KÖZGAZDASÁGI SZEMLE, LXXI. ÉVF., 2024. ÁPRILIS (353–378. o.)

HERMANN ZOLTÁN–KERTESI GÁBOR–VARGA JÚLIA

A teszteredmények társadalmi egyenlőtlensége és az általános iskolai szegregáció

Tanulmányunk a különböző társadalmi hátterű diákok általános iskolai teljesítményében megnyilvánuló különbségek és a szegregáció alakulását mutatja be a 2006 és 2019 közötti időszakban. Számításaink szerint az adott időszakban jelentős mértékben nőtt a különböző társadalmi helyzetű diákok iskolák közötti elkülönülése, és különösen az alacsony státuszú (az alsó jövedelmi ötödbe tartozó) diákok szegregációja erősödött. Az időszak második felében a teszteredmények társadalmi helyzet szerinti különbségei is növekedtek. A szegregáció és a tanulói teljesítmények társadalmi státusz szerinti egyenlőtlenségei között is összefüggést találtunk: azokban a járásokban és városokban, ahol erősebb a szegregáció, jellemzően nagyobbak a teszteredmények társadalmi egyenlőtlenségei is.*[*]
Journal of Economic Literature (JEL) kód: I21, I24, C19.

Bevezetés

A magyar közoktatási rendszer teljesítményének egyik alapvető jellemzője, hogy az egyéni eredmények szóródása nemzetközi összehasonlításban is kiugróan nagy mértékben tükrözi a tanulók közötti társadalmi különbségeket. A 2015. évi PISA-felmérés[1] adatai szerint például a természettudományos teszteredmények szóródását a felmérésben részt vevő országok közül Magyarországon magyarázta legnagyobb arányban a diákok társadalmi háttere (*OECD* [2018]). Ugyanezt a tendenciát

[*] Tanulmányunkat a magyar társadalomtudomány kiemelkedő alakja, a nemrégiben elhunyt *Csapó Benő* emlékének ajánljuk.

A tanulmány rövidített változata megjelent a Munkaerőpiaci tükör, 2022 (Társadalmi egyenlőtlenségek és mobilitás) című kötetben.

[1] PISA: *Programme for International Student Assessment* – a nemzetközi tanulói teljesítmény mérésének OECD-programja.

Hermann Zoltán, KRTK Közgazdaságtudományi Intézet, BCE Közgazdaságtan Intézet (e-mail: hermann.zoltan@krtk.hun-ren.hu).
Kertesi Gábor, KRTK Közgazdaságtudományi Intézet (e-mail: kertesi.gabor@krtk.hun-ren.hu).
Varga Júlia, KRTK Közgazdaságtudományi Intézet (e-mail: varga.julia@krtk.hun-ren.hu).
A kézirat első változata 2024. január 24-én érkezett szerkesztőségünkbe.
DOI: https://doi.org/10.18414/KSZ.2024.4.353

figyelhettük meg már a kilencvenes években is: az 1999. évi TIMSS-mérések[2] szerint a családi háttér hatása a matematikai teszteredményekre több mint ötven ország közül Magyarországon volt a negyedik legerősebb (*Schütz és szerzőtársai* [2008]).

A magyar közoktatás egy másik alapvető vonása az általános iskolai oktatásban tapasztalható nagyfokú szegregáció. A különböző családi hátterű diákok elkülönülése nemzetközi összehasonlításban is jelentős mértékű (*Csapó és szerzőtársai* [2009]). Számos érv szól amellett, hogy a közoktatási rendszer e két jellemzője egymással is szorosan összefügg. A hátrányos helyzetű diákok szegregált oktatása többféle áttételen keresztül növelheti a tanulói teljesítmények társadalmi különbségeit; az iskolák, illetve a diákok közötti különbségek pedig erős ösztönzőt jelentenek a szülők számára, hogy a tágabb lakóhelyi környezetükben elérhető jobb minőségű és diák-összetételű iskolákba írassák a gyereküket, amivel maguk is hozzájárulnak a diákok közti társadalmi elkülönülés újratermelődéséhez (*Kertesi–Kézdi* [2005]).

Ez a tanulmány a társadalmi háttér szerinti teljesítménykülönbségek és a szegregáció időbeli változását mutatja be leíró jelleggel a 2006 és 2019 közötti időszakban. Elemzésünk alapját egy folytonos társadalmi státuszmutató képezi, amelyet az Országos Kompetenciamérés tanulói kérdőívének adatai alapján hoztunk létre, a korábbi gyakorlattól némileg különböző eljárást alkalmazva, melynek fontosabb részleteit az adatokról szóló részben ismertetjük.

A társadalmi háttér szerinti teljesítménykülönbségeket egy egyszerű mutatóval, a társadalmi háttér és a tesztpontszámok közötti kapcsolat erősségével mérjük. A tanulók társadalmi helyzet szerinti iskolai elkülönülését az előzőkben hivatkozott társadalmi státuszmutató eloszlásának alsó, illetve felső kvintiliséhez tartozó – a legrosszabb, illetve a legjobb társadalmi helyzetben levő (szegény, illetve jobb módú) – gyerekek iskolai elkülönülésével mérjük. A szegregáció vizsgálatánál két kérdést veszünk szemügyre: Milyen mértékben szorulnak ki a szegény gyermekek a társadalom nem szegény többségének iskoláiból? Milyen mértékben különülnek el a társadalom felső rétegének gyermekei mindenki mástól olyan iskolákban, amelyek csak korlátozott mértékben hozzáférhetők a kevésbé jómódú gyerekek számára? A diákok iskolák közötti társadalmi elkülönülését emellett egy általánosabb mutatóval, a társadalmi státuszmutató iskolák közötti varianciájának a teljes variancián belüli részarányával is mérjük. Tanulmányunk utolsó részében a tanulói teljesítmények társadalmi meghatározottságának és az iskolák közötti társadalmi elkülönülésnek az összefüggéseit vizsgáljuk, kihasználva a járások, illetve városok közötti keresztmetszeti heterogenitásokat.

Az iskolai szegregációt kiterjedt szociológiai és közgazdasági irodalom vizsgálja Magyarországon. Ez az irodalom nagyobb részben az etnikai szegregációt és kisebb részben a legszegényebb (halmozottan hátrányos helyzetű) diákok szegregációját vizsgálta (*Havas–Liskó* [2006], *Kertesi–Kézdi* [2009]). A szegregáció időbeli változására vonatkozóan kevesebb adat áll rendelkezésre (*Kertesi–Kézdi* [2012], *Varga* [2022*a*]). Az itt következő beszámoló több társadalmi csoportra – a társadalmi

[2] TIMSS: *Trends in International Mathematics and Science Study* – nemzetközi összehasonlító teljesítménymérés matematikából és természettudományokból.

hierarchia alsó és felső ötödére – koncentrálva és a vizsgált időszak minden évében azonos mérési eljárást alkalmazva vizsgálja a gyerekek iskolák közötti elkülönülését. Az időbeli változások így, reményeink szerint, pontosabban megragadhatók, tekintve, hogy a mérési eljárásunk nincs kitéve olyan torzításoknak, mint a hátrányos vagy a halmozottan hátrányos helyzet időben változó definíciói, amelyektől a korábbi elemzések nem függetleníthették magukat.[3]

Adatok és módszerek

Elemzésünk az Országos Kompetenciamérés 2006 és 2019 közötti hullámainak adataira épül. A kompetenciamérés a 6., 8. és 10. évfolyamra járó tanulók matematikai és szövegértési készségeit méri fel standardizált tesztek alapján, néhány kivételtől (elsősorban speciális nevelési igényű diákoktól) eltekintve teljeskörűen. A diákok és szüleik egy háttérkérdőívet is kitöltenek, amely fontos információkat tartalmaz a család társadalmi jellemzőiről.

Az elemzési mintát azok a hatodikos és nyolcadikos diákok alkotják, akik *a)* kitöltötték a háttérkérdőívet legalább egyszer a hatodikos, nyolcadikos vagy tizedikes mérés keretében, vagy *b)* kitöltötték a matematikai és szövegértési tesztek közül legalább az egyiket hatodik vagy nyolcadik évfolyamon. A két évfolyamon az elemzési minta csaknem 2,3 millió diákot foglal magában, ezen belül a nyolcadikosok száma megközelíti az 1,25 milliót.

A tanulói háttérkérdőív adatai (a szülők iskolázottsága, a lakásban található könyvek száma, egyes tartós fogyasztási cikkek, különböző iskolai és iskolán kívüli állami támogatások megléte stb.) alapján létrehoztunk egy folytonos társadalmi státuszindexet. (Az index létrehozásához felhasznált változók listáját a *Függelék A1. táblázata* tartalmazza.) A felhasznált változók aggregálásához a súlyokat a szövegértés teszteredményein becsült regressziós modell paramétereiből nyertük. A becslés a 2006–2019-es időszak egyesített keresztmetszeti mintáján, a standardizált szövegértés pontszámain történt. Az így kapott nyers társadalmi háttérindex értékeit ezután évenként és évfolyamonként standardizáltuk, az átlag és a szórás így minden évben 0 és 1 közötti érték. A társadalmi háttérindex létrehozását a *Függelék A részében* mutatjuk be részletesen.

A társadalmi háttérindex létrehozásának legnagyobb nehézségét a hiányzó értékek nagy száma jelentette. A háttérkérdőívet évente hozzávetőlegesen a diákok ötöde nem tölti ki. Ezt a problémát nagyobbrészt úgy hidaltuk át, hogy kihasználtuk az Országos Kompetenciamérés adatainak paneltulajdonságait – hogy minden diákot háromszor,[4] 6., 8. és 10. évfolyamos korában mérnek fel –, és a hiányzó adatokat az időben stabilnak tekinthető változók (például a szülők iskolázottsága) esetében egy másik naptári év másik évfolyamán meglevő adatokkal pótoltuk.

