

2

Koulutussosiologista tutkimusta suomalaisilla rekisteripohjaisilla aineistoilla

Elina Kilpi-Jakonen & Jani Erola¹

Luvussa käydään läpi rekisteripohjaisen koulutussosiologisen tutkimuksen historiaa Suomessa ja tutkimusaiheita, joita rekisteriaineistoilla on analysoitu. Keskustelemme rekisteripohjaisen tutkimuksen vahvuuksista ja heikkouksista sekä tulevaisuuden näkymistä. Yksi vahvuuksista on mahdollisuus analysoida yksilöiden sosiaalisia suhteita: linkittää yksilöitä vanhempiinsa ja muihin sukulaisiin sekä samaa koulua käyviin tai samassa naapurustossa asuviin. Toinen vahvuus on mahdollisuus pitkittäisanalyysiin, joissa samoja henkilöitä pystytään tutkimaan elinkaaren yli. Molempia vahvuuksia on hyödynnetty monipuolisesti suomalaisessa rekisteritutkimuksessa. Luvussa käydään läpi käytännön esimerkki siitä, miten rekisteriaineistoja yhdistetään ja analysoidaan. Tässä esimerkissä tutkimme koulutuksen ylisukupolvisuutta ja sitä, missä määrin vanhempien koulutustaso selittää ylioppilaiden korkeakouluttautumista, selittääkö eroa ylioppilaskirjoitusten äidinkielen kokeen arvosanat ja minkälaista ajallista muutosta on tapahtunut. Ylisukupolvinen eriarvoisuus siirtymissä lukioista korkeakouluun väheni 1970-luvulla syntyneiden keskuudessa, mutta on sen jälkeen pysynyt tasaisena. Eriarvoisuutta on edelleen havaittavissa, vaikka äidinkielen arvosana vakioitaisiin. Tämä viittaa siihen, että perhetaustaerot eivät selity kyky- tai osamiserioilla. Luvun lopuksi keskustellaan rekisteritutkimuksen tulevaisuudesta: kysely- ja rekisteriaineistojen laajamittaisemmasta linkittämisestä ja rekisteritutkimuksen infrastruktuurin parantamisesta.

¹ Kilpi-Jakonen ja Erola ovat osallistuneet luvun kirjoittamiseen yhtäläisellä panoksella.

Rekisteripohjaisen tutkimuksen pitkä perinne koulutussosiologiassa

Ensimmäinen yksilötason rekisteridataa käyttänyt koulutussosiologinen tutkimus julkaistiin jo vuonna 1984. Hermusen Tilastokeskukselle kirjoittamassa raportissa ”Oppilaiden sosiaalinen tausta” tutkittiin sosiaalisen perhetaustan yhteyttä opiskelijoiden koulutusvalintoihin. Tutkimusta varten yhdistettiin rekisteri- ja väestölas-kentätietoja oppilaista ja heidän vanhemmistaan.

Tutkimusta voidaan pitää yhtenä ensimmäisistä moderneista rekisteritutkimuksista kahdesta syystä. Ensinnäkin, kun useammassa rekisterissä samat yksilöt voidaan identifioida luotettavasti henkilötunnuksen avulla, voidaan eri rekistereiden tietosisällöt myös yhdistää yksilötasolla. Vaikka rekisteritietoja on hyödynnetty tutkimuksessa yhtä pitkään kuin väestörekistereitä on ollut olemassa, vasta tämä mahdollisuus teki rekistereitä hyödyntävästä tutkimuksesta sisällöllisesti ja analyttisesti kilpailukykyistä kyselyaineistoja hyödyntävien tutkimusten kanssa. Suomessa koko väestön kattava henkilötunnus oli otettu käyttöön 1971, sitä edeltävä sosiaaliturvatunnus jo 1964. Toiseksi yksilötason tunnistetietoja sisältävät rekisterit antoivat mahdollisuuden muodostaa aineistoja myös ihmisten välisistä sosiaalisista yhteyksistä sellaisella tarkkuudella ja luotettavuudella, joka ei aiemmin ollut mahdollista millään muulla aineistolla. Rekistereiden sisältämät tiedot esimerkiksi perheenjäsenistä, luokka- tai työtovereista tarjoavat edelleen luotettavampaa ja tarkempaa tietoa kokonaisten populaatioiden sosiaalisista siteistä kuin oikeastaan mikään muu tutkimuseettisesti hyväksyttävä tiedonkeruutapa.

Varhaisessa koulutussosiologisessa tutkimuksessa nojattiin vahvasti rekisteriaineistojen korkealle laadulle. Myös Hermusen käyttämät tutkimusmenetelmät olivat kuvailevia, eikä tulosten tieteellisestä ja yhteiskunnallisesta relevanssista käyty kirjassa kovin laajaa keskustelua. Silti jo tämän tutkimuksen johtopäätös oli sama kuin lukuisissa sitä seuranneissa: koulutuksen laajentaminen ei ollut poistanut perhetaustaeroja koulutukseen osallistumisessa sen paremmin kuin koulutustason tai -alan valinnassa. Isoaho, Kivinen ja Rinne toistivat Hermusen tutkimuksen vuonna 1990 ja Kivinen ja Rinne laajensivat aikaisempia tutkimuksia sarjalla uusia julkaisuja vuonna 1995. Fokus uusissa julkaisuissa säilyi koulutuksen sosiaalisen periytymisen tutkimisessa. Valkonen, Pensola ja Jalovaara (1996) käyttivät myös rekisteriaineistoja tutkimuksessaan peruskoulu-uudistuksen vaikutuksesta sosiaaliseen periytymiseen Opetushallituksen vuonna 1996 julkaisemassa kirjassa. Tutkimustulosten tieteellinen ja yhteiskunnallinen tulkinta laajeni merkittävästi näiden tutkimusten myötä, vaikka menetelmälliset lähestymistavat säilyivät edelleen nykypäivän kriteerein arvioituna yksinkertaisina.

Voidaan siis sanoa, että koulutussosiologinen tutkimus oli suomalaisen rekisteritutkimuksen ensimmäisessä aallossa 1980-luvulta 1990-luvulle hyvin keskeisessä roolissa. On silti silmiinpistävää, että edistysaskeleita otettiin tänä aikana lähinnä rekisteriaineistojen innovatiivisessa hyödyntämisessä, ei niinkään käytettyjen menetelmien osalta. Vaikka tilastollisella mallintamisella oli muiden sosiaalisten erojen tutkimuksessa Suomessakin pitkät perinteet (esim. Pöntinen, 1983), koulutusso-

siologiset tutkimukset perustuivat pitkään ristiintaulukoinneille, näistä laskettujen vetosuhteiden vertailuille sekä ryhmien välisten keskiarvojen vertailuille. Tämän suhteen kontrasti oli suuri verrattuna vaikkapa väestötieteelliseen tutkimukseen, jossa rekisteriaineistojen hyödyntäminen myös aloitettiin varhain mutta jossa aineistoja analysoitiin yksilötason datana suhteellisen vaativia tilastomenetelmiä hyödyntäen (esim. Valkonen ym., 1990).

2000-luvulle tultaessa kiinnostus koulutussosiologisiin tutkimusaiheisiin laajeni taloustieteisiin. Esimerkiksi Uusitalon (1999) väitöskirjassa *“Essays in economics of education”* tutkittiin rekisteriaineistojen avulla samoja koulutuksen sosiaaliseen eriarvoisuuteen ja ylisukupolvisiin tekijöihin liittyviä aiheita kuin oli jo joitakin vuosia tutkittu koulutussosiologian alalla. Menetelmällisesti väitöskirja oli kuitenkin huomattavasti edellä koulutussosiologisia vastineitaan Suomessa. Uusitaloa on seurannut suuri määrä rekistereitä hyödyntäviä koulutuksen taloustieteen tutkijoita, joiden monen tuloksia voidaan pitää relevantteina koulutussosiologiselle tutkimukselle (esim. Pekkarinen, 2008; Pekkarinen ym., 2009; Pekkala Kerr ym., 2013; Karhunen & Uusitalo, 2017; Tervonen ym., 2018).

Koulutussosiologisessa tutkimuksessa käytettyjen menetelmien osalta tilanne kuitenkin muuttui nopeasti 2000-luvun ensimmäisenä vuosikymmenenä. Laajoja ja laadukkaita yksilötason rekisteripohjaisia aineistoja alettiin analysoida edistynein tilastomenetelmin. Ensimmäisenä myös menetelmällisesti modernina koulutussosiologisena tutkimuksena voidaan pitää Kauppinen väitöskirjaa vuodelta 2004 sekä väitöskirjalle perustuvia korkeatasoisissa kansainvälisissä joulnealeissa julkaistuja artikkeleita (Kauppinen, 2007, 2008). Kauppinen tutki perhetaustan ja asuinalueen merkitystä helsinkiläisnuorten toisen asteen kouluttautumiseen käyttäen monitasomalleja. Samoihin aikoihin rekistereitä ovat alkaneet käyttää myös muut sosiologit. Tämän luvun kirjoittajien ensimmäiset kontribuutiot löytyvät heti seuraavilta vuosilta. Erola tutki vuoden 2009 artikkelissaan koulutuksen merkitystä sosiaalisen liikkuvuuden muutokseen Suomessa 1970-luvulta 2000-luvulle ja Kilpi-Jakonen väitöskirjassaan 2010 (Kilpi, 2010) maahanmuuttajataustaisten nuorten koulutusta.

2010-luvulle tultaessa suomalaisen rekisteripohjaisen koulutussosiologisen tutkimuksen – ja aihetta käsittelevien tutkimushankkeiden – määrä moninkertaistui nopeasti. Tänä päivänä tutkimusala on tieteelliseltä tasoltaan kansainvälisesti tunnustettua huippuluokkaa ja sitä julkaistaan sosiologian alan kärkeijoulnealeissa. Tämä perustuu paitsi tutkijoiden osaamiselle, myös sille, että rekistereitä hyödyntämällä voidaan vastata tutkimuskysymyksiin, joihin vastaaminen muilla aineistoilla olisi erittäin vaikeaa. Suomalaisia rekistereitä hyödyntäen on tutkittu tutkimusaiheita, joiden luotettava tutkiminen olisi käytännössä mahdotonta monissa muissa yhteiskunnissa – kuten koulutukseen uudelleenhakeutuminen aikaisemmassa haussa hylkäämisen jälkeen (Heiskala ym., 2023) tai sitä, miten etäisyys koulupaikkaan on yhteydessä sukupuolelle epätyypilliseen toisen asteen koulutukseen (Prix ym., 2024).

Perhetaustan mittaaminen

Koulutussosiologisessa rekisteritutkimuksessa on usein keskitytty perhetaustan ja koulutuksen välisen yhteyden selvittämiseen. Paitsi että perhetaustan mukaiset erot ovat eriarvoisuuden tutkimisen näkökulmasta keskeisiä, selittää aiheen suosiota myös rekistereistä saatavan perhetaustatiedon luotettavuus.

Kyselytutkimuksessa perhetaustatietoa pyydetään yleensä vastaajalta itseltään. Tämä laajalti käytössä oleva käytäntö asettaa kuitenkin omat haasteensa tutkimuksellisesti: esimerkiksi PISA-aineistossa 15-vuotiaat systemaattisesti aliarvioivat vanhempiensa keskiasteen koulutusta ja yliarvioivat korkea-asteen suorittamista (Lehti & Laaninen, 2021). Vääristymä voi syntyä monesta eri syystä: että vastaajina toimivat lapset eivät yksinkertaisesti tiedä vanhempiensa koulutustasoa tai että he raportoivat sen tarkoituksella tai tarkoittamattaan väärin. Perhetaustatiedon virheellisuuden lisäksi kyselyistä saatavan perhetaustatiedon luotettavuutta rajoittaa myös sosioekonomisesti ja sukupuolen mukaan valikoitunut vastaajakato ja osittainen tietojen puuttuminen. Ei olekaan yllätys, että jo Hermusen vuoden 1984 tutkimusta motivoidaan sillä, että koulutustilastoja ja väestölaskentatietoja yhdistämällä voidaan luotettavalla tavalla arvioida henkilön kotitautaa.

