INVEST

INVEST Working Papers 19/2020

Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa

Hannu Lehti
Markus Laaninen

23.9.2020

The Inequalities, Interventions, and New Welfare State (INVEST) aims at increasing wellbeing of Finnish society during childhood, youth and early adulthood and preventing psychosocial risks compromising such development through innovative interventions. Based on cutting-edge research on the conditions and mechanisms involved at different periods of development, INVEST will evaluate and develop various universal and targeted interventions to improve the efficiency of the current welfare state institutions at critical points of the early life course. INVEST aims at providing a new model for the welfare states that is more equal, better targeted to problem groups, more anticipatory as well as economically and socially sustainable. INVEST is a Flagship project of the Academy of Finland.

Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa

Lehti Hannu, Valtiotieteiden tohtori, erikoistutkija, Turun yliopisto, hpleht@utu.fi
Laaninen Markus, Valtiotieteiden maisteri, Tohtorikoulutettava, Turun yliopisto, majlaan@utu.fi

Yhteyshenkilö: Hannu Lehti, hpleht@utu.fi

Tiivistelmä

Perhetaustan yhteys oppimistuloksiin Suomessa PISA- ja rekisteriaineistojen valossa

Aikaisempien tutkimusten mukaan oppimistulosten eriarvoisuus vanhemman taustan mukaan on kasvanut Suomessa viime vuosina. Tutkimusnäyttö perustuu vahvasti koulukyselyaineistoihin kuten PISA-aineistoon, jossa vanhempien koulutuksen ja ammattiasema on kysytty lapsilta. Tällaiset kyselyaineistot ovat alttiita vastausharhan aiheuttamalle mittausvirheelle. Tässä tutkimuksessa osoitamme, että PISA-aineistot sisältävät suurta mittausvirhettä vanhempien koulutuksen osalta. Siksi vanhemman ammattiasema on luotettavampi mittari, kun halutaan tutkia kotitaustan yhteyttä oppimistuloksiin PISA-aineistolla. Kun mittausvirhe ja negatiivinen valikoituminen vuosien välillä huomioidaan, oppimistulosten eriarvoisuus ei ole kasvanut Suomessa viime vuosina.

Avainsanat: oppimistulosten eriarvoisuus, kotitausta, mittausvirhe, ylisukupolvisuus, PISA-aineisto

Abstract

Family background and learning outcomes in Finland according to PISA- and register data

According to previous studies, the inequality of learning outcomes according to family background has increased in Finland in recent years. The evidence is strongly based on school surveys such as PISA data, where children are asked about their parental educational and occupation status. However, such surveys are prone to measurement error due to reporting bias. We show that the PISA data contain a large measurement error in terms of parental education. Therefore, parental occupational status is a more reliable indicator of family background when association between family background and learning outcomes are studied with PISA data. When the measurement error and the negative selection are taken into account, the results show that the inequality of learning outcomes has not increased in Finland in recent years.

Key words: Inequality of learning outcomes, family background, measurement error, intergenerational effects, PISA data

Johdanto

Perhetausta on yhteydessä oppimistuloksiin ja koulumenestykseen. Korkeammin koulutettujen vanhempien lapset menestyvät keskimäärin paremmin opinnoissa ja kouluttautuvat pidemmälle kuin matalammin koulutettujen lapset (Barone 2006; Breen & Jonsson 2005; Volante ym. 2019; Chmielewski 2019; Bol ym. 2014). Suomessa perhetaustan yhteys oppimistuloksiin on kuitenkin ollut kansainvälisesti vertaillen heikkoa ja Suomea onkin pidetty oppimistulosten tasa-arvon mallimaana (Arinen & Karjalainen 2007; Chmielewski 2019; Kupari 2005; Barone 2006; Linnakylä & Sulkunen 2002). Kuitenkin viimeaikaisten kansainvälisten oppimisen arviointien, kuten PISA-, TIMSS- ja PIRLS-tutkimusten, mukaan lasten oppimistulosten erot vanhempien taustan osalta olisivat kasvaneet viime vuosina (Leino ym. 2019; Salmela-Aro & Chmielewski 2019; Vettenranta ym. 2016). Tutkimustulokset on tuotu esiin mediassa narratiivina, jossa maalataan kuvaa oppimisen eriarvoisuuden kasvusta ja tasa-arvon romuttumisesta. Huoli oppimistulosten eriarvoisuuden kasvusta on lisääntynyt.

Koska tieto sosioekonomisen taustan yhteyksistä oppimiseen kertoo Suomen koulutusjärjestelmän tasa-arvoisuudesta, on tärkeää tutkia luotettavasti, ovatko perhetaustan mukaiset erot oppimisessa todella kasvaneet. Esimerkiksi yhtenä syynä Suomen korkeisiin PISA-tuloksiin on aikaisemmin pidetty sitä, että oppimistulokset eroavat vähän perhetaustan mukaan (OECD 2010). Pienet erot oppilaiden välillä näyttäisivät olevan Suomen etu kansainvälisissä oppimisen arvioinneissa.

Oppimistulosten eriarvoisuuden tarkastelussa on tärkeää, että lasten perhetausta on mitattu luotettavasti. Mahdollinen mittausvirhe perhetaustassa voi vääristää tuloksia ja johtaa vääriin tulkintoihin koulutusjärjestelmän lisääntyneestä epätasa-arvosta (kts. Engzell & Johnsson 2015). Suomen PISA-tutkimuksissa tiedot vanhempien koulutuksesta sekä ammattiasemasta pohjautuvat lasten raportoihin tietoihin. Siksi vanhemman koulutuksen ja ammattiaseman osalta aineistossa saattaa esiintyä vastausharhasta johtuvaa mittausvirhettä. Tällä tarkoitetaan sitä, että lapset luokittelevat väärin oman vanhempansa koulutuksen ja ammattiaseman. Mikäli tutkimuksissa käytetään lapsilta saatuja vastauksia perheen sosioekonomisesta asemasta, niiden luotettavuutta pitää arvioida. Aikaisemmat PISA-aineistoilla tehdyt tutkimukset perhetaustan yhteydestä oppimiseen eivät ole ottaneet huomioon mahdollista mittausvirhettä, johon kiinnitämme tässä tutkimuksessa huomiota hyödyntämällä toisen asteen hakurekisteriä niiden ikäluokkien osalta, jotka osallistuivat PISA-tutkimuksiin.

Tarkastelemme kuinka paljon PISA-aineisto sisältää mittausvirhettä äidin ja isän koulutuksen sekä ammattiaseman osalta verrattuna toisen asteen hakurekisteriaineiston tietoihin koko ikäluokasta. Tutkimme, onko oppimistulosten eriarvoisuudessa vanhemman koulutuksen ja ammattiaseman mukaan eroja PISA- ja rekisteriaineistolla mitattuna. Lisäksi tutkimme sekä PISA- että rekisteriaineistolla, ovatko oppimistulosten erot todella kasvaneet perhetaustan mukaan viime vuosina. Tarkastelemme PISA-aineistolla oppimistuloksia lukutaidossa ja matematiikassa. Rekisteriaineistolla tarkastelemme peruskoulun päästötodistuksen äidinkielen ja matematiikan arvosanoja sekä lukuaineiden arvosanojen keskiarvoja.