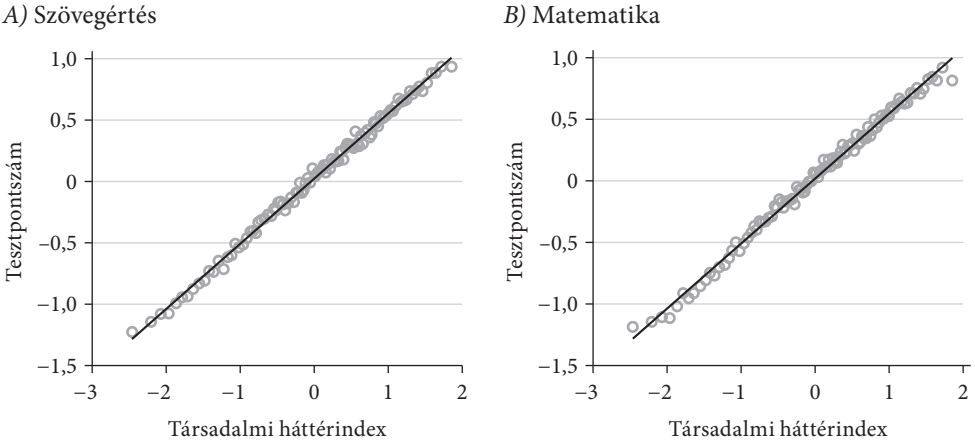
[3] A roma etnicitás mérése a 2006 és 2019 közötti kompetenciamérések háttérkérdőívi adatai alapján az időszak egészében hasonló. Időben változó definíciókkal akkor szembesülünk, ha az időbeli összehasonlítást kiterjesztjük 1992-re, illetve az 1980 és 1992 közötti időszakra is.

[4] Az évfolyamismétlő tanulók háromnál többször is előfordulhatnak a kompetenciamérésekben.

Az adathiány kisebb részét egy imputálási eljárás segítségével pótoltuk. Olyan diákok esetében, akikre nézve semmilyen forrásból nem állt rendelkezésre kérdőívi háttéradat, de volt teszteredmény, a megfigyelt teszteredmények és a *mindkét* adattal rendelkezők teszteredményének és társadalmi háttérindexének együttes eloszlása alapján imputáltunk háttérindexértékeket. A társadalmi háttérindexnek ezt a változatát használjuk a szegregáció mérése során, míg a teszteredmények társadalmi háttér szerinti egyenlőtlenségének elemzésekor, értelemszerűen, nem használunk imputált értékeket.

A tanulói teljesítmények társadalmi háttér szerinti egyenlőtlenségét a teszteredmények és a társadalmi háttérindex közötti kapcsolat erősségével mérjük. Évenként és évfolyamonként kétváltozós regressziókat becsülünk, ahol a függő változó a tesztpontszám, a jobb oldali változó pedig a társadalmi háttérindex. Mindkét változót évenként és évfolyamonként standardizáltuk. Az egyenlőtlenség mérőszáma a társadalmi háttérindex standardizált regressziós együtthatója, amely azt mutatja meg, hogy egy szórásegységgel magasabb társadalmi státuszérték hány szórásegységgel magasabb tesztpontszámmal jár együtt. Egy lineáris regressziós együttható nyilvánvalóan akkor ragadja meg jól a társadalmi háttér szerinti egyenlőtlenséget, ha a státusz és a tesztpontszám közötti összefüggés lineáris, azaz ha a státuszértékek a teljes eloszlás mentén, annak valamennyi szegmensében hasonló arányú tesztpontszám-különbségeket jeleznek előre. Az *1. ábra* a 2019-ben 8. évfolyamra járó tanulók példáján mutatja be, hogy az összefüggés valóban lineáris (az összefüggés a többi év valamennyi évfolyamán hasonló mintázatú).

1. ábra
A társadalmi státusz és a tesztpontszám összefüggése, 2019
(státuszindex szerinti percentilisekben mérve, 8. évfolyamos tanulók)



Megjegyzés: a társadalmi háttérindex és a tesztpontszám standardizált értékei szórásegységekben mérve.

Az iskolai elkülönülést kétféle mutatószámmal mérjük. Az egyik az iskolák közötti variancia részaránya a társadalmi háttérindex teljes varianciáján belül. Minél nagyobb ez az

arány, annál nagyobb mértékben különülnek el különböző iskolákban a különböző társadalmi státuszú diákok, és annál homogénebb az iskolák belső összetétele.

A másik mérőszám a szegregációs index, amelyet a társadalmi háttérindex alsó, illetve felső kvintiliséhez tartozó szegény, illetve jobb módú gyerekek iskolák közötti elkülönülésének mérésére számoltunk ki. Az iskolák közötti szegregációs index azt mutatja meg, hogy egy adott területi egységen belül egy társadalmi csoport és a komplementere közötti véletlenszerű kontaktusok valószínűsége hány százalékkal alacsonyabb annál, mint amekkora abban a hipotetikus esetben lenne, ha adott társadalmi csoporthoz tartozó tanulók eloszlása egyenletes lenne az adott területi egység iskoláiban. Intuitív megfogalmazással: a szegregációs index azt mutatja meg, hogy a különböző társadalmi csoportok között lehetséges kontaktusok hány százaléka hiúsul meg e tanulók iskolák közötti elkülönülése következtében.[5]

A szegregációs mutatók 0 és 1 közötti értékeket vehetnek fel; 0 a szegregáció teljes hiányára, 1 pedig tökéletes szegregációra utal. A külső variancia részaránya és a szegregációs index értéke közvetlenül nem hasonlítható össze egymással, időbeli trendjeik vagy keresztmetszeti szóródásuk azonban igen. A szegregáció mérése rendkívül érzékeny arra, hogy mérőszámainkat milyen kiterjedésű területi egységre számoljuk ki. Minél nagyobb területi egységre – akár az egész országra – számolunk szegregációs mutatókat, annál inkább befolyásolja a mérőszámok értékét a különböző társadalmi csoportok eltérő földrajzi elhelyezkedése. Minél inkább egybeesnek ezek a földrajzi egységek a valódi iskolapiacokkal – amelyeken belül a diákok lakóhelyük és az iskola között többségében ingáznak –, annál jobban ki tudjuk küszöbölni a szegregációs indexek értékéből a lakóhelyi elhelyezkedés földrajzi-társadalmi különbségeiből adódó összetételhatásokat.

Az iskolai szegregáció mérése során az Országos Kompetenciamérésben megfigyelt telephelyeket tekintjük iskolának. Az elemzésben a speciális nevelési igényű diákok nem szerepelnek. Kihagyásukkal mind a szegregáció mértékét, mind pedig a teszteredményekben tükröződő társadalmi egyenlőtlenségek mértékét inkább alulbecsüljük. Mérési eredményeinket a 8. évfolyamra járó diákokra nézve mutatjuk be. A 8. évfolyamos diákokra vonatkozó becslések a teljes intézményi kínálatot átfogják: ezen az évfolyamon már a hatosztályos gimnáziumok hatása is érvényesül.

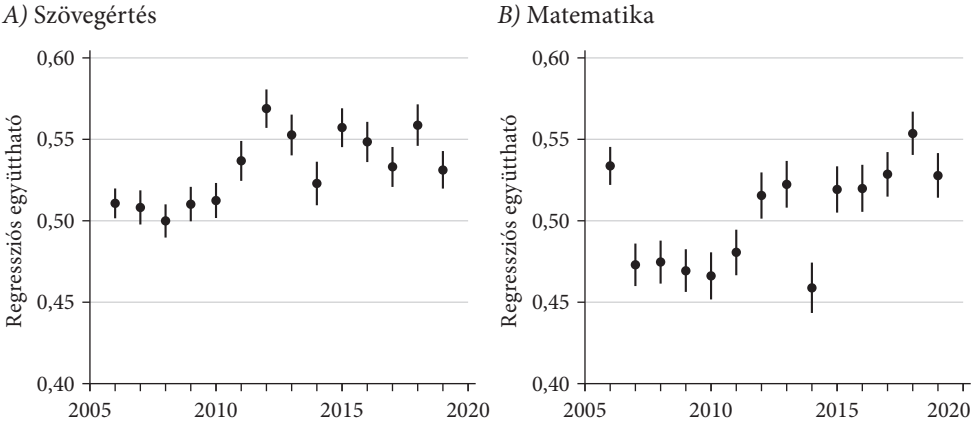
A teszteredmények társadalmi egyenlőtlenségének időbeli alakulása

A *2. ábra* a teszteredmények társadalmi egyenlőtlenségének időbeli alakulását mutatja be a 2006 és 2019 közötti időszakban a mindenkori 8. évfolyamos tanulók mintáján, ismételt keresztmetszeti becslések segítségével, amelyekben a standardizált teszteredményeket regresszáltuk a társadalmi háttérindex standardizált értékein. Az együtthatók értékeit a 95 százalékos konfidenciasávokkal együtt közöljük.

[5] A szegregációs index képletszerű levezetését lásd *Clotfelter* [1999], [2004], illetve *Kertesi–Kézdi* [2009] írásában.

A becsült együtthatók értékei többségében szignifikánsan magasabbak az időszak második felében, mint az időszak elején, ami az egyenlőtlenségek növekedését jelzi. Matematikából egy szórásegységnyi társadalmi státuszkülönbség 2007 és 2011 között 0,47–0,48 szórásegységnyi tesztpontszámtöbblettel járt együtt, míg a 2012–2019-es időszakban 0,52–0,53 szórásegységnyivel, ami jelentős, 10 százalékos növekedést jelent. Szövegértésben is hasonló változást látunk. A 6. évfolyamos diákok esetében kevésbé egyértelműen tapasztalható az egyenlőtlenségek növekedése (lásd a *Függelék B1. ábráját*), ami azzal függhet össze, hogy a hatodik évfolyamra járó tanulóknál a kisgimnáziumok szerepe mérsékeltebb.