Rekistereiden kautta voidaan myös arvioida lapsuuden perheympäristön muita resursseja – tai niiden puutetta – poikkeuksellisella tarkkuudella. Esimerkiksi Suomea ja Yhdysvaltoja koskevassa tutkimuksessa selvitettiin, miten isovanhempien, setien ja tätien kouluttautuminen voi kompensoida vanhempien matalaa koulutusta lasten kouluttautumisen selittäjänä (Erola ym., 2018). Tutkimuksen yhdysvaltalaisaineistona käytettiin kyselypohjaista, erittäin laadukkaana pidettyä *Panel study on income dynamics* -aineistoa, jonka avulla perheiden jäsenten elämänkulkua eri sukupolvien läpi on mahdollista seurata 1960-luvun lopulta alkaen.

Suomen kohdalla vastaava asetelma toteutettiin rekistereitä hyödyntäen. Tutkimuksen tulos oli selvä: kompensaatiolla oli selvä yhteys kouluttautumiseen kummassakin maassa, mutta moninkertaisesti vahvempi Yhdysvalloissa. Yhdysvaltalaisaineiston rajoitteiden vuoksi, jotka kiteytyvät suhteellisen pieneen tutkittavien joukkoon, jolta oli saatavissa tiedot kolmen sukupolven (ja laajennetun perheen) koulutustasosta, tutkimuksessa käytettiin suhteellisen yksinkertaista setien ja tätien koulutusindikaattoria: (aineistossa olevan) korkeimmin koulutetun sedän tai tädin koulutusta vuosina. Sen sijaan aiemmassa vain Suomea koskeneessa tutkimuksessa laajennetun perheen kouluttautumisen merkitystä oli ollut mahdollista testata monin verroin monipuolisemmin, esimerkiksi laajennetun perheen kouluttautumisen kokonaismääränä, sen keskimääräisenä tasona perheverkostossa ja korkeimpana saavutettuna koulutuksena kaikkien laajennetun perheen jäsenten joukossa (Lehti & Erola, 2017).

Rekistereiden avulla erityyppisten perheiden resurssien merkitystä voidaan vertailla yhdessä ja erikseen tavoilla, jotka eivät olisi mahdollisia muita aineistolähteitä käyttäen. Esimerkkeinä voidaan mainita tutkimukset, joissa on analysoitu vanhempien työttömyyden ja sosiaaliturva-asiakkuuden (Lehti, Erola & Karhula,

2019), vanhempien eron (Kailaheimo-Lönnqvist ym., 2025) tai kuoleman (Kailaheimo-Lönnqvist & Erola, 2020; Kailaheimo-Lönnqvist & Kotimäki, 2020; Prix & Erola, 2017) yhteyttä lasten koulutukseen. Rekistereiden käyttäminen on erityisen suositeltavaa tutkittaessa koulutuksen roolia ylisukupolvisessa luokka- ja tuloliikkuvuudessa, koska perhetaustan luotettava mittaaminen on myös näissä kysymyksissä avainasemassa. Erityisesti tulojen mittaaminen on – molempien sukupolvien osalta – luotettavampaa rekisterien avulla verrattuna kyselypohjaisiin tutkimuksiin (esim. Golsteyn & Hirsch, 2019).

Perhetaustan mittaaminen ei kuitenkaan ole edes rekistereiden avulla täysin ongelmaton. Esimerkiksi tietoa ulkomailta suoritetuista tutkinnoista ei pääsääntöisesti saada suomalaisista rekistereistä, mikä on merkittävä rajoite maahanmuuttajien kouluttautumista tutkittaessa. Tämä puute ei kuitenkaan ole estänyt maahanmuuttajien koulutuksen rekisteripohjaista tutkimusta Suomessa (esim. Kilpi, 2010; Kilpi-Jakonen 2011, 2012, 2017). Osaa perheiden resursseista, kuten kulttuurista pääomaa, on myös vaikea mitata rekistereiden avulla. Muista pohjoismaista löytyy kuitenkin erinomaisia esimerkkejä siitä, miten kulttuurista pääomaa voi rekistereiden avulla lähestyä, kuten eri kouluaineiden testitulosten avulla (Andersen & Hansen, 2012) tai kirjastokäyttäytymisen perusteella (Blaabæk, 2023).

Rekisteritutkijoilla on lisäksi menetelmällinen keino tutkia myös mittaamattomien perhetaustatekijöiden vaikutusta niin sanottujen sisarusmallien avulla. Malleja varten tarvitaan tietoa vähintään kahdelta sisarukselta esimerkiksi koskien heidän koulutustaan. *Sisaruskorrelaatioiden* avulla voidaan tutkia perhetaustan kokonaisvaikutuksia haluttuihin lasten lopputulemiin. Oletus on, että kaikkea, mikä tekee sisarukset samankaltaiseksi, voidaan pitää perhetaustaan liittyvänä. Tätä samankaltaisuutta mittaa sisarusten jakama varianssi, joka voidaan suhteuttaa kokonaisvarianssin määrään. Menetelmää käyttäen on muun muassa vertailtu perhetaustan kokonaisvaikutusta koulutuksessa Suomessa, Saksassa, Norjassa, Ruotsissa, Isossa-Britanniassa ja Yhdysvalloissa (Grätz ym., 2021), ja sitä on käytetty myös maahanmuuttotastaisten nuorten aikuisten koulutuksen tutkimiseen Suomessa (Hyvärinen & Erola, 2011). Toisaalta osassa tutkimuksissa on pyritty käyttämään sisarusia erottavia tekijöitä eri perhetaustatekijöiden kausaalisten vaikutusten arviointiin. Niin sanottujen *kiinteiden vaikutusten sisarusmallien* ajatuksena on vakioida kaikki myös havaitsemattomat perhetaustatekijät kontrolloimalla muuttujia, joiden suhteen sisarukset poikkeavat toisistaan. Esimerkiksi Lehti, Erola ja Karhula (2019) hyödyntävät lasten ikää vanhempien työttömyyden kokemisessa päätelläkseen vanhempien työttömäksi jäämisen kausaalista vaikutusta lasten kouluttautumiseen. Kailaheimo-Lönnqvist ja Erola (2020) hyödyntävät samankaltaista asetelmaa vanhemman kuoleman kausaalivaikutuksen identifioimiseksi.

Koulutussiirtymien seuraaminen elämänkaareissa

Viimeksi mainitut vanhempien työttömyyttä ja kuolemaa koskevat tutkimukset hyödynsivät toista rekisteriaineistojen vahvuutta, mahdollisuutta seurata yksilöitä läpi elämänkaaren. Erityisen hyödyllistä tämä on koulutussiirtymien tutkimisen näkökulmasta. Koulutuksellisten nivelvaiheiden keskeinen merkitys lopullisen koulutustason määräytymiselle ymmärretään hyvin koulutustutkimuksessa. Näiden luotettava tutkiminen asettaa tutkimusaineistoille suuria vaatimuksia: yksilöitä pitää voida seurata pitkään, usein jopa koko koulutusuran ajan, ensin varhaiskasvatukseen, tämän jälkeen peruskouluun, toiselle asteelle peruskoulutuksesta ja lopulta korkea-asteelle ja jopa jatko-opintoihin. Vaikka tällaisia siirtymiä huomioivia kyselypohjaisia aineistoja löytyykin useista maista, on vastaajakato niissä lähes aina merkittävä ongelma jo pelkän elämänvaiheen vuoksi (esim. Bingley & Martinello, 2017; Siddiqui ym., 2019). Varsinkin koulutusuran loppuvaiheen siirtymät tapahtuvat usein samaan aikaan, kun nuoret siirtyvät asumaan itsenäisesti eikä heitä enää tavoiteta kyselyihin vastaamista varten. Rekisteriaineistoissa, joista poistutaan pääosin maastamuuttamisen ja kuoleman kautta, tällainen kato ei ole koskaan ongelma.

Paljon julkisuutta saaneita esimerkkejä rekistereiden avulla toteutetuista koulutussiirtymiä koskeneista tutkimuksista ovat koskeneet varhaiskasvatukseen osallistumisen yhteyttä myöhäisempään koulutukseen (Hiilamo ym., 2018; Karhula ym., 2017). Ehkäpä laajimmin tätä rekisteriaineistojen ominaisuutta ovat kuitenkin hyödyntäneet Härkönen ja Sirniö (2020), jotka tarkastelivat perhetaustan merkitysten vaiheittaista muutosta toisen ja korkea-asteen koulutussiirtymien läpi. Tutkimus antaa erityisen tarkan kuvan perhetaustan mukaisen eriarvoisuuden muodostumisesta koulutusuran eri vaiheissa. Syntymäkohortit 1960–1985 kattava tutkimus antaa tietoa myös tällaisen eriarvoisuuden ajallisesta muutoksesta.

Paitsi että rekistereiden perusteella voidaan seurata toteutuneita siirtymiä, ne mahdollistavat myös toteutumatta jääneiden siirtymien tutkimisen esimerkiksi hakerekistereiden perusteella. Melko varhainen esimerkki rekistereiden tällaisen ominaisuuden hyödyntämisestä on Rinteen ja kumppanien (2008) tutkimus *”Yliopiston porteilla: aikuiset ja nuoret hakijat ja sisäänpäässeet 2000-luvun alun Suomessa”*. Tuore esimerkki on Heiskalan ja muiden (2023) jo aiemmin mainittu tutkimus, jossa tarkasteltiin perhetaustan merkitystä uudelleen hakemiselle sen jälkeen, kun aiemmassa yliopistohaussa on tullut hylätyksi. Tällaisen tutkimuksen toteuttamisessa keskeinen etu on, että rekisterit antavat mahdollisuuden huomioida valikoivuus siinä, ketkä jatkavat hakemista, kuten myös siinä, ketkä eivät koskaan hakeudu korkea-koulutukseen osaamistasostaan ja kyvyistään riippumatta (esim. Heiskala, Erola & Kilpi-Jakonen, 2021).

Mahdollisuus koulutusurien erityisen pitkään seurantaan on välttämätöntä myös, kun tutkitaan jatko-opintoihin valikoitumista (Helin ym., 2019; Helin ym., 2023), aikuiskoulutusta (Kilpi-Jakonen & Sirniö, 2020; Nori & Vanttaja, 2018; Pöyliö & McMullin, 2025) tai koulutuksen tuottoja työmarkkinoilla ja työmarkkinariskejä (esim. Heiskala, Erola & McMullin, 2021; Isopahkala-Bouret & Nori, 2021; Lehti &

Karonen, 2020; Prix, 2013; Sipilä ym., 2011; Sirniö ym., 2016; taloustieteen puolelta esimerkiksi Silliman & Virtanen, 2022). Kaikkien näiden aiheiden tutkiminen luotettavalla tavalla on vaikeaa ja välillä jopa mahdotonta ilman korkeatasoisia rekisteriaineistoja.

