyvin kuvan 13 perusteella.



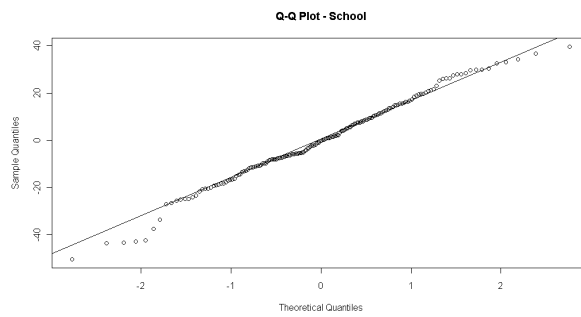
Kuva 9: Standardoitujen jäännösten jakauma



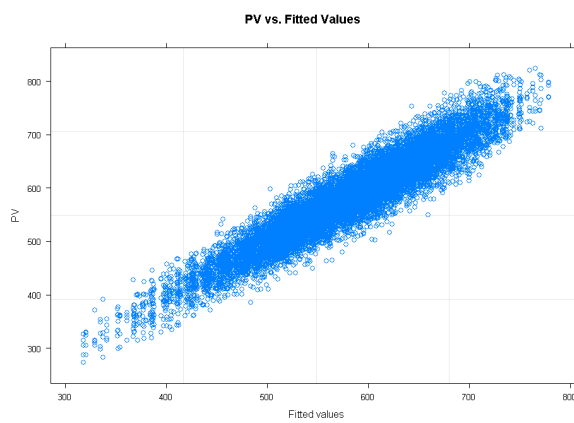
Kuva 10: Standardoidut jäännökset vs. sovitteet



Kuva 11: Jäännösten Q-Q -plot



Kuva 12: Kouluun liittyvän satunnaisvaikutuksen Q-Q -plot



Kuva 13: Alkuperäiset PV:t vs. sovitteet

6 Yhteenveto

Tässä pro gradu -tutkielmassa tarkasteltiin aineistosta saatujen sekä aineiston pohjalta luotujen oppilaiden taustoja kuvaavien muuttujien vaikutusta oppilaiden yhteiskunnalliseen osaamiseen. Oppilaiden osaamista kuvaavat Plausible Value -arvot on generoitu posteriorijakaumista, jotka on saatu taustatietojen ja koevastausten perusteella. Käytettävien taustamuuttujien määrä saatiin pienemmäksi pääkomponenttianalyysin avulla mallintamisen ja tulosten selkeyttämiseksi. Mallinnuksessa käytettiin koulu- ja oppilasvaihtuksen huomioivaa varianssikomponenttimallia, jotta havaintojen mahdollinen korreloiminen sekä koulujen että oppilaiden sisällä saatiin huomioitua.

Osa aineistosta poistettiin taustamuuttujiin liittyvän puuttuvan tiedon takia ja vaikka tästä huolimatta aineiston koko jäi suureksi, niin huomattavan osan puuttuminen aineistosta voi vaikuttaa heikentävästi tuloksien luotettavuuteen. Aineiston laatua voidaan pitää korkeatasoisena otannan, Plausible Value -muuttujien ja käytettyjen menetelmien osalta. Tehtävien määrän tulisi olla riittävän pieni, jotta yläkouluiikäisten keskittymiskyky säilyisi kaikkien kysymysten osalta. Aineiston luotettavuutta on mahdollisesti heikentänyt kysymysten ja tehtävien suuri määrä ja niihin vastaamiseen kulunut aika, joka oli enintään 45 minuuttia sekä kyselylomakkeessa että koeviikkossa. Mallinnusta olisi voinut parantaa ottamalla huomioon tyttöjen ja poikien mahdollinen erivarianssisuus.

