



Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa

HANNU LEHTI & MARKUS LAANINEN

Aikaisempien tutkimusten mukaan perhetaustan mukainen oppimistulosten eriarvoisuus on voimistunut Suomessa viime vuosina. Tutkimusnäyttö pohjaa vahvasti koulukyselyaineistoihin, kuten PISA-aineistoon, jossa tieto perhetaustasta perustuu lasten vastauksiin. Artikkelit osoittaa, että PISA-aineistoissa vanhempien koulutukseen liittyy suurta mittausvirhettä. Kun lisäksi huomioidaan muutokset vanhempien koulutusjakaumissa, havaitaan, ettei oppimistulosten eriarvoisuus ole voimistunut Suomessa viime vuosina.

.....
English summary at the end of the article

Johdanto

Vaikka perhetausta on yhteydessä oppimistuloksiin ja koulumenestykseen kaikkialla (Barone 2006; Volante ym. 2019), Suomessa perhetaustan yhteys oppimistuloksiin on ollut kansainvälisesti vertaillen heikko (Chmielewski 2019). Suomea onkin pidetty oppimistulosten tasa-arvon mallimaana (Arinen & Karjalainen 2007; Barone 2006; Kupari 2005; Linnakylä & Sulkunen 2002). Kuitenkin viimeaikaisten PISA-aineistojen hyödyntävien tutkimusten mukaan perhetaustan ja lasten oppimistulosten yhteys olisi vahvistunut viime vuosina (Leino ym. 2019; Salmela-Aro & Chmielewski 2019; Vettenranta ym. 2016; Chmielewski 2019). Mediassa tutkimustulokset on tuotu esiin narratiivina, jossa maalataan kuvaa koulutusjärjestelmän tasa-arvon romuttamisesta. Huoli oppimistulosten eriarvoisuuden kasvusta on lisääntynyt.

Tutkimus on osa Suomen Akatemian rahoittamaa Eriarvoisuuden, interventioiden ja uuden hyvinvointiyhteiskunnan lippulavahanketta (INVEST) (päättönumero: 320162).

Tieto perhetaustan yhteydestä oppimiseen kertoo Suomen koulutusjärjestelmän tasa-arvoisuudesta. Onkin tärkeää tutkia luotettavasti, onko perhetaustan yhteys oppimiseen vahvistunut. Esimerkiksi yhtenä syynä Suomen korkeisiin PISA-tuloksiin on pidetty sitä, että oppimistulosten erot perhetaustan mukaan ovat pieniä (OECD 2010). Pienet osaamiserot eri taustoista tulevien oppilaiden välillä on myös kansainvälisissä vertailussa osoittautunut Suomen koulutusjärjestelmän eduksi (Martins & Veiga 2010; Barone 2006).

Oppimistulosten eriarvoisuuden tarkastelussa on tärkeää, että lasten perhetausta on mitattu luotettavasti. Mahdollinen mittausvirhe perhetaustassa voi vääristää tuloksia ja johtaa väärin tulokintoihin koulutusjärjestelmän epätasa-arvosta (kts. Engzell & Johnsson 2015). Yleisimpänä perhetaustan mittarina on käytetty vanhempien koulutusta, joka kytkeytyy perheen inhimilliseen ja kulttuuriseen pääomaan ja edeltää vanhemman ammattia ja tuloja (Erola ym. 2016). Suomessa vanhempien koulutuksen on osoitettu ennustavan ammattiasemaa tai kotitalouden tuloja voi-

makkaammin lasten koulutusta ja yhteiskunnallista asemaa (Erola ym. 2016; Lehti 2020; Karhula & Sirniö 2019).

PISA (*Programme for International Students Assessment*) on OECD:n (*Organisation for Economic Co-operation and Development*) jäsenmaiden yhteinen tutkimusohjelma, jonka avulla selvitetään 15-vuotiaiden nuorten osaamista matematiikassa, äidinkielessä sekä luonnontieteissä. Tutkimus on tehty kolmen vuoden välein vuodesta 2000 lähtien. Suomen PISA-tutkimuksissa tiedot vanhempien koulutuksesta pohjautuvat lasten raportointiin tietoihin. Siksi vanhemman koulutuksen osalta aineistossa saattaa esiintyä vastausharhasta johtuvaa mittausvirhettä. Tällä tarkoitetaan sitä, että lasten vastausten perusteella vanhempien koulutus luokitellaan väärin. Mikäli tutkimuksissa käytetään lapsilta saatuja vastauksia perhetaustasta, niiden luotettavuutta tulisi arvioida. Aikaisemmat PISA-aineistoilla tehdyt tutkimukset perhetaustan yhteydestä oppimiseen eivät ole ottaneet huomioon mahdollista vanhempien koulutukseen liittyvää vastausharhaa.

Tässä tutkimuksessa tarkastelemme, sisältävätkö PISA-aineistot vastausharhasta johtuvaa mittausvirhettä vanhemman koulutusmuuttujan osalta ja kuinka suurta mahdollinen virhe on. Vertaamme PISA-aineistojen vanhempien koulutusjakaumia koko ikäluokan rekisteritietoihin. Lisäksi tutkimme, onko oppimistulosten eriarvoisuudessa vanhemman koulutuksen ja ammattiaseman mukaan eroja PISA- ja rekisteriaineistojen välillä ja miten näiden aineistojen mukaan oppimistulosten eriarvoisuus on kehittynyt viime vuosina. Havainnollistamme myös, miksi eriarvoisuuden ajallisen kehityksen tarkastelussa on tärkeää huomioida vanhempien koulutustason muutokset. PISA-aineistoilla hyödynnämme oppimistuloksia lukutaidossa ja matematiikassa. Rekisteriaineistolla tarkastelemme perusopetuksen päättötodistuksen äidinkielen ja matematiikan arvosanoja sekä lukuaineiden arvosanojen keskiarvoja.

Vastausharha koulukyselyissä

Vastausharha aiheuttaa aineistoon virhettä. Vastausharha voi syntyä, kun vastaaja ei vastaa totuuden mukaisesti kyselyyn, koska vastausta voidaan pitää yhteiskunnallisesti epäsuotavana (Krumpal 2013). Vastausharha voi johtua myös siitä, että yksilöillä ei ole tietoa oikeasta vastauksesta. He saat-

tavat silti vastata, koska vastaaminen on psykologisesti suotuisempaa kuin vastaamatta jättäminen. (Tourangeau & Yan 2007)

Kysyttäessä vanhemman koulutusta vastausharha voikin syntyä siitä, että lapset eivät tiedä tarkkaan vanhemman koulutustasoa. Tällöin oppilaat soveltavat heuristista arviointia, vastaavat sosiaalisesti hyväksytyllä tavalla tai jättävät vastaamatta kysymykseen (Schulz 2006). Koska yksilöillä on taipumus ennemmin yli- kuin aliarvioida omaa taustaansa (Dunning 2011), lapset todennäköisesti yliarvioivat vanhempien koulutuksen. Toisaalta lapset voivat tietää paremmin vanhempien ammatin ja pyrkivät ammatin avulla arvioimaan vanhempien koulutusta (Engzell & Jonsson 2015).

Aikaisemman tutkimuksen mukaan yleisin huomio verrattaessa lasten ja vanhempien vastauksia koskien perhetaustaa on, että lapset raportoivat vanhempien ammatin luotettavammin kuin koulutuksen (Engzell & Johnsson 2015; Jerrim & Micklewright 2014; Schulz 2006; Lien ym. 2001; Looker 1989). Vanhempien koulutustason vastausharha voi olla seurausta monimutkaisista koulutusjärjestelmistä (Schulz 2006). Lapset tietävät vanhempiensa koulutuksen hyvin, mikäli vanhemmilla on perus- tai korkea-asteen tutkinto, mutta muita tutkintoja lasten on vaikeampi hahmottaa (Kreuter ym. 2010; Schulz 2006). Vastausharhassa ei ole havaittu eroja isän ja äidin koulutustietojen välillä (Jerrim & Micklewright 2014; Schulz 2006; Lien ym. 2001).