Luotettavuus, valikoituminen ja kausaaliset asetelmat

Rekistereiden ylivoimainen vahvuus on, että niiden avulla voidaan tutkia koko Suomessa asuvaa väestöä. Tästä on erityistä hyötyä silloin, kun tutkimuksessa halutaan tutkia melko pienten väestöryhmien kouluttautumista. Otosaineistoissa monien tällaisten ryhmien koko jää niin pieneksi, ettei tilastollisen analyysin tekeminen olisi mahdollista – tilastollisen epävarmuuden indikaattorit, kuten estimaattien luottamusvälit ja virhetermit, jäävät yksikertaisesti liian suuriksi. Kyselyaineistojen vastaamisen virheet ja epävarmuudet vieläpä kasvattavat niiden otoskokovaatimuksia, kun taas rekisteriaineistoissa monet tällaiset virhelähteet puuttuvat tyystin. Voidaankin yleistäen väittää, että tyypilliset virheet rekisteriaineistoissa (kuten virheelliset luokittelut ja kategorisoinnit) pikemminkin pienentävät kuin kasvattavat luottamusvälejä ja virhetermejä. Tämä johtuu siitä, että tyypillisesti tällaiset virheet vähentävät vaihtelua muuttujassa pikemminkin kuin lisäävät sitä.

Mahdollisuus tutkia pieniä ryhmiä on ollut hyödyllistä esimerkiksi maahanmuuttajaryhmien koulutusta vertaillen (ks. esim. Kilpi-Jakonen, 2011, 2012, 2017; taloustieteen puolelta Ansala ym., 2020). Maahanmuuttajia on ollut Suomessa historiallisesti vähän ja monet kansallisuusryhmät ovat niin pieniä, ettei niiden tutkiminen ole yksinään järkevää edes kokonaisaineistoja hyödyntäen. Jopa monien maahanmuuttajien pääryhmien tutkiminen on Suomessa otosaineistoilla vaikeaa ryhmien suhteellisen pienen koon vuoksi, vaikka tieto kerätäänkin useissa kyselytutkimuksissa (esimerkiksi European Social Survey tai Survey of Health, Ageing and Retirement in Europe). Muita esimerkkejä pienten osaryhmien tutkimisesta ovat esimerkiksi ei-ylioppilastaustaiset yliopistoon hakijat (Haltia ym., 2018) sekä aiemmin mainitut tutkimukset sukupuolelle epätyypillisistä koulutusvalinnoista (Prix & Kilpi-Jakonen, 2022; Prix ym., 2024) ja jatkokoulutukseen valikoitumisesta (Helin ym., 2019; Helin ym., 2023). Myös eriarvoisuus eri yliopistoihin ja hyvinkin tarkasti määritellyille koulutusaloille/-ohjelmiin hakemisessa ja valikoitumisessa on ollut alusta lähtien suomalaisen koulutussosiologisen tutkimuksen mielenkiinnon kohteena (esim. Hermunen, 1984; Kalalahti & Nori, 2023; Kivinen ym., 2001; Kivinen ym., 2012; Nori, 2011; taloustieteen puolelta esim. Suhonen, 2014).

Yleisesti kyselytutkimuksiin vastaaminen on vähentynyt dramaattisesti viime vuosikymmeninä (Jabkowski & Cichocki, 2025). Paitsi että tämä rajoittaa kyselyaineistojen kokoa, kärsii luotettavuus myös siitä, että vastaajavalikoituminen samalla vahvistuu. Toisinaan valikoitumista on siinä määrin, että havaintojen yleistäminen vastaajista koko kohdepopulaation edes painottamista hyödyntäen muuttuu kyseenalaiseksi. Erityisen vähän kyselyihin vastaavat matalasti koulutetut nuoret miehet,

toisin sanoen juuri se ryhmä, jonka kouluttautumisen puutteista tarvittaisiin erityisen kipeästi tietoa. Terveysongelmat ovat usein myös syy jättää vastaamatta kyselyihin (tai jopa jäädä alkuperäisten otosten ulkopuolelle), mutta rekisteriaineistojen avulla on voitu myös tutkia, miten lapsen terveysongelmat ovat yhteydessä kouluttautumiseen ja missä määrin ne selittävät koulutuksen ylisukupolvista periytymistä (Mikkonen ym., 2020, 2021).

Valikoituminen on ongelma kaikessa tilastollisessa analyysissä. Koko väestön kattavan rekisteriaineiston kohdalla tutkijalla on kuitenkin joukko keinoja, joilla valikoitumisharhaa voi yrittää huomioida. Tavallisin tapa on koettaa tilastollisen mallintamisen keinoin arvioida, missä määrin kolmansien tekijöiden mukainen valikoituminen selittää kahden muun tekijän havaittua yhteyttä. Esimerkiksi McMullin ja kumppanit (2021) tutkivat lapsena paikkakunnan muuttamisen yhteyttä lasten kouluttautumiseen. Tutkimuksessa näiden välillä havaittiin negatiivinen yhteys. Tämä selittyi osittain valikoitumisella muuttamiseen. Paikkakunta vaihtui lapsuuden aikana yleisimmin vanhempien eron ja työttömyyttä seuranneen uuden työllistymisen seurauksena, joista kummallakin oli aikaisemman tutkimuksen perusteella negatiivinen vaikutus lapsen kouluttautumiseen.

Toisinaan tutkimusasetelmat mahdollistavat vahvempien kausaalisten päätelmien tekemisen syy- ja seuraussuhteista. Koulutuksen tutkimisen kohdalla kuuluisin esimerkki tällaisesta kvasikokeellisesta asetelmasta on Pekkarisen ja kumppanien (2009) tutkimus peruskoulu-uudistuksen vaikutuksesta lasten tuloihin aikuisuudessa. Koska peruskoulu-uudistus toteutettiin portaittain alueellisesti pohjoisesta etelään edeten, pystyivät tutkijat hyödyntämään tätä tutkittavasta ilmiöstä riippumattonta vaihtelua instrumenttina uudistuksen kausaali vaikutuksen identifiointiin. Lahtinen ja muut (2024) käyttivät vastaavaa asetelmaa tutkiakseen, miten peruskoulutuksen laajentaminen on vaikuttanut geneettisen perimän ja koulutuksen väliseen yhteyteen. Lehti, Erola ja Karhula (2019) hyödynsivät kausaali vaikutusten havaitsemisen instrumenttina saman perheen lasten eri ikää vanhemman jäädessä työttömäksi ja Kailaheimo-Lönnqvist ja Erola (2020) sisarusten eri ikää vanhemman kuollessa. Tarkiainen ja kumppanit (2024) tutkivat sisarusten välisiä eroja muun muassa vanhempien taloudellisissa resursseissa, käydyissä kouluissa ja asutuissa naapurustoissa. Lehti, Erola ja Tanskanen (2019) puolestaan käyttivät isovanhempien ja lastenlasten elinkaarien päällekkäisyyden vaihtelevaa pituutta instrumenttina sille, missä määrin lasten oli mahdollista ”altistua” isovanhempien vaikutuksille isovanhempien kausaalisen vaikutuksen identifioimiseksi. Taloustieteen puolella toisen asteen oppilaitosten sisäänpääsyrajoja on käytetty identifioimaan koulutuksen vaikutuksia myöhempään elämään (esimerkiksi Tervonen ym., 2018; Silliman & Virtanen, 2022).

Monien tutkimuskysymysten kohdalla kausaalisen vaikutuksen identifioimista kiinnostavampaa on historiallisten muutosten luotettava havainnointi. Tämä on tärkeää, jos halutaan arvioida niin institutionaalisen ympäristön kuin väestörakenteen muutoksien yhteyttä kouluttautumiseen. Esimerkiksi koulutusjärjestelmän muutokset toteutuvat usein pikemminkin sarjana uudistuksia, jotka kohdistuvat osin tai

kokonaan päällekkäin toistensa kanssa pikemminkin kuin toisistaan varmasti eroteltavina muutoksina. Koulutusrekistereiden pitenevät aikasarjat tarjoavat tällaisiin analyyseihin yhä enemmän mahdollisuuksia. Rekistereiden erinomaisia aikasarjaominaisuuksia hyödynsivät jo Kivinen ja kumppanit (2001) tutkiessaan koulutusekspansion yhteyttä koulutuksen periytymiseen Suomessa 1980- ja 1990-luvuilla, Erola (2009) tarkastellessaan koulutuksen yhteyttä luokkaliikkuvuuden muutokseen Suomessa 1935–1975 syntyneillä kohorteilla (myös Ruggera ym., 2024; tuloliikkuvuuden osalta Sirniö ym., 2017), Karhunen ja Uusitalo (2017) tutkiessaan koulutusmahdollisuuksien tasa-arvon muutosta 1944–1990-syntyneillä kohorteilla sekä Härkönen ja Sirniö (2020) tutkiessaan perhetaustan yhteyttä koulutuksen nivelvaiheiden siirtymiin 1960–1985 syntyneillä.

Empiirinen esimerkki: ylisukupolvinen eriarvoisuus korkeakoulutuksessa

Esimerkkinä rekistereiden käytöstä koulutussosiologisessa tutkimuksessa tarkastelemme seuraavaksi, millaisessa yhteydessä perhetausta on korkeakoulutuksen suorittamiseen erityisesti ylioppilaiden keskuudessa ja missä määrin ylioppilaskirjoituksissa pärjääminen voisi selittää tätä yhteyttä. Jotta esimerkki toimisi suuntaa antavana ohjeistuksena rekistereiden käytöstä, käymme suhteellisen tarkasti läpi, minkälaisia päätöksiä teemme valitessamme tutkittavien joukon ja yhdistäessämme eri aineistoja.

Tutkimuskysymys linkittyy yhtäältä kysymykseen siitä, miten koulutuksen eriarvoisuus on muuttunut yli ajan Suomessa, jota ovat aiemmin tutkineet taloustieteen puolelta Karhunen ja Uusitalo (2017) ja sosiologian puolelta Kivinen ja kollegat (2001, 2007, 2012) sekä Härkönen ja Sirniö (2020). Toisaalta tutkimuskysymys linkittyy myös kysymykseen perhetaustan niin sanotuista primaari- ja sekundaariefekteistä eli siitä missä määrin ylisukupolvinen eriarvoisuus koulutussiirtymissä selittyy nuorten erilaisilla oppimistuloksilla/arvosanoilla (ns. primaariefekti) ja missä määrin erilaisista perheistä tulevien nuorten erilaisilla koulutusvalinnoilla (ns. sekundaariefekti). Tätä on kansainvälisesti tutkittu hyvin paljon kyselyaineistoilla (esim. Jackson, 2013), mikä toimii hyvin silloin, kun koulutussiirtymiä mitataan ajallisesti suhteellisen pian osaamisen/arvosanojen mittaamisen jälkeen. Kuitenkin oikeastaan vain rekisteripohjaisilla aineistoilla voidaan luotettavasti mitata asiaa, jos kiinnostuksen kohteena on koulutustaso aikuisena. Aiempaa suomalaista rekisteritutkimusta aiheesta on liittyen peruskoulun päättötodistuksen keskiarvoon ja siirtymiin toiselle asteelle ja korkeakoulutukseen (Kilpi-Jakonen, 2011; Kilpi-Jakonen ym., 2016; Heiskala, Erola & Kilpi-Jakonen, 2021).

Aineisto ja menetelmät

Otoksen määrittely

Ylioppilaskirjoitusten tulokset ovat olemassa rekistereissä vuodesta 1967 alkaen. Aloitamme tarkastelumme kuitenkin vasta vuoden 1970 ylioppilaskirjoituksista. Tämä on toisaalta siksi, että arvosanajärjestelmä muuttui vuonna 1970, mutta myös siksi, että rekisterissä vuodelle 1969 on huomattavasti vähemmän kirjoittajia kuin muina vuosina, joten rekisterissä on mitä ilmeisimmin puuttuvia tietoja. Koska 1970 ylioppilaksi kirjoittaneet ovat pitkälti aikaisintaan 1952 syntyneitä, aloitamme seurantamme tästä syntymäkohortissa.