Demokraattisuus, osallistuminen koulun toimintaan, kodin sosioekonominen asema ja lailliseen protestointiin osallistuminen vaikuttivat vasteeseen positiivisesti. Osallistuminen järjestötoimintaan, olosuhteet koulussa, isänmaallisuus ja laittomaan protestointiin osallistuminen vaikuttivat vasteeseen negatiivisesti. Järjestötoimintaan osallistumisen negatiivinen vaikutus voi johtua esimerkiksi siitä, että aktiivisuus voi olla enemmänkin aatteellista kuin tiedonhankintaan suuntautuvaa. Parempien kouluolosuhteiden voisi olettaa vaikuttavan positiivisesti vasteeseen, mutta mallinnus antoi päinvastaisen tuloksen. Mahdollisesti huonommat kouluolosuhteet voivat motivoida hankkimaan tietoa asioiden parantamiseksi ja näin vaikuttaa testitulokseen. Testissä huonommin menestyneet oppilaat voivat olla tyytyväisiä kouluolosuhteisiin, koska eivät välitä asiasta tai eivät koe tarvetta muutokseen. Paremmin menestyneet oppilaat voivat vastaavasti nähdä tarvetta muutokseen, mutta eivät saa sitä.

Kiinnostuksella politiikkaan ja suhtautumisella tasa-arvoon on negatiivinen

vaikutus vasteeseen niillä oppilailla, joiden kotona puhutaan ensisijaisesti muuta kuin testissä käytettävää kieltä ja positiivinen vaikutus niillä, joiden kotona puhutaan ensisijaisesti testissä käytettävää kieltä. Tulos voi johtua eroista kulttuureissa.

Koulutuksen kehittämiseksi on tärkeää tietää tämänhetkinen tiedon ja ymmärryksen taso sekä millaisia arvoja ja kiinnostuksen kohteita nuorilla on. Myös näihin vaikuttavat tekijät on hyvä tiedostaa, jotta opetus saataisiin mahdollisimman hyvin vastaamaan tämänhetkisiä tarpeita. Tästä syystä koulutuskimukset ja niistä saatujen aineistojen analysointi yhteiskunnan kannalta hyödyllistä.

Viitteet

- [1] ADAMS, R. J. (2002): *Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.).* Technical report for the OECD Programme for International Student Assessment, OECD, Paris.
- [2] ADAMS, R. J., WILSON, M. R., WANG, W. C. (1997): *The multidimensional random coefficients multinomial logit.* Applied Psychological Measurement, 21, 1-24.
- [3] ADAMS, R. J., WILSON, M. R., WU, M. L. (1997): *Multilevel item response modeling: An approach to errors in variables regression.* Journal of Educational and Behavioral Statistics, 22, 47-76.
- [4] BOCK, R. D., AITKEN, M. (1981): *Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm.* Psychometrika, 46, 443-459, Springer, New York.
- [5] DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B. (1977): *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, Series B, 39, 1-38.
- [6] GELMAN, A., CARLIN, J., STERN, H., RUBIN, D. (1995): *Bayesian data analysis.* First edition, Chapman and Hall, New York.
- [7] GOLDSTEIN, H. (2011): *Multilevel Statistical Models.* 4th edition, Wiley, Chichester.
- [8] JOLLIFFE, I. T. (2002): *Principal Component Analysis.* 2nd edition, Springer, New York.
- [9] MCCULLAGH, P., NELDER, J. (1989): *Generalized Linear Models.* Chapman and Hall, New York.
- [10] MCCULLOCH, C. E., SEARLE, S. R. (2001): *Generalized, linear, and mixed models.* Wiley, New York.
- [11] OECD (2009): *PISA Data Analysis Manual: SPSS Second Edition.* OECD, Paris.
- [12] PINHEIRO, J. C., BATES, D. M. (2000): *Mixed-Effects Models in S and S-PLUS.* Springer, New York.

- [13] R DEVELOPMENT CORE TEAM (2008): *R: A language and environment for statistical computing* [computer software]. R Foundation for Statistical Computing, Vienna. URL <http://www.R-project.org>.
- [14] RASCH, G. (1960): *Probabilistic models for some intelligent and attainment tests*. Nielsen & Lydiche, Copenhagen.
- [15] SCHULZ, W., AINLEY, J., FRAILLON, J., KERR, D., LOSITO, B. (2010): *ICCS 2009 International Report*. International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.
- [16] SCHULZ, W., AINLEY, J., FRAILLON, J. (EDS.) (2011): *ICCS 2009 Technical Report*. International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.
- [17] SUONINEN, A., KUPARI, P., TÖRMÄKANGAS, K. (2010): *Nuorten yhteiskunnalliset tiedot, osallistuminen ja asenteet*. Kansainvälisen ICCS 2009 -tutkimuksen päätulokset. Koulutuksen tutkimuslaitos, Jyväskylän yliopistopaino, Jyväskylä.
- [18] VOLODIN, N., ADAMS, R. J. (2002): *The estimation of polytomous item response models with many dimensions*. Unpublished report, University of Melbourne, Victoria, Australia.
- [19] WARM, T. A. (1989): *Weighted likelihood estimation of ability in item response theory*. *Psychometrika*, 54(3), 427-450, Springer, New York.
- [20] WU, M. L. (2005): *The Role of Plausible Values in Large-Scale Surveys*. *Studies in Educational Evaluation* 31, 114-128, Elsevier, Amsterdam.
- [21] WU, M. L., ADAMS, R. J., WILSON, M. R., HALDANE, S. (2007): *ACER ConQuest: General item response modelling software* [computer program]. Camberwell, Victoria, Australia: Australian Council for Educational Research.