Tutkimukset ovat osoittaneet, että vanhempien koulutuksen sekä lasten oppimistulosten yhteys on vahvempi, kun lasten ilmoittamien tietojen sijaan hyödynnetään vanhempien raportoimaa tietoa koulutuksesta (Engzell & Johnsson 2015). Kysely- ja rekisteriaineistoja vertailevassa tutkimuksessa havaittiinkin, että vanhemmat raportoivat koulutuksensa lapsiaan luotettavammin. Sen sijaan sekä vanhempien että oppilaiden ilmoittaman ammattiaseman ja oppimistulosten yhteyden vahvuus on samansuuruista. (Engzell & Johnsson 2015; Jerrim & Micklewright 2014.)

Lasten raportoiman koulutuksen käyttäminen perhetaustan mittarina voi johtaa perhetaustan merkityksen aliarvioimiseen oppimistuloksia tarkasteltaessa. Ammattiasema sen sijaan vaikuttaa luotettavammalta perhetaustan mittarilta.

Perhetaustan yhteys nuorten oppimistuloksiin Suomessa

Aikaisempien tutkimusten mukaan perhetaustalla on yhteys oppimistuloksiin kaikissa yhteiskunnissa mutta yhteyden voimakkuus vaihtelee yhteiskuntien välillä (Bol ym. 2014; Chmielewski 2019; Barone 2006). Perhetaustan merkitys lasten oppimistuloksiin sekä myöhempään kouluttautumiseen on Suomessa kuitenkin heikompi kuin monissa muissa länsimaissa. (Chmielewski 2019; Lehti 2020; Martins & Veiga 2010; Barone 2006).

Suomen PISA-raporttien mukaan korkeampi perhetausta on yhteydessä parempiin oppimistuloksiin riippumatta siitä, mitataanko osaamista lukutaidossa, matematiikassa vai luonnontieteessä (Arinen & Karjalainen 2007; Kupari ym. 2013; Kupari 2005; Leino ym. 2019; Linnakylä & Sulkunen 2002; Vetterranta ym. 2016). Vuosien 2015 ja 2018 Suomen PISA-tulosten mukaan perhetaustan yhteys oppimistuloksiin on vahvistunut ja siten kuilu eri perhetaustoista tulevien oppilaiden välillä syventynyt. Lisäksi perhetaustan ja PISA-tulosten yhteyden vahvuus oli kansainvälisesti vertaillen keskitasoa. (Leino ym. 2019; Vetterranta ym. 2016)

Useita kansainvälisiä oppimistulosarviointejä tarkastelevan tutkimuksen mukaan oppimistulosten eriarvoisuus on voimistunut Suomessa viimeisten vuosikymmenten aikana (Salmela-Aro & Chmielewski 2019). Tulos pohjautuu kuitenkin vahvasti vuosien 2000–2015 PISA-aineistoihin.

Ongelmana Katariina Salmela-Aron ja Anna Chmielewskin (2019) tutkimuksessa on, että perhetaustaa mitataan luokitellulla muuttujalla ja näin eriarvoisuuden kasvu saattaa selittyä negatiivisella valikoitumisella vanhempien koulutusjakauman muuttuessa. Koulutuksen laajenemisen vuoksi nuorempien PISA-kohorttien vanhemmat ovat koulutetumpia kuin vanhempien PISA-kohorttien. Matalasti koulutettujen joukko siis vähenee ja valikoituu ajan myötä (ks. Karhunen & Uusitalo 2017). Kun verrataan esimerkiksi korkeasti ja matalasti koulutettujen perheiden lasten oppimistuloksia vuosien välillä, voi eriarvoisuuden kasvu selittyä negatiivisella valikoitumisella.

Salmela-Aro ja Chmielewski (2019) käyttävät vanhempien koulutusjakauman 10., 50. sekä 90. persentiilien välisiä oppimistulosten ”suorituskuiluja”, jotta koulutustason muutos tulisi huomioi-

duksi. Ongelmana on se, että persenteileihin perustuva suorituskuilu on tarkoitettu muodostettavaksi jatkuvista muuttujista. Jos vanhemman koulutusta mitataan luokitellulla muuttujalla, ei persentiilien tarkastelu ole välttämättä mielekäsää. Salmela-Aro ja Chmielewski (2019) toteavatkin artikkelissaan, että esimerkiksi vuoden 2015 PISA-aineistossa 90. ja 50. persenteileissä vanhemman koulutus jakaantui samaan koulutustasoon (ISCED 5A). Tällöin he eivät voineet laskea 90. ja 50. persentiilien välistä suorituskuilua.

Havainnollistamme seuraavaksi persenteileihin perustuvan suorituskuilun ongelmaa. Liitetaulukossa 1a (ks. www.julkari.fi) on vanhempien (isän tai äidin) korkeimman koulutusluokan (ISCED) prosenttijakaumat sekä kertymäfunktio PISA-aineistoissa. Kertymäfunktio (kumulatiivinen %) osoittaa persenteilit, joihin eri koulutusluokat asettuvat. Taulukkoon on myös lihavoitu ne vanhemman koulutusluokat, jotka ovat 10. ja 90. persentiilien kohdalla. Nämä koulutusluokat eivät pysy samansuuruisina PISA-aineistojen välillä. Vuonna 2000 vanhemman koulutusluokkiin (ISCED 1 ja ISCED 4–6), joihin 10. ja 90. persenteilit asettuivat, kuului 10,7 ja 35,9 prosenttia oppilaista. Sen sijaan vuonna 2015 10. ja 90. persentiilien koulutusluokat sisälsivät 9,3 ja 52,5 prosenttia oppilaista. Koulutustason laajenemisen johdosta 90. persenteiliin asettuvan vanhempien koulutusluokan suhteellinen koko siis vaihtelee PISA-aineistoissa huomattavasti. Tällöin persenteileihin perustuvan suorituskuilun mittaaminen ei kuvaa oikein eriarvoisuuden kehitystä PISA-aineistojen välillä. Sen sijaan tulisi verrata samankokoisten ryhmien välistä suorituskuilua.

Vanhemman koulutus voidaan operationalisoida luotettavaksi mittariksi muuttamalla se koulutusvuosiksi. Jatkuva muuttuja ei ole herkkä valikoitumiselle, ja sitä voidaan käyttää ajallisen muutoksen tarkastelussa (ks. Härkönen & Sirniö 2020).

Tutkimusasetelma

Ensimmäiseksi arvioimme oppilaiden ilmoittamien vanhempien koulutusta koskevien tietojen luotettavuutta PISA-aineistoissa vertaamalla vanhempien koulutusjakaumia PISA- ja rekisteriaineistoissa. Ensimmäinen tutkimuskysymyksemme on seuraava:

- Sisältävätkö PISA-aineistot mittausrvirheitä vanhempien koulutusmuuttujien osalta, ja kuinka suurta virhe on verrattuna rekisteritietoihin?

Koska emme voi yhdistää PISA-aineistoja rekistereihin, on mahdollista, että mittausrvirhe johtuu vastausharhan sijaan siitä, että otos on valikoitunut vanhempien koulutuksen mukaan (otosharha). Palaamme keskusteluun otosharhan mahdollisuudesta artikkelin johtopäätöksissä.

Seuraavaksi tutkimme oppimistulosten eriarvoisuutta vanhempien koulutuksen ja ammattiaseman mukaan. Tarkastelemme PISA-aineistoilla matematiikan ja lukutaidon oppimistuloksia sekä rekisteriaineistolla peruskoulun päättötodistuksen matematiikan ja äidinkielen arvosanoja. Lopuksi tarkastelemme rekisteriaineistoilla, miten vanhemman koulutus ja ammattiaseman yhteys peruskoulun päättötodistuksen lukuaineiden keskiarvoihin on kehittynyt viime vuosina. Tutkimuskysymyksemme ovat seuraavat:

- Millaiselta vanhemman koulutuksella ja ammatilla mitattu oppimistulosten eriarvoisuus näyttää PISA- ja rekisteriaineistolla mitattuna?
- Ovatko oppimistulosten erot vanhemman koulutuksen ja ammattiaseman mukaan kasvaneet Suomessa?