2. ábra
A társadalmi háttér hatása a 8. évfolyamon mért tesztpontszámokra, 2006–2019 (kétváltozós standardizált regressziós együtthatók)



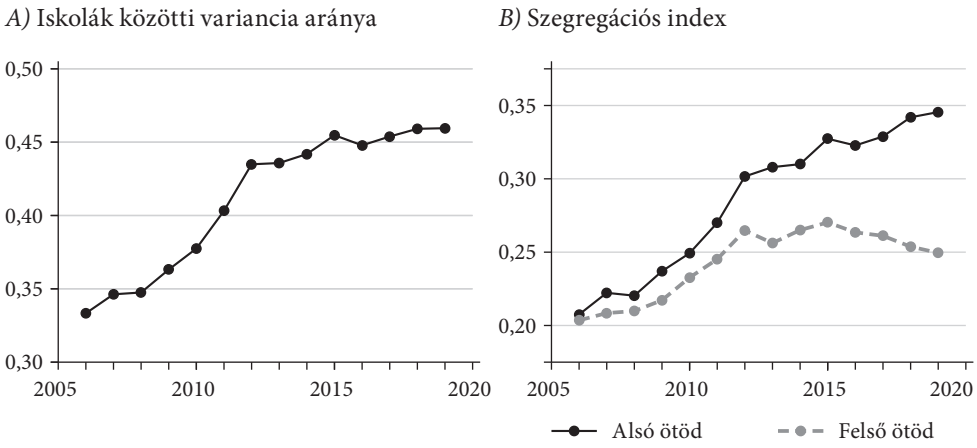
Megjegyzés: minden pont egy kétváltozós regressziós modell eredményét mutatja, ahol a függő változó a standardizált tesztpontszám, a jobb oldali változó pedig az évenként és évfolyamonként standardizált társadalmi háttérindex. Az ábra a becsült együtthatókat mutatja, 95 százalékos konfidenciaintervallumokkal (iskolaszinten klaszterezett standard hibák alapján).

A szegregáció időbeli alakulása

A *3. ábrán* a szegregációs mutatók alakulását követhetjük nyomon, Magyarországot egyetlen nagy iskolapiacnak tekintve. A szegregációt az iskolák közötti variancia részarányával mérve azt tapasztaljuk, hogy a társadalmi státusz szerinti elkülönülés az időszak egészében folyamatosan nőtt, bár a növekedés üteme 2015 után valamennyire lelassult. Az időszak egészét tekintve azonban a növekedés jelentős mértékű: az iskolák közötti variancia aránya 0,33-ról 0,46-ra nőtt, ami csaknem negyven százalékos növekedést jelent. Az elkülönülés általános növekedése mögött azonban rendkívül heterogén folyamatok állnak. Az alsó és felső ötödbe tartozó diákok szegregációs indexei 2012-ig nagyjából hasonló meredekségű trend szerint emelkedtek, 2012 után azonban az alacsony státuszú diákok szegregációja tovább nőtt, miközben a felső

ötöd szegregációs indexe lényegében változatlan maradt. A társadalmi helyzet szerinti növekvő elkülönülés mögött elsősorban a szegény diákok erősödő szegregációja áll. Ez a változás drámai mértékű: az alsó ötöd szegregációs indexe bő egy évtized alatt több mint másfélszeresére emelkedett.

3. ábra
Az iskolák közötti társadalmi elkülönülés mérőszámainak alakulása országosan, 2006–2019 (8. évfolyamos tanulók)



A szegregáció növekedése azonban nemcsak abból adódhat, hogy egy-egy térségen belül nő a szegények és gazdagok iskolák közötti elkülönülése, hanem abból is, hogy a szegények és gazdagok területi koncentrációja nő, azaz a fejlett térségek lakóhelyeiről egyre inkább kiszorulnak a szegény családok, a leszakadó területekről pedig egyre inkább eltünedeznek a jobb módúak.

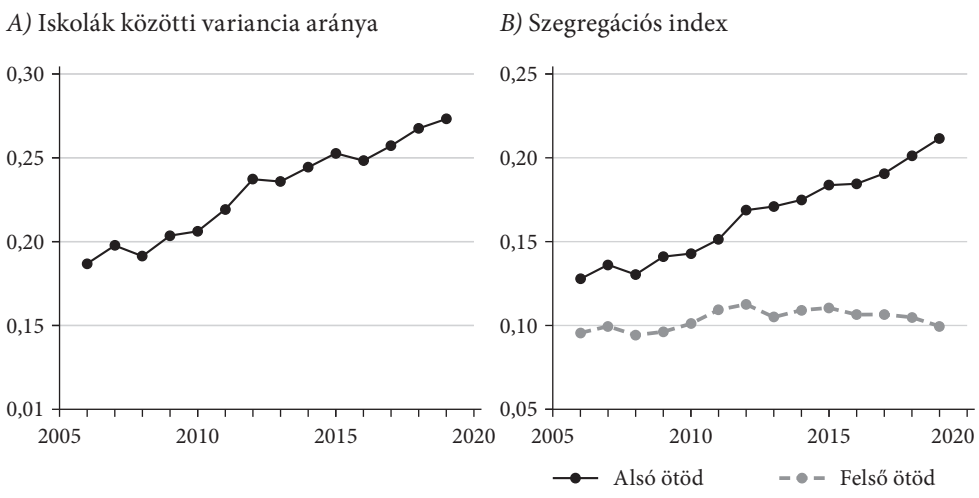
Az iskolás korú gyerekeket nevelő szegény és jobb módú családok növekvő mértékű földrajzi elkülönülése adatokkal igazolható. Ha a *3. ábrán* látható számításokban az iskolák helyett a járásokat választjuk az elkülönülés alapegységének, akkor meg tudjuk becsülni azt is, hogy a tanulók iskolák közötti szegregációjának mekkora része származik abból, hogy társadalmi összetételét tekintve egyre homogénebb tanulói populáció él, illetve jár iskolába az ország különböző járásaiban. A tanulók iskolák közötti, országos szinten mért szegregációs indexén (S) belül elkülöníthető egy, a tanulók járások közötti földrajzi elkülönülését mérő komponens ($S^{járás}$).[6] Bármelyik szegregációs mérőszámot tekintjük is, a földrajzi elkülönülés jelentősége 2006 és 2019 között igazolhatóan nőtt: ennek az összetevőnek a súlya az iskolák közti variancia mérőszámában ($Var^{járás}/Var$) a 2006. évet jellemző 37 százalékos részarányról 45 százalékra, a szegregációs indexek közül ($S^{járás}/S$) az alsó ötöd indexében 35 százalékról 40 százalékra, a felső ötöd indexében pedig 30 százalékról 37 százalékra emelkedett. (A *Függelék B2. ábrája* mutatja a teljes idősorokat.)

[6] A dekompoziciós eljárásról lásd *Clotfelter* [1999], illetve *Varga* [2022*b*] 69–70. o.

Ezeket a jelentős összetételhatásokat a számításainkból úgy tudjuk kiküszöbölni, ha a valódi iskolapiacokhoz – tipikus iskolai ingázási körzetekhez – hasonló, homogénebb lakosságú földrajzi körzetekre (járásokra és városokra)[7] számolunk átlagos szegregációs mutatókat. A járások és még inkább a városok olyan iskolapiacoknak tekinthetők, amelyekben észszerű (fél-egy óránál nem több időt igénylő), olcsó tanulói bérlettel hozzáférhető napi tömegközlekedési ingázási távolságon belül vannak az iskolák a gyerekek lakóhelyeitől, így a közlekedési hálózat *elvileg* bárki számára elérhetővé teszi a körzet teljes iskolakínálatát. Ha a különböző családi hátterű tanulók ezt a könnyen elérhető iskolakínálatot nagyon különböző mértékben képesek a maguk javára fordítani, az az iskolai szegregáció mechanizmusainak tulajdonítható.

A *4. ábra* a járásokra és többiskolás városokra kiszámított szegregációs mérőszámok átlagainak alakulását mutatja be. Mivel az iskolapiacok belső működésére vagyunk kíváncsiak, *súlyozatlan* országos átlagokat számolunk, hogy megmutathassuk, hogyan alakul egy tipikusnak tekinthető magyarországi iskolai ingázási körzeten (járáson, illetve városon) belül a tanulók iskolák közötti elkülönülésének időbeli trendje, az egyes körzetek méretétől *függetlenül*.

4. ábra
Az iskolák közötti társadalmi elkülönülés mérőszámainak alakulása, 2006–2019
(a járások, illetve városok szintjén mérve, súlyozatlan átlagok, 8. évfolyamos tanulók)

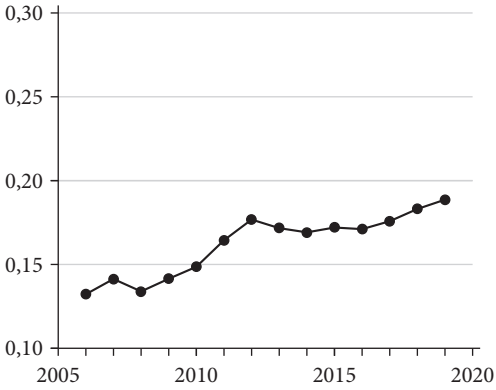


[7] Budapestet összevontan egyetlen járásnak tekintjük, a városok közül csak azok jönnek számításba, ahol legalább két iskola működik. Egyiskolás városokban nem beszélhetünk iskolák közötti szegregációról. A városok száma 2008-ban összesen 306 volt, ezek közül 174 városban működött a teljes időszakban legalább két iskola.

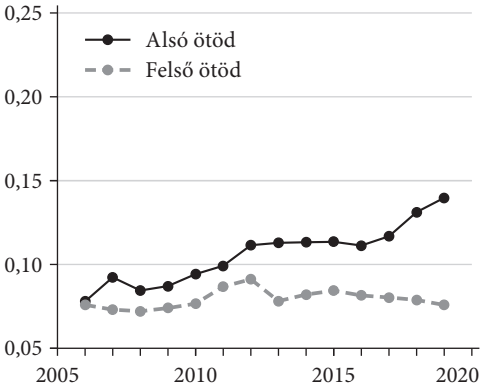
A 4. ábra folytatása

Városok

C) Iskolák közötti variancia aránya



D) Szegregációs index



Megjegyzés: az egyes járásokra, illetve városokra számolt szegregációs mutatók adott évre vonatkozó *súlyozatlan* átlagai. Budapest egyetlen járásként szerepel ($N = 175$). Csak a teljes időszakban legalább két iskolával rendelkező városokra számoltunk szegregációs mutatókat ($N = 174$).