Liitteet

Liite 1: Esimerkkitehtävä 1

Esimerkkitehtävä 1

Julkisella keskustelulla tarkoitetaan ihmisten välistä avointa mielipiteiden vaihtoa. Julkista keskustelua voidaan käydä lehtien palstoilla, televisiossa, radiossa, internet-foorumeissa tai julkisissa kokouksissa. Julkisessa keskustelussa voidaan käsitellä paikallisia, alueellisia, kansallisia tai kansainvälisiä kysymyksiä.

Millä tavoin julkinen keskustelu voi hyödyttää yhteiskuntaa?

Mainitse kaksi erilaista tapaa.

1. _____

2. _____

KOODAUSOHJE

Koodi 2: Yhteiskunnallisen tietämyksen taso 3

Viittaa kahteen eri hyötykategoriaan viidestä alla mainitusta.

- parempi tieto tai ymmärrys jonkin aiheen tai tilanteen asiasisällöstä
- tarjoaa ratkaisuja ongelmiin TAI foorumin, jolta ratkaisuja voi tulla
- yhteiskunnallisen yhteisymmärryksen lisääntyminen, erilaisuuden hyväksyminen tai turhautumisen väheneminen
- lisää ihmisten uskoa tai motivaatiota osallistua yhteiskuntansa asioihin
- edustaa/toteuttaa ihmisten ilmaisuvapauden periaatetta

Koodi 1: Yhteiskunnallisen tietämyksen taso 2

Viitattu vain yhteen edellä mainituista viidestä hyötykategoriasta (myös kun mainittu eri seikat, mutta samasta kategoriasta).

Liite 2: Esimerkkitehtävät 2, 3 ja 4

Esimerkkitehtävä 2

Mikko osti itselleen uudet kengät. Myöhemmin hän sai tietää, että yhtiö, joka kengät oli valmistanut, käyttää tehtaassaan lapsityövoimaa ja maksaa näille hyvin pientä palkkaa. Mikko sanoo, ettei hän enää pane noita kenkiä jalkaansa.

Miksi Mikko kieltäytyy pitämästä uusia kenkiään?

- Hän arvelee, että lasten tekemät kengät eivät kestä kovin pitkään.
- Hän ei halua tukea kengät valmistanutta yhtiötä. (*)
- Hän ei halua tukea lapsia, jotka valmistivat kengät.
- Häntä suututtaa, että hän maksoi kengistä enemmän kuin mikä niiden arvo on.

(Yhteiskunnallisen tietämyksen taso 1)

Esimerkkitehtävä 3

Mikko haluaa muiden ihmisten kieltäytyvän ostamasta kyseisiä kenkiä.

Kuinka hän voi parhaiten yrittää vaikuttaa asiaan?

- ostamalla itse kaikki kengät, jotta kukaan muu ei voi ostaa niitä
- palauttamalla kengät kauppaan ja pyytämällä rahansa takaisin
- tukkimalla kaupan sisäänkäynnin, niin että ihmiset eivät pääse sinne
- kertomalla muille siitä, miten kyseiset kengät valmistetaan (*)

(Yhteiskunnallisen tietämyksen taso alle 1)

Esimerkkitehtävä 4

Mikä on ammattiliittojen päätarkoitus? Niiden päätarkoitus on...

- parantaa valmistettavien tuotteiden laatua.
- nostaa tehtaiden tuotannon määrää.
- parantaa työntekijöiden olosuhteita ja palkkoja. (*)
- luoda tasapuolisempi verojärjestelmä.