Lukumme kirjoitusajankohtana uusimmat käytössämme olevat koulutustiedot ovat vuodelta 2022. On tavanomaista, että rekistereissä olevat tiedot saadaan tutkimuskäyttöön 1,5–2 vuotta rekisteröinnin jälkeen. Suomessa valmistutaan korkeakouluista edelleen verrattain vanhoina, joten tutkintotietoja ei välttämättä kannata tutkia aivan nuorilta aikuisilta. Toisaalta, jotta saamme mukaan myös suhteellisen nuoria kohortteja, emme myöskään halua odottaa valmistumista kovin pitkälle aikuisuuteen. Tässä yhteydessä päätämme tutkia korkeakoulusta valmistumista viimeistään 30-vuotiaana, mikä on muutama vuosi enemmän kuin mediaanivalmistumisikä suomalaisista korkeakouluista. Päätämme siis seurantamme vuonna 1992 syntyneisiin, jotka ovat 30-vuotiaita vuonna 2022.

Meillä on vuositason tiedot korkeimmasta koulutuksesta vuodesta 1987 alkaen. Tästä syystä käytämme vuoden 1985 väestölaskentaa vanhimpien kohorttien tutkintotietoihin. Vuosina 1952–1954 syntyneiltä koulutustieto poimitaan siis hieman vanhempana kuin tasan 30-vuotiaana, samoin kuin vuonna 1956 syntyneiltä (vuoden 1987 tieto eli 31-vuotiaana).

Rajaamme tutkittavien joukon Suomessa syntyneisiin suomalaistaustaisiin (Tilastokeskuksen syntyp2-muuttujan avulla). Tämä tarkoittaa sitä, että vähintään yksi henkilön vanhemmista on Suomessa syntynyt. Teemme tämän erityisesti sen takia, että maahanmuuttajavanhempien koulutustiedot ovat rekistereissä puutteelliset. Suomessa ylioppilaksi kirjoittaneita maahanmuuttotaustaisia henkilöitä on tutkimissamme kohorteissa vielä suhteellisen pieni määrä. Tämä joukko kasvaa vuosi vuodelta, joten heidän tutkimisensa tulee yhä relevantimmaksi ja aineistojen puolesta myös helpommaksi. Rajaamme ulkopuolelle myös henkilöt, joilla ei ole elossa olevaa (tai Suomessa asuvaa) vanhempaa sinä vuonna, kun mittaamme vanhempien koulutustason (ks. alla). Tutkimuksen otoksen ulkopuolelle jää myös Suomen ulkopuolella 30-vuotiaana asuvat ja alle 30-vuotiaana kuolleet.

Selitettävä muuttuja: korkeakoulutuksen suorittaminen

Tutkimme korkeakoulutuksen suorittamista 30-vuotiaaksi mennessä. Tämä määritellään tutkinnon suorittamisena yliopistossa tai ammattikorkeakoulussa. Suomalaisissa rekistereissä koulutustieto on perusasteen jälkeen suoritetuista tutkinnoista. Tilastokeskuksen FOLK perus -aineistossa korkein tutkinto määritellään yhden

numeron tarkkuudella (ututku_aste), kun taas vuoden 1985 väestölaskennassa ja varsinaisissa tutkintorekistereissä (EDUC tutkinto -aineisto) se on saatavilla kuuden numeron tarkkuudella. Käytämme tässä tutkimuksessa tätä tarkempaa kuusinumeroista muuttujaa. Sen avulla luokittelemme pois korkeakoulutettujen joukosta ammatillisen korkea-asteen tutkinnot (pääasiassa erilaisia insinöörejä) sekä muutamat muut tutkinnot, jotka ututku_aste-muuttujassa määrittyvät korkeakoulutettuihin (luokkaan 6) mutta eivät ole varsinaisia korkeakoulututkintoja (ks. <https://koulutusluokitus.fi/>). Käytämme tätä samaa kuusinumeroista muuttujaa ja luokitusta myös vanhempien koulutuksen luokitteluun. Koska nämä ammatillisen korkea-asteen insinöörit ovat suurehko ryhmä erityisesti vanhempien keskuudessa, sillä on absoluuttisten prosenttiosuuksien analyysissä jonkin verran väliä, luokitellaanko heidät korkeakoulutettuihin vai ei. Mallien selityksasteen perusteella heidät kannattaa luokitella ei-korkeakoulutettuihin.

Tässä yhteydessä on myös relevanttia huomauttaa, että yleisemmällä tasolla Tilastokeskuksen hallinnollisia syitä varten tekemät luokitukset eivät aina ole (koulutus)sosiologisesti mielekkäimpiä luokituksia, joten tutkijan on myös syytä pohdita luokitusten sopivuutta omaan tutkimukseensa. Esimerkiksi Tilastokeskuksen käyttämä sosioekonomista asemaa mittaava sose-muuttuja ei ole sosiologiseen tutkimukseen perustuva. Tämän vuoksi sosioekonomisesta tai luokka-asemista kiinnostuneet tutkijat muodostavat aineistoihin paremmin tutkimuskäyttöön sopivat luokituksensa itse.

Selittävät muuttujat: vanhempien koulutus ja ylioppilaskirjoitusten arvosanat

Vanhempien koulutustiedon otamme siltä vuodelta, kun henkilö on 18-vuotias. Vanhimmissa kohorteissa käytämme vuosien 1970, 1975, 1980 ja 1985 väestölaskentoja siten, että tiedot otetaan lähimmältä väestölaskentavuodelta ennen 18. ikävuotta (eli 14–18-vuotiaana). Kuten yllä mainittu, käytämme myös vanhempien koulutuksen luokitteluun kuusinumeroista koulutuskoodia ja muokkaamme kaksiluokkaisen muuttujan, joka erottelee sen perusteella, onko vähintään yksi vanhemmista suorittanut korkeakoulututkinnon. Koska koulutustiedot on vain perusasteen jälkeen suoritetuista tutkinnoista, on aineistoa muokatessa tärkeä varmistua siitä, tarkoitetaako puuttuva tieto sitä, että vanhempi on suorittanut vain perusasteen vai sitä, että vanhempi ei ole aineistossa lainkaan. Tästä syystä käytämme apumuuttujana vanhempien pääasiallista toimintaa (ptoim1): vanhemmat, joilla on tieto pääasiallisesta toiminnasta muttei tietoa tutkinnoista luokitellaan ei-korkeakoulutetuiksi. Vuoden 1975 osalta joudumme vielä turvautumaan tietoon vanhempien syntymävuodesta, koska tietoa pääasiallisesta toiminnasta ei ole (ja toisaalta emme voi käyttää tätä muuttujaa vuodelle 1970, koska tässä väestölaskennassa on henkilöitä, joilla on puuttuva syntymävuosi, vaikka he ovat aineistossa muuten).

Jälleen huomautamme tästä yleisestä asiasta rekisteriaineistojen suhteen: joissain muuttujissa puuttuva tieto on oikeasti relevantti. Koulutuksen lisäksi tämä on esimerkiksi työttömyys- ja työllisyyskuukausissa, joissa puuttuva arvo voi vastata tutkimuksellisesti nollaa.

Jotta saamme linkitettyä vanhempien koulutustiedot lapsiin, meidän täytyy ensin linkittää vanhempien pseudo-henkilönumerot lasten pseudo-henkilönumeroihin (FOLK lapsi - vanhemmat -aineisto). Kohorttirajauksemme alkaen 1952-syntyneistä on myös siinä mielessä perusteltu, että lasten ja vanhempien linkitykset ovat saatavilla tästä aineistosta kohtuullisen kattavasti vasta 1950-luvulla tai myöhemmin syntyneille. Koulutuksen ylisukupolvista periytyvyyttä ei voi suomalaisilla rekisteriaineistoilla siis ylipäänsä tarkastella kovinkaan paljoa vanhemmilla kohorteilla koko väestön tasolla, koska linkitykset vanhempien koulutukseen käyvät haastavammiksi. Karhunen ja Uusitalo (2017) kuitenkin aloittavat seurantansa 1944 syntyneistä ja Kivinen ja kumppanit (2007, 2012) analysoivat 1946 syntyneitä.

Kuten yllä mainittu, analysoimme ylioppilaskirjoituksia vuodesta 1970 alkaen. Koska ylioppilaskokeita voi suorittaa eri-ikäisenä (myös varsinaiseksi ylioppilaaksi valmistumisen jälkeen), päätämme tässä tutkimuksessa rajata kokeet korkeintaan 25-vuotiaana suoritettuihin (tässä yhteydessä vuoteen 2017 asti). Käsittelemme siis ylioppilastutkintorekistereitä siten, että linkitämme kokeen suorittamisvuodelta (tai lähimmältä aiemmalta väestölaskentavuodelta) tiedon syntymävuodesta. Koska tämä ei vielä anna kaikille syntymävuotta, teemme sen vielä toiseen kertaan suorittamisvuotta edeltävältä vuodelta (tai mahdolliselta aiemmalta väestölaskentavuodelta). Tästä huolimatta on pieni määrä kokeiden suorittaneilta, joille emme pysty luotettavalla tavalla määrittelemään kokeen suorittamisikää.

Käytämme tässä tutkimuksessa pelkästään äidinkielen arvosanaa; tämä voi olla suomen, ruotsin, suomi toisena kielenä, ruotsi toisena kielenä, inarinsaame, pohjoissaame tai koltansaame. Äidinkieli on ainut koe, joka on pysynyt pakollisena koko tarkastelemamme ajanjakson ajan. Henkilöille, joille löytyy useampi hyväksytty arvosana, käytämme korkeinta arvosanaa (kuitenkin alle 25-vuotiaana suoritetusta kokeesta). Ylioppilaskokeiden arvosanat harmonisoimme siten, että vuodesta 1996 käytössä ollut eximia yhdistetään laudaturiin. Arvosana on siis mallisammme viisi-luokkaisena muuttujana.

Menetelmät

Tässä esimerkissä käyttämämme menetelmät ovat suhteellisen yksinkertaisia. Esitämme ensiksi kuvioiden avulla koulutustason kehityksen yli ajan (syntymävuoteen perustuen). Tämän jälkeen analysoimme saman vanhempien koulutustason perusteella, jotta pääsemme käsiksi ylisukupolviseen eriarvoisuuteen. Analysoimme sitten vanhempien koulutustason mukaista eroa korkeakoulutuksen suorittamisen todennäköisyydessä äidinkielen ylioppilaskirjoitusten suorittaneiden keskuudessa; ensiksi kokonaisuudessaan (syntymävuoden mukaan) ja sitten vakioimalla äidinkielen ylioppilaskirjoituksen arvosanat. Esitämme myös nämä tulokset kuviona.

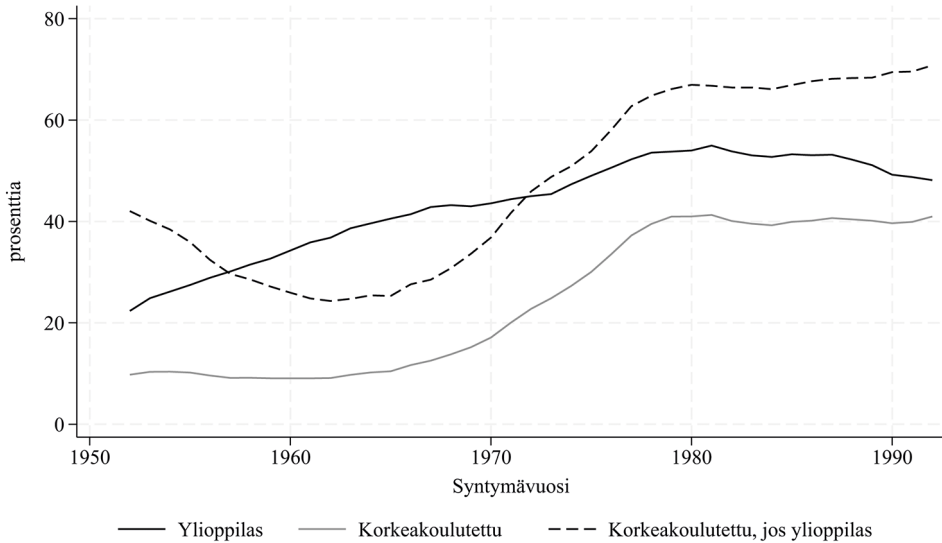
Tarkemmin ottaen mallinamme korkeakoulutuksen suorittamista lineaarisella todennäköisyydshallilla, jossa ensimmäisessä vaiheessa selittävinä tekijöinä ovat vanhempien koulutustaso, henkilön oma syntymävuosi ja näiden välinen interaktio. Toisessa vaiheessa lisäämme malliin arvosanat. Kuviota varten tuotamme näistä malleista ennustetut todennäköisyydet, joista esitämme koulutusryhmien välisen

erotuksen. Koska korkeakoulutuksen suorittaminen on binäärinen muuttuja, mallinnuksen voisi tehdä myös logistisella regressiomallilla. Analyysimme kuitenkin kuuluu mallien välinen vertailu ja interaktio, joten lineaarinen todennäköisyysmalli on suoraviivaisempi tapa analyysillemme.