A területi koncentráció változásának hatásától megtisztított becsléseink megerősítik a *3. ábra* alapján megfogalmazott megfigyeléseket: a szegregáció növekedése nem pusztán összetétel-változások következménye; a szegregáció az iskolapiacokon *belül* is valóban jelentős mértékben és folyamatosan növekedett. Ezek mögött a változások mögött azonban – mint az országos trendek bemutatásánál hangsúlyoztuk – társadalmilag heterogén folyamatok állnak: miközben a jómódú családok gyerekeinek iskolai elkülönülése a társadalom többi részétől az időszak egészében mérsékeltnek és nagyjából állandónak tekinthető, addig a szegény tanulók egyre nagyobb mértékű kiszorulása a nem szegény többség iskoláiból a magyar iskolarendszer egyik végzetes fejleményének tekinthető: olyan jelentős társadalmi változásnak, amely beláthatatlan ideig rontani fogja a szegény gyerekek felnőttkori életesélyeit.

A *4. ábrán* látható összefüggések robusztusak; akkor is fennállnak, ha a szegény és a jobb módú tanulók megkülönböztetését más definíció alapján végezzük el. Amint arról a *Függelék B3. ábrája* alapján meggyőződhetünk, a járások, illetve városok szintjén mért szegregációs indexek akkor is a *4. ábrához* hasonló, divergens trendeket követik, ha az alsó és felső ötödöket nem egységesen, az *országos* eloszlás alapján, hanem a járásokon, illetve városokon *belüli egyedi eloszlások* alapján határozzuk meg.

A szegregáció mértéke és a teszteredmények egyenlőtlensége közötti összefüggés

A szegregáció és a tanulói teljesítmények társadalmi státusz szerinti egyenlőtlensége között is van kapcsolat. A szegregált iskolák növelhetik a szegény diákok lemaradását, a nagyobb egyenlőtlenségek pedig fokozhatják a magasabb társadalmi státuszú

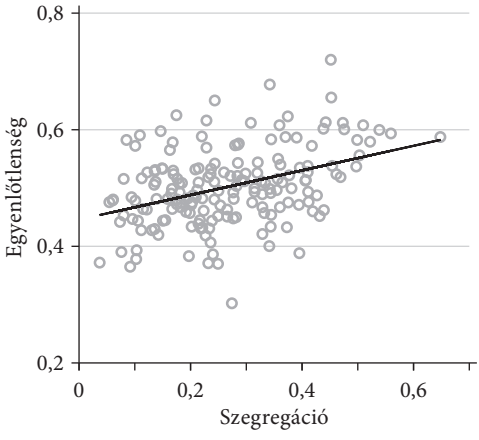
szülők elkülönülés iránti igényét. Empirikusan ezt az összefüggést a keresztmetszeti területi különbségek alapján mutatjuk be. *Card–Rothstein* [2007] megközelítését alkalmazva megvizsgáljuk, hogy azokban a térségekben, ahol erősebb a szegregáció, jellemzően nagyobbak-e a társadalmi státusz szerinti iskolai teljesítménybeli különbségek. Egyszerű kétváltozós összefüggéseket mutatunk be, amelyek későbbi részletesebb kutatások alapjául szolgálhatnak.

Az *5. ábra* alapjául szolgáló számítások megfigyelési egységei a járások és a többiskolás városok. Ezekre a területi egységekre számoltunk a 2017 és 2019 közötti időszak minden egyes évére szegregációs mérőszámokat, valamint kétváltozós teszt–háttérindex regressziókat, amelyeknek regressziós együtthatóival mérjük az illető területi egységet jellemző egyenlőtlenségeket. A szegregációs és egyenlőtlenségi mérőszámokat minden járásra és minden városra, továbbá a 2017 és 2019 közötti időszak minden egyes évére kiszámoltuk, majd az így kiszámított értékeket a 2017–2019-es időszak éveire átlagoltuk, hogy méréseink véletlenszerű hibáit csökkentsük. Az *5. ábra* paneljeinek vízszintes tengelyén a szegregáció mértékét az iskolák közötti variancia részarányával mérjük. A függőleges tengelyen pedig az egyenlőtlenségi mutatók értékeit (a regressziós paraméterértékek átlagait) ábrázoljuk. Az *5. ábra A)* és *B)* panelje a járások, *C)* és *D)* panelje pedig a városok szintjén mutatja be a szegregáció mértéke és a teszteredmények egyenlőtlensége közti összefüggést a matematika és a szövegértés teszteredményeire vonatkozóan. Az összefüggést a két változóra illesztett lineáris regressziós egyenesek mutatják.

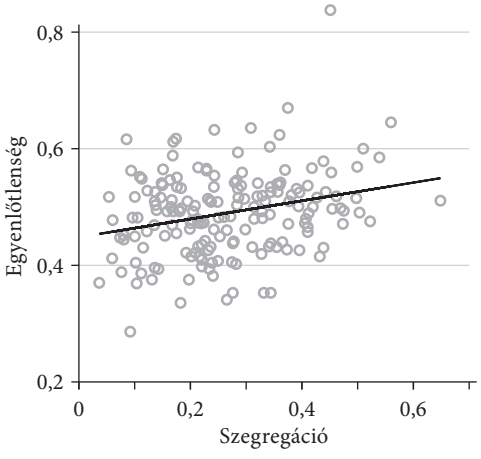
5. ábra
A szegregáció és a teszteredmények társadalmi egyenlőtlenségének összefüggése, 2017–2019 (a járások, illetve városok szintjén mérve, 8. évfolyamos tanulók)

Járások

A) Szövegértés



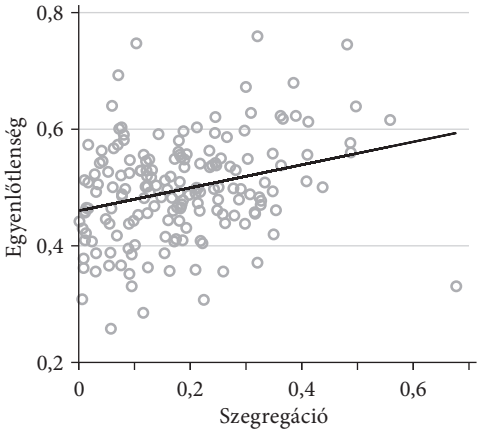
B) Matematika



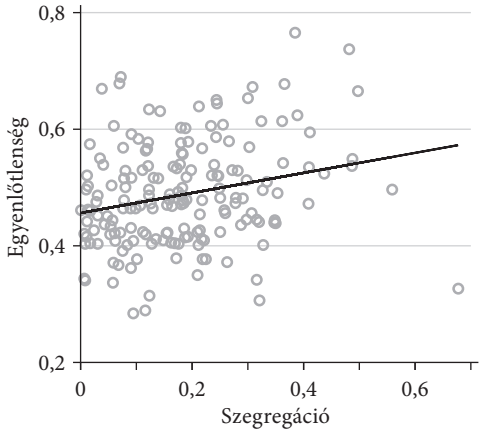
Az 5. ábra folytatása

Városok

C) Szövegértés



D) Matematika



Szegregáció: a társadalmi háttérindex iskolák közötti varianciájának aránya a teljes varianción belül. *Egyenlőtlenség*: a mutatót kétváltozós regressziós modellekből becsültük, amelyekben a függő változó a standardizált tesztpontszám, a jobb oldali változó pedig az évenként és évfolyamonként standardizált társadalmi háttérindex.
Megjegyzés: az ábrán az egyes járásokra és városokra a 2017–2019 közti időszak minden egyes évére számolt értékek súlyozatlan átlagai szerepelnek. Budapest összevontan egyetlen járásként szerepel ($N = 175$). Csak a teljes időszakban legalább két iskolával rendelkező városokat vettük figyelembe ($N = 174$).

Az eredmények azt mutatják, hogy az erősebb szegregáció a teszteredmények nagyobb egyenlőtlenségével jár együtt mind a járások, mind a városok szintjén. Az összefüggés minden esetben szignifikáns, erősségét tekintve pedig közepesnek mondható. Nyilvánvalóan mindkét tényezőt a járások és városok számos egyéb, itt figyelmen kívül hagyott jellemzője is befolyásolhatja. A korrelációs együtthatók (p-értékek) a járások esetében a szövegértésre 0,39 (0,0000), a matematikára pedig 0,27 (0,0004); ugyanezek a számok a városokra: 0,29 (0,0001) és 0,24 (0,0014).

Összegzés

Tanulmányunkban a társadalmi háttér szerinti iskolai teljesítménykülönbségek és a szegregáció időbeli változását vizsgáltuk egyszerű leíró eszközökkel. Az eredmények alapján három megállapítást tehetünk. Először, 2006 és 2019 között jelentős mértékben nőtt az alacsony státuszú (az alsó ötödbe tartozó) diákok iskolai szegregációja. Másodszor, növekedtek a teszteredmények társadalmi helyzet szerinti különbségei. Harmadszor, azokban a járásokban és városokban, ahol erősebb szegregációt mértünk, jellemzően nagyobbnak találtuk a teszteredmények társadalmi egyenlőtlenségeit is.

Az e megfigyelések mögötti folyamatok megértése további részletes elemzéseket igényel, pusztán ezekből a megállapításokból nem vonhatunk le messzemenő következtetéseket. Különösen igaz ez a szegregáció és a teszteredmények egyenlőtlenségei közötti összefüggésre: a megfigyelt keresztmetszeti korreláció alapján nem állíthatjuk, legfeljebb további kutatások számára igazolandó hipotézisként fogalmazhatjuk meg, hogy a szegregáció növekedése is hozzájárulhatott az iskolai egyenlőtlenségek növekedéséhez. Mindezen fenntartások mellett kimondhatjuk: a szegény tanulók iskolai szegregációjának drámai mértékű növekedése az oktatási rendszeren túlmutató, nagy és megoldásra váró társadalmi problémákra hívja fel a figyelmet.

Hivatkozások

Card, D.–Rothstein, J. [2007]: Racial segregation and the black-white test score gap. Journal of Public Economics, Vol. 91. No. 11–12. 2158–2184. o. https://doi.org/10.1016/j.jpubeco.2007.03.006.

Clotfelter, C. T. [1999]: Public school segregation in metropolitan areas. Land Economics, Vol. 5. No. 4. 487–504. o. https://doi.org/10.2307/3147061.

Clotfelter, C. T. [2004]: After Brown: The rise and retreat of school desegregation. Princeton University Press, Princeton–Oxford.

Csapó Benő–Molnár Gyöngyvér–Kinyó László [2009]: A magyar oktatási rendszer szelektivitása a nemzetközi összehasonlító vizsgálatok eredményeinek tükrében. Iskolakultúra, 19. évf. 3–4. sz. 3–13. o.

Havas Gábor–Liskó Ilona [2006]: Szegregáció a roma tanulók általános iskolai oktatásában. Megjelent: *Havas Gábor–Liskó Ilona*: Óvodától a szakmáig. Új Mandátum Könyvkiadó, Budapest, 40–127. o.

Kertesi Gábor–Kézdi Gábor [2005]: Általános iskolai szegregáció I. Okok és következmények. Közgazdasági Szemle, 52. évf. 4. sz. 317–355. o.

Kertesi Gábor–Kézdi Gábor [2009]: Általános iskolai szegregáció Magyarországon az ezredforduló után. Közgazdasági Szemle, 61. évf. 11. sz. 959–1000. o.

Kertesi Gábor–Kézdi Gábor [2012]: Ethnic segregation between Hungarian schools. Long-run trends and geographic distribution. Statisztikai Szemle, 90. évf. 16. sz. 18–45. o.

OECD [2018]: Equity in education. Breaking down barriers to social mobility. OECD, Párizs.

Schütz, G.–Ursprung, H. W.–Wössmann, L. [2008]: Educational policy and equality of opportunity. Kyklos, Vol. 61. No. 2. 279–308. o. https://doi.org/10.1111/j.1467-6435.2008.00402.x.

Varga Júlia (szerk.) [2022*a*]: A közoktatás indikátorrendszere, 2021. KRTK Közgazdaságtudományi Intézet, Budapest.

Varga Júlia (szerk.) [2022*b*]: Technikai útmutató „A közoktatás indikátorrendszere, 2021" című kiadványhoz. KRTK Közgazdaságtudományi Intézet, Budapest.

Függelék

A) A társadalmi háttérindex létrehozása

Miért van szükség társadalmi háttérindexre?

Társadalmi háttérindex mutatóra azért van szükség, mert a lehetséges alternatív megoldás, hogy a gyerekek társadalmi státuszát szüleik iskolai végzettségével mérjük, hosszabb időbeli összehasonlítások céljaira korlátozottan alkalmas. Ugyanaz az iskolai végzettség egymástól távolabbi időpontokban nagyon eltérő társadalmi helyzeteket jelenthet. Például: amíg az 1990-ben született tanulók 0–8 osztályos végzettségnél magasabb iskolai végzettséget nem szerzett szülei a társadalom nagyméretű és meghatározó csoportját alkották, addig a 2005-ös születési kohorsz hasonló iskolai végzettségű szülei egy méretében jóval kisebb és marginálisabb társadalmi csoportot képeznek. A 0–8 osztályt végzett szülők gyermekeit tizenöt év távlatában már *nem* tekinthetjük egymással jól összehasonlítható társadalmi csoportoknak.

Időbeli összehasonlításokhoz időben egységes és stabil mércét szabad használni. Ilyen mérce lehet például az egy főre jutó családi jövedelem, pontosabban: az e jövedelemmutató alapján számított *relatív* jövedelmi helyzet. A társadalom mindenkori jövedelemeloszlásában az egyes decilis- vagy percentilisértékek mindig ugyanazt a társadalmi csoportot jelenítik meg: a legalsó decilis a mindenkori legszegényebbeket, a legfelső a mindenkori leggazdagabbakat. Az azonos módon meghatározott társadalmi csoportok *időben változó* tulajdonságainak összehasonlítása értelmes feladat. A szülők iskolázottsága közvetlenül nem alakítható át a *relatív* társadalmi helyzetet pontosan mérő, időben egységes mutatóvá, mivel ez részletes és folytonos skálán mérő változót igényel, amilyen a jövedelem.

A legtöbb társadalmi célú adatgyűjtésben, különösen az adminisztratív célú adatgyűjtésekben, nem áll rendelkezésre jövedelemadat. Számos adatfelvétel esetében azonban, második legjobb megoldásként, rendelkezésre állnak a családok relatív jövedelmi helyzetével jól korrelált családi háttérinformációk. Az Országos Kompetenciamérés előnyös tulajdonsága, hogy viszonylag sok, jövedelemmel korrelált háttéradatot tartalmaz a tanulók családjairól. Ezekből az adatokból létrehozható egy szintetikus mutató – a társadalmi háttérindex –, amely a *relatív jövedelmi helyzet közelítő változójaként* folytonos és monoton módon jellemzi a társadalom hierarchikus különbségeit. Az *A) Függelék* végén egy speciális mintán megmutatjuk, hogy az általunk létrehozott mutató erősen együtt mozog az érintett tanulók családjainak egy főre jutó jövedelmével: így jó közelítő jövedelemmutatónak tekinthető.

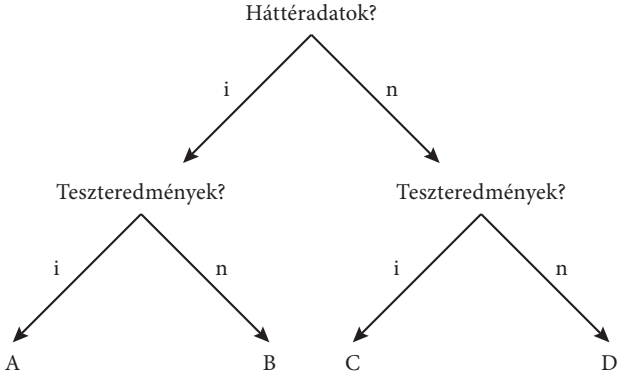
Az általunk alkalmazott eljárás, *alapelveit* tekintve, nem tartalmaz újdonságokat. Hasonló elven készít társadalmi háttérindexeket az Oktatási Hivatal a kompetenciamérés eredményeinek, illetve az OECD a PISA-vizsgálat eredményeinek értékelésekor. Eljárásunk *részletei* azonban, megítélésünk szerint, több újdonságot tartalmaznak, amelyekre a kifejtés során felhívjuk a figyelmet.

A társadalmi háttérindex létrehozásának lépései

Elsőként egy egyesített keresztmetszeti adatbázist hoztunk létre a 2006 és 2019 közötti Országos Kompetenciamérések 6., 8. és 10. évfolyamos tanulóinak olvasási és matematikai teszteredményeiből, illetve családi háttéradataiból. Az adatfelvétel a 2006. és 2007. évben csak a 8. évfolyamos tanulókat mérte fel, a 2008. évtől kezdve azonban a három évfolyam mindegyikére kiterjedt. 2008 óta minden tanuló egyéni szintű azonosítót kap, ami lehetővé teszi, hogy az egyéni tanulói teljesítményeket időben, évfolyamok között kövessük. Ezekre az elsődleges információkra támaszkodva próbáljuk a teljeskörűség igényével a közoktatás 6., 8. és 10. évfolyamain megjelenő tanulókat egy folytonos társadalmi mutatószámmal jellemezni.

Egy teljeskörűségre igényt tartó társadalmi háttérmutató létrehozásának legnagyobb nehézségét a hiányzó információk viszonylag nagy száma jelenti. A háttérkérdőívet évente a diákok ötöde nem tölti ki, illetve a diákok 7–10 százaléka – a sajátos nevelési igényű diákok többsége – nem ír teszteket [lásd lejjebb az *A4. táblázat* (2) és (3) oszlopát]. A kompetenciamérés mérési keretéhez tartozó diákokat az adathiányok szempontjából négy halmazra bonthatjuk (lásd *A1. ábra*). E négy halmaz sajátosságai meghatározzák, milyen eljárást követhetünk.

A1. ábra
A kompetenciamérés mérési keretéhez tartozó tanulók részhalmazai



Számítási eljárásunk az alábbi logikát követi.

Első lépés: adott év adott évfolyamára járó diákok hiányzó háttéradatait feltöltjük azon naptári évek adataival, amelyekben ugyanazon diák szolgáltatott családi háttéradatokat.

Második lépés: a családi háttéradatokkal rendelkező tanulók ($A \cup B$) adataiból háttérindexet készítünk.

Harmadik lépés: a családi háttéradatokkal nem rendelkező tanulók (C) esetében más információkat bevonva háttérindexértékeket imputálunk.

Negyedik lépés: egyesítjük az $A \cup B$ és C halmazokra számított háttérindexadatokat.

A hiányzó családi háttéradatok pótlása

A társadalmi háttérindex létrehozásához a tanulói kérdőív[8] családi háttérinformációiból az *A1. táblázatban* felsorolt adatokat használtuk.

A1. táblázat
A társadalmi háttérindex létrehozásához felhasznált változók

Változó	Kódértékek
Az anya, apa legmagasabb iskolai végzettsége	0–8 osztály szakiskola érettségi főiskola egyetem
Iskolai szociális transzferek – kedvezményes étkezés – ingyenes étkezés – ingyenes tankönyv	 van, nincs van, nincs van, nincs
Iskolán kívüli szociális transzferek – gyermekvédelmi kedvezmény	 van, nincs
Családi vagyontárgyak – mobiltelefon – számítógép – személygépkocsi – fürdőszoba	Hány van belőlük? 0, 1, 2, 3+ 0, 1, 2, 3+ 0, 1, 2, 3+ 0, 1, 2, 3+
Családi nyaralás	hányszor? soha egyszer-kétszer többször minden nyáron
Otthoni internetkapcsolat	van, nincs
A lakásban levő könyvek száma	kevesebb mint egypolcnyi (körülbelül 0–50) egypolcnyi (körülbelül 50) 2-3 könyvespolcnyi (maximum 150) 4-6 könyvespolcnyi (maximum 300) 2 könyvszekrényre való (300–600) 3 vagy több könyvszekrényre való (600–1000) 1000-nél több
Családjának életszínvonala	átlag alatti átlagos átlag fölötti
Környéken lakó családok többsége	szegény átlagos jómódú

[8] https://www.oktatas.hu/kozneveles/meresek/kompetenciameres/hatterkerdoivek.

A hiányzó adatok problémáját első lépésben úgy hidaltuk át, hogy kihasználtuk az Országos Kompetenciamérés adatainak paneltulajdonságait, nevezetesen azt, hogy az évismétlő diákok különböző naptári évek azonos évfolyamán többször szerepelnek, illetve azt, hogy alapesetben minden diákot többször (6., 8. és 10. évfolyamos korában) is felmérnek; a hiányzó adatokat így egy másik naptári évből származó családi háttéradatokkal pótolhatjuk.[9] Az adatpótlás során először az évismétlő diákok hiányzó háttéradatait töltöttük föl a szomszédos naptári évek megfelelő adataival, ezt követően pedig a más évfolyamokon meglevő hasonló adatokat használtuk ugyanerre a célra.

Társadalmi háttérindex számítása a háttéradatokkal rendelkező diákokra

Minden több dimenzióból összetett mutató képzésének alapkérdése, hogy a mutató alapjául szolgáló egyedi indikátorokat milyen súlyozási eljárással aggregáljuk. Mi abból az elvből indultunk ki, hogy az egyes indikátorokat a többi indikátorhoz képest olyan mértékben célszerű figyelembe venni, amekkora társadalmi jelentősége van az illető mutatónak a többihez képest. Tartalmi szempontból ez azt jelenti, hogy az egyes indikátorok relatív fontosságát a súlyozásnál nem valamilyen *ad hoc* választás alapján döntjük el, hanem egy mindenki által fontosnak tekintett társadalmi kimeneti változóra való tekintettel. Ilyen fontos kimeneti változó az Országos Kompetenciamérésben egyedül a teszteredmény lehet.

Technikailag ezt azt jelenti, hogy az aggregáláshoz szükséges súlyokat egy olyan regressziós egyenlet paramétereiből vesszük, amelyben a teszteredményeket a releváns családi háttéradatokon (az *A1. táblázatban* szereplő változókon) regresszáljuk. Mivel a szövegértési teszteredmények a matematikai teszteredményeknél erősebben függnek össze a társadalmi háttérváltozókkal, ezért bal oldali változónak a szövegértési teszteredményeket választottuk. Ilyen becsléseket nyilvánvalóan a diákoknak csak arra a halmazára lehet készíteni, akik *egyaránt* rendelkeznek családi háttéradatokkal és olvasásteszteredményekkel (*A* halmaz). Ugyanakkor az ezekből a becslésekből származó paramétereket *súlyokként* a családi háttéradatokkal rendelkező diákok azon részhalmazára is használhatjuk, akiknek nincsenek teszteredményeik (*B* halmaz).

A súlyok képzéséhez az alábbi egyenletet becsüljük a kompetenciamérésben részt vevő 6. és 8. évfolyamos diákok *A* részhalmazán a 2006 és 2019 közötti adatfelvételek egyesített keresztmetszeti adatain.

$$T_{itg} = \alpha + \sum_k \beta^k X^k_{itg} + \varepsilon_{itg}, \qquad i \in A, \qquad t = 2006 - 2019, \qquad g = 6|8. \tag{1}$$

A függő változó a *t*-edik év *g*-edik évfolyamára járó tanulók (*i*) szövegértési teszteredményeinek (*T*) standardizált értéke (a standardizálást évenként és évfolyamonként külön-külön elvégezve), *X*-szel jelölve az *A2. táblázatban* felsorolt családi háttérinformációk *k* elemű vektorát.

[9] Adathiányok esetén az *A1. táblázatban* szereplő valamennyi információt megpróbáltuk pótolni. Egyedül az internetkapcsolatra vonatkozó kérdésnél nem pótoltuk a hiányzó adatokat.

Az (1) egyenletet a teljes időszak egyesített keresztmetszeti adatain becsültük meg annak ellenére, hogy a családi háttér – elsősorban a szülői iskolázottság – szerinti összetétel időben változott, és így a megfigyelt jellemzők teszteredményekkel való összefüggése is változhatott valamelyest. A súlyok így a családi háttérjellemzőknek a teszteredményekhez való *átlagos* hozzájárulását tükrözik a teljes időszakban. Egy időben egységes és stabil társadalmi helyzetet tükröző mutató időben változatlan súlyokat követel meg.

Az (1) egyenlet becsléséből kizártuk a 10. évfolyamos tanulókat, továbbá a másik két évfolyamról a sajátos nevelési igényű diákok többségét.[10] A 10. évfolyamos tanulókat abból a megfontolásból hagytuk ki, hogy elkerüljük egy társadalmilag előzetesen erősen szelektált sokaság részvételét a becsléseinkben. Az érintett életkori csoport egy része 16 évesnél idősebbként túl van a kötelező iskoláztatási korhatáron, és korai lemorzsolódottként nincs jelen a kompetenciamérés mérési keretében. A lemorzsolódás a családi háttérváltozóknak is függvénye lehet, ráadásul időben változó mértékben, ami torzíthatná az (1) egyenletből számított paraméterek értékét.

Az *A2. táblázat* mutatja az (1) egyenlet becslési eredményeit.

A2. táblázat
Az (1) regressziós egyenlet becslési eredményei (függő változó: olvasás-szövegértés teszteredmények)

Változók	Együttható	Standard hiba
Anya iskolai végzettsége		
Szakiskola	0,173***	(0,003)
Érettségi	0,339***	(0,004)
Főiskola	0,452***	(0,005)
Egyetem	0,569***	(0,006)
Apa iskolai végzettsége		
Szakiskola	0,147***	(0,003)
Érettségi	0,291***	(0,004)
Főiskola	0,390***	(0,005)
Egyetem	0,506***	(0,006)
Kedvezményes étkezés	–0,003	(0,003)
Ingyenes étkezés	–0,101***	(0,004)
Ingyenes tankönyv	–0,038***	(0,003)
Rendszeres gyermekvédelmi kedvezmény	–0,060***	(0,003)
Mobiltelefonok száma	0,054***	(0,002)
Számítógépek száma	0,014***	(0,001)
Személygépkocsik száma	–0,012***	(0,001)
Fürdőszobák száma	0,017***	(0,002)

[10] Kivételt képeznek az úgynevezett beilleszkedési, tanulási és magatartási zavarokkal rendelkező sni diákok (btm-tanulók), akiket bevontunk az elemzésbe, de ezt a jellemzőjüket kontrollváltozóval nem jelöltük meg. A btm-tanulókat 2010-től lehet a kompetenciamérésben megkülönböztetni.

Az A2. táblázat folytatása

Változók	Együttható	Standard hiba
Családi nyaralás az elmúlt négy évben		
Egyszer-kétszer	0,032***	(0,003)
Többször	0,036***	(0,003)
Minden nyáron	0,074***	(0,003)
Otthoni internetkapcsolat	0,093***	(0,003)
Lakásban levő könyvek száma		
kb. 50	0,157***	(0,003)
50–150	0,305***	(0,003)
150–300	0,405***	(0,004)
300–600	0,491***	(0,004)
600–1000	0,564***	(0,004)
1000+	0,600***	(0,006)
Családjának életszínvonala		
Átlagos	–0,023***	(0,003)
Átlag fölötti	–0,060***	(0,004)
A környéken lakó családok többsége		
Átlagos	0,095***	(0,004)
Jómódú	0,010**	(0,005)
Az anya iskolai végzettsége (hiányzó érték)	0,223***	(0,009)
Az apa iskolai végzettsége (hiányzó érték)	0,164***	(0,006)
Kedvezményes étkezés (hiányzó érték)	–0,121***	(0,005)
Ingyenes étkezés (hiányzó érték)	–0,125***	(0,005)
Ingyenes tankönyv (hiányzó érték)	–0,070***	(0,006)
Gyermekvédelmi kedvezmény (hiányzó érték)	0,034***	(0,005)
Mobiltelefonok száma (hiányzó érték)	0,368***	(0,012)
Számítógépek száma (hiányzó érték)	–0,033***	(0,009)
Személygépkocsik száma (hiányzó érték)	–0,182***	(0,007)
Fürdőszobák száma (hiányzó érték)	0,101***	(0,009)
Családi nyaralás (hiányzó érték)	–0,077***	(0,005)
Otthoni internetkapcsolat (hiányzó érték)	–0,071***	(0,004)
Lakásban levő könyvek száma (hiányzó érték)	0,108***	(0,011)
Családjának életszínvonala (hiányzó érték)	0,071***	(0,007)
Környéken lakó családok (hiányzó érték)	0,120***	(0,008)
Konstans	–1,114***	(0,016)
A megfigyelések (tanulók) száma	2 119 484	
R^2	0,287	

Függő változó: évenként és évfolyamonként standardizált teszteredmények.
Minta: 6. és 8. évfolyamos tanulók 2006 és 2019 közötti egyesített keresztmetszeti mintája.
Megjegyzés: zárójelben iskolaszinten klaszterezett standard hibák.
*** $p < 0{,}01$, ** $p < 0{,}05$, * $p < 0{,}1$.

Az *A2. táblázat* paraméterei alapján számolunk társadalmi háttérindex-értékeket ($\hat{S}_{itg}$) azokra a diákokra, akiknek vannak egyedi családi háttéradatai ($A \cup B$ halmaz) – a 6., 8. és 10. évfolyamra járó tanulókat egyaránt beleértve. A (2) egyenletben a jobb oldali változók paraméterei fölött kalappal jelöltük azt a tényt, hogy a számításhoz felhasznált paraméterek az *A2. táblázatban* szereplő *becsült* értékek.

$$\hat{S}_{itg} = \hat{\alpha} + \sum_k \hat{\beta}^k X^k_{itg}, \quad i \in (A \cup B), \quad t = 2006 - 2019, \quad g = 6|8|10. \tag{2}$$

Társadalmi háttérindex imputálása háttéradatokkal nem rendelkező diákokra

A diákok nem elhanyagolható részére még ezzel együtt sem tudtuk a társadalmi háttérindex értékeit kiszámolni. Ezek azok a tanulók, akikhez még az adatfeltöltések után sem találtunk családi háttéradatokat, viszont rendelkeznek teszteredményekkel (*C* halmaz).

Ezekhez a diákokhoz a teszteredmény és a családi háttérindex *A* halmazon megfigyelhető *együttes eloszlásának tulajdonságaiból* és a *C* halmazon mért *tényleges* teszteredményekből lehetséges társadalmiháttérindex-értékeket rendelni. Ehhez elegendő azzal a feltételezéssel élni, hogy az *A* halmazon megfigyelt összefüggés a családi háttérindex ($\hat{S}_{itg}$) és a teszteredmények (T^r_{itg}, T^m_{itg}) között a *C* halmazon is ugyanúgy érvényes. Az összefüggés mintázatát az *A* halmazon a (3) egyenlet írja le. A kontrollváltozók közé felvettük a nem (*L*: lány) és a halmozottan hátrányos helyzet (*H*) mutatóját, *r* és *m* felső indexek jelzik az olvasásteszt, illetve a matematikateszt eredményeit. A teszteredmények mutatóit naptári évenként és évfolyamonként standardizáltuk.[11] A sajátos nevelési igényű tanulók ebben a becslésben sem szerepelnek, és a 10. évfolyamos diákokat ugyanazon oknál fogva kihagytuk a becslésből, mint ami miatt kihagytuk őket az (1) egyenlet becsléséből. Az együttes eloszlás tulajdonságait a (3) egyenlet paraméterei írják le.

$$\hat{S}_{itg} = \delta + \sum_l \theta^l T^l_{itg} + \kappa L_i + \lambda H_i + \mu_{itg}, \quad i \in A, \quad t = 2006 - 2019, \quad g = 6|8, \quad l = r|m. \tag{3}$$

A becsléseket három különböző időszakra kellett elkészítenünk, tekintve, hogy a 2006–2007 közötti időszakban nem állt rendelkezésre információ a halmozottan hátrányos helyzetű tanulókról, a 2008–2013-as és a 2014–2019-es időszakok között pedig erősen megváltozott a halmozottan hátrányos helyzet hivatalos definíciója. Becslési eredményeinket az *A3. táblázatban* közöljük.

Az *A3. táblázatban* megbecsült paraméterek felhasználásával imputáljuk a társadalmi háttérindex értékeit a teszteredményekkel igen, de primer családi háttéradatokkal nem rendelkező diákokra (a *C* halmaz elemeire). Az imputálást nem végezzük el a sajátos nevelési igényű tanulókra, a társadalmi háttérindex és a teszteredmények *A* halmazon megfigyelt együttes eloszlása ugyanis nem feltétlenül esik egybe azzal az eloszlással, amelyet – ha tényadatokkal rendelkeznénk – az ő esetükben megfigyelhetnénk.

[11] A becslésben minden olyan *C* halmazbeli tanulót szerepeltettünk, akinek volt legalább egyfajta teszteredménye. A hiányzó változóértékeket ez esetben is külön (hiányzó) változóval jelöltük meg.

A3. táblázat
A (3) regressziós egyenlet becslési eredményei (függő változó: társadalmi háttérindex)

	2006–2007	2008–2013	2014–2019
	(1)	(2)	(3)
Matematikateszt	0,122***	0,090***	0,103***
	(0,003)	(0,002)	(0,002)
Olvasásteszt	0,173***	0,164***	0,164***
	(0,003)	(0,002)	(0,002)
Lány	–0,041***	–0,021***	–0,014***
	(0,002)	(0,001)	(0,001)
Halmozottan hátrányos helyzetű		–0,657***	–0,631***
		(0,004)	(0,006)
Matematikateszt (hiányzó érték)	0,026	–0,014	–0,050**
	(0,073)	(0,024)	(0,022)
Olvasásteszt (hiányzó érték)	–0,128	–0,090***	–0,039
	(0,080)	(0,024)	(0,043)
Konstans	0,027***	0,100***	0,120***
	(0,004)	(0,003)	(0,004)
A megfigyelések (tanulók) száma	177 231	1 022 310	920 381
R^2	0,301	0,440	0,407

Függő változó: a (2) egyenlet által becsült háttérindexértékek.
Minták: 6., 8. évfolyamos tanulók 2006–2007, 2008–2013, 2014–2019 közötti egyesített keresztmetszeti mintái. *Megjegyzés:* zárójelben iskolaszinten klaszterezett standard hibák.
*** $p<0{,}01$, ** $p<0{,}05$, * $p<0{,}1$.

A (4) egyenlet mutatja imputálási eljárásunk eredményét: a társadalmi háttérindex imputált értékeit tartalmazó változót $\left(\hat{S}^{*}_{itg}\right)$, amelyet a C halmaz elemeire számoltunk ki. A (4) egyenletben a jobb oldali változók paraméterei fölött kalappal jelöltük azt a tényt, hogy az imputáláshoz felhasznált paraméterek az *A3. táblázatban* szereplő *becsült* értékek. A 10. évfolyamos diákok esetében a C halmazra nem imputálunk változóértékeket.

$$\hat{S}^{*}_{itg}=\hat{\delta}+\sum_{l}\hat{\theta}^{l}T^{l}_{itg}+\hat{\kappa}L_{i}+\hat{\lambda}H_{i},\qquad i\in C,\qquad t=2006-2019,\qquad g=6|8,\qquad l=r|m. \quad (4)$$

Az $A\cup B$ és C halmazokra számított háttérindexadatok egyesítése

Eljárásunk utolsó lépéseként egyetlen változóban egyesítjük az $A\cup B$ és C halmazokra számított háttérindexadatokat [lásd (5) képlet].[12]

[12] A 10. évfolyamos tanulók esetében: $SES_{itg}=\hat{S}_{itg}$, $i\in(A\cup B)$, $g=10$.

$$SES_{itg} = \begin{cases} \hat{S}_{itg}, & i \in (A \cup B) \\ \hat{S}^{*}_{itg}, & i \in C \end{cases}, \quad g = 6|8. \tag{5}$$

Az ebből nyert társadalmi háttérindex (*SES*) a mérési keretben szereplő tanulók csaknem teljes halmazát lefedi. Az eljárás végén mért adathiányok átlagosan a diákok nagyjából 4-5 százalékát teszik ki, ami a kiinduló helyzetben mért 20 százalékhoz képest jelentős eredménynek tekinthető *(A4. táblázat).*

A4. táblázat
Az adathiányok mértéke az eljárás előtt és után (az adott évfolyam tanulóinak százalékában)

Év	Évfolyam	Nincs olvasási teszteredmény	Nincs társadalmi háttér-információ		
			kitöltetlen háttérkérdőívek	adatfeltöltés után	adatfeltöltés és imputálás után
	(1)	(2)	(3)	(4)	(5)
2006	8.	7,9	22,4	22,4	8,5
2007	8.	6,4	15,8	15,8	6,3
2008	6.	6,8	18,7	5,5	3,0
	8.	7,4	23,4	9,7	4,0
	10.	9,2	26,2	25,1	8,6
2009	6.	7,2	19,1	5,6	3,2
	8.	7,8	23,0	9,8	4,4
	10.	9,6	25,2	23,1	8,4
2010	6.	7,2	16,2	5,2	3,3
	8.	7,7	19,7	4,5	2,6
	10.	10,4	23,0	9,4	4,5
2011	6.	7,2	16,4	5,2	3,2
	8.	8,1	20,3	4,2	2,4
	10.	11,0	22,5	7,4	3,7
2012	6.	6,9	15,7	5,0	3,2
	8.	8,2	19,7	3,7	2,3
	10.	11,4	22,7	4,0	2,3
2013	6.	7,2	14,9	4,6	3,0
	8.	8,8	19,1	3,8	2,4
	10.	12,6	22,5	3,3	2,0
2014	6.	7,5	15,1	4,7	3,0
	8.	8,5	18,6	3,4	2,1
	10.	12,6	21,4	2,6	1,5

Az A4. táblázat folytatása

Év	Évfolyam	Nincs olvasási teszteredmény	Nincs társadalmi háttér-információ		
			kitöltetlen háttérkérdőívek	adatfeltöltés után	adatfeltöltés és imputálás után
	(1)	(2)	(3)	(4)	(5)
2015	6.	6,4	13,9	4,6	2,7
	8.	7,6	17,3	3,1	1,9
	10.	9,0	18,7	2,3	1,2
2016	6.	6,9	16,1	7,5	3,5
	8.	8,2	19,1	3,2	1,8
	10.	9,9	19,2	2,4	1,4
2017	6.	6,6	14,9	7,5	3,5
	8.	8,1	19,2	3,4	1,9
	10.	9,9	18,7	2,1	1,2
2018	6.	6,5	16,1	15,4	5,9
	8.	8,3	20,0	6,4	2,7
	10.	8,5	18,0	2,3	1,2
2019	6.	6,9	17,3	17,0	6,2
	8.	9,1	21,7	6,1	2,7
	10.	10,4	19,5	2,3	1,3

Az adathiány többsége azokból a tanulókból adódik, akiknek eleve nem voltak sem teszteredményei, sem családi háttéradatai (*D* halmaz), illetve a sajátos nevelési igényű diákokból, akiknek egy részét a szövegben megadott okok miatt kihagytuk a becslésekből.