(Yhteiskunnallisen tietämyksen taso 2)

Liite 3: Kyselylomake esimerkki 1

SINÄ ITSE

K1 Milloin olet syntynyt?

(Kirjoita kuukausi ja vuosi)

_____ 19____
kuukausi vuosi

K2 Oletko tyttö vai poika?

tyttö poika

₁ ₂

K2b Mihin seuraavista ryhmistä katsot lähinnä kuuluvasi?

(Merkitse rasti vain yhteen ruutuun)

Suomenkieliseen kantaväestöön ₁

Suomenruotsalaiseen väestöön ₂

Saamelaisiin ₃

Romaneihin ₄

Maahanmuuttajiin ₅

Paluumuuttajiin ₆

Johonkin muuhun ₇

Liite 4: Kyselylomake esimerkki 2

K9 Minkä tasoinen koulutus isälläsi tai miespuolisella huoltajallasi on?

Jos et ole varma oikeasta vaihtoehdosta, kysy neuvoa kokeen valvojalta.

(Merkitse rasti vain yhteen ruutuun)

Yliopisto- tai ammattikorkeakoulututkinto
(esim. maisteri, kandidaatti, lääkäri,
erikoissairaanhoitaja, diplomi-insinööri,
amk-insinööri, ekonomi, tradenomi) ₁

Erikoisammattitutkinto lukion tai ammatti-
koulun jälkeen (esim. opistotutkinto,
sairaanhoitaja, opistoinsinööri, yo-
merkonomi, merkonomi) ₂

Lukiotutkinto tai keskiasteen ammatillinen
tutkinto (esim. ylioppilas, ammattikoulu,
perus- tai lähihoitaja, teknikko, merkantti) ₃

Perus- tai keskikoulu ₄

Kansa- tai kansalaiskoulu ₅

Muutama vuosi kansakoulua ₆

Liite 5: Pääkomponentit sekä niiden sisältämät muuttujat ja lataukset (WLE tarkoittaa, että muuttuja on WLE-pistemäärämuuttuja)

POLIT	Kiinnostus politiikkaan	lataus
intpols	kiinnostus politiikkaan (WLE)	0,802
inpolef	koettu poliittisten asioiden tuntemus (WLE)	0,794
infpact	osallistuminen politiikkaan lähitulevaisuudessa (WLE)	0,724
citeff	arvio omasta yhteiskunnallisesta kyvykkyydestä (WLE)	0,701
elecpart	aikomus äänestää vaaleissa (WLE)	0,658
polpart	aikomus osallistua politiikkaan (WLE)	0,634
poldisc	politiikasta keskustelu koulun ulkopuolella (WLE)	0,633
citcon	suhtautuminen kansalaisvelvollisuuksiin (WLE)	0,608
parint	vanhempien kiinnostus politiikkaan	0,526
suppart	jonkin puolueen kannattaminen	0,476
RGHT	Suhtautuminen tasa-arvoon	lataus
ethrght	suhtautuminen etnisten ryhmien oikeuksiin (WLE)	0,841
immrgh	suhtautuminen maahanmuuttajien oikeuksiin (WLE)	0,818
geneql	suhtautuminen sukupuolten tasa-arvoon (WLE)	0,717
citsoq	suhtautuminen toimintaan hyvien asioiden puolesta (WLE)	0,606
SOSEC	Kodin sosioekonominen asema	lataus
hiscd	vanhempien korkein koulutustaso	0,803
hisei	vanhempien korkein ammattiasema	0,789
homelit	kotona olevien kirjojen lukumäärä	0,607
sisced	odotettu koulutustaso	0,588
DEM	Demokraattisuus	lataus
valparts	oppilaiden osallistuminen koulujen päätöksentekoon (WLE)	0,792
demval	demokraattisten arvojen kannattaminen (WLE)	0,749
intrust	luottamus yhteiskunnallisiin instituutioihin (WLE)	0,535
COND	Olosuhteet koulussa	lataus
stutrel	oppilaiden ja opettajien suhteet (WLE)	0,762
opdisc	luokkahuonekeskustelujen koettu avoimuus (WLE)	0,728
studinf	oppilaan mielipiteen huomioiminen koulussa (WLE)	0,624