Käyttämämme luottamusvälit on laskettu samaan tapaan kuin ne laskettaisiin otosaineistoissa. Nämä luottamusvälit ovat kokonaisaineistoa käytettäessä konservatiivisempia kuin niiden tarvitsisi olla, koska aineiston koko on suuri, eikä rekisteriaineistossa ole kaikkia samoja virhevarianssin lähteitä kuin kyselyaineistoissa. Ei ole konsensusta siitä, miten estimaattien epävarmuutta pitäisi tarkkaan ottaen arvioida rekisteriaineistoja käytettäessä. On kuitenkin selvää, että tutkijat haluavat hyödyntää jonkinlaisia epävarmuuden estimaatteja. Liian konservatiivisten epävarmuuden mittareiden käyttäminen on useimpien tutkijoiden mielestä parempi vaihtoehto kuin se, ettei niitä raportoitaisi lainkaan.

Tulokset

Aloitamme siis puhtaasti kuvailevilla tuloksilla kohorteista kokonaisuudessaan. Kuviossa 1 esitetään kolme koulutustulosta: ylioppilaaksi viimeistään 25-vuotiaana kirjoittaneiden osuus (joka siis mitataan äidinkielen kokeen hyväksytysti suorittamisena), korkeakoulututkinnon viimeistään 30-vuotiaana suorittaneiden osuus ja korkeakoulututkinnon suorittaneiden osuus ylioppilaista. Koulutuksen ekspansio näkyy molemmilla koulutusasteilla. 1950-luvun alkupuolella syntyneistä vain hie-
man yli viidesosa kirjoitti ylioppilaaksi ja vain noin kymmenesosa suoritti korkea-
koulututkinnon (30-vuotiaaksi mennessä). Koulutusekspansio 1950- ja 1960-luvulla
syntyneiden joukossa on erityisen vahvaa ylioppilaaksi kirjoittamisessa – ja tämä
jatkui vielä 1970-luvulla syntyneiden joukossa. Korkeakouluekspansio taas alkoi
1960-luvun loppupuolella syntyneillä ja jatkui erityisen vahvana 1970-luvulla syn-
tyneiden keskuudessa. 1980-luvun alkupuolella syntyneiden jälkeen ylioppilaiden
osuus on lähtenyt laskuun, ja korkeakoulututkinnon suorittaneiden osuus on pysy-
nyt tasaisena; molempien tutkintojen osalta 1981-syntyneet ovat korkeimmin kou-
luttautunut kohortti (55 % ylioppilaita ja 41 % korkeakoulututkinnon suorittaneita).
Tämän analyysin nuorimmassa kohortissa 48 prosenttia on suorittanut ylioppilas-
tutkinnon.



Kuvio 1. Ylioppilaat ja korkeakoulutetut syntymävuoden perusteella.

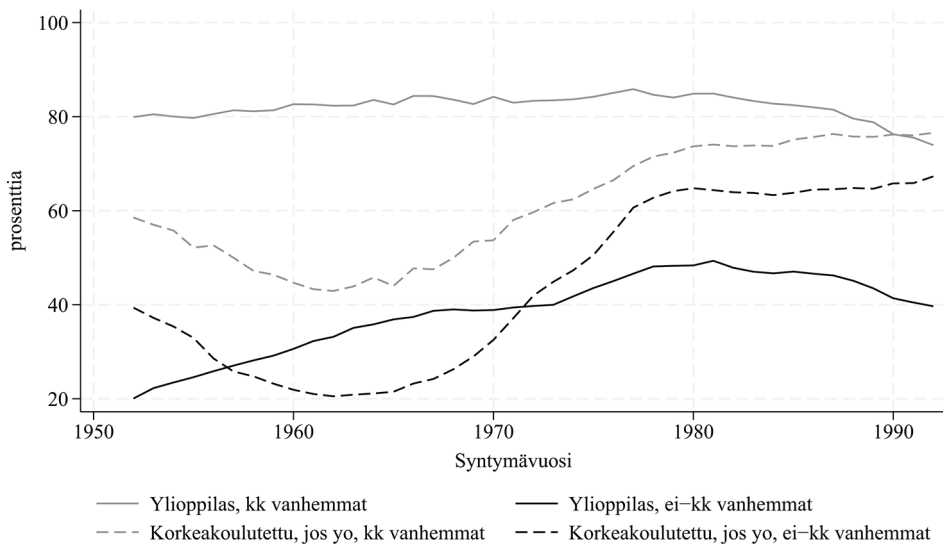
Nämä trendit näkyvät mielenkiintoisena kehityksenä korkeakoulutuksen suorittamisen todennäköisyydessä ylioppilaiden keskuudessa (kuvio 1). 1952 syntyneistä ylioppilaista 42 prosenttia suoritti korkeakoulututkinnon 30-vuotiaaksi mennessä. Tämä osuus kuitenkin laski ripeään tahtiin, kun ylioppilaiden osuus kasvoi mutta korkeakoulutettujen ei. Alimmillaan osuus oli 24 prosenttia 1962-syntyneiden keskuudessa, josta sitten korkeakouluekspansion myötä tämä lähti kasvuun. 1980-luvun alkupuolella syntyneistä ylioppilaista 66 prosenttia suoritti korkeakoulututkinnon. Ylioppilaiden osuuden lähdettyä laskuun, tämä osuus on hieman vielä kasvanut ja 1992 syntyneistä ylioppilaista 71 prosenttia oli suorittanut korkeakoulututkinnon vuoteen 2022 mennessä.

Sivuhuomiona tässä kohden voimme todeta, että vain neljä prosenttia niistä henkilöistä, jotka eivät olleet suorittaneet ylioppilastutkintoa suoritti kuitenkin korkeakoulututkinnon 30-vuotiaaksi mennessä. Tässä on kuitenkin suuri ajallinen muutos: 1950- ja 1960-luvuilla syntyneillä tämä osuus oli alle prosentin mutta 1970-luvun lopulla syntyneistä lähtien noin 10 prosenttia. Kuitenkin jos väestön koulutustasoa halutaan nostaa yli 50 prosenttiin, tulisi ammatillisen tutkinnon suorittaneiden korkeakouluopintovalmiuksia ja -mahdollisuuksia parantaa huomattavan paljon – mitä nykyiset koulutusleikkaukset eivät tue eikä myöskään ylioppilastutkintoon yhä vahvemmin nojaava korkeakoulutuksen valintauudistus (Karhunen ym., 2022) – tai siirtää toisen asteen koulutuksen painopistettä selkeästi enemmän lukioon päin.

Siirrymme seuraavaksi näiden trendien tarkasteluun vanhempien koulutustason mukaan. Kuviossa 2 esitellään ajalliset trendit ylioppilaiden osuudessa sekä korkeakoulututkinnon suorittaneiden osuudessa ylioppilaiden joukosta sen perusteella, onko vähintään yksi vanhempi korkeakoulutettu vai ei. Ylioppilastutkinnon

suorittaneiden osuus niiden keskuudessa, joilla on vähintään yksi korkeakoulutettu vanhempi, on pysynyt yllättävänkin tasaisena koko tarkastellun ajanjakson ajan. Jo 1950-luvun alussa syntyneillä neljä viidestä kirjoitti ylioppilaaksi ja tässä tapahtui vain hyvin pientä kasvua 1960-luvun loppupuolella ja 1970-luvulla syntyneiden keskuudessa. Hieman dramaattisempi muutos on tapahtunut tämän jälkeen, kun osuus on lähtenyt laskuun: ”huippuvuodesta”, jolloin 86 prosenttia kirjoitti ylioppilaaksi (1977 syntyneet), on tultu alas 74 prosenttiin (1992 syntyneet). Lukioekspansio kosketti siis erityisesti ei-korkeakoulutettujen vanhempien lapsia, joilla osuus kasvoi viidesosasta lähes puoleen, ”huippuvuoden” ollessa 1981 syntyneillä (49 %). Tämän jälkeen laskua on ollut myös tässä ryhmässä ja nuorimmassa kohortissa vain 40 prosenttia kirjoitti ylioppilaaksi.

Lukioekspansioon taustalla on myös tämän ryhmän vanhempien koulutustason kasvu: ryhmä kattaa niin peruskoulutetut kuin opistotason koulutuksen saaneet vanhemmat, joten tämän ryhmän sisällä on ollut tänä aikana myös vahvaa keskimääräisen koulutustason kasvua. Kokonaisuudessaan ryhmien välinen ero ylioppilaaksi kirjoittaneiden osuudessa on kaventunut lähes koko tarkastellun ajanjakson ajan: 1952 syntyneillä ryhmien välinen ero oli 60 prosenttiyksikköä, 1970-luvun loppupuolella ja 1980-luvun alkupuolella syntyneiden keskuudessa se oli 36–37 prosenttiyksikköä, minkä jälkeen se on pysynyt 34–35 prosenttiyksikössä.

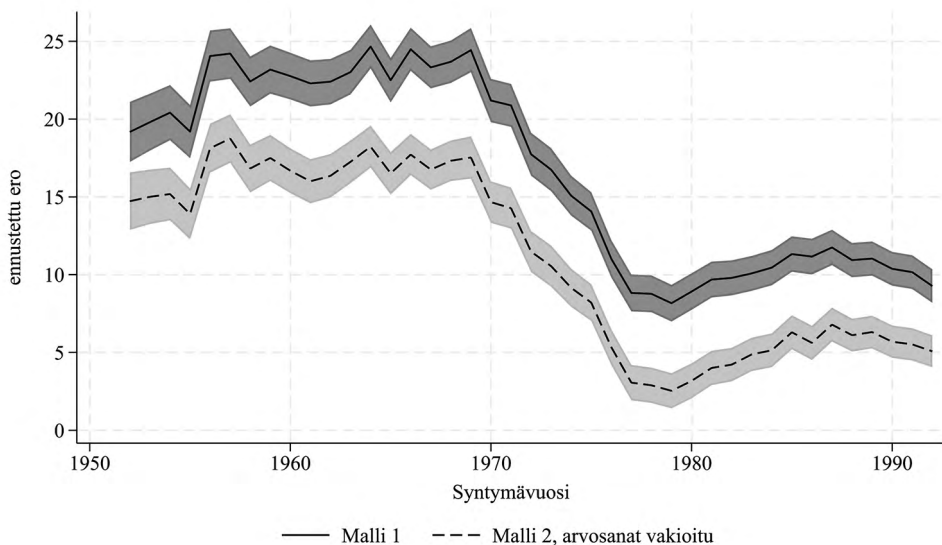


Kuvio 2. Koulutus syntymävuoden ja vanhempien koulutustason perusteella.