Koska käytämme eri vuosien PISA-aineistoja ja vastaavien ikäluokkien päättötodistuksen arvosanoja rekistereistä, pystymme tarkastelemaan molemmilla aineistoilla, ovatko erot oppimistuloksissa kasvaneet vanhemman koulutuksen ja ammattiaseman mukaan. PISA-testit ja perusopetuksen päättötodistuksen arvosanat mittaavat hieman eri asioita. Kouluarvosanojen ja PISA-testin tulosten yhteys on kuitenkin havaittu vahvaksi Suomessa (Linnakylä & Sulkunen 2002). Tässä tutkimuksessa emme voi tarkastella PISA-kokeiden ja peruskoulun päättötodistuksen yhteyttä, koska se vaatisi PISA-aineistojen yhdistämistä rekistereihin.

Aikaisemmin ei ole tutkittu, miten perhetaustaa mittaavien muuttujien erilainen operationalisointi vaikuttaa perhetaustan ja lasten oppimistulosten yhteyden voimakkuuteen. Tässä tutkimuksessa tarkastelemme operationalisoinnista johtuvia eroja hyödyntämällä sekä jatkuvia että luokiteltuja muuttujia.

Aineistot

PISA-aineistot

Käytämme tutkimuksessa PISA-aineistoja vuosilta 2000, 2003, 2006, 2009, 2012 ja 2015. PISA-aineistot ovat saatavilla OECD:n internet-sivuilta (www.oecd.org/pisa).

Suomen PISA-tutkimuksen kohdepopulaatio koostuu helmikuun ja tammikuun välisenä aikana syntyneistä oppilaista, jotka ovat tutkimuksen aikana noin 15-vuotiaita. Suomessa kohdejoukon suuruus on ollut noin 60 000 oppilasta. Otanta on kaksivaiheinen. Ensin valitaan tutkimukseen osallistuvat koulut ja tämän jälkeen kouluista kokeisiin osallistuvat oppilaat. Otannassa huomioidaan koulujen alueellinen sekä kokoon perustuva edustavuus. (Nissinen ym. 2018.)

Joinakin vuosina Suomen PISA-tutkimuksissa on poimittu yliotos ruotsinkielisistä ja maahanmuuttajataustaisista nuorista. Yliotoksesta johtuvaa vääristymää korjataan analyysivaiheessa painokertoimilla. Vuoden 2012 Suomen PISA-tutkimuksen otokseen valittiin kaikki koulut, joissa oli vähintään viisi maahanmuuttajataustaista oppilasta (Kupari ym. 2013). Tämä on herättänyt keskustelua, koska oletettavasti näissä kouluissa kantaväestön sosiaalinen tausta on matalampi, mikä voi heijastua koetuloksiin (kts. Kivinen & Hedman 2017). Tämä tulee huomioida vuoden 2012 PISA-tuloksia tulkittaessa.

PISA-tutkimusten kohdejoukko sisältää pienen määrän oppilaita, jotka ovat rekisteriaineiston ikäluokkia vuoden nuorempia (tammikuussa syntyneet). Jotta PISA- ja rekisteriaineistot vastaavat samoja ikäluokkia, jätimme PISA-aineistoista pois tammikuussa syntyneet (n. 4–8 % aineistosta riippuen). Vuoden 2000 PISA-aineistossa matematiikan tulokset voidaan havaita vain puolelle osallistuneista oppilaista (Väljærvi & Linnakylä 2002), minkä vuoksi keskityimme kyseisen vuoden osalta vain lukutaidon tuloksiin.

PISA-tutkimuksessa oppilaat suorittavat kaksituntisen kokeen. Tämän jälkeen he vastaavat taustakyselyyn, jossa selvitetään esimerkiksi nuorten kotitaustaa sekä asenteita oppimiseen. Taustakyselyssä vanhempien koulutusastetta tiedustellaan valmiiden vastausvaihtoehtojen avulla. Vanhempien ammatti kysytään kahdella avovastauksella. Ensimmäiseen avovastaukseen oppilaat kirjoittavat vanhempiensa päätyön ja toiseen mitä työssä tehdään. Näiden vastausten perusteella PISA-aineistoihin määritellään vanhempien ammattinimike.

Rekisteri-aineistot

Tutkimuksen rekisteriaineistot pohjautuvat FOLK-henkilöaineistoon, joka on kokonaisaineisto Suomessa asuvista henkilöistä vuosilta 1987–2016. Henkilöaineistojen lisäksi käytämme rekisteritietoja toisen asteen yhteishakurekisteristä (Tilastokeskus 2021). Aineistoista saamme PISA-tutkimukseen osallistuneiden ikäluokkien tiedot perusopetuksen päättötodistuksen arvosanoista, vanhempien koulutuksesta ja ammatista. Valitsimme lapsista tarkasteluun vain ne ikäluokat, jotka osallistuivat PISA-tutkimukseen (vuosina 1984, 1987, 1990, 1993, 1996 ja 1999 syntyneet).

Vanhempien koulutustiedot perustuvat FOLK-henkilöaineiston tutkintomoduuliin, jossa on tiedot 15 vuotta täyttäneen väestön suorittamista tutkinnoista (Tilastokeskus 2021). Tiedot vanhempien koulutuksesta otettiin samalta vuodelta, kun lapset osallistuivat PISA-tutkimukseen. Koska pieneltä määrältä maahanmuuttajataustaisista lapsista puuttui tieto molempien vanhempien koulutuksesta (ks. liitetaulukko 2a, puuttuvat tiedot, www.julkari.fi), teimme kuvailevat analyysit erikseen vain Suomessa syntyneille lasten äideille. Suomessa syntyneiden lasten osalta tutkintomoduuli ei sisältänyt puuttuvia tietoja äidin koulutuksesta (ks. liitetaulukko 3, www.julkari.fi). Nämä tulokset eivät eronneet niistä tuloksista, jotka esittelemme tulososiossa (tulokset saa pyydettyä kirjittajilta).

Tiedot vanhempien ammatista ovat peräisin FOLK-henkilöaineiston työssäkäyntimoduulista, jossa on tiedot 18–74-vuotiaille palkansaajille ja yrittäjille (Tilastokeskus 2021). Ammattitiedot ovat saatavilla vuosilta 1990, 1993, 1995, 2000 ja 2004–2016, joten työssäkäyntimoduuli sisältää puuttuvia tietoja tietyiltä vuosilta. Vanhemmat voivat myös olla työttöminä tai työmarkkinoiden ulkopuolella. Koska suomalaista rekisteriaineistoa hyödyntävä tutkimus on osoittanut, että vanhempien ammatti ei muutu enää merkittävästi lapsen syntymän jälkeen (Erola ym. 2016), otimme vanhempien ammatin ensimmäiseltä mahdolliselta vuodelta, kun se on saatavissa lapsen ollessa alle 17-vuotias. Vanhempien ammatti oli saatavilla noin 90 prosentille, kun lapset olivat 10–16-vuotiaita.¹ Liitetaulukoissa 4a–4c (ks. www.julkari.fi) esitetään ISCO-luokitukseen perustuvat vanhempien ammattien jakaumat.

¹ 80 prosentille vanhemman ammattiasema oli saatavilla samana vuonna kuin syntymäkohortti osallistui PISA- testiin.

Koska vanhempien ammateissa on puuttuvia rekisteritietoja, emme arvioi PISA-aineistojen vanhempien ammattien edustavuutta, emmekä tarkastele siihen kohdistuvaa mittausvirhettä. Hyödynnämme vanhempien ammattiasemaa regressioanalyysissä toistaaksemme tulokset vanhempien koulutuksen ja lasten PISA-tulosten ja arvosanojen yhteydestä. Näissä analyyseissä vakioidaan vanhempien ammatin puuttuvat tiedot. Kuten aiemmin todettiin, koulukyselyaineistoissa vanhempien ammatti on ollut vanhempien koulutusta luotettavampi mittari. Lisäksi kahden eri perhetaustaa mittaavan muuttujan käyttäminen lisää tulosten luotettavuutta.