Liite 6: Koulujen sisäkorrelaatiokertoimen laskeminen

```
# ICC VALITTU MALLI:
> m.val = lme(PV ~ GEND+LANG+POLIT+RGHT+SOSEC+DEM+COND+ATTCNT+PARTSCHL+PARTCOM+ILLPROT+LEGPROT
+ +GEND:SOSEC+GEND:ATTCNT+GEND:LEGPROT
+ +LANG:POLIT+LANG:RGHT+LANG:ILLPROT+LANG:LEGPROT,data=data2,random= 1|ID/STUDENT,method="ML");
> summary(m.val)
:
Random effects:
Formula: ~1 | ID
(Intercept)
StdDev: 20.74870

Formula: ~1 | STUDENT %in% ID
(Intercept) Residual
StdDev: 53.33357 30.68038
:
> 20.74870^2/(20.74870^2+53.33357^2+30.68038^2)
[1] 0.1021066
```

Liite 7: stutrel-muuttujan (oppilaiden ja opettajien suhteet) taustalla olevat kysymykset

K18 Missä määrin olet samaa tai eri mieltä seuraavien sinua ja kouluasi koskevien väittämien kanssa?

(Merkitse rasti vain yhteen ruutuun kullakin rivillä)

	Täysin samaa mieltä	Samaa mieltä	Eri mieltä	Täysin eri mieltä
a) Useimmat opettajani kohtelevat minua oikeudenmukaisesti	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b) Oppilaat tulevat hyvin toimeen useimpien opettajien kanssa	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
c) Useimmat opettajat ovat kiinnostuneita oppilaiden hyvinvoinnista	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e) Useimmat opettajani todella kuuntelevat, mitä sanottavaa minulla on	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f) Jos tarvitsen lisäapua, saan sitä opettajiltani	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Liite 8: opdisc-muuttujan (luokkahuonekeskustelujen koettu avoimuus) taustalla olevat kysymykset

K16 Kun tavallisilla oppitunneilla käsitellään poliittisia ja yhteiskunnallisia asioita, kuinka usein tapahtuu seuraavaa?

(Merkitse rasti vain yhteen ruutuun kullakin rivillä)

	<i>Ei koskaan</i>	<i>Harvoin</i>	<i>Toisi- naan</i>	<i>Usein</i>
b) Opettajat kannustavat oppilaita tekemään itsenäisiä päätöksiä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Opettajat rohkaisevat oppilaita ilmaisemaan mielipiteitään	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Oppilaat ottavat esille ajankohtaisia poliittisia tapahtumia keskusteltavaksi luokassa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Oppilaat ilmaisevat luokassa mielipiteitään, vaikka ne eroaisivatkin luokan enemmistön mielipiteistä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f) Opettajat kannustavat oppilaita keskustelemaan asioista erilaisia mielipiteitä edustavien ihmisten kanssa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g) Selittäessään asioita tunnilla opettajat tarkastelevat niitä eri puolilta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Liite 9: studinf-muuttujan (oppilaan mielipiteen huomioiminen koulussa) taustalla olevat kysymykset

K17 Missä määrin oppilaiden mielipiteet otetaan koulussasi huomioon seuraavista asioista päätettäessä?

(Merkitse rasti vain yhteen ruutuun kullakin rivillä)

	<i>Paljon</i>	<i>Jossain määrin</i>	<i>Vähän</i>	<i>Ei lainkaan</i>
a) Millä tavoin tunneilla opetetaan	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b) Mitä tunneilla opetetaan	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
c) Opetus-/oppimateriaalit	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
d) Lukujärjestys	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e) Luokan säännöt	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f) Koulun säännöt	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Liite 10: Valitun mallin muuttujien väliset korrelaatiot