Tarkasteltaessa korkeakoulututkinnon suorittaneita ylioppilaiden keskuudessa, ajalliset trendit vanhempien koulutustason mukaan näyttävät suhteellisen samankaltaisilta (kuvio 2). Molemmissa ryhmissä näkyy se, että lukioekspansio tapahtui ajallisesti huomattavasti ennen korkeakouluekspansiota: korkeakoulutuksen suorittaneiden osuus korkeakoulutettujen vanhempien ylioppilaslapsista oli 58 prosenttia

1952 syntyneiden keskuudessa ja laski 43 prosenttiin 1962 syntyneiden keskuudessa, ei-korkeakoulutetuilla vanhemmilla vastaavat luvut olivat 40 prosenttia ja 21 prosenttia. Korkeakouluekspansio antoi kuitenkin matalammin koulutettujen vanhempien lapsille mahdollisuuden kiiä kiinni eroa ryhmien välillä. Tultaessa 1980-luvulla syntyneisiin, korkeakoulutettujen vanhempien ylioppilaslapsista 74 prosenttia suoritti korkeakoulututkinnon, mikä on noussut pikkuhiljaa vielä 77 prosenttiin, ja ei-korkeakoulutettujen vanhempien ryhmässä vastaavat luvut ovat 64 prosenttia ja 67 prosenttia.

Tämä vanhempien koulutustasojen välinen erotus on mallinnettuna esitetty kuviossa 3. 1952 syntyneiden kohdalla erotus oli 19 prosenttiyksikköä, mistä se lukioekspansion myötä kasvoi itse asiassa noin 22–24 prosenttiyksikköön 1956–1969 syntyneiden keskuudessa. Vasta tämän jälkeen erot alkoivat kaventua ja olivat alimmillaan 8 prosenttiyksikköä 1979 syntyneillä. 1980-luvulla ja 1990-luvun alussa syntyneillä erot ovat liikkuneet 9–11 prosenttiyksikön välillä. Tässä kohden voi myös selkeästi huomata, että koulutuseriarvoisuutta on huomattavasti enemmän ylioppilaaksi kirjoittamisessa kuin siirtymässä ylioppilaasta korkeakoulututkinnon suorittamiseen.



Kuvio 3. Mallinnettu ero korkeakoulutuksen suorittamisen todennäköisyydessä ylioppilaiden keskuudessa vanhempien koulutustason perusteella syntymävuosittain. Toisessa mallissa vakioitu äidinkielen ylioppilaskirjoitusten arvosanat. 95 prosentin luottamusvälit estimaattien ympärillä.

Mallissa 2 huomioidaan äidinkielen ylioppilaskirjoitusten arvosanat korkeakoulututkinnon suorittamista ennustavana tekijänä (kuvio 3). Kaikissa kohorteissa arvosanojen huomioiminen vähentää koulutusryhmien välistä eroa noin 4–7 prosenttiyksikköä. Suurimmillaan selitetty ero on prosenttiyksikköinä mitattuna 1960-luvun

loppupuolella ja aivan 1970-luvun alussa syntyneiden keskuudessa ja pienimmillään 1990-luvun alkupuolella syntyneiden keskuudessa – ja toisaalta myös 1950-luvun alkupuolella syntyneiden keskuudessa. Jos arvosanojen selitysosuus suhteutetaan eron suuruuteen, ajallista muutosta on ollut huomattavasti enemmän. 1950-luvulla syntyneiden keskuudessa arvosanojen huomioiminen selitti eroista noin neljänneksen, 1960-luvulla syntyneiden keskuudessa tämä alkoi nousta ja saavutti kolmasosan 1970-luvun alkupuolella syntyneiden keskuudessa, nousten jo kahteen kolmasosaan 1970-luvun loppupuolella syntyneiden keskuudessa. Tästä se tasaantui noin puoleen 1980-luvun alun ja puolivälin tienoilla ja on pysynyt noin 45 prosentissa 1980-luvun loppupuolella ja 1990-luvun alkupuolella syntyneiden keskuudessa.

Kokonaisuudessa tuloksemme osoittavat siis, että koulutuseriarvoisuus Suomessa (tutkimillamme koulutusasteilla) väheni siirryttäessä 1960-luvulla syntyneistä 1970-luvulla syntyneisiin. Kaikilla tutkimillamme tavoilla koulutuseriarvoisuus oli matalimmillaan 1979–1981 syntyneiden keskuudessa: näissä kohorteissa oli pienin ero ylioppilastutkinnon suorittamisessa ja siitä korkeakoulututkintoon siirtymisessä, minkä lisäksi ylioppilaskokeen arvosanat selittivät suhteessa suurimman osan kouluttautumiseroista. Niin kutsuttu perhetaustan sekundaariefekti oli siis näissä kohorteissa sekä absoluuttisesti että suhteellisesti pienin. Hieman mittarista riippuen erot ovat tämän jälkeen joko kasvaneet tai pysyneet kohtuullisen tasaisina ja pieniä kaventumisen merkkejä on aivan nuorimmissa tutkimissamme kohorteissa (1992 syntyneet).

Kuten Karhunen ja Uusitalo (2017) huomauttavat, yksi syy havaitsemallemme koulutuseriarvoisuuden heikentymiselle on vanhempien koulutuksesta käyttämämme kaksiluokkainen muuttuja. Matalasti koulutettujen vanhempien koulutusjakauma on muuttunut huomattavasti ajan myötä, mikä tarkoittaa sitä, että ns. matalasti koulutetut vanhemmat ovat nuorimmissa kohorteissa huomattavasti paremmin koulutettuja (lähinnä toisen asteen ja opistotason koulutuksen saaneita) kuin vanhimmissa kohorteissa (pitkälti kansakoulun suorittaneita). Tekemämme lisäanalyysit vanhempien koulutuksesta kolmiluokkaisena muuttujana osoittavat samankaltaisen eriarvoisuuden laskun 1970-luvulla syntyneiden keskuudessa, mutta toisaalta vahvemman kasvun 1980-luvulla syntyneissä.

Aiemmat tutkimukset ovat siis keskittyneet ylisukupolvisen eriarvoisuuteen koko väestön keskuudessa eikä pelkästään lukion suorittaneiden keskuudessa. Karhunen ja Uusitalo (2017) analysoivat suurin piirtein samoja kohortteja (1944–1990) ja yliopisto-opintojen aloittamista 24-vuotiaaksi mennessä huomioiden vanhempien koulutuksen neliluokkaisena mittarina. Heidän tuloksensa osoittivat koulutuseriarvoisuuden pysyneen suhteellisen tasaisena 1950-luvulla syntyneiden keskuudessa ja lähteneen pieneen nousuun 1960-luvun loppupuolelta lähtien. Vanhempien tuloilla mitattuna ylisukupolvinen eriarvoisuus oli säilynyt hyvinkin tasaisena. Härkösen ja Sirniön (2020) tutkimuksessa mitattiin myös koulutusta 30-vuotiaana hieman suppeammalla kohorttirajauksella (1960–1985) mutta huomioiden sekä vanhempien että lasten koulutustaso viisiluokkaisina (koulutusvuosiksi muunnettuina). Heidän tuloksensa osoittivat suhteellisen tasaista koulutuseriarvoisuuden kasvua koko

ajanjakson yli. Yhteistä meidän tuloksillemme on havainto siitä, että koulutuseriarvoisuuden kannalta lukio-opinnot/ylioppilastutkinto ovat vahva pullonkaula.

Rekisteripohjaisen koulutussosiologisen tutkimuksen tulevaisuus

Olemme esimerkeissämme keskittyneet erityisesti Tilastokeskuksen kautta saataviin rekistereihin, jotka lähtökohtaisesti kattavat koko väestön, vaikkakin huomattava osa mainitsemistamme tutkimuksista on toteutettu otosaineistoilla, jotka Tilastokeskus on rekistereistä tuottanut. Ennen etäkäyttöympäristön (FIONA) käyttöönottoa, otosaineistojen käyttö oli ainoa tapa tehdä rekisteripohjaista tutkimusta. Jonkin verran koulutussosiologista tutkimusta on myös tehty linkittämällä paikallisempia rekistereitä muihin rekistereihin. Tällaisia ovat esimerkiksi kouluista saatujen kurssivalintojen ja -arvosanojen linkittäminen ylioppilastutkintotietoihin (Kupiainen ym., 2018; Kupiainen ym., 2023; Ouakrim-Soivio ym., 2017) tai avoimen yliopiston opiskelijoiden tietojen linkittäminen erilaisiin Tilastokeskuksen rekistereihin (Jauhiainen ym., 2007). Koulutussosiologisiin tutkimuksiin soveltuvien rekisteriaineistojen kehitys on viime vuosina edennyt vahvasti, kun esimerkiksi KOSKI-rekisteri, joka kattaa esi- ja perusopetukseen osallistumisen vuodesta 2020 alkaen, on tullut tutkijoiden käyttöön.

Kaikkea tutkimuksessa tarvittavaa tietoa ei ole mahdollista saada rekistereistä, minkä vuoksi kyselyiden tekeminen on välttämätöntä myös tulevaisuudessa. Kun vastaaminen varsinkin laajoihin kyselytutkimuksiin kuitenkin vähenee, nousevat toisiinsa linkitetyt kysely- ja rekisteriaineistojen yhdistelmät entistä tärkeämmiksi tietolähteiksi. Kun kyselyyn vastaamisen kynnys on korkea, kannattaa kaikki rekistereistä saatava tarvittava tieto vastaajista kerätä niiden kautta. Tämänkaltaisia tutkimuksia on alkanut viimeisen vuosikymmenen aikana ilmestyä yhä useammin. Esimerkkeinä tästä ovat Hotulaisen ja kumppanien (2016) tutkimus, jossa oppilaiden oppimista ja hyvinvointia tutkittiin heidän ollessaan seitsemännellä ja yhdeksännellä luokalla, minkä jälkeen seuranta jatkettiin rekisteritiedoilla toisen asteen valinnoista, sekä Pulkkinen ja Rautopuron (2022) tutkimus, jossa vuoden 2015 ja 2018 PISA-aineistoja linkitettiin yhteisvalintarekistereihin. Myös vanhempia kyselyaineistoja on linkitetty rekisteritietoihin esimerkiksi Ojalan ja muiden (2021) tutkimuksessa, jossa tutkittiin nuorten ajankäyttöä ja sosiaalisia siteitä vuoden 1979 ajankäyttötutkimuksella ja sen yhteyttä kouluttautumiseen 40 vuoden ikään mennessä. Otospohjaisesti kerättyjä geenitietoja on myös linkitetty laajasti rekisteriaineistoihin ja tutkittu perimän roolia ylisukupolvisessa koulutuseriarvoisuudessa (Lahtinen ym., 2023) ja peruskoulu-uudistuksen vaikutusta perimän vaikutukseen, kuten yllä jo mainittu (Lahtinen ym., 2024).

Suomi on muiden pohjoismaiden tapaan rekisteritutkimuksen edelläkävijämaita. Rekistereitä on meillä viime vuosina hyödynnetty runsaasti ja usein innovatiivisella tavalla koulutussosiologisessa tutkimuksessa. Rekistereiden edut on kuitenkin huomattu myös muualla, ja pääosassa Euroopan maista niitä voikin jo käyttää tut-

kimukseen. Tätä lukua kirjoitettaessa parhaita edellytyksiä rekisteritutkimukseen ei enää löydykään yhdestäkään pohjoismaista, vaan monelle ehkä yllätyksenä Hollannista, jossa aineistojen käytettävyyteen on panostettu erityisen vahvasti viime vuosina. Rekistereiden tuoma kilpailuetu suomalaiselle tutkimukselle ei enää lähivuosina ole itsestäänselvyys, vaan korkean tieteellisen tason säilyttäminen edellyttää myös tiedepoliittista tukea. Samaan aikaan tarve rekistereiden tarjoamalle informaatiolle kasvaa tutkijayhteisön ulkopuolella tietopohjaisen päätöksenteon työkaluna (ks. Kauko ym., 2024). Muun muassa näistä syistä rekisteritutkimuksen infrastruktuurin kehittämishanke FIRE on nostettu kansalliselle tutkimusinfrastruktuurien tiekartalle. Myös koulutustarpeisiin on herätty: ensimmäisiä rekisteriaineistojen käyttöön keskittyviä kursseja on jo järjestetty muun muassa Helsingin ja Turun yliopistoissa. Opetustarjonnan odotetaan laajenevat huomattavasti lähivuosina. Haasteista huolimatta koulutussosiologisen rekisteritutkimuksen tulevaisuus näyttääkin varsin valoisalta.