Lasten matematiikan ja äidinkielen arvosanat sekä perusopetuksen päättötodistuksen keskiarvo ovat peräisin toisen asteen yhteishakurekisteristä. Vuosittain noin 4–5 prosentilla peruskoulun päättävistä oppilaista oli puuttuvia tietoja rekisterissä (liitetaulukko 5, www.julkari.fi). Rekisteristä puuttuu tieto arvosanoista, jos lapsi ei osallistunut toisen asteen hakuun, haki oppilaitokseen toisen asteen haun ulkopuolella esimerkiksi ulkomaille tai kotipaikkakunta on Ahvenanmaa. Poistimme nämä oppilaat aineistosta ennen analyysia. Toisen asteen hakurekisteri sisälsi paljon puuttuvia tietoja äidinkielen arvosanoista vuonna 2009 ja matematiikan arvosanoista vuonna 2012. Kun tarkastelemme perhetaustan yhteyttä oppimistuloksiin, nämä vuodet on jätetty pois analyyseistä.

Muuttajat ja menetelmät

Tutkimuksen selitettävänä muuttujina tarkastelemme rekisteriaineistolla perusopetuksen päättötodistuksen matematiikan ja äidinkielen arvosanoja sekä lukuaineiden keskiarvoa. PISA-aineistoilla tarkastelemme lukutaidon ja matematiikan koetuloksista johdettuna PV-lukuarvoja (*Plausible values*) (ks. Nissinen ym. 2018). Analyysijä varten standardoimme kaikki selitettävät muuttajat, joten muuttajien keskiarvo on 0 ja keskihajonta 1.

Selittäviä muuttujia, vanhempien koulutusta ja ammattiasemaa, mitataan sekä luokitelluilla että jatkuvilla muuttujilla. Sekä PISA- että rekisteriaineistoissa vanhempien koulutus on ISCED 2011-luokituksena (vuoteen 2011 asti ISCED 1997-luokitus). Muodostimme ISCED-luokituksista kolme muuttujaa. Koulutus vuosissa määriteltiin seuraavasti: 9 vuotta (ISCED 1–2 tai vähemmän), 12 vuotta (ISCED 3–4), 14,5 vuotta (ISCED 5B)

ja 16,5 vuotta (ISCED 5A tai korkeampi). Kaksi-
luokkainen koulutusmuuttuja sisältää luokat keskiaste tai alempi (ISCED 4 tai alempi) sekä alin korkea-aste tai korkeampi (ISCED 5B tai ylempi). Kolmeluokkainen koulutusmuuttuja sisältää luokat perusaste tai alempi (ISCED 1–2), keskias-
te (ISCED 3–4) sekä korkea-aste (ISCED 5B–6).

Vanhempien ammattimuuttuja perustuu ISCO-08-luokitukseen (vuoteen 2008 asti ISCO-88-luokitus). ISCO-luokitukselta on johdettu kaksi muuttujaa. Vanhemman ammattiasemaa mittaamme ISEI-indeksillä, joka on jatkuva muuttuja. ISEI-indeksi on johdettu ISCO-luokitukselta ja sen arvot vaihtelevat 16 ja 90 pisteen välillä (Ganzeboom ym. 1992). Kaksiluokkainen vanhemman ammattia kuvaava muuttuja sisältää luokat ”ISCO 1–3” (asiantuntijat, erikoisasiantuntijat ja johtajat) sekä ”ISCO 4–9 ja 0” (toimisto- ja palvelutyöntekijät, palvelu- ja myyntityöntekijä, maanviljelijät, metsätyöntekijät, rakennus-, korjaus- ja valmistustyöntekijät, prosessi- ja kuljestyöntekijät sekä muut työntekijät).

Muodostimme selittävät muuttujat sekä äideille että isille. Lisäksi teimme erikseen muuttujat kuvaamaan vanhempien korkeinta koulutusta ja ammattiasemaa. Näiden muuttujien arvot vastaavat sen vanhemman tietoja, jolla on korkeampi koulutus tai ammattiasema.

Ensimmäisen tutkimuskysymyksen osalta tarkastelemme mittausvirhettä vertaamalla äidin ja isän koulutusjakautumia PISA- ja rekisteriaineistojen välillä. Käytämme tarkastelussa kolmeluokkaisen koulutusmuuttujaa. Esitämme tulokset kuviona, joka näyttää, kuinka monta prosenttiyksikköä äidin ja isän koulutuksen jakaumat eroavat PISA-aineistoissa verrattuna rekisteriaineistoon. Vanhempien koulutuksen prosenttijakaumat sekä tapausmäärät ovat liitteenä (liitetaulukot 2a–2c, ks. www.julkari.fi).

Toisen ja kolmannen tutkimuskysymyksen osalta standardoimme vanhemman koulutuksen vuosissa ja ISEI-indeksin. Tulokset vanhemman jatkuvien muuttujien ja lasten oppimistulosten yhteydestä esitämme siis korrelaatiokertoimena². Estimaatti kertoo, kuinka monta keskihajontaa oppimistulokset muuttuvat, kun vanhemman muuttuja kasvaa yhdellä keskihajonnalla. Estimaattia on käytetty ylisukupolvisissa tutkimuksissa osoittamaan korrelaation voimakkuus vanhemman ja

lapsen muuttujan välillä (Martins & Veiga 2010). Koska käytämme koulutusta ja ammattiasemaa jatkuvina muuttujina, vanhempien koulutusjakautuman muutos tulee huomioiduksi analyyseissa. Raportoimme päätulokset vain vanhempien (isän tai äidin) korkeinta koulutusta ja ammattiasemaa käyttäen. Liitekuvioiden 1a–1d raportoimme tulokset erikseen äideille ja isille (ks. www.julkari.fi).

Kuten aiemmin mainitsimme, vanhemman koulutuksen ja oppimistulosten yhteyttä on mitattu aikaisemmin luokitellulla muuttujalla (esim. Salmela-Aro & Chmielewski 2019). Tässä tutkimuksessa havainnollistamme tätä operationalisoinnin ongelmaa käyttämällä vanhemman koulutuksesta ja ammatista muodostettuja kaksiluokkaisia muuttujia. Tarkastelemme osaamiskuilun standardia poikkeamaa verratessamme korkeasti (korkea-aste) ja matalasti (keskiaste tai alempi) koulutettujen vanhempien sekä korkeassa (ISCO 1–3) ja matalassa (ISCO 4–9 ja 0) ammatissa toimivien vanhempien lasten oppimistuloksia. Analyysimenetelmänä käytämme lineaarista regressioanalyysiä (OLS).

Hyödynnämme PISA-aineistojen analyyseissä Stata-ohjelmalle kirjoitettua Repest-pakettia, jonka avulla aineistojen hierarkkisuus, otanta-asetelmat sekä PV-lukuarvojen käyttö huomioidaan analyyseissä (Avvisati & Keslair 2014).

Vanhempien koulutuksen mittausvirhe PISA-aineistoissa

Ensimmäisenä tutkimustehtävänä selvitämme, sisältyykö vanhempien koulutusmuuttujiin mittausvirhettä ja kuinka suurta virhe on. Tarkastelemme, kuinka monta prosenttiyksikköä PISA-aineistojen äidin ja isän koulutusjakaumat poikkeavat rekisteriaineiston vastaavista jakaumista.

Kuviosta 1 havaitaan, että vuosien 2003–2015 PISA-aineistoissa sekä äidin että isän koulutus arvioidaan korkeammaksi verrattuna rekisteriaineistoihin. Erot johtuvat pääosin korkeakoulutettujen vanhempien liian suuresta osuudesta sekä keskias-
teen tutkinnon suorittaneiden liian pienestä osuudesta PISA-aineistoissa. Myös perusasteen tutkinnon suorittaneiden vanhempien osuudet arvioidaan PISA-aineistoissa liian pieniksi. Poikkeukse-
na on vuosi 2000, joka näyttää kaikkein tasapainoisimmalta verrattuna rekisteriaineistoon. Tuoloin peruskoulutettujen vanhempien osuus on yliarvioitu ja puolestaan korkeakoulutettujen van-

2 Vastaa Pearsonin korrelaatiota. Tulokset ovat myös yhteneväisiä järjestykorrelaation (Spearman) kanssa.

hempien osuus aliarvioitu. Vuonna 2000 toisen asteen tutkinnon suorittaneiden vanhempien osuus vastasi hyvin rekisteriaineiston tietojä. Tämä johtuu todennäköisesti siitä, että vuonna 2000 vanhempien koulutus kysyttiin karkeammalla luokittelulla kuin muina vuosina. Oppilailta kysyttiin, ovatko vanhemmat suorittaneet perusasteen, toisen asteen tai korkea asteen tutkinnon. Muina vuosina annettiin useampia vastausvaihtoehtoja korkea-asteen tutkinnolle.