	GEND	LANG	POLIT	RGHT	SOSEC	DEM	COND	ATTCNT	PARTSCHL	PARTCOM
GEND	1	-0.03	0.05	-0.10	0.04	0.12	0.03	0.14	-0.09	-0.12
LANG	-0.03	1	0.29	0.28	0.01	-0.01	0.00	-0.05	0.03	0.02
POLIT	0.05	0.29	1	-0.17	0.00	-0.02	-0.05	0.02	-0.03	-0.05
RGHT	-0.10	0.28	-0.17	1	-0.01	-0.10	-0.02	-0.01	0.01	0.04
SOSEC	0.04	0.01	0.00	-0.01	1	-0.04	0.07	0.07	-0.08	0.07
DEM	0.12	-0.01	-0.02	-0.10	-0.04	1	-0.07	-0.26	-0.08	0.04
COND	0.03	0.00	-0.05	-0.02	0.07	-0.07	1	-0.09	-0.01	-0.04
ATTCNT	0.14	-0.05	0.02	-0.01	0.07	-0.26	-0.09	1	0.02	-0.02
PARTSCHL	-0.09	0.03	-0.03	0.01	-0.08	-0.08	-0.01	0.02	1	-0.19
PARTCOM	-0.12	0.02	-0.05	0.04	0.07	0.04	-0.04	-0.02	-0.19	1
ILLPROT	0.02	0.26	0.12	0.37	0.05	-0.01	0.04	0.01	0.05	0.03
LEGPROT	-0.03	-0.25	-0.43	-0.37	-0.02	-0.01	0.02	-0.01	-0.02	0.02
GEND:SOSEC	0.00	-0.03	-0.02	0.01	-0.69	0.00	-0.05	-0.03	0.01	-0.02
GEND:ATTCNT	0.04	0.03	0.00	-0.01	-0.02	0.05	-0.02	-0.60	0.01	-0.01
GEND:LEGPROT	0.04	0.01	0.05	-0.03	0.10	-0.01	0.06	0.04	-0.01	-0.05
LANG:POLIT	-0.01	-0.28	-0.97	0.16	-0.05	0.01	0.02	-0.04	-0.02	0.02
LANG:RGHT	0.01	-0.28	0.17	-0.98	-0.01	0.01	-0.01	0.02	0.00	-0.03
LANG:ILLPROT	0.02	-0.26	-0.11	-0.37	-0.04	0.02	-0.02	0.00	-0.05	-0.03
LANG:LEGPROT	-0.01	0.25	0.42	0.38	0.01	-0.01	-0.03	0.01	0.00	-0.01

	ILLPROT	LEGPROT	GEND:SOSEC	GEND:ATTCNT	GEND:LEGPROT	LANG:POLIT	LANG:RGHT	LANG:ILLPROT	LANG:LEGPROT
GEND	0.02	-0.03	0.00	0.04	0.04	-0.01	0.01	0.02	-0.01
LANG	0.26	-0.25	-0.03	0.03	0.01	-0.28	-0.28	-0.26	0.25
POLIT	0.12	-0.43	-0.02	0.00	0.05	-0.97	0.17	-0.11	0.42
RGHT	0.37	-0.37	0.01	-0.01	-0.03	0.16	-0.98	-0.37	0.38
SOSEC	0.05	-0.02	-0.69	-0.02	0.10	-0.05	-0.01	-0.04	0.01
DEM	-0.01	-0.01	0.00	0.05	-0.01	0.01	0.01	0.02	-0.01
COND	0.04	0.02	-0.05	-0.02	0.06	0.02	-0.01	-0.02	-0.03
ATTCNT	0.01	-0.01	-0.03	-0.60	0.04	-0.04	0.02	0.00	0.01
PARTSCHL	0.05	-0.02	0.01	0.01	-0.01	-0.02	0.00	-0.05	0.00
PARTCOM	0.03	0.02	-0.02	-0.01	-0.05	0.02	-0.03	-0.03	-0.01
ILLPROT	1	-0.55	-0.03	0.00	0.02	-0.12	-0.37	-0.98	0.54
LEGPROT	-0.55	1	0.02	0.01	-0.16	0.42	0.38	0.54	-0.97
GEND:SOSEC	-0.03	0.02	1	0.02	-0.18	0.03	0.00	0.03	-0.01
GEND:ATTCNT	0.00	0.01	0.02	1	-0.05	0.00	0.01	0.01	0.00
GEND:LEGPROT	0.02	-0.16	-0.18	-0.05	1	-0.04	0.02	-0.03	0.01
LANG:POLIT	-0.12	0.42	0.03	0.00	-0.04	1	-0.16	0.12	-0.43
LANG:RGHT	-0.37	0.38	0.00	0.01	0.02	-0.16	1	0.37	-0.39
LANG:ILLPROT	-0.98	0.54	0.03	0.01	-0.03	0.12	0.37	1	-0.54
LANG:LEGPROT	0.54	-0.97	-0.01	0.00	0.01	-0.43	-0.39	-0.54	1