Lähteet

- Andersen, P. L., & Hansen, M. N. (2012). Class and cultural capital—The case of class inequality in educational performance. *European Sociological Review*, 28(5), 607–621. <https://doi.org/10.1093/esr/jcro29>
- Ansala, L., Hämmäläinen, U., & Sarvimäki, M. (2020). Age at arrival, parents and neighborhoods: understanding the educational attainment of immigrants' children. *Journal of Economic Geography*, 20(2), 459–480. <https://doi.org/10.1093/jeg/lbz017>
- Bingley, P., & Martinello, A. (2017). Measurement error in income and schooling and the bias of linear estimators. *Journal of Labor Economics*, 35(4), 1117–1148. <https://doi.org/10.1086/692539>
- Blaabæk, E. H. (2023). Stratification in parents' selection of developmentally appropriate books for children: register-based evidence from Danish public libraries. *European Societies*, 25(1), 37–65. <https://doi.org/10.1080/14616696.2022.2087100>
- Erola, J. (2009). Social mobility and education of Finnish cohorts born 1936—75: Succeeding while failing in equality of opportunity? *Acta Sociologica*, 52(4), 307–327. <https://doi.org/10.1177/0001699309348701>
- Erola, J., Kilpi-Jakonen, E., Prix, I., & Lehti, H. (2018). Resource compensation from the extended family: Grandparents, aunts, and uncles in Finland and the United States. *European Sociological Review*, 34(4), 348–364. <https://doi.org/10.1093/esr/jcy021>
- Golsteyn, B. H., & Hirsch, S. (2019). Are estimates of intergenerational mobility biased by non-response? Evidence from the Netherlands. *Social Choice and Welfare*, 52(1), 29–63. <https://doi.org/10.1007/s00355-018-1138-0>
- Grätz, M., Barclay, K. J., Wiborg, Ø. N., Lyngstad, T. H., Karhula, A., Erola, J., Präg, P., Laidley, T., & Conley, D. (2021). Sibling similarity in education across and within societies. *Demography*, 58(3), 1011–1037. <https://doi.org/10.1215/00703370-9164021>
- Haltia, N., Jauhiainen, A., & Isopahkala-Bouret, U. (2018). Epätavallista reittiä yliopistoon: Ilman ylioppilastutkintoa yliopistoon hakeneiden taustat ja sisäänpääsy koulutukseen. Teoksessa R. Rinne, N. Haltia, S. Lempinen & T. Kaunisto (toim.), *Eriarvoistuva maailma – tasa-arvoistava koulu?* (ss. 223–244). Suomen kasvatustieteellinen seura.
- Helin, J., Koerselman, K., Nokkala, T., Tohmo, T., & Viinikainen, J. (2019). Equal access to the top? Measuring selection into Finnish academia. *Social Inclusion*, 7(1), 90–100. <https://doi.org/10.17645/si.v7i1.1620>
- Helin, J., Jokinen, J., Koerselman, K., Nokkala, T., & Räikkönen, E. (2023). It runs in the family? Using sibling similarities to uncover the hidden influence of family background in doctoral education and academic careers. *Higher Education*, 86(1), 1–20. <https://doi.org/10.1007/s10734-022-00897-7>
- Heiskala, L., Erola, J., & Kilpi-Jakonen, E. (2021). Compensatory or multiplicative advantage? Parental resources, school achievement and access to higher education in Finland. *European Sociological Review*, 37(2), 171–185. <https://doi.org/10.1093/esr/jcaa046>

- Heiskala, L., Erola, J., & McMullin, P. (2021). Formal differentiation at upper secondary education in Finland: Subject-level choices and stratified pathways to socio-economic status and unemployment. *Longitudinal and Life Course Studies*, 12(3), 323–343. <https://doi.org/10.1332/175795921X16137561576439>
- Heiskala, L., Kilpi-Jakonen, E., Sirniö, O., & Erola, J. (2023). Persistent university intentions: Social origin differences in stopping applying to university after educational rejection(s). *Research in Social Stratification and Mobility*, 85, 100801. <https://doi.org/10.1016/j.rssm.2023.100801>
- Hermunen, H. (1984). *Oppilaiden sosiaalinen tausta 1980*. Tilastokeskus.
- Hiilamo, H., Merikukka, M., & Haataja, A. (2018). Long-term educational outcomes of child care arrangements in Finland. *Sage Open*, 8(2). <https://doi.org/10.1177/2158244018774823>
- Hotulainen, R., Rimpelä, A., Hautamäki, J., Karvonen, S., Kinnunen, J. M., Kupiainen, S., Lindfors, P., Minkkinen, J., Pere, L., Thuneberg, H., Vainikainen, M. P., & Wallenius, T. (2016). *Osaaminen ja hyvinvointi yläkoulusta toiselle asteelle: Tutkimus metropolialueen nuorista*. Helsingin yliopisto. <http://hdl.handle.net/10138/301729>
- Hyvärinen, S., & Erola, J. (2011). Perhetaustan vaikutus toisen polven maahanmuuttajien kouluttautumiseen Suomessa. Mahdollisuuksien tasa-arvo stressitestissä. *Yhteiskuntapolitiikka*, 76(6), 644–657.
- Härkönen, J., & Sirniö, O. (2020). Educational transitions and educational inequality: A multiple pathways sequential logit model analysis of Finnish birth cohorts 1960–1985. *European Sociological Review*, 36(5), 700–719. <https://doi.org/10.1093/esr/jcaa019>
- Isoaho, H., Kivinen, O., & Rinne, R. (1990). *Nuorten koulutus ja kotitausta*. Tilastokeskus.
- Isopahkala-Bouret, U., & Nori, H. (2021). On a level playing field? The effect of parents' education on master's degree graduates' labour market success in Finland. *Advances in Applied Sociology*, 11(12), 626–643. <https://doi.org/10.4236/aasoci.2021.1112053>
- Jabkowski, P., & Cichocki, P. (2025). Survey response rates in European comparative surveys: a 20-year decline irrespective of sampling frames or survey modes. *Quality & Quantity*, 59(Suppl 1), 635–655. <https://doi.org/10.1007/s11135-024-01993-9>
- Jackson, M. (2013). Introduction: how is inequality of educational opportunity generated? The case for primary and secondary effects. Teoksessa M. Jackson (toim.), *Determined to succeed?: Performance versus choice in educational attainment* (ss. 1–33). Stanford University Press.
- Jauhiainen, A., Nori, H., & Alho-Malmelin, M. (2007). Various portraits of Finnish open university students. *Scandinavian Journal of Educational Research*, 51(1), 23–39. <https://doi.org/10.1080/00313830601079017>
- Kailaheimo-Lönnqvist, S., & Erola, J. (2020). Child's age at parental death and university education. *European Societies*, 22(4), 433–455. <https://doi.org/10.1080/14616696.2020.1719179>
- Kailaheimo-Lönnqvist, S., Jalovaara, M., & Myrskylä, M. (2025). Parental separation and children's education—Changes over time? *European Journal of Population*, 41(1), 5. <https://doi.org/10.1007/s10680-024-09721-7>

- Kailaheimo-Lönnqvist, S., & Kotimäki, S. (2020). Cause of parental death and child's health and education: The role of parental resources. *SSM-Population Health*, 11, 100632. <https://doi.org/10.1016/j.ssmph.2020.100632>
- Kalalahti, M., & Nori, H. (2023). Lääketieteen opintoihin valikoituminen ja lääkärin ammatin periytyminen. Teoksessa M. Kalalahti, H. Silvennoinen, J. Varjo & M. Vilkmann (toim.), *Koulutus ja yhteiskuntaluokka* (ss. 29–49). Kasvatussosiologian vuosikirja 4. Tampere University Press. <https://doi.org/10.61201/tup.865.c1245>
- Kallio, J. M., Kauppinen, T. M., & Erola, J. (2016). Cumulative socio-economic disadvantage and secondary education in Finland. *European Sociological Review*, 32(5), 649–661. <https://doi.org/10.1093/esr/jcw021>
- Karhula, A., Erola, J., & Kilpi-Jakonen, E. (2017). Home sweet home? Long-term educational outcomes of childcare arrangements in Finland. Teoksessa H.-P. Blossfeld, N. Kulic, J. Skopek & M. Triventi (toim.), *Childcare, early education and social inequality – A cross-national perspective* (ss. 268–283). Edward Elgar Publishing.
- Karhunen, H., Pekkarinen, T., Suhonen, T., & Virkola, T. (2022). *Opiskelijavalintauudistuksen seurantatutkimuksen loppuraportti*. VATT Muistiot 67. Valtion taloudellinen tutkimuskeskus.
- Karhunen, H., & Uusitalo, R. (2017). 50 vuotta koulutusmahdollisuuksien tasa-arvoa. *Yhteiskuntapolitiikka*, 82(3), 296–303.
- Kauko, J., Penttinen, J., Rötä, M., & Voutilainen, T. (2024). *Sujuva kasvatus ja koulutustietojen hyödyntäminen. Selvitysryhmän loppuraportti*. Opetus ja kulttuuriministeriö. <https://urn.fi/URN:ISBN:978-952-263-926-4>
- Kauppinen, T. M. (2004). *Asuinalueen ja perhetaustan vaikutukset helsinkiläisnuorten keskiasteen tutkintojen suorittamiseen* [Väitöskirja, Helsingin yliopisto]. Helda. <http://hdl.handle.net/10138/12027>
- Kauppinen, T. M. (2007). Neighborhood effects in a European city: Secondary education of young people in Helsinki. *Social Science Research*, 36(1), 421–444. <https://doi.org/10.1016/j.ssresearch.2006.04.003>
- Kauppinen, T. M. (2008). Schools as mediators of neighbourhood effects on choice between vocational and academic tracks of secondary education in Helsinki. *European Sociological Review*, 24(3), 379–391. <https://doi.org/10.1093/esr/jcn016>
- Kilpi, E. (2010). *The education of children of immigrants in Finland* [Väitöskirja, Oxfordin yliopisto]. ORA. <https://ora.ox.ac.uk/objects/uuid:9ae6dfc5-bda6-4d6f-8780-c97abab350e9>
- Kilpi-Jakonen, E. (2011). Continuation to upper secondary education in Finland: Children of immigrants and the majority compared. *Acta Sociologica*, 54(1), 77–106. <https://doi.org/10.1177/0001699310392604>
- Kilpi-Jakonen, E. (2012). Does Finnish educational equality extend to children of immigrants? Examining national origin, gender and the relative importance of parental resources. *Nordic Journal of Migration Research*, 2(2), 167–181. <https://doi.org/10.2478/V10202-011-0039-4>
- Kilpi-Jakonen, E. (2017). Maahanmuuttajien lasten koulutussiirtymät toisen asteen koulutuksen läpi ja korkeakouluihin. *Kasvatus*, 48(3), 217–231.

- Kilpi-Jakonen, E., Erola, J., & Karhula, A. (2016). Inequalities in the haven of equality? Upper secondary education and entry into tertiary education in Finland. Teoksessa H.-P. Blossfeld, S. Buchholz, J. Skopek & M. Triventi (toim.), *Models of secondary education and social inequality – An international comparison* (ss. 181–195). Edward Elgar Publishing.
- Kilpi-Jakonen, E., & Sirniö, O. (2020). Formal adult education and labour market inequalities in Finland. Teoksessa J. Schrader, A. Ioannidou & H.-P. Blossfeld (toim.), *Monetäre und nicht monetäre Erträge von Weiterbildung. Edition ZfE, vol 7* (ss. 83–105). Springer VS.
https://doi.org/10.1007/978-3-658-25513-8_4
- Kivinen, O., Ahola, S., & Hedman, J. (2001). Expanding education and improving odds? Participation in higher education in Finland in the 1980s and 1990s. *Acta Sociologica*, 44(2), 171–181. <https://doi.org/10.1177/000169930104400205>
- Kivinen, O., Hedman, J., & Kaipainen, P. (2007). From elite university to mass higher education: Educational expansion, equality of opportunity and returns to university education. *Acta Sociologica*, 50(3), 231–247. <https://doi.org/10.1177/0001699307080929>
- Kivinen O., Hedman J., & Kaipainen P. (2012). Koulutusmahdollisuuksien yhdenvertaisuus Suomessa. Eriarvoisuuden uudet ja vanhat muodot. *Yhteiskuntapolitiikka*, 77(5), 559–566.
- Kivinen, O., & Rinne, R. (1995a). Korkeakoulutuksen kastijako. *Janus*, 2, 97–116.
- Kivinen, O., & Rinne, R. (1995b). Koulutuserojen pysyvyys. *Sociologia*, 32(2), 90–105.
- Kivinen, O., & Rinne, R. (1995c). Koulutuksen periytyvyys: nuorten koulutus ja tasa-arvo Suomessa. Tilastokeskus.
- Kupiainen, S., Marjanen, J., & Ouakrim-Soivio, N. (2018). *Ylioppilas valintojen pyörteissä: lukio-opinnot, ylioppilastutkinto ja korkeakoulujen opiskelijavalinta*. Suomen ainedidaktinen tutkimusseura. <http://hdl.handle.net/10138/231687>
- Kupiainen, S., Ouakrim-Soivio, N., & Hanska, J. (2023). Finnish matriculation examination's exam in Social Studies – an appropriate gatekeeper and competence support? *Journal of Social Science Education*, 22(2). <https://doi.org/10.11576/jsse-5942>
- Lahtinen, H., Korhonen, K., Martikainen, P., & Morris, T. (2023). Polygenic prediction of education and its role in the intergenerational transmission of education: Cohort changes among Finnish men and women born in 1925–1989. *Demography*, 60(5), 1523–1547. <https://doi.org/10.1215/00703370-10963788>
- Lahtinen, H., Martikainen, P., Korhonen, K., Morris, T., & Myrskylä, M. (2024). Educational tracking and the polygenic prediction of education. *Sociological Science*, 11(8), 186–213. <https://doi.org/10.15195/v11.a8>
- Lehti, H., & Erola, J. (2017). How do aunts and uncles compensate for low parental education in children's educational attainment? Teoksessa J. Erola & E. Kilpi-Jakonen (toim.), *Social inequality across the generations* (ss. 89–111). Edward Elgar Publishing.
- Lehti, H., Erola, J., & Karhula, A. (2019). The heterogeneous effects of parental unemployment on siblings' educational outcomes. *Research in Social Stratification and Mobility*, 64, 100439. <https://doi.org/10.1016/j.rssm.2019.100439>
- Lehti, H., Erola, J., & Tanskanen, A. O. (2019). Tying the extended family knot—Grandparents' influence on educational achievement. *European Sociological Review*, 35(1), 29–48. <https://doi.org/10.1093/esr/jcy044>

- Lehti, H., & Karonen, E. (2020). Koulutuksen laajentuminen, koulutuseri-inflaatio ja palkkatulojen heterogeenisyys eri yliopistoaloilla 1990- ja 2000-luvulla. *Sosiologia*, 57(2), 130–150. <https://journal.fi/sosiologia/article/view/124555>
- Lehti, H., & Laaninen, M. (2021). Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa. *Yhteiskuntapolitiikka*, 86(5–6), 520–532.
- McMullin, P., Karhula, A., Kilpi-Jakonen, E., & Erola, J. (2021). Geographical mobility and children's non-completion of upper secondary education in Finland and Germany: Do parental resources matter? *British Educational Research Journal*, 47(6), 1587–1610. <https://doi.org/10.1002/berj.3745>
- Mikkonen, J., Remes, H., Moustgaard, H., & Martikainen, P. (2020). Evaluating the role of parental education and adolescent health problems in educational attainment. *Demography*, 57(6), 2245–2267. <https://doi.org/10.1007/s13524-020-00919-y>
- Mikkonen, J., Remes, H., Moustgaard, H., & Martikainen, P. (2021). Early adolescent health problems, school performance, and upper secondary educational pathways: A counterfactual-based mediation analysis. *Social Forces*, 99(3), 1146–1175. <https://doi.org/10.1093/sf/soaa048>
- Nori, H. (2011). *Keille yliopiston portit avautuvat? Tutkimus suomalaisiin yliopistoihin ja eri tieteenaloille valikoitumisesta 2000-luvun alussa* [Väitöskirja, Turun yliopisto]. UTUPub. [https://urn.fi/URN:ISBN 978-951-29-4526-9](https://urn.fi/URN:ISBN%20978-951-29-4526-9)
- Nori, H., & Vanttaja, M. (2018). Pitkä tie yliopistoon: Toisen mahdollisuuden käyttäjät. *Aikuiskasvatus*, 38(4), 276–290. <https://doi.org/10.33336/aik.88373>
- Ojala, S., Kan, M.-Y., & Oinas, T. (2021). Teenage time use and educational attainment in adulthood in Finland. *Social Sciences*, 10(9), 343. <https://doi.org/10.3390/socsci10090343>
- Ouakrim-Soivio, N., Kupiainen, S., & Marjanen, J. (2017). Toimivatko oppilas- ja opiskelija-arvioinnin kriteerit? Oppiaineiden välinen ja sukupuolen mukainen vaihtelu perusopetuksen ja lukion päättöarvosanoissa ja arvosanojen yhteys nuorten oppiainevalintoihin. Teoksessa V. Britschgi & J. Rautopuro (toim.), *Kriteerit puntarissa* (ss. 81–117). Kasvatusalan tutkimuksia 74. Suomen kasvatus-tieteellinen seura.
- Pekkala Kerr, S., Pekkarinen, T., & Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3), 577–602. <https://doi.org/10.1086/669493>
- Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment. *The Scandinavian Journal of Economics*, 110(4), 807–825. <https://doi.org/10.1111/j.1467-9442.2008.00562.x>
- Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, 93(7–8), 965–973. <https://doi.org/10.1016/j.jpubeco.2009.04.006>
- Prix, I. (2013). More or different education? The joint impact of educational level and fields of study on earnings stratification in Finland, 1985–2005. *Acta Sociologica*, 56(3), 265–284. <https://doi.org/10.1177/0001699312474381>
- Prix, I., & Erola, J. (2017). Does death really make us equal? Educational attainment and resource compensation after paternal death in Finland. *Social Science Research*, 64, 171–183. <https://doi.org/10.1016/j.ssresearch.2016.10.012>

- Prix, I., & Kilpi-Jakonen, E. (2022). Not in a class of one's own: social origin differentials in applying to gender-(a) typical fields of study across the educational hierarchy. *European Sociological Review*, 38(6), 920–941. <https://doi.org/10.1093/esr/jcac007>
- Prix, I., Sirniö, O., & Saari, J. (2024). Better close to home? Geographical and socioeconomic constraints on gendered educational transitions at the upper secondary level. *Research in Social Stratification and Mobility*, 89, 100879. <https://doi.org/10.1016/j.rssm.2023.100879>
- Pulkkinen, J., & Rautopuro, J. (2022). The correspondence between PISA performance and school achievement in Finland. *International Journal of Educational Research*, 114, 102000. <https://doi.org/10.1016/j.ijer.2022.102000>
- Pöntinen, S. (1983). *Social mobility and social structure: A comparison of Scandinavian countries* (Vol. 20). Finnish Society of Sciences and Letters.
- Pöyliö, H., & McMullin, P. (2025). Participation in formal adult education and family life – a gendered story. *European Sociological Review*, 41(2), 248–264. <https://doi.org/10.1093/esr/jcae032>
- Rinne, R., Haltia, N., Nori, H., & Jauhiainen, A. (2008). *Yliopiston porteilla: Aikuiset ja nuoret hakijat ja sisäänpäässeet 2000-luvun alun Suomessa*. Suomen kasvatustieteellinen seura.
- Ruggera, L., Kilpi-Jakonen, E., & Erola, J. (2024). Intergenerational social mobility of cohorts born from 1951 to 1980 in Finland. *Acta Sociologica*, 68(1), 102–120. <https://doi.org/10.1177/00016993241265597>
- Siddiqui, N., Boliver, V., & Gorard, S. (2019). Reliability of longitudinal social surveys of access to higher education: The case of next steps in England. *Social Inclusion*, 7(1), 80–89. <https://doi.org/10.17645/si.v7i1.1631>
- Silliman, M., & Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*, 14(1), 197–224. <https://doi.org/10.1257/app.20190782>
- Sipilä, N., Kestilä, L., & Martikainen, P. (2011). Koulutuksen yhteys nuorten työttömyyteen: Mihin peruskoulututkinto riittää 2000-luvun alussa? *Yhteiskuntapolitiikka*, 76(2), 121–134.
- Sirniö, O., Kauppinen, T. M., & Martikainen, P. (2016). Income trajectories after graduation: An intergenerational approach. *Advances in Life Course Research*, 30, 72–83. <https://doi.org/10.1016/j.alcr.2016.04.001>
- Sirniö, O., Kauppinen, T. M., & Martikainen, P. (2017). Cohort differences in intergenerational income transmission in Finland. *Acta Sociologica*, 60(1), 21–39. <https://doi.org/10.1177/0001699316649649>
- Suhonen, T. (2014). Field-of-study choice in higher education: Does distance matter? *Spatial Economic Analysis*, 9(4), 355–375. <https://doi.org/10.1080/17421772.2014.961533>
- Tarkiainen, L., Kemppainen, T., Lahtinen, H., Bernelius, V., & Martikainen, P. (2024). The effect of cumulative childhood exposure to neighbourhood socioeconomic disadvantage on school performance – A register-based study on neighbourhoods, schools, and siblings. *European Sociological Review*, 40(3), 403–416. <https://doi.org/10.1093/esr/jcad058>

- Tervonen, L., Kortelainen, M., & Kanninen, O. (2018). Eliittilukioiden vaikutukset yliopisto-opintojen aloittamiseen ja koulutusalan valintaan. *Yhteiskuntapolitiikka*, 83(4), 374–386.
- Uusitalo, R. (1999). *Essays in economics of education* [Väitöskirja, Helsingin yliopisto]. Helda. <http://urn.fi/URN:ISBN:951-45-8704-9>
- Valkonen, T., Martelin, T., & Rimpelä, A. (1990). *Eriarvoisuus kuoleman edessä: Sosioekonomiset kuolleisuuserot Suomessa 1971–85*. Tilastokeskus.
- Valkonen, T., Pensola, T., & Jalovaara, M. (1996). Peruskoulu-uudistuksen vaikutus sukupuolten, alueiden ja sosiaaliryhmien välisiin eroihin perusasteen jälkeisessä koulutuksessa. Teoksessa R. Jakku-Sihvonen, A. Lindström & S. Lipsanen (toim.), *Toteuttaako peruskoulu tasa-arvoa?* (ss. 316–335). Opetushallitus.