Vastausharha koulukyselyissä

Vastausharha on kattotermi kyselyissä esiintyvälle mittausvirheelle. Vastausharha syntyy, kun vastaaja ei vastaa totuuden mukaisesti kyselyyn, koska vastausta voidaan pitää yhteiskunnallisesti epäsuotavana (Krumpal 2013). Otokseen voi syntyä vastausharhaa myös siksi, että yksilöillä ei ole tietoa oikeasta vastauksesta. He saattavat silti vastata, koska vastaaminen on psykologisesti suotuisampaa kuin vastaamatta jättäminen. (Tourangeau & Yan 2007)

Kysyttäessä vanhemman koulutusta tai ammattiasemaa vastausharha voikin syntyä siitä, että lapset eivät tiedä tarkkaan vanhemman koulutustasoa tai ammattiasemaa. Tällöin oppilaat soveltavat heuristista arviointia, vastaavat sosiaalisesti hyväksytyllä tavalla tai jättävät vastaamatta kysymykseen (Schulz 2006). Koska yksilöillä on taipumus enemmän yliarvioida kuin aliarvioida omaa taustaansa (Dunning 2011), lapset todennäköisesti yliarvioivat vanhempien koulutuksen ja ammattiaseman. Toisaalta lapset voivat tietää paremmin vanhemman 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). Lapset tietävät vanhempien koulutuksen paremmin, mikäli vanhemmilla on perus- tai korkea-asteen tutkinto, mutta muut koulutusasteet ovat vaikeampia tietää (Kreuter ym. 2010; Schulz 2006). Vastausharha vanhempien koulutuksessa voi olla seurausta monimutkaisista koulutusjärjestelmistä (Schulz 2006).

Vanhempien koulutuksen sekä lasten oppimistulosten yhteys on vahvempi, kun lasten ilmoittamien tietojen sijaan hyödynnetään vanhempien raportoimaa tietoa koulutuksesta. Sen sijaan sekä vanhempien että oppilaiden ilmoittaman ammattiaseman ja oppimistulosten yhteyden vahvuus on samansuuruista. (Engzell & Johnsson 2015; Jerrim & Micklewright 2014) Kysely- ja rekisteriaineistoja vertailevassa tutkimuksessa havaittiinkin, että vanhemmat raportoivat koulutuksensa lapsiaan luotettavammin (Engzell & Johnsson 2015). Lasten raportoiman koulutuksen käyttäminen perhetaustan mittarina voi johtaa perhetaustan merkityksen aliarvioimiseen oppimistuloksia tarkasteltaessa, mutta ammattiasema sen sijaan vaikuttaa luotettavalta perhetaustan mittarilta. Aikaisemman tutkimuksen mukaan sekä äidin että isän koulutuksen ja ammattiaseman vastausharhassa ei ole ollut eroja (Jerrim & Micklewright 2014; Schulz 2006; Lien ym. 2001).

Edellä esitetyt tutkimukset pystyivät hyödyntämään vanhempien vastauksia lasten vastausten lisäksi. Vaikka vanhempien raportoimat tiedot ovat yleensä luotettavampia, ne sisältävät myös mittausvirhettä (Engzell & Johnsson 2015; Jerrim & Micklewright 2014; Kreuter ym. 2010).

Kansainvälisissä laajoissa oppimistulosarvioinneissa, kuten PISA-tutkimuksissa, oppilaat täyttävät taustakyselyn, jossa selvitetään esimerkiksi vanhempien ammattia sekä koulutusta. Osassa PISA-tutkimuksen maista kysely tehdään myös oppilaiden vanhemmille, mutta Suomessa vanhempainkyselyä ei ole toteutettu.

Perhetaustan yhteys nuorten oppimistuloksiin Suomessa

Aikaisempi tutkimus on osoittanut selvästi, että vanhemman taustalla on yhteys oppimistuloksiin kaikissa yhteiskunnissa mutta yhteyden voimakkuus vaihtelee yhteiskuntien välillä (Bol ym. 2014; Chmielewski 2019; Barone 2006). Tutkimusten mukaan Suomessa vanhemman sosioekonomisella asemalla ja koulutuksella on vähemmän merkitystä lasten oppimistuloksiin ja koulutukseen verrattuna muihin maihin (Chmielewski 2019; Lehti 2020).

Vuosien 2000 ja 2003 PISA-tutkimusten Suomen raporttien mukaan sekä vanhempien ammatillinen asema, että koulutus olivat positiivisessa yhteydessä lasten oppimistuloksiin. Perhetaustan yhteys PISA-tuloksiin oli kuitenkin Suomessa kansainvälisesti vertaillen heikkoa. (Linnakylä & Sulkunen 2002; Kupari 2005) Vuodesta 2006 alkaen perhetaustaa on mitattu PISA-aineistoihin valmiiksi lasketulla ESCS-indeksillä (Economic Social and Cultural Status Index), joka yhdistää tiedot vanhempien koulutuksesta, ammatillisesta asemasta sekä perheen omistamista asioista. Myös vuodesta 2006 lähtien tulokset ovat samansuuntaisia. Korkeampi perhetausta on yhteydessä parempiin tuloksiin PISA-testeissä riippumatta siitä, mitataanko osaamista lukutaidossa, matematiikassa vai luonnontieteessä (Arinen & Karjalainen 2007; Sulkunen ym. 2010; Kupari ym. 2013).

Vuoden 2015 Suomen PISA-tulosten mukaan perhetaustan merkitys oppimiseen on kasvanut ja siten kuilu eri perhetaustoista tulevien oppilaiden välillä syventynyt. Lisäksi perhetaustan yhteys PISA-tuloksiin oli kansainvälisesti vertaillen enää keskitasoa. (Vettenranta ym. 2016) Myös vuoden 2018 Suomen PISA-tulosten mukaan perhetaustan mukaiset erot oppimistuloksissa ovat jatkaneet kasvuaan (Leino ym. 2019).

Useita kansainvälisiä oppimistulosarviointeja tarkastelevan tutkimuksen mukaan vanhempien koulutuksen mukainen oppimistulosten eriarvoisuus on kasvanut Suomessa viimeisten vuosikymmenien aikana. Tutkimuksessa tarkasteltiin vanhempien koulutuksen mukaista 90. ja 10. persentiilin välistä kuilua lasten oppimistuloksissa. Vuosien 1984 ja 2005 välillä syntyneiden ikäluokkien kohdalla vanhempien koulutuksen mukainen oppimistulosten eriarvoisuus näyttää kasvaneen lineaarisesti. Tulos pohjautuu kuitenkin vahvasti vuosien 2000, 2003, 2006, 2009, 2012 ja 2015 PISA-tutkimuksiin. (Salmela-Aro & Chmielewski 2019) Ilman PISA-aineistoja vanhempien koulutuksen mukainen oppimistulosten eriarvoisuus ei näytä kasvaneen kyseisessä tutkimuksessa. Lisäksi verrattuna useisiin muihin länsimaihin eriarvoisuuden kasvu oppimisessa näyttäytyy Suomessa vähäiseltä (Chmielewski 2019).

Ongelmana edellä mainituissa tutkimuksissa on se, että perhetaustaa on tarkasteltu kategorisella muuttujalla (Salmela-Aro & Chmielewski 2019; Leino ym. 2019). Kun esimerkiksi vanhemman koulutusta mitataan kategorisella muuttujalla, eriarvoisuuden kasvu saattaa selittyä negatiivisella valikoitumisella, koska vanhempien koulutuksen jakauma ei ole muuttumaton yli ajan. Koulutuksen laajenemisen johdosta nuorempien PISA-kohorttien vanhemmat ovat koulutetumpia kuin vanhempien kohorttien vanhemmat. Tämän takia matalasti koulutettujen joukko vähenee ja valikoituu yli ajan (ks. Karhunen & Uusitalo 2017). Kun verrataan esimerkiksi korkea- ja matalasti koulutettujen perheiden lasten oppimistuloksia vuosien välillä, voi eriarvoisuuden kasvu selittyä tällä negatiivisellä valikoitumisella.