Kuvion 1 perusteella voidaan todeta, että PISA-aineistot sisältävät suurta mittausvirhettä vanhemman koulutuksen osalta. Vuosina 2003–2015 mittausvirhe korkeakoulutetuilla vanhemmilla on noin 18–31 prosenttiyksikköä. Äidin koulutus sisältää isän koulutusta vähemmän mittausvirhettä, mutta erot eivät ole suuria. Tulos viittaa siihen, että eriarvoisuus vanhemman koulutuksen mukaan saattaa olla suurempaa kuin PISA-aineistoilla ollaan arvioitu, koska havaittu virhe johtaa eriarvoisuuden aliarvioimiseen.

Vanhempien koulutuksen ja ammattiaseman yhteys lasten oppimistuloksiin

Seuraavaksi tarkastelemme vanhempien koulutuksen ja ammattiaseman yhteyttä lasten oppimistuloksiin PISA- ja rekisteriaineistoissa. Lisäksi tutkimme, miten vanhemman taustan yhteys oppimistulosten on kehittynyt vuosina 2000–2015.

Kuviossa 2 esitetään vanhempien korkeimman koulutuksen sekä ammattiaseman yhteys lukutaidon pisteisiin (PISA) sekä perusopetuksen päättötodistuksen äidinkielen arvosanoihin (rekisteri). Kuvioista nähdään, että rekisteriaineistossa vanhempien koulutuksen yhteys oppimistuloksiin on vahvempaa kuin PISA-aineistoissa riippumatta siitä, mitataanko koulutusta luokiteltuna tai jatkuvana muuttujana. Sen sijaan vanhempien ammattiaseman sekä oppimistulosten yhteys on lähes yhtä suurta. Erot vanhempien koulutuksen yhteydestä oppimistuloksiin johtuvat todennäköisesti PISA-aineiston vanhempien koulutuksen mittausvirheestä (ks. kuvio 1). Jos mittausvirhe johtuu vastausharhasta, lapset yliarvioivat vanhempien koulutuksen. Siten korkeammin koulutettujen vanhempien lasten ryhmään sekoittuu matalammin koulutettujen lapsia, joiden PISA-testien tulokset ovat keskimäärin heikompia. Siksi PISA-aineistot aliarvioivat vanhempien koulutuksen yhteyttä oppimistuloksiin.

Kuviossa 3 esitetään vanhempien korkeimman koulutuksen ja ammattiaseman yhteys matematiikan pisteisiin (PISA) sekä perusopetuksen päättötodistuksen matematiikan arvosanoihin (rekisteri). Tulokset ovat samansuuntaisia lukutaidon ja äidinkielen tulosten kanssa. Vanhempien koulutuksen ja matematiikan yhteys on vahvempaa rekisteri- kuin PISA-aineistoissa. Poikkeuksena on vuosi 2015, jolloin yhteys on yhtä vahvaa. Tämä johtuu siitä, että rekisteriaineistossa matematiikan arvosanojen yhteys vanhempien koulutukseen on heikentynyt, kun PISA-aineistoissa matematiikan pisteiden ja vanhempien koulutuksen yhteys on puolestaan voimistunut. Kun tarkastellaan vanhempien ammattiaseman yhteyttä matematiikan osaamiseen, huomataan, että yhteys on lähes yhtä vahvaa molemmissa aineistoissa (kuvio 3, paneelit C ja D). Poikkeuksena tässäkin on vuosi 2015, jolloin PISA-aineistossa yhteys on hieman vahvempaa, kuin rekisteriaineistossa (kuvio 3, paneeli D).

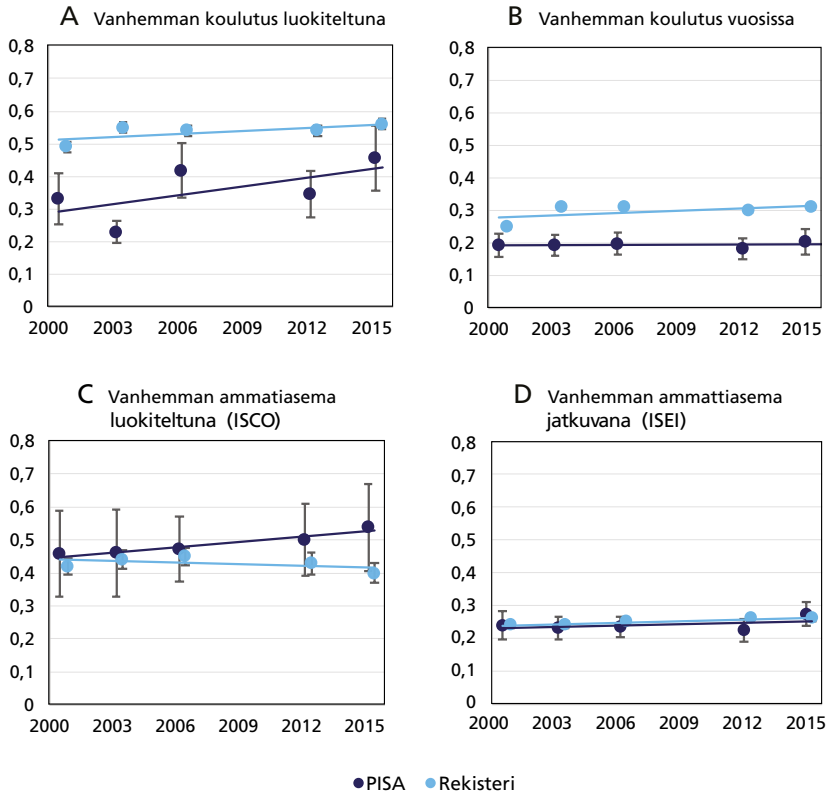
Kuvioista 2 ja 3 voidaan havaita, että PISA-aineistoissa oppimistulosten eriarvoisuus on kasvanut lukutaidossa ja matematiikassa, kun tarkastellaan luokiteltujen muuttujien tuloksia (paneelit A ja C). Muutos ei ole kuitenkaan tilastollisesti merkitsevä ammattiaseman osalta. Tulokset ovat yhteneväisiä aikaisemman tutkimuksen kanssa, jonka mukaan oppimistulosten eriarvoisuus on voimistunut Suomessa (Salmela-Aro & Chmielewski 2019). Sen sijaan rekisteriaineistoissa ei havaita eriarvoisuuden voimistuneen, vaikka yhteyksiä tarkastellaan luokitelluilla muuttujilla (Paneelit A ja C).

Koska luokiteltu muuttuja on altis jakauman muutoksille vuosien välillä, tulos ei välttämättä kerro todellisesta eriarvoisuuden kasvusta vaan valikoitumisesta. Lisäksi PISA-aineistoissa havaittu vanhempien koulutusmuuttujaan sisältyvä mittausvirhe vaikuttaa tuloksiin. Tästä seuraa, että matalammin koulutettujen ryhmä (perus- ja keskiaste) valikoituu koulutusrakenteen muutoksen sekä vastausharhan johdosta. Valikoitumisen ja mittausvirheen ongelmia voidaan havainnollistaa vertailemalla, miten äitien ja isien koulutusjakaumissa tapahtuvat erilaiset muutokset ovat yhteydessä oppimistuloksiin. Liitetaulukossa 2b ja 2c raportoimme koulutusjakaumat äideille ja isille. Liittekuvioissa 1a–1d raportoimme erikseen äitien sekä isien koulutuksen ja ammatin yhteydet oppimistuloksiin. (Ks. www.julkari.fi.)

Liitetaulukosta 2b nähdään, että äitien koulutusjakauma muuttuu PISA-aineistoissa vuosien



Kuvio 1. Vanhempien koulutuksen poikkeamat prosenttisyksikköinä PISA-aineistoissa 2000-2015 verrattuna koko ikäluokan rekisteritietoihin.



Kuvio 2. Vanhempien korkeimman koulutuksen ja ammattiaseman yhteys lukutaidon (PISA) ja äidinkielen (rekisteri) osaamiseen 2000–2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.