Az adatfeltöltés és imputálás utáni hiány aránya a 2006. és 2007. évben azért viszonylag magasabb, mert a felmérés ezekben az években csak a 8. évfolyamos diákokra terjedt ki, és emiatt nem volt honnan feltölteni a hiányzó adatokat, az időszak *eleji* (2008–2009. évi) felmérések 10. évfolyamos diákjainál, illetve az időszak *végi* (2018–2019. évi) felmérések 6. évfolyamos diákjainál pedig azért, mert az előbbieknek még nem voltak *korábbi*, az utóbbiaknak pedig még nincsenek *későbbi* évfolyamokról származó háttéradatai. Eljárásunk, ha a kompetenciamérést egy állandósult állapotában – a 2010 és 2017 közötti időszak egy tetszőleges évében – figyeljük meg, az adathiányt nagyon alacsony (3 százalék körüli) értékre képes leszorítani.

A társadalmi háttérindex mutató validálása a Tárki Életpálya-felmérésének adatain

Az általunk létrehozott mutató nagyon erősen együtt mozog az érintett tanulók családjainak egy főre jutó jövedelmével: a *relatív jövedelmi helyzet* egyfajta *közelítő változójának* tekinthető. Ennek az állításnak az empirikus alátámasztására bemutatunk egy számítást

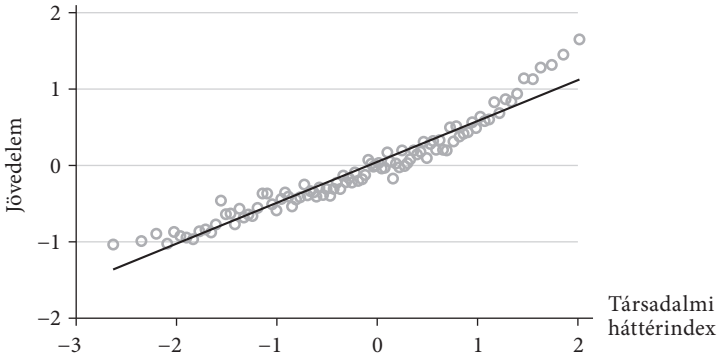
a Tárki Életpálya-felmérésének adatain,[13] amelyben *egyszerre* rendelkezésünkre állnak az Országos Kompetenciamérés családi háttéradatai, illetve – sok egyéb fontos információ mellett – a megfigyelt tanulók családjainak jövedelemadatai is.

Az Életpálya-felmérés a rendszerváltás utáni Magyarország egyik legnagyobb országosan reprezentatív társadalomtudományi adatfelvétele volt, amely tízezer, 2006 májusában 8. évfolyamra járó és az adott tanév kompetenciamérésén részt vevő tanuló életét követte nyomon középiskolai pályafutása során. Az Életpálya-felmérés referenciasokaságába tartoztak azok a sajátos nevelési igényű tanulók is, akikkel 2006-ban kivételesen egy egyszerűsített olvasás-szövegértés tesztet is kitöltettek. Az Életpálya-felmérés 2006 őszétől 2012 nyaráig tartó hat kérdezési hulláma összesen hat tanévet fogott át. Az adatfelvétel 2006. évi első hullámához a felmérésben részt vevő tanulók kompetenciaeredményeit és tanulói háttérkérdőívének adatait is hozzárendelték. Mindegyik adatfelvételi hullám során kérdeztek a család/háztartás anyagi helyzetére vonatkozó információkat.

Az *A2. ábrán* látható grafikon vízszintes tengelyén az Életpálya-felmérésben részt vevő tanulók társadalmi háttérindexének standardizált mutatója szerepel (szórásegységekben mérve), függőleges tengelyén pedig az érintett családok egy főre jutó jövedelmének standardizált mérőszáma (szintén szórásegységekben mérve). A jövedelemmutatót az Életpálya-felmérés két változójából konstruáltuk meg: a háztartási kiadások 2006. évi, illetve a háztartási jövedelmek 2007. évi értékei alapján,[14] főkomponens-elemzés segítségével.

Az *A2. ábrán* látható pontok a társadalmi háttérindex percentilisértékeihez rendelt átlagos jövedelemértékeket tartalmazzák.

A2. ábra
A társadalmi háttérindex és az egy főre jutó családi jövedelem összefüggése a Tárki Életpálya-felmérés adatain, 2006 (nyolcadik évfolyamos tanulók)



Vízszintes tengely: a társadalmi háttérindex standardizált értékeiből képzett percentilisek.
Függőleges tengely: a háttérindex standardizált percentilisértékeihez rendelt jövedelemmutató átlagos értékei (a mutató szórásegységeiben mérve).
Jövedelemmutató: a háztartási kiadások 2006. évi és a háztartási jövedelmek 2007. évi értékei alapján számolva, főkomponens-elemzéssel. $N \approx 9000$ tanuló.
Forrás: Tárki Életpálya-felmérés, 2006, 2007.

[13] https://adatbanktest.tarki.hu/cgi-bin/katalogus/tarkifo_hun.pl?sorszam=TDATA-H86.
[14] Mindkét indikátort fogyasztási egységekre vetítve és 2010. évi forintértéken mértük.

A két indikátor *egyéni* szintű korrelációja erős (0,573), nagyon erősen szignifikáns (0,00001, $N \approx 9000$); és a percentilisek szintjén aggregálva, jól illeszkedik. Noha az alsó és a felső háttérindexértékekkel némileg alulbecsüljük a tényleges jövedelmeket, az összefüggés az értelmezési tartomány azon szakaszain is monoton. Mindezekből a megfigyelésekből azt a következtetést szűrjük le, hogy a társadalmi háttérindex a relatív jövedelmi helyzet elég jó közelítő mutatójának tekinthető.

A társadalmi háttérindex használatával kapcsolatos további megfontolások

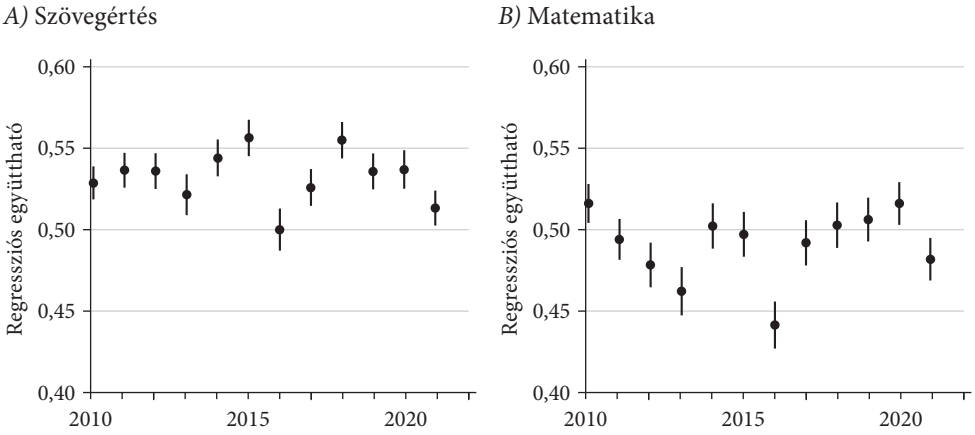
Amennyiben a társadalmi háttérindex mutatóját olyan kutatási célokra használjuk, amelyekben a teszteredményben mért tanulói teljesítmény a kimeneti változó, akkor az adatpótlásban csak a primer adatokból való adatfeltöltésig mehetünk el, és csak $A \cup B$ halmazra nézve számolhatunk mutatóértékeket. A teszteredményekből és a tesztpontszám–háttérindex együttes eloszlásából történő imputálás nem megengedett, a C halmaz kiegészítő információit nem lehet bevonni a változóképzésbe.[15] Az imputálás során ugyanis olyan változóértékek keletkeznek, amelyek maguk is függvényei a teszteredményeknek, ilyenformán a teszteredmények magyarázatára nem használhatók. Ezt a változóképzési korlátozást érvényesítettük a főszöveg *1.*, *2.* és *5. ábrájához* vezető számításokban.

Más kutatási célok esetében, ahol a teszteredményeket *nem* használjuk kimeneti változóként, és ahol nagyon fontos szempont, hogy adataink a célsokaságot minél *teljeskörűbben* lefedjék – például, ha a társadalmi háttérindex alapján akarunk szegregációs mutatókat minél kisebb területi egységekre számolni (mint például a főszöveg *4. ábrájához* vezető számításokban) –, az eljárás valamennyi lépését végigkövethetjük.

[15] Az *A4. táblázat* (4) és (5) oszlopának összehasonlítása mutatja, hogy ez mekkora adatveszteséget jelent.

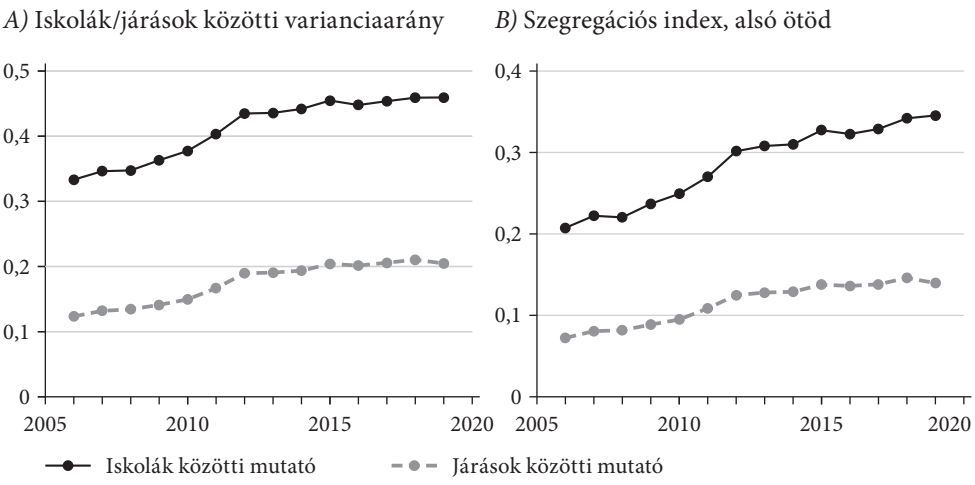
B) Kiegészítő eredmények

B1. ábra
A társadalmi háttér hatása a 6. évfolyamon mért tesztpontszámokra, 2