Tämän vuoksi esimerkiksi Salmela-Aro ja Chmielewski (2019) käyttävät persentiileihin perustuvia ”suorituskuiluja”, koska niiden jakauma pysyy vuosien välillä samana. Ongelmana kuitenkin on se, että persentiileihin perustuva suorituskuilu on tarkoitettu muodostettavaksi jatkuvista muuttujista, esimerkiksi tuloista. Jos vanhemman koulutusta ei ole mitattu jatkuvana vaan kategorisena muuttujana, kuten Suomen PISA-aineistossa, persentiileihin perustuva kuilu ei välttämättä toteudu. Salmela-Aro ja Chmielewski (2019) toteavatkin artikkelissaan, että esimerkiksi PISA 2015 -aineistossa 90. ja 50. persentiileissä vanhemman koulutus jakaantui samaan luokkaan (ISCED 5) eikä sitä voinut käyttää heidän tutkimuksessaan suorituskuilun mittarina.

Persentiileihin perustuva luokittelu on ongelmallista myös siksi, että Suomen PISA-aineistoissa koulutusluokkien prosenttijakaumat eivät pysy samanlaisena vuosien välillä. Täten persentiileihin perustuvaa suorituskuilua ei voida luotettavasti niin, että esimerkiksi 90/10 jakauma pysyisi samana jokaisen vuoden välillä. Havainnollistaaksemme tätä ongelmaa olemme muodostaneet Suomen PISA-aineistoille desiileihin perustuvan luokittelun vuosittain. Liitetaulukko 1 näyttää, että esimerkiksi vuonna 2003 alimpaan desiiliin kuuluu 31 prosenttia tapauksista kymmenen prosentin sijaan. Lisäksi se, kuinka paljon ylempi ja alempi desiili sisältävät tapauksia prosentuaalisesti, vaihtelee vuosien välillä. Luotettavinta olisi käyttää vanhemman koulutusta jatkuvana muuttujana muuttamalla se koulutusvuosiksi. Jatkuva muuttuja ei ole herkkä valikoitumiselle ja sitä voidaan käyttää yli ajan (ks. Härkönen & Sirniö 2020).

Aikaisemmat tutkimukset eivät ole myöskään huomioineet, miten kotitausta on järkevää operationalisoida. Usein vanhemman sosioekonomista taustaa mitataan summamuuttujalla, joihin on yhdistetty koulutuksen ja ammattiaseman lisäksi muita kodin muita taustamuuttujia (ks. esim. Arinen & Karjalainen 2007). Summamuuttujien ongelmana voidaan pitää, sitä etteivät ne kerro, mikä tekijä vanhemman taustassa on varsinaisesti yhteydessä lasten oppimistuloksiin. Summamuuttujat saattavat myös lisätä mittausvirhettä, koska kaikki summamuuttujan osamuuttujat voivat sisältää mittausvirhettä. Esimerkiksi vanhempien ja lasten ilmoittaman kodin kirjojen määrän yhteys on heikko ja sen on näytetty sisältävän huomattavaa raportointiharhaa (Jerrim & Micklewright 2014; Engzell 2018). Summamuuttujien tulkinta saattaa muodostua myös hankalaksi.

Tutkimusasetelma

Tarkastelemme ensin, kuinka yhteneväisiä äidin ja isän koulutuksen sekä ammattiaseman jakaumat ovat PISA- ja rekisteriaineistoissa. Arvioimme PISA-tutkimusten taustakyselyn luotettavuutta vanhempien koulutuksesta ja ammattiasemasta verrattuna koko ikäluokkaan. Ensimmäinen tutkimuskysymyksemme on:

1. Sisältääkö PISA-aineisto mittausvirhettä vanhemman koulutuksen ja ammattiaseman mukaan ja kuinka iso mittausvirhe on?

Koska emme voi yhdistää PISA-aineistoja rekistereihin, on mahdollista, että PISA-aineiston mittausvirhe ei johdu vastausharhasta vaan siitä, että otos on valikoitunut vanhemman taustan mukaan (otosharha). Palaamme keskusteluun otosharhan mahdollisuudesta artikkelin lopussa. Aikaisempaan tutkimukseen perustuen oletamme, että vanhemman

ammattiasema sisältää vähemmän mittausvirhettä kuin vanhemman koulutus. Oletamme myös, että äidin sekä isän koulutus ja ammattiasema arvioidaan yhtä hyvin.

Seuraavaksi tutkimme oppimistulosten eriarvoisuutta vanhemman koulutuksen ja ammattiaseman mukaan. Tarkastelemme PISA-aineistoilla matematiikan ja lukutaidon oppimistuloksia sekä rekisteriaineistolla peruskoulun päästötodistuksen matematiikan ja äidinkielen arvosanoja. Lopuksi havainnoimme rekisteriaineistolla peruskoulun päästötodistuksen lukuaineiden keskiarvoa. Tutkimuskysymyksemme ovat:

2. Millaiselta vanhemman koulutuksella ja ammattiasemalla mitattu oppimistulosten eriarvoisuus näyttäytyy PISA- ja rekisteriaineistolla mitattuna?
3. Ovatko oppimistulosten erot vanhemman koulutuksen ja ammattiaseman mukaan kasvaneet Suomessa?

Koska käytämme PISA-aineistoja eri mittausvuosilta ja vastaavien ikäluokkien päästötodistuksen arvosanoja rekistereistä, pystymme havaitsemaan molemmilla aineistoilla, ovatko erot oppimistuloksissa kasvaneet vanhemman taustan mukaan. Emme kuitenkaan väitä, että PISA-testin tulokset ja peruskoulun päästötodistuksen arvosanat mittaavat samoja asioita. Tässä tutkimuksessa sitä ei voida myöskään testata, koska se vaatisi PISA-tulosten yhdistämistä rekistereihin. Aikaisemman tutkimuksen mukaan kouluarvosanojen ja PISA-testin tulosten on havaittu Suomessa korreloivan keskenään. Esimerkiksi tutkimuksessa, jossa äidinkielen arvosana kysyttiin oppilailta, havaittiin, että sen yhteys on vahva PISA-kokeiden lukutaidon pisteiden kanssa (Linnakylä & Sulkunen 2002).

Koska aikaisempi tutkimus ei ole huomioinut muuttujien määrittelyä ja mittarien rakentamista tutkimme myös, miten vanhempien mittarin operationalisointi vaikuttaa vanhemman muuttujan yhteyden voimakkuuteen vuosien välillä. Teemme kaikki mallit sekä jatkuvilla että luokitelluilla muuttujilla. Tarkastelemme myös eroavatko äitien ja isien yhteydet lasten oppimistuloksiin (ks. Liitekuviot 1-4).

Aineistot

Käytämme tutkimuksessa PISA-aineistoja vuosilta 2000, 2003, 2006, 2009, 2012 ja 2015 sekä samoilta vuosilta toisen asteen hakurekisterin tietoja niistä ikäluokista (1984, 1986, 1990, 1993, 1996, 1999), jotka osallistuivat PISA-tutkimukseen. Haimme näille samoilta ikäluokille toisen asteen rekistereistä matematiikan ja äidinkielen arvosanat sekä päästötodistuksen keskiarvon.

Rekisteriaineisto sisältää kaikki Suomessa asuvat henkilöt ja heidän lapsensa vuosilta 1987-2016. Täten saamme rekisteritiedot lasten vanhempien osalta ja voimme yhdistää vanhemmat lapsiinsa. Rekisteriaineiston tiedot vanhempien koulutuksen osalta otettiin samana vuonna, kun lapset osallistuivat PISA-testiin. Rekisteriaineiston vanhemman koulutus siis vastaa PISA-aineistojen vuosia täydellisesti.

Rekisteriaineisto sisältää vanhemman ammatin osalta tietyiltä vuosilta puuttuvia arvoja. Tämän takia emme saaneet mitattua ammattiasemaa kaikilta lapsilta täysin samana vuonna jolloin lapsi osallistui PISA-testiin. Jos ammattiasemaa ei löytynyt juuri sinä vuonna, kun

lapsi täytti 16-vuotta ja osallistui PISA-testiin, otimme ammattiaseman vuoden nuorempana aina 13-vuotiaaksi asti. Kuitenkin noin 80 prosentilla ammattiasema löytyi juuri sinä vuonna, kun lapset osallistuivat PISA-testiin. Siten ammattiaseman vastaavuus PISA-aineiston kanssa on luotettava. Suomalaisella rekisteriaineistoilla tehty tutkimus on myös osoittanut, että vanhemman ammattiasema ei muutu enää merkittävästi tämän ikäisillä lapsilla (Erola ym. 2016).

Koska toisen asteen rekistereissä oli paljon puuttuvia arvoja äidinkielen osalta vuonna 2009 ja matematiikan vuonna 2012, jätämme nämä vuodet pois analyyseistä, kun tarkastelemme vanhemman taustan yhteyttä oppimistuloksiin. Koska rekisteriaineisto ei sisällä mittausvirhettä, joka liittyy kyselyaineistojen sisältämään vastaus- ja otosarhaan, sen antamia tuloksia voidaan pitää luotettavampina kuin PISA-aineiston tuloksia.

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 suoritettu kolmen vuoden välein vuodesta 2000 lähtien. PISA-tutkimusten kohdejoukko sisältää myös oppilaita, jotka ovat kohdekohorttia vuoden nuorempia (tammikuussa syntyneet). Jotta PISA- ja rekisteriaineistot vastaavat samaa kohorttia, jätimme kohdekohorttia nuoremmat pois analyyseistä (noin 4-8 prosenttia riippuen aineistoista). Vuoden 2000 PISA-aineistossa vain puolet osallistujista vastasi matematiikan testeihin (Välijärvi & Linnakylä 2002), minkä vuoksi keskitymme kyseisen vuoden osalta vain lukutaidon tuloksiin.

Muuttajat ja menetelmät

Selittävinä muuttujina käytämme sekä PISA- että rekisteriaineistoilla vanhempien koulutusta ja ammattiasemaa. Luotettavuuden takaamiseksi muokkasimme muuttujat aineistojen välillä vastaamaan täysin toisiaan. Sekä PISA- että rekisteriaineistoissa vanhempien koulutusmuuttujaa tarkastellaan ISCED 2011 (ennen vuotta 2011 ISCED 1997) luokitukseen pohjalta ja vanhemman ammattiasemaa ISCO-08-luokituksella (ennen vuotta 2008 ISCO-88) sekä ISEI-indeksillä.

Kuvailevassa osiossa, jossa tarkastelemme vanhemman koulutuksen ja ammattiaseman mittausvirhettä PISA-aineistossa, luokittelimme vanhemman koulutuksen kolmeen luokkaan. Ammattiaseman luokittelimme ISCO-luokituksen pohjalta kuuteen eri luokkaan. Teimme näin selkeyden vuoksi. Liian moniluokkainen ammattiasema-muuttuja tekisi tulosten tulkinnasta monimutkaista. Alkuperäinen ISCO-08 luokitus sisältää yhteensä 436 ammattiluokkaa ja kymmenen pääluokkaa. Kuusi pääluokkaa säilytti hyvin alkuperäisen kymmenen luokkaisen ISCO-luokituksen taitoihin perustuvan ammattirakenteen. (International Labor Office 2012). Luokittelimme vanhemman ammattiaseman seuraavasti: 1. Johtajat (ISCO 1), 2. Asiantuntijat (ISCO 2-3), 3. Toimisto- ja palvelutyöntekijät (ISCO 4-5), 4. Maanviljelijät (ISCO 6), 5. Manuaaliset alat (ISCO 7-8) ja 6. Muut työntekijät (ISCO 9). Vanhemman koulutuksen luokittelimme kolmeen luokkaan: perusaste (ISCED 2 tai matalampi), toinen aste (ISCED 3-4) ja korkea-aste (ISCED 5B tai korkeampi).

Mittausvirhettä tarkastelemme vertaamalla PISA-kyselyn vanhemman koulutuksen ja ammattiaseman jakaumaa koko ikäluokan jakaumaan rekisteriaineistossa. Esitämme tulokset kuviaina, jotka näyttävät, kuinka monta prosenttiyksikköä vanhemman koulutuksen ja ammattiaseman jakaumat eroavat PISA-aineistossa verrattuna rekisteriaineistoon. Liitteissä esitämme mittausvuosittain vanhemman koulutuksen (liitetaulukot 2a-2c) ja ammattiaseman (liitetaulukot 3a-3c) prosenttijakaumat sekä tapausten määrät, joista mittausvirhe laskettiin.

Kun tarkastelemme vanhemman koulutuksen ja ammattiaseman yhteyttä lasten oppimistuloksiin, käytämme sekä jatkuvia että kaksiluokkaisia muuttujia. Vanhemman ammattiasemaa mittaamme yleisesti sosiologiassa käytetyllä ISEI-indeksillä, joka mittaa vanhemman sosioekonomista asemaa jatkuvana muuttujana (Ganzeboom ym. 1992). ISEI-indeksi vaihtelee 16 ja 90 pisteen välillä. Vanhemman koulutuksen vuosissa teimme ISCED -luokista seuraavasti: 9 vuotta (peruskoulu tai alempi), 12 vuotta (toinen aste ja erikoisammattitutkinto), 14,5 vuotta (alin korkea-aste) ja 16,5 vuotta (korkea-aste tai ylempi). Lopuksi standardoimme kaikki vanhempien jatkuvat muuttujat

Jatkuvien muuttujien lisäksi tarkastelemme oppimistulosten suorituskuilua kaksiluokkaisilla vanhemman koulutus ja ammattiasema muuttujilla. Vanhemman koulutusta mitataan kaksiluokkaisella dummy-muuttujalla, joka kertoo, kuinka suuri ero oppimistuloksissa on korkea-asteen ja toisen asteen tai vähemmän koulutusta saaneiden vanhempien lapsilla. Vanhemman ammattiaseman osalta vertaamme osaamiskuilua ISCO 1-3 luokkien (asiantuntijat, erikoisasiantuntijat ja johtajat) ja sitä alempien välillä (ISCO 4-9 ja 0).

Käytämme analyyseissa vanhempien (isän tai äidin) korkeinta koulutusta ja ammattiasemaa. Raportoimme kuitenkin tulokset liitteissä erikseen äideille ja isille (ks. Liitekuviot 1-4).

Rekisteriaineistolla analysoimme peruskoulun päättötodistuksen matematiikan, äidinkielen sekä lukuaineiden keskiarvoa. PISA-aineistoilla tehdyillä analyyseilla tarkastelemme lukutaidon ja matematiikan testeistä johdettuina PV-lukuarvoja (Plausible values) (kts. Nissinen ym. 2018). Kaikissa PISA-aineistojen analyyseissa käytämme oppilaspainoja sekä koetulosten keskivirheiden laskentaan tarkoitettuja replikaatiopainoja hyödyntämällä STATA-ohjelmalle tehtyä Repest-pakettia (Avvisati & Keslair 2014).

Koska standardoimme vanhemman koulutuksen vuosissa ja ISEI-indeksin, esitämme tulokset vanhemman jatkuvien muuttujien ja lasten oppimistuloksien välillä korrelaatiokertoimena (beta-estimaatti). 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 (ks. Björklund & Jäntti 1997)

Kategoriset muuttujat näyttävät osaamiskuilun standardin poikkeamana korkeakoulutettujen ja matalasti koulutettujen vanhempien lasten välillä sekä korkean ammattiaseman ja muiden ammattiasemien välillä. Kuten aiemmin totesimme, aiemmin varsinkin vanhempien koulutustaustaa on mitattu luokitelluilla muuttujilla (Salmela-Aro & Chmielewski 2019). Tämä voi johtaa valikoitumisen ongelmaan koulutusjakauman muuttuessa ajassa. Haluammekin tässä tutkimuksessa havainnollistaa tätä ongelmaa. Analyysimenetelmänä käytämme lineaarista regressiota (OLS).

Vanhempien koulutus ja ammattiasema PISA- ja rekisteriaineistoissa

Ensimmäisenä tutkimustehtävänä selvitämme, sisältävätkö PISA-aineistot mittausvirhettä vanhemman koulutuksen ja ammattiaseman mukaan ja kuinka suurta mittausvirhe on. Tarkastelemme ensin äidin ja isän koulutuksen prosenttijakaumien eroja PISA- ja rekisteriaineistoissa.

Kuviosta 1 havaitaan, että vuosien 2003-2015 PISA-aineistoissa oppilaat arvioivat sekä äidin että isän koulutuksen korkeammaksi verrattuna koko ikäluokkaan rekisteriaineistossa. Erot johtuvat pääosin korkeakoulutettujen vanhempien liian suuresta osuudesta ja perusasteen sekä etenkin toisen asteen tutkinnon suorittaneiden vanhempien liian pienestä osuudesta. Poikkeuksena on vuosi 2000, joka näyttää kaikkein tasapainoisimmalta verrattuna rekisteriaineistoon. Tuolloin peruskoulutettujen vanhempien osuus on yliarvioitu ja puolestaan korkeakoulutettujen vanhempien osuus aliarvioitu. Vuonna 2000 toisen asteen tutkinnon suorittaneiden vanhempien osuus vastasi hyvin rekisteriaineiston tietoja. Tämä johtuu todennäköisesti siitä, että vuonna 2000 vanhempien koulutus kysyttiin karkeammalla luokittelulla kuin muina vuosina. Oppilailta kysyttiin, ovatko vanhemmat suorittaneet perusasteen, toisen asteen tai toisen asteen jälkeisen tutkinnon, kun muina vuosina annettiin useampia vastausvaihtoehtoja joista valita.

Kuvio 1. Vanhempien koulutuksen erot prosenttiyksikköinä PISA- ja rekisteriaineistoissa 2000-2015.

Kuvion 1 perusteella voidaan todeta, että PISA-aineistot sisältävät suurta mittausvirhettä vanhemman koulutuksen osalta. Vuosina 2003-2015 mittausvirhe korkeakoulutetuilla

vanhemmilla on noin 19 - 32 prosenttiyksikköä. Äidin koulutus sisältää vähemmän mittausrvirhettä mutta erot eivät ole suuria äitien ja isien koulutuksen mittausrvirheen välillä. Tulos asettaa kyseenalaiseksi PISA-kyselyllä tehdyt aikaisemmat tutkimukset oppimistulosten eroista vanhemman koulutuksen mukaan varsinkin, kun mittausrvirhe ei ole systemaattista vuosien välillä.

Kuviosta 2 havaitaan, että PISA-kyselyiden äidin ja isän ammattiasema vastaa melko hyvin rekisteriaineiston tietoja. Suurimmat erot isän ammattiaseman suhteen johtuvat johtajien liian suurena osuutena sekä asiantuntijoiden liian pienenä osuutena (ks. myös liitetaulukko 3c). Mittausrvirhe on noin 10 prosenttiyksikön luokkaa (vuosina 2000, 2003, 2009 ja 2012) ollen selvästi pienempi verrattuna isän koulutukseen. Jos kyseessä on lasten raportointiharha, lapset sekoittavat asiantuntijat johtajiin. Tätä voidaan pitää pienenä virheenä, koska luokat ovat lähellä toisiaan. Vuonna 2015 lapset lisäksi yliarvioivat manuaaliset alat ja puolestaan aliarvioivat asiantuntijatehtävät, mutta mittausrvirhe jää pieneksi ollen noin kuusi prosenttiyksikköä. Äitien ammattien osalta johtajien ja asiantuntijoiden osuudet yliarvioituvat sekä toimisto- ja palvelutyöntekijöiden osuudet alirvioituvat varsin selvästi vuosina 2000 ja 2009. Vuosina 2003, 2006, 2012 ja 2015 PISA-aineistot vastaavat kuitenkin hyvin rekisteritietoja äidin ammattiaseman suhteen. Kuvioiden 1 ja 2 perusteella näyttää siltä, että mittausrvirhettä on vähemmän vanhemman, ja varsinkin äitien, ammatin kuin koulutuksen suhteen. Siten tulokset vastaavat aikaisempia tutkimuksia.

Kuvio 2. Vanhempien ammattiaseman erot prosenttiyksikköinä PISA- ja rekisteriaineistoissa 2000-2015.

Vanhempien koulutuksen ja ammattiaseman yhteys lasten oppimistuloksiin

Seuraavaksi tarkastelemme oppimistulosten eriarvoisuutta vanhemman koulutuksen ja ammattiaseman mukaan PISA- ja rekisteriaineistoissa. Lisäksi tutkimme ovatko erot oppimistuloksissa kasvaneet vuosina 2000-2015.

Kuviossa 3 esitetään vanhemman koulutuksen sekä ammattiaseman yhteys lukutaitoon (PISA) sekä äidinkielen arvosanoihin (rekisterit). Kuviosta nähdään, että rekisteriaineistossa vanhempien koulutuksen sekä oppimistulosten yhteys on vahvempaa kuin PISA-aineistoissa riippumatta siitä mitataanko vanhempien koulutus luokiteltuna tai jatkuvana. Sen sijaan vanhempien ammattiaseman sekä oppimistulosten yhteys on lähes yhtä suurta. Erot johtuvat todennäköisesti PISA-aineiston sisältämästä mittausvirheestä vanhemman koulutuksen osalta (ks. Kuvio 1). Tämä mittausvirhe voi johtua raportointiharhasta. Kun lapset yliarvioivat vanhemman koulutusta, sekoittuu korkeammin koulutettujen vanhempien

lasten ryhmään matalammin koulutettujen lapsia, joiden PISA-testien tulokset ovat keskimäärin alhaisempia. Näin ollen PISA-aineisto aliarvioi vanhemman koulutuksen yhteyttä lasten oppimistuloksiin. Kuviosta nähdäänkin, että rekisteriaineistossa erot ovat suuremmat vanhemman koulutuksen mukaan.

Kuviossa 4 esitetään vanhemman koulutuksen ja ammattiaseman yhteys matematiikan osalta PISA-aineistolla ja rekistereillä. Tulokset ovat samansuuntaisia kuin lukutaidon ja äidinkielen tuloksia tarkasteltaessa. Vanhempien koulutuksen sekä matematiikan oppimistulosten yhteys on vahvempaa rekisteriaineistossa kuin PISA-aineistoissa. Poikkeuksena on vuosi 2015 jolloin yhteys on yhtä suurta. Tämä johtuu siitä, että rekisteriaineistolla erot vanhemman taustan mukaan ovat vähentyneet, kun taas PISA-aineistossa kasvaneet. Kun tarkastellaan, vanhemman ammattiaseman yhteyttä matematiikan osaamiseen huomataan, että PISA- että rekisteriaineistoissa yhteys on lähes yhtä suurta (ks. Paneeli C ja D). Poikkeuksena on vuosi 2015 jolloin PISA-aineistossa vanhemman ammattiaseman yhteys matematiikan oppimistuloksiin on selvästi rekisteriaineistoa vahvempaa. Ero liittyy todennäköisesti siihen, että vuonna 2015 oppilaat aliarvioivat isän ammattiaseman PISA-aineistossa mutta ei äidin, vaikka toisina vuosina isän ammattiasema yliarvioidaan (ks. kuvio 2). Koska oppilaat arvioivat äidin ammattiaseman paremmin kuin isän, antaa äidin ammattiasema luotettavammat tulokset (ks. Kuvio 2 ja liitetaulukko 3b). Kun yhteyttä matematiikkaan tarkastellaan pelkästään äidin ammattiaseman suhteen, on yhteyden harha pienempää PISA- ja rekisteriaineistojen välillä (Liitekuvio 3).

Kuvioista 3 ja 4 voidaan havaita, että PISA-aineiston mukaan oppimistulosten eriarvoisuus on kasvanut lukutaidossa ja matematiikassa, kun tarkastellaan tuloksia luokiteltujen muuttujien osalta (Paneeli A ja C). Muutos ei ole kuitenkaan tilastollisesti merkitsevä ammattiaseman osalta. Tulokset ovat yhteneväisiä aikaisemman tutkimuksen kanssa, jonka mukaan oppimistulosten eriarvoisuus olisi kasvanut Suomessa (Salmela-Aro & Chmielewski 2019).

Koska luokiteltu muuttuja on altis valikoitumiselle, tulos ei välttämättä kerro todellisesta eriarvoisuuden kasvusta, vaan siitä, että vanhemman koulutuksen jakauma muuttuu yli ajan ja matalasti koulutettujen ryhmä valikoituu negatiivisesti. Matematiikan osalta

rekisteriaineiston analyysit osoittavat, että erot olisivat pienentyneet vuosien välillä. Valikoitumisen takia eroja on luotettavampaa tarkastella jatkuvilla muuttujilla.

Kuviosta 3 nähdään, että vanhemman koulutuksen ja ammattiaseman yhteys lasten lukutaitoon (PISA) ja äidinkielen arvosanaan (rekisteri), ei ole merkittävästi kasvanut Suomessa, kun käytetään jatkuvia muuttujia (Paneeli B ja D). Vaikka äidinkielen oppimistulosten eron kasvu vuosien 2000 ja 2015 välillä on tilastollisesti merkitsevä rekisteriaineistossa vanhemman koulutusvuosien osalta (Paneeli B), muiden vuosien välillä ero ei ole enää tilastollisesti merkitsevä. Korrelaatiokerroin vaihtelee välillä 0,24 - 0,29.

Jatkuvien muuttujien osalta nähdään, etteivät oppimistulosten erot ole juurikaan kasvaneet Suomessa. Täytyy kuitenkin huomata, että rekisteri ja PISA-aineistot antavat ristiriitaisen kuvan matematiikan osalta. Rekisterit näyttävät, että perhetaustan mukaiset erot ovat pienentyneet ja PISA-aineistot, että erot ovat kasvaneet. Koska PISA-aineisto sisältää huomattavaa mittausvirhettä, eikä virhe ole systemaattista yli ajan, voidaan rekistereillä tehtyjä analyyseja pitää luotettavampina. Täten oppimistulosten eriarvoisuus Suomessa ei ole kasvanut viime vuosina äidinkielen osalta ja matematiikan osalta jopa vähentynyt.

Kuvio 3. Vanhempien koulutuksen ja ammattiaseman yhteys lukutaidon (PISA) ja äidinkielen (rekisteri) osaamiseen 2000-2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.

Kuvio 4. Vanhempien koulutuksen ja ammattiaseman yhteys matematiikan osaamiseen PISA- ja rekisteriaineistoissa 2003, 2006, 2009 ja 2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.

Perhetaustan yhteys peruskoulun päättötodistuksen lukuaineiden keskiarvoon

Lopuksi tarkastelemme vanhemman 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 5 on esitetty vanhemman koulutuksen (paneeli A) sekä ammattiaseman (paneeli B) yhteys peruskoulun päättötodistuksen keskiarvoon jatkuvilla muuttujilla mitattuna (koulutus vuosissa ja ISEI). Kuviosta havaitaan, että yhteys on pysynyt lähes samana koko tarkastelujaksolla paitsi vuonna 2000 yhteys on hieman heikompi koulutuksen osalta ja vuosina 2000 sekä 2003 vanhemman ammatin (ISEI) mukaan. Viime vuosina eriarvoisuus vanhemman taustan mukaan ei ole kuitenkaan muuttunut peruskoulun keskiarvojen osalta. Kuvio 5 näyttää myös, että vanhemman koulutuksella on ammattiasemaa hieman voimakkaampi yhteys oppimistuloksiin. Ero on kuitenkin hyvin pieni (esimerkiksi vuonna 2015 korrelaatiokerroin: $0,35 - 0,31 = 0,04$). Voidaankin sanoa, että vanhemman koulutus ja ammattiasema arvioivat yhtä hyvin lasten oppimistuloksia.

Kuvio 5. Vanhempien korkeimman koulutuksen sekä ammattiaseman yhteys peruskoulun päättötodistuksen keskiarvoon 2003-2015. Y-akselilla korrelaatiokerroin ja x-akselilla mittausvuosi.

Yhteenveto

Tästä tutkimuksessa halusimme selvittää, sisältävätkö PISA-aineistot mittausvirhettä vanhemman taustamuuttujien osalta. Vertasimme Suomen PISA-aineistojen vanhempien koulutuksen sekä ammattiaseman jakaumia rekistereihin niissä ikäluokissa, jotka osallistuivat PISA-testeihin. Lisäksi tarkastelimme, onko oppimistulosten eriarvoisuus lisääntynyt Suomessa tutkimalla vanhempien koulutuksen sekä ammattiaseman yhteyttä lasten oppimistuloksiin yli ajan sekä PISA- että rekisteriaineistolla. Lopuksi havainnoimme, ovatko erot vanhemman koulutuksen ja ammattiaseman mukaan kasvaneet Suomessa, kun tarkastellaan toisen asteen päättötodistuksen keskiarvoja.

Tulokset osoittavat, että varsinkin vanhemman koulutuksen osalta PISA-testin kyselyaineisto sisältää huomattavaa mittausvirhettä. Poikkeuksena on vuosi 2000 jolloin oppilaiden raportoima vanhempien koulutus on hyvin linjassa rekisteritietojen kanssa. Syynä tähän voi olla, että vuoden 2000 PISA-tutkimuksen taustakyselyssä ei eritelty toisen asteen jälkeisiä tutkintoja. Aikaisemman tutkimuksen mukaan lapsilla on vaikeuksia tietää vanhempien toisen asteen jälkeisiä tutkintoja (Schulz 2006).

Toisin kuin vanhempien koulutuksen, vanhempien ammattiaseman mittausvirhe on pieni tai mitätön varsinkin äitien kohdalla. Kun oppimistulosten eroja halutaan tutkia PISA-aineistolla, vanhemman - mieluiten äidin - ammattiasemaa tulisi käyttää mittaamaan kotitaustaa.

Tulokset osoittavat myös, että oppimistulosten eriarvoisuus vanhemman koulutuksen tai ammattiaseman mukaan ei ole kasvanut viime vuosina Suomessa. Rekisteriaineistossa vanhempien koulutuksen yhteys oppimistuloksiin on vahvempaa kuin PISA-aineistoissa. Sen sijaan vanhempien ammattiaseman yhteys oppimistuloksiin on saman suuruista molemmissa aineistoissa. Ero näyttää johtuvan PISA-aineistojen vanhempien koulutusta koskevasta mittausvirheestä. Kun oppilaat yliarvioivat vanhempiensa koulutuksen, se johtaa vanhempien koulutuksen sekä lasten oppimistulosten yhteyden aliarviointiin.

Rekisteriaineiston perusteella vanhempien koulutuksen tai ammattiaseman yhteys peruskoulun päättötodistuksen keskiarvoon on pysynyt saman suuruisena viime vuosina. Tulos kumoaa aikaisempien, pääosin PISA-aineistoihin pohjautuvien, tutkimukset, joiden mukaan perhetaustan mukainen oppimisen eriarvoisuus olisi kasvanut Suomessa varsinkin viime vuosina (Salmela-Aro & Chmielewski 2019; Vettenranta ym. 2016; Leino ym. 2019).

Tulostemme perusteella erot aikaisempaan tutkimukseen johtuvat osaltaan PISA-aineistojen mittausvirheestä varsinkin vanhemman koulutuksen suhteen sekä siitä, että aikaisemmassa tutkimuksessa vanhemman tausta on operationalisoitu väärin kategorisena muuttujana. Kun oppimistuloksia verrataan PISA-testivuosien välillä siten, että vanhemman koulutusta ja ammattiasema mitataan jatkuvalla muuttujalla, myös PISA-aineistoihin perustuvien tulosten mukaan perhetaustan mukaiset erot oppimisessä eivät ole kasvaneet (lukutaito) tai ovat kasvaneet hyvin vähän (matematiikka). Rekisteriaineistoissa ei esiinny vastausharhasta johtuvaa mittausvirhettä, minkä vuoksi niitä voidaan pitää kyselyaineistoja luotettavampina tutkittaessa perhetaustan yhteyttä oppimistuloksiin.

Peruskoulun päättötodistuksen arvosanat mittaavat opetussuunnitelman mukaista oppimista ja PISA-testit puolestaan myöhemmän elämän kannalta tärkeiden avaintaitojen osaamista. Vaikka kouluarvosanat sekä PISA-tulokset mittaavat jokseenkin eri asioita, on näiden havaittu korreloivan keskenään. Esimerkiksi Suomessa äidinkielen arvosanan sekä PISA-koetulosten lukutaidon pisteiden välinen yhteys on havaittu vahvaksi (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).

On syytä huomata, että tutkimuksessa havaittu 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. Vanhempien ammattiaseman yhdenmukaisuus PISA- ja rekisteriaineistoissa viittaavatkin siihen, että otantavirheen sijaan PISA-aineistoissa esiintyy suurta

vastausharhaa oppilaiden vanhempien koulutuksen suhteen. Lapset eivät siis joko tiedä vanhempiensa koulutusta tai sitten he vastaavat tarkoituksella väärin. Aikaisempi tutkimus on näyttänyt, että lapset, jotka saavat parempia oppimistuloksia tietävät tarkemmin vanhemman koulutuksen kuin ne, joiden oppimistulokset ovat heikompia (Jerrim & Micklewright 2014). Jatkossa PISA-aineistot pitäisi yhdistää rekisteritietoihin, jotta saataisiin täysi varmuus mittausvirheen lähteestä.

Tämä tutkimus osoittaa, että PISA-tutkimusten 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 tulee hyödyntää enemmän ennen kuin lähdetään tekemään johtopäätöksiä koulutusjärjestelmän uudistamisesta kasvavan eriarvoisuuden torjumiseksi.

Kirjallisuus

Arinen, Pekka & Karjalainen, Tommi (2007) PISA 2006 Ensituloksia. Opetusministeriön julkaisuja 2007:38.

Avvisati, Francesco & Keslair, François (2014) REPEAT: 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.

Björklund, Anders, & Jäntti, Markus (1997) Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, 87(5), 1009-1018.

Breen, Richard & Jonsson, Jan (2005) Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology* 31/2005, 223-43.

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 (2018) What Do Books in the Home Proxy For? A Cautionary Tale. *Sociological Methods & Research*, April 2019. doi:[10.1177/0049124119826143](https://doi.org/10.1177/0049124119826143).

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.

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 OCCUPATIONSISCO-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.

Karhunen, Hannu & Uusitalo, Roope (2017) 50 vuotta koulutusmahdollisuuksien tasa-arvoa. *Yhteiskuntapolitiikka* 82 (3), 296-303.

Kreuter, Frauke & Eckman 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 julkaisuja 2013:20.

Kupari, Pekka (2005) Kotitausta näkyy matematiikan koetuloksissa. Teoksessa Pekka Kupari & Jouni Välijärvi (toim.) Osaaminen kestäväällä pohjalla: PISA 2003 Suomessa. Jyväskylä: Gummerus Oy, 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, Nina & Hiltunen, Jenna & Lintuvuori, Meri & Lähteinen, Suvi & Lämsä, Joni & Nissinen, Kari & Nissinen, Virva & Puhakka, Eija & Pulkkinen, Jouni & Rautopuro, Juhani & Sirén, Marjo & Vainikainen, Mari-Pauliina & Vetterranta, Jouni (2019) PISA 2018 Ensituloksia - Suomi parhaiden joukossa. Opetus- ja kulttuuriministeriön julkaisu 2019:40.

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*. Kirjapaino Oma Oy: Jyväskylä, 9-40.

Looker, Dianne (1989) Accuracy of proxy reports of parental status characteristics. *Sociology of Education* 62 (4), 257–276.

Nissinen, Kari & Rautopuro, Juhani & Puhakka, Eija (2018) PISA-tutkimuksen Metodologiasta. Teoksessa Juhani Rautopuro & Kalle Juuti (toim.) *PISA pintaa syvemmältä – PISA 2015 Suomen pääraportti*. 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 päättöarvioinnin kriteerit? Oppilaiden saamat arvosanat ja opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina. Raportit ja selvitykset 9/2013, Opetushallitus: Helsinki.

Salmela-Aro, Katariina & Chmielewski, Anna (2019) Socioeconomic Inequality and Student Outcomes in Finnish Schools. In Louis Volante & Sylke Schnepf & John Jerrim & Don Klinger (ed.) *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.

Sulkunen, Sari & Välijärvi, Jouni & Arffman, Inga & Harju-Luukkainen, Heidi & Kupari, Pekka & Nissinen, Kari & Puhakka, Eija & Reinikainen, Pasi (2010) *PISA 2009 Ensituloksia*. Opetus- ja kulttuuriministeriön julkaisu 2010:21.

Tourangeau, Roger & Yan, Ting (2007) Sensitive questions in surveys. *Psychological bulletin* 133 (5), 859.

Vetterranta, 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.

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) Tulevaisuuden osaajat - PISA 2000 Suomessa Kirjapaino Oma Oy: Jyväskylä.

Perhetaustan yhteys koulumenestykseen Suomessa PISA- ja rekisteriaineistojen valossa

Liite taulukot ja kuvat

Liitetaulukko 1. Vanhemman koulutuksen (kategorinen) prosenttiosuudet desiiittain.

Desiilit	PISA 2000	PISA 2003	PISA 2006	PISA 2009	PISA 2012	PISA 2015
1.	11 %	31 %	19 %	11 %	12 %	11 %
2.	11 %	-	4 %	38 %	9 %	29 %
3.	28 %	-	35 %	-	24 %	-
4.	-	34 %	-	-	-	53 %
5.	14 %	-	-	47 %	50 %	-
6.	-	-	42 %	-	-	-
7.	36 %	35 %	-	-	-	-
8.	-	-	-	-	-	-
9.	-	-	-	-	-	-
10.	-	-	-	4 %	5 %	8 %

Liitetaulukko 2a. Vanhemman koulutuksen prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Vanhempi	P	R	P	R	P	R	P	R	P	R	P	R
Perusaste	21,7	10,1	10,0	7,6	6,0	5,8	3,6	4,8	2,9	4,1	1,6	3,6
Toinen aste	42,3	43,9	22,8	42,2	16,9	41,5	15,7	39,1	17,5	36,9	17,7	35,7
Korkea-aste	35,9	46,0	67,2	50,2	77,1	52,7	80,7	56,1	79,5	59,0	80,7	60,7
Ei tietoja	4,8	0,0	1,0	0,0	1,5	0,0	1,3	0,0	2,8	0,0	1,6	0,0
N	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Koulutuksen prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitetaulukko 2b. Äidin koulutuksen prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Äiti	P	R	P	R	P	R	P	R	P	R	P	R
Perusaste	30,8	22,8	16,5	17,3	11,5	14,1	8,4	11,6	6,2	10,0	4,3	9,0
Toinen aste	42,5	43,6	25,5	43,8	19,2	44,2	19,6	42,9	21,7	41,2	22,1	40,2
Korkea-aste	26,6	33,7	60,1	39,0	72,3	41,8	74,3	45,5	75,9	48,8	75,6	50,9
Ei tietoja	6,2	0,0	2,1	0,0	2,9	0,0	2,2	0,0	3,8	0,0	2,0	0,0
N	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Koulutuksen prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitetaulukko 2c. Isän koulutuksen prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Isä	P	R	P	R	P	R	P	R	P	R	P	R
Perusaste	33,5	29,1	21,8	25,9	17,4	23,2	13,3	20,5	12,6	19,0	8,2	17,5
Toinen aste	40,9	40,6	26,8	42,1	23,0	44,2	20,3	44,9	25,4	44,9	26,1	45,8
Korkea-aste	25,7	30,4	51,4	32,0	59,6	32,6	66,4	34,5	62,0	36,1	65,8	36,8
Ei tietoja	8,7	0,0	5,3	0,0	4,8	0,0	4,6	0,0	6,9	0,0	3,3	0,0
<i>N</i>	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Koulutuksen prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitetaulukko 3a. Vanhemman ammattiaseman prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Vanhempi	P	R	P	R	P	R	P	R	P	R	P	R
Johtajat	19,7	8,2	19,6	8,5	15,3	12,4	24,6	13,4	20,5	8,9	9,4	10,2
Asiantuntijat	26,2	45,1	40,2	47,6	41,3	45,4	47,8	47,2	45,5	52,7	48,8	53,9
Toimisto- ja palvelutyö	17,3	27,1	25,4	25,4	27,2	26,6	16,5	25,8	23,6	26,4	30,8	25,0
Maanviljelijät	9,1	6,0	4,5	5,7	4,6	4,2	3,1	3,5	2,5	3,1	2,7	2,6
Manuaaliset alat	11,8	11,8	9,1	11,1	9,4	9,9	6,9	8,6	6,9	7,7	7,1	7,1
Muut työntekijät	5,7	1,7	1,2	1,6	2,2	1,4	1,1	1,4	1,0	1,2	1,2	1,1
Ei tietoja	1,9	11,6	1,4	16,5	2,7	6,4	1,3	4,8	3,4	3,0	4,2	4,3
<i>N</i>	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Ammattien prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitetaulukko 3b. Äidin ammattiaseman prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Äiti	P	R	P	R	P	R	P	R	P	R	P	R
Johtajat	5,7	2,4	5,8	2,5	5,7	3,5	8,4	4,1	7,3	2,9	3,6	3,3
Asiantuntijat	44,4	35,0	37,7	38,3	39,9	38,5	52,3	41,1	46,1	43,8	44,1	46,0
Toimisto- ja palvelutyö	32,3	39,7	40,1	38,3	38,2	38,9	26,1	38,0	34,5	38,3	40,8	37,0
Maanviljelijät	4,9	5,9	4,0	5,6	4,0	4,0	2,9	3,2	2,5	3,0	2,7	2,6
Manuaaliset alat	5,9	7,0	6,0	6,7	4,8	6,1	4,9	5,4	4,5	5,5	3,6	5,1
Muut työntekijät	6,8	10,0	6,4	8,6	7,4	9,0	5,4	8,1	5,0	6,5	5,3	6,1
Ei tietoja	6,7	14,1	5,4	21,6	9,4	8,2	5,0	6,6	8,6	6,0	9,5	6,9
<i>N</i>	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Ammattien prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitetaulukko 3c. Isän ammattiaseman prosenttiosuudet vuosina 2000-2015 PISA- ja rekisteriaineiston mukaan

	2000		2003		2006		2009		2012		2015	
Isä	P	R	P	R	P	R	P	R	P	R	P	R
Johtajat	17,4	6,7	17,3	6,8	13,2	10,3	20,3	11,0	17,1	7,1	7,4	8,3
Asiantuntijat	27,4	34,0	27,8	35,5	26,7	32,4	28,7	34,0	28,4	38,7	33,3	39,3
Toimisto- ja palvelutyö	7,7	8,2	7,2	8,5	7,5	8,0	8,0	8,6	11,3	11,1	13,0	11,0
Maanviljelijät	9,0	8,4	7,0	8,1	7,6	6,6	5,5	5,9	5,3	5,5	5,1	4,7
Manuaaliset alat	35,5	37,6	36,6	36,0	39,6	37,1	33,5	34,9	36,1	34,0	39,4	33,0
Muut työntekijät	3,1	5,0	4,0	5,0	5,5	5,5	4,1	5,7	1,8	3,6	2,0	3,6
Ei tietoja	8,6	15,0	6,3	19,7	14,0	10,4	6,4	9,2	10,4	8,3	13,9	9,1
<i>N</i>	4470	60170	5312	54605	4384	59014	5346	58473	8076	55047	5422	52363

P = PISA, R = Rekisteriaineisto, PISA-aineisto painotettu.

Huom. Ammattien prosenttiosuudet laskettu ilman puuttuvia arvoja

Liitekuvio 1. Äidin koulutuksen ja ammattiaseman yhteydet lukutaitoon (PISA) ja äidinkielen arvosanaan (rekisteri). Y-akseli korrelaatiokerroin ja x-akseli mittausvuosi.

Liitekuvio 2. Isän koulutuksen ja ammattiaseman yhteydet lukutaitoon (PISA) ja äidinkielen arvosanaan (rekisteri). Y-akseli korrelaatiokerroin ja x-akseli mittausvuosi.

Liitekuvio 3. Äidin koulutuksen ja ammattiaseman yhteydet matematiikan osaamiseen (PISA) ja arvosanaan (rekisteri). Y-akseli korrelaatiokerroin ja x-akseli mittausvuosi.

Liitekuvio 4. Isän koulutuksen ja ammattiaseman yhteydet matematiikan osaamiseen (PISA) ja arvosanaan (rekisteri). Y-akseli korrelaatiokerroin ja x-akseli mittausvuosi.