

Eriytyvät koulutuspolut – osaamisen ja vanhempien sosioekonomisen aseman yhteys oppilaan koulutusvalintoihin

3. Menetelmät

Osaamisryhmien muodostaminen (Karvin aineisto)

Karvin pitkittäisaineiston oppilaat jaettiin arvioinnin kolmessa mittauspisteessä (3., 6. ja 9. luokka) kolmeen suunnilleen yhtä suureen osaamisryhmään (alin, keskimäinen ja ylin osaamiskolmannes). Ryhmäjaon avulla tarkasteltiin osaamisen pysyvyyttä sekä oppilaiden suhteellisen aseman muutoksia perusopetuksen aikana.

Osaamisryhmät muodostettiin kussakin mittauspisteessä matematiikan arviointitulosten sekä matematiikan ja äidinkielen opettaja-arvioiden perusteella. Kuudennella ja yhdeksännellä luokalla opettaja-arvioina käytettiin oppilaiden arvosanoja. Kolmannella luokalla opettajat arvioivat oppilaiden osaamista asteikolla 0–2 (0 = heikompaa kuin kuvauksessa, 1 = kuvauksen tasoista, 2 = huomattavasti parempaa kuin kuvauksessa).

Matematiikan arviointitulokset ja opettaja-arviot standardoitiin erilaisten mitta-asteikkojen vuoksi erikseen kussakin mittauspisteessä (keskiarvo = 0, keskihajonta = 1). Standardoiduista muuttujista muodostettiin keskiarvona yhdistelmämuuttuja, jota käytettiin oppilasryhmien muodostamisessa.

Puuttuvien tietojen käsittely (Karvin aineisto)

Karvin pitkittäisaineisto käsitti 3 502 oppilasta. Tilastokeskuksen rekisteritietoja puuttui osalta oppilaista: 971 oppilaalta (27,7 %) puuttui tieto toisen asteen koulutuksesta ja 823 oppilaalta (23,5 %) tieto huoltajan sosioekonomisesta asemasta 9. luokan aikana. Opettaja-arvioissa, matematiikan ja äidinkielen arvosanoissa sekä asennekyselyissä puuttuvia tietoja oli 0,1–6,1 %.

Puuttuvia tietoja käsiteltiin moni-imputoinnilla (multiple imputation) (vanBuuren, 2018). Imputointi toteutettiin chained equations -menetelmällä, jonka avulla tuotettiin 30 imputoitua aineistoa. Jatkuville muuttujille imputoinnissa käytettiin predictive mean matching -menetelmää, järjestysasteikollisille muuttujille ordinaalista regressiota ja luokitteleville muuttujille multinomiaalista logistista regressiota. Imputointimallit sisälsivät oppilaan taustatietoja, kuten sukupuoli, asenteet matematiikkaa kohtaan, arvosanat, sekä aiemmat osaamismittaukset eli matematiikan arviointitulokset sekä opettajien antamat arviot oppilaiden osaamisesta. Imputointi toteutettiin R-ohjelmiston mice-kirjastolla (van Buuren, 2011). Analyysissä imputoitujen aineistojen tulokset yhdistettiin Rubin'in sääntöjen mukaisesti (Rubin, 1987), ja raportoidut estimaatit perustuvat näihin yhdistettyihin tuloksiin.

Tilastolliset analyysit

Peruskouluajan siirtymiä osaamisryhmien välillä ja niistä toisen asteen koulutukseen kuvattiin virtauskaavioiden eli Sankey-kuvioiden avulla. Lisäksi aineistosta laskettiin osaamisryhmittäiset osuudet eri toisen asteen koulutuspoluille siirtyneistä oppilaista. Virtauskuvioita käytettiin havainnollistamaan myös

opiskelijoiden siirtymiä toiselta asteelta korkea-asteen koulutukseen ja sen jälkeisiin sosioekonomisiin asemiin.

Kuvailevan analyysin lisäksi osaamisryhmien ja oppilaan taustatekijöiden yhteyksiä tarkasteltiin multinomiaalisilla logistisilla regressiomalleilla. Selittävinä muuttujina käytettiin oppilaan sukupuolta sekä huoltajan sosioekonomista asemaa. Multinomiaalista regressiota käytettiin myös toisen asteen koulutusvalintojen ja oppilaan taustatekijöiden välisen yhteyden tarkastelemiseen. Mallissa selittävinä muuttujina olivat perusopetuksen yhdeksännen luokan osaamisryhmä, oppilaan sukupuoli ja huoltajan sosioekonominen asema. Mallien perusteella laskettiin myös ennustetut todennäköisyydet eri koulutuspoluille siirtymisestä.

Peruskoulun jälkeisen koulutuspolun yhteyttä myöhempään sosioekonomiseen asemaan tarkasteltiin niin ikään multinomiaalisella logistisella regressiolla. Selitettävänä muuttujana oli oppilaan sosioekonominen asema vuonna 2023. Selittäjinä olivat toisen asteen koulutusvalinta, korkea-asteen koulutus, sukupuoli sekä huoltajan sosioekonominen asema.

Tulkinnan helpottamiseksi tulokset esitetään pääasiassa prosenttiosuuksina. Osa prosenttiosuuksista perustuu havaittuihin jakaumiin (laskettu aineistosta) ja osa multinomiaalisten regressiomallien perusteella laskettuihin ennustettuihin todennäköisyyksiin. Regressiomallien varsinaiset tulostaulukot löytyvät raportin liitteistä 1–4. Kaikki analyysit toteutettiin R-ohjelmistolla käyttäen nnet- (Venables & Ripley, 2002), emmeans- (Lenth & Piaskowski, 2026) ja marginaaleffects (Arel-Bundock, Greifer & Heiss, 2024)-paketteja.