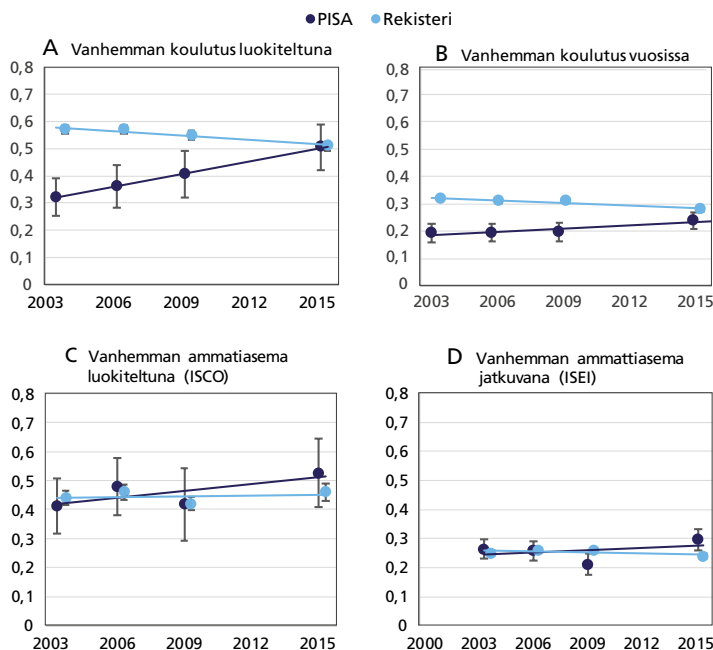
välillä siten, että matalammin koulutettujen (perus- ja keskiaste) osuus pienenee merkittävästi. Äideille erikseen tehdyistä analyyseistä (liitekuviot 1a ja 1c) nähdäänkin, että PISA-aineistoissa oppimistulosten ja äidin koulutuksen yhteyden vahvuus voimistuu, kun yhteyttä tarkastellaan luokitellulla äidin koulutusmuuttujalla.

Sen sijaan matalammin koulutettujen (peruskoulu ja keskiaste) isien osuudet muuttuvat vain hyvin vähän PISA-aineistoissa (liitetaulukko 2c). Liitekuviosta 1b ja 1d havaitaan, että oppimistulosten ja luokitellun isän koulutus -muuttujan yhteyden vahvuus ei voimistu. Yhteyden vahvuus näyttää jopa heikentyvän, koska mittausvirhe isän koulutusmuuttujassa kaventaa koulutusryhmien välisiä oppimiseroja. Äideillä on keskimäärin isä korkeampi koulutus, minkä vuoksi pääanalyysit (kuviot 2 ja 3) kuvastavat vahvemmin äitien koulutuksen yhteyttä oppimistuloksiin. Valikoitumisen takia yhteyksiä on luotettavampaa tutkia jatkuvilla muuttujilla.

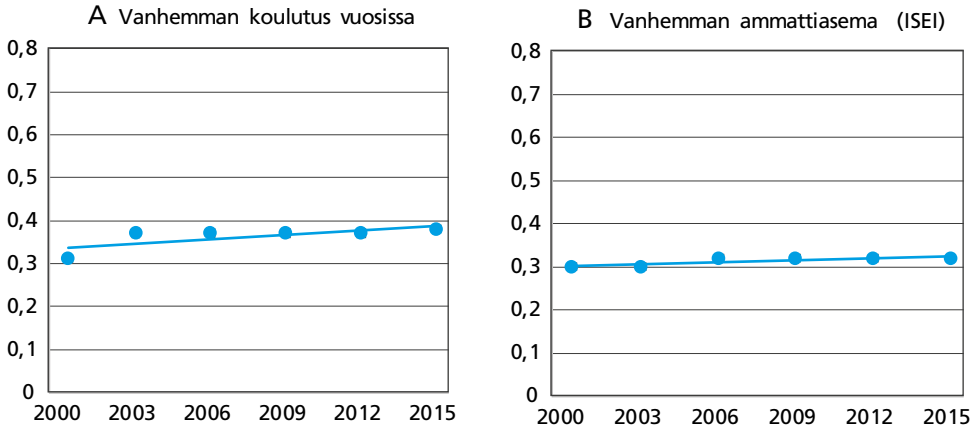
Jatkuvien muuttujien osalta havaitaan, että kummankaan aineiston perusteella ei voida sanoa eriarvoisuuden voimistuneen tai heikentyneen

merkittävästi. Kuviosta 2 nähdään, että vanhemman koulutuksen ja ammattiaseman yhteys lasten lukutaidon pisteisiin PISA-testeissä ja äidinkielen arvosanaan perusopetuksen päättötodistuksessa ei ole voimistunut (paneelit B ja D). Rekisteriaineiston mukaan vanhemman koulutusvuosien ja äidinkielen arvosanan yhteys voimistuu tilastollisesti merkittävästi vuosien 2000 ja 2003 välillä (paneeli B). Kuitenkin vuodesta 2003 eteenpäin vuosien väliset erot eivät ole enää tilastollisesti merkittäviä. Korrelaatiokerroin on 0,25 vuonna 2000 ja muina vuosina 0,30–0,31. PISA-aineistoissa vanhemman koulutuksen ja lukutaidon korrelaatiokerroin vaihtelee välillä 0,18–0,20.

Kuviosta 3 nähdään, että rekisteri ja PISA-aineistot antavat ristiriitaisen kuvan vanhempien koulutusvuosien yhteydestä matematiikan tuloksiin. Rekisteriaineistojen mukaan eriarvoisuus on heikentynyt ja PISA-aineistojen mukaan puolestaan hieman voimistunut. Muutos on kuitenkin tilastollisesti merkittävä vain rekisteriaineiston osalta. Matematiikan tulosten sekä vanhempien ammattiaseman yhteys ei ole merkittävästi muuttunut vuosina 2003–2015.



Kuvio 3. Vanhempien korkeimman koulutuksen ja ammattiaseman yhteys matematiikan osaamiseen PISA- ja rekisteriaineistoissa 2003, 2006, 2009 ja 2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.



Kuvio 4. Vanhempien korkeimman koulutuksen sekä ammattiaseman yhteys peruskoulun päättötodistuksen keskiarvoon 2003–2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.

Perhetaustan yhteys peruskoulun päättötodistuksen lukuaineiden keskiarvoon

Lopuksi tarkastelemme vanhempien koulutuksen ja ammattiaseman yhteyttä peruskoulun päättötodistuksen lukuaineiden keskiarvoon. Näin voimme selvittää, miten oppimistulosten erot perhetaustan mukaan ovat kehittyneet kokonaisuudessaan. Lukuaineiden keskiarvoja on tärkeää tutkia, koska toisen asteen haku perustuu usein lukuaineiden keskiarvoihin.

Kuviossa 4 on esitetty vanhempien koulutuksen (paneeli A) sekä ammattiaseman (paneeli B) yhteys perusopetuksen päättötodistuksen lukuaineiden keskiarvoon jatkuvilla muuttujilla mitattuna. Kuviosta havaitaan, että yhteys on pysynyt lähes samansuuruisena koko tarkastelujaksolla. Vuonna 2000 yhteys on hieman heikompi koulutuksen ja vuosina 2000 ja 2003 vanhemman ammattiaseman osalta. Viime vuosina perhetaustan yhteys perusopetuksen päättötodistuksen lukuaineiden keskiarvoon ei ole muuttunut. Vanhemman koulutuksella on ammattiasemaa hieman voimakkaampi yhteys oppimistuloksiin. Ero on kuitenkin pieni (esim. vuonna 2015 korrelaatiokertoimien erotus oli $0,38 - 0,32 = 0,06$). Vanhempien koulutuksen ja ammattiaseman yhteys nuorten perusopetuksen päättötodistuksen lukuaineiden keskiarvoon on siis lähes yhtä vahvaa.

Johtopäätökset

Tästä tutkimuksesta selvitimme, sisältävätkö PISA-aineistot mittausvirhettä oppilaiden vanhempien koulutuksen osalta. Vertasimme Suomen PISA-aineistojen oppilaiden vanhempien koulutusjakaumia rekisteritietojen jakaumiin niissä ikäluokissa, jotka osallistuvat PISA-testeihin. Lisäksi tarkastelimme, onko oppimistulosten eriarvoisuus voimistunut Suomessa tutkimalla vanhempien koulutuksen sekä ammattiaseman yhteyttä lasten oppimistuloksiin sekä PISA- että rekisteriaineistolla vuosina 2000–2015. Lopuksi havainnoimme, ovatko erot perusopetuksen päättötodistuksen lukuaineiden keskiarvoissa kasvaneet vanhemman koulutuksen ja ammattiaseman mukaan.

Tulostemme mukaan PISA-aineistot sisältävät huomattavaa mittausvirhettä vanhempien koulutuksen osalta. Poikkeuksena on vuosi 2000, jolloin oppilaiden raportoima vanhempien koulutus on hyvin linjassa rekisteritietojen kanssa. Syynä voi olla se, että vuoden 2000 PISA-tutkimuksen taustakyselyssä ei eritelty vanhempien toisen asteen jälkeisiä tutkintoja. Aikaisemman tutkimuksen mukaan lapsilla onkin vaikeuksia erottaa toisen asteen jälkeisiä tutkintoja toisistaan (Schulz 2006).

Vanhempien koulutuksen yhteys nuorten oppimistuloksiin on vahvempaa rekisteri- kuin PISA-aineistoissa. Sen sijaan vanhempien ammat-

tiaseman yhteys oppimistuloksiin on molemmissa aineistoissa yhtä vahvaa. Erot rekisteri- ja PISA-aineistojen tuloksissa johtuvat todennäköisesti PISA-aineistojen vanhempien koulutusta koskevista mittausvirheistä. Kun oppilaat yliarvioivat kyselyssä vanhempiensa koulutuksen, se johtaa vanhempien koulutuksen sekä lasten oppimistulosten yhteyden aliarviointiin. Tulokset ovat linjassa aikaisempien oppimistuloksia sekä perhetaustaa koskevien tutkimusten kanssa (Engzell & Johnson 2015; Jerrim & Micklewright 2014).

Tarkastelimme oppimistulosten eriarvoisuuden muutosta sekä jatkuvilla että luokitelluilla perhetaustan muuttujilla. Jatkuvien muuttujien avulla pystyimme huomioimaan vanhempien koulutus- ja ammattijakaumissa tapahtuneet ajalliset muutokset, kuten korkeakoulutuksen laajentumisen. Luokitelluilla muuttujilla pyrimme havainnollistamaan niiden käyttöön liittyvät valikoitumista koskevat ongelmat.

Jatkuvilla muuttujilla tarkasteltuna oppimistulosten eriarvoisuus ei näytä voimistuneen tilastollisesti merkittävästi PISA- tai rekisteriaineistoissa. Eriarvoisuus on jopa hieman heikentynyt rekisteriaineiston matematiikan arvosanoissa. Analysoimme erikseen myös perusopetuksen päättötodistuksen lukuaineiden keskiarvoa, joka on tärkeä toisen asteen yhteishaussa. Tässäkään tarkastelussa emme havainneet eriarvoisuuden voimistuneen viime vuosina.

Tuloksemme eivät ole linjassa aikaisempien PISA-aineistoihin perustuvien tutkimusten kanssa, joiden mukaan oppimistulosten eriarvoisuus on voimistunut Suomessa viime vuosina (Leino ym. 2019; Salmela-Aro & Chmielewski 2019; Vetenranta ym. 2016). Koska PISA-aineisto sisältää huomattavaa mittausvirhettä vanhempien koulutuksen suhteen, pidämme rekistereillä tehtyjä analyysejä luotettavimpina. Rekisteriaineistoissa ei esiinny vastausharhasta johtuvaa mittausvirhettä.

PISA- ja rekisteriaineistot eivät varsinaisesti mitata samoja asioita. Perusopetuksen päättötodistuksen arvosanat mittaavat opetus suunnitelman mukaista oppimista ja PISA-testit puolestaan myöhemmän elämän kannalta tärkeiden avaintaitojen osaamista. Vaikka kouluarvosanat sekä PISA-tulokset mittaavat jokseenkin eri asioita, on niiden

havaittu korreloivan keskenään. Esimerkiksi Suomessa äidinkielen arvosana sekä PISA-kokeiden lukutaidon pisteet korreloivat vahvasti (Linnakylä & Sulkunen 2002). Tulee kuitenkin huomioida, että opettajien antamiin kouluarvosanoihin voi vaikuttaa päättöarvioinnin kriteerien lisäksi koulun sekä luokan osaamistaso (Harju-Luukkanen ym. 2016; Ouakrim-Soivio 2013).

Tutkimuksessa havaittu vanhemman koulutukseen liittyvä mittausvirhe voi johtua myös otosharhasta. Suomen PISA-tutkimusten otanta vaikuttaa kuitenkin luotettavalta (Nissinen ym. 2018). Lisäksi tällöin otantavirhe olisi systemaattista ja sitä esiintyisi useammassa PISA-tutkimuksessa. Todennäköisempää on, että oppilaat eivät joko tiedä vanhempiensa koulutusta tai sitten he vastaavat tarkoituksella väärin. Aikaisempi tutkimus on osoittanut, että ne oppilaat, joiden oppimistulokset ovat parempia, tietävät tarkemmin vanhempiensa koulutuksen kuin ne oppilaat, joiden oppimistulokset ovat heikompia (Jerrim & Micklewright 2014). Saman suuntainen tulkinta voidaan tehdä tämän tutkimuksen perusteella, koska PISA-aineistoilla eriarvoisuus aliarvioidaan verrattuna rekisteritietoihin. Siten heikommin menestyvät oppilaat yliarvioivat vanhempiensa koulutuksen.

Suomen PISA-tutkimuksissa taustakysely voitaisiin tulevaisuudessa tehdä myös vanhemmille, koska vanhempien vastauksiin sisältyy vähemmän virhettä kuin lasten vastauksiin (Engzell & Johnson 2015). Toinen vaihtoehto – ja ehkä helpompi tapa – olisi pyytää perheiltä lupa PISA-aineistojen yhdistämiseksi rekisteritietoihin, jotta oppimistulosten eroista perhetaustan mukaan saataisiin luotettavampaa tietoa.

Tämä tutkimus osoittaa, että kyselytutkimusten aineistojen luotettavuuden tarkastelua ei voida sivuuttaa, kun tutkimuksessa hyödynnetään lasten täyttämän taustakyselyn tietoja. Suomessa voidaan hyödyntää yhä enemmän laadukkaita koko väestöön perustuvia rekisteriaineistoja. Rekisteripohjaista tutkimustietoa tuleekin hyödyntää enemmän ennen kuin lähdetään tekemään johtopäätöksiä koulutusjärjestelmän uudistamisesta kasvavan eriarvoisuuden torjumiseksi.

Saapunut 3.2.2021
Hyväksytty 21.4.21

KIRJALLISUUS

- Arinen, Pekka & Karjalainen, Tommi (2007) PISA 2006. Ensituloksia. Opetusministeriön julkaisuja 2007:38. Helsinki.
- Avvisati, Francesco & Keslair, François (2014) RE-PEST: Stata module to run estimations with weighted replicate samples and plausible values. Statistical Software Components S457918, Boston College Department of Economics 6.1.2020. (luettu 9.9.2020). <https://ideas.repec.org/c/boc/bocode/s457918.html>
- Barone, Carlo (2006) Cultural capital, ambition and the explanation of inequalities in learning outcomes: A comparative analysis. *Sociology* 40 (6), 1039–1058.
- Bol, Thijs & Witschge, Jacqueline & Van de Werfhorst, Herman & Dronkers, Jaap (2014) Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces* 92 (4), 1545–1572.
- Chmielewski, Anna (2019) The global increase in the socioeconomic achievement gap, 1964 to 2015. *American Sociological Review* 84 (3), 517–544.
- Dunning, David (2011) The Dunning–Kruger effect: On being ignorant of one’s own ignorance. *Advances in experimental social psychology* 44/2011, 247–296.
- Engzell, Per & Johnsson, Jan (2015) Estimating Social and Ethnic Inequality in School Surveys: Biases from Child Misreporting and Parent Nonresponse. *European Sociological Review* 31 (3), 312–325.
- Erola, Jani & Jalonen, Sanni & Lehti, Hannu (2016) Parental education, class and income over early life course and children’s achievement. *Research in Social Stratification and Mobility* 44/2016, 33–43.
- Ganzeboom, Harry & De Graaf, Paul & Treiman, Donald (1992) A standard international socio-economic index of occupational status. *Social science research* 21 (1), 1–56.
- Harju-Luukkanen, Heidi & Vettenranta, Jouni & Ouakrim-Soivio, Najat & Bernelius, Venla (2016) Differences between students’ PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data. *Education Inquiry* 7 (4), 463–479.
- Kivinen, Osmo & Hedman, Juha & (2017) Monistelliset PISA-tulokset ja niiden koulutuspoliittiset tulkinnat. *Politiikka: Valtiotieteellisen yhdistyksen julkaisu* 59 (4), 250–263.
- Härkönen, Juho & Sirniö, Outi (2020) Educational Transitions and Educational Inequality: A Multiple Pathways Sequential Logit Model Analysis of Finnish Birth Cohorts 1960–1985. *European Sociological Review* 2020, 1–20.
- International Labor Office (2012). *International Standard Classification of Occupations. ISCO-08*. Geneva. <https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf> (luettu 9.9.2020)
- Jerrim, John & Micklewright, John (2014) Socio-economic Gradients in Children’s Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background? *European Sociological Review* 30 (6), 766–781.
- Karhula, Aleksi & Sirniö, Outi (2019) Ylisukupolvinen eriarvoisuus Suomessa. Poikkitieteellinen katsaus yhteiskuntatieteellisiin tutkimuksiin. *Yhteiskuntapolitiikka* 84 (5–6), 572–683.
- Karhunen, Hannu & Uusitalo, Roope (2017) 50 vuotta koulutusmahdollisuuksien tasa-arvoa. *Yhteiskuntapolitiikka* 82 (3), 296–303.
- Kreuter, Frauke & Eckmann Stephanie & Maaz Kai & Watermann Rainer (2010) Children’s Reports of Parents’ Education Level: Does it Matter Whom You Ask and What You Ask About? *Survey Research Methods* 4 (3), 127–138.
- Krumpal, Ivar (2013) Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity* 47 (4), 2025–2047.
- Kupari, Pekka & Välijärvi, Jouni & Andersson, Leif & Arffman, Inga & Nissinen, Kari & Puhakka, Eija & Vettenranta, Jouni (2013) PISA 2012 Ensituloksia. Opetus- ja kulttuuriministeriön julkaisu 2013:20. Helsinki.
- Kupari, Pekka (2005) Kotitausta näkyy matematiikan koetuloksissa. Teoksessa Pekka Kupari & Jouni Välijärvi (toim.) Osaaminen kestäväällä pohjalla: PISA 2003 Suomessa. Jyväskylä: Jyväskylän yliopisto, 115–126.
- Lehti, Hannu (2020) The role of a kin in educational and status attainment. *Akateeminen väitöskirja*. Turku: Turun yliopisto.
- Leino, Kaisa & Ahonen, Arto & Hienonen, Ninja & Hiltunen, Jenna & Lintuvuori, Meri & Lähteinen, Suvi & Lämsä, Joni & Nissinen, Kari & Nissinen, Virva & Puhakka, Eija & Pulkkinen, Jonna & Rautopuro, Juhani & Sirén, Marjo & Vainikainen, Mari-Paoliina & Vettenranta, Jouni (2019) PISA 2018 Ensituloksia – Suomi parhaiden joukossa. Opetus- ja kulttuuriministeriön julkaisu 2019:40. Helsinki.
- Lien, Nanna & Friestad, Christine & Klepp, Knut-Inge (2001) Adolescents’ proxy reports of parents’ socioeconomic status: how valid are they? *Journal of Epidemiology and Community Health* 55/2001, 731–737.
- Linnakylä, Pirjo & Sulkunen, Sari (2002) Millainen on Suomalaisten nuorten lukutaito? Teoksessa Jouni Välijärvi & Pirjo Linnakylä (toim.) Tulevaisuuden osaajat – PISA 2000 Suomessa. Jyväskylä: Jyväskylän yliopisto, 9–40.
- Looker, Dianne (1989) Accuracy of proxy reports of parental status characteristics. *Sociology of Education* 62 (4), 257–276.
- Martins, Lurdes & Veiga, Paula (2010) Do inequalities in parents’ education play an important role in PISA students’ mathematics achievement test score disparities? *Economics of Education Review* 29 (6), 1016–1033.
- Nissinen, Kari & Rautopuro, Juhani & Puhakka, Eija (2018) PISA-tutkimuksen metodologiasta. Teoksessa Juhani Rautopuro & Kalle Juuti (toim.) PISA pintaa syvemmältä – PISA 2015 Suomen pää-

- raportti. Suomen kasvatustieteellinen seura: Jyväskylä, 345–378.
- OECD (2010) Strong Performers and Successful Reformers in Education Lessons from PISA for the United States. OECD Publishing. <http://dx.doi.org/10.1787/9789264096660-en> (luettu 9.9.2020)
- Ouakrim-Soivio, Najat (2013) Toimivatko päättöarviointin kriteerit? Oppilaiden saamat arvosanat ja opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina. Raportit ja selvitykset 9/2013. Helsinki: Opetushallitus.
- Salmela-Aro, Katariina & Chmielewski, Anna (2019) Socioeconomic Inequality and Student Outcomes in Finnish Schools. In Louis Volante & Sylke Schnepf & John Jerrim & Don Klinger (eds.) Socioeconomic Inequality and Student Outcomes. Education Policy & Social Inequality, vol 4. Springer: Singapore, 153–168.
- Schulz, Wolfram (2006) Measuring the Socio-economic Background of Students and its Effect on Achievement in PISA 2000 and PISA 2003. Annual Meetings of the American Educational Research Association (AERA), San Francisco.
- Sirin, Selcuk (2005) Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research* 75 (3), 417–453.
- Tilastokeskus (2021) Taika-tutkimusaineistokatalogi. <https://taika.stat.fi/> (luettu 21.1.2021)
- Tourangeau, Roger & Yan, Ting (2007) Sensitive questions in surveys. *Psychological bulletin* 133 (5), 859.
- Vettenranta, Jouni & Välijärvi, Jouni & Ahonen, Arto & Hautamäki, Jarkko & Hiltunen, Jenna & Leino, Kaisa & Lähteinen, Suvi & Nissinen, Kari & Nissinen, Virva & Puhakka, Eija & Rautopuro, Juhani & Vainikainen, Mari-Pauliina (2016) PISA 15 – Ensituloksia. Opetus- ja kulttuuriministeriön julkaisu 2016:41. Helsinki.
- Volante, Louis & Schnepf, Sylke & Jerrim, John & Klinger, Don (eds.) (2019) Socioeconomic inequality and student outcomes: Cross-national trends, policies, and practices. Springer: Singapore.
- Välijärvi, Jouni & Linnakylä Pirjo (toim.) (2002) Tulovaisuuden osajat – PISA 2000 Suomessa. Jyväskylä: Jyväskylän yliopisto.

ENGLISH SUMMARY

Hannu Lehti & Markus Laaninen: Family background and learning outcomes in Finland in the light of PISA and register data (Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa)

Earlier research suggests that inequality in learning outcomes based on family background has increased in Finland in recent years. The evidence is strongly based on school surveys such as PISA data, where information on family background is based on children's reports. However, such surveys are prone to measurement errors due to reporting bias.

We show that PISA data contain a large measurement error in terms of parental education. When this error and changes in the distribution of parental education are taken into account, the results indicate that there has in fact been no increase in inequality in learning outcomes in recent years.

Keywords: inequality of learning outcomes, family background, measurement error, intergenerational effects, PISA data.

KIRJOITTAJAT

Lehti, Hannu, VTT, erikoistutkija, Turun yliopisto (hpleht@utu.fi)

Laaninen, Markus, VTM, tohtorikoulutettava, Turun yliopisto (majlaan@utu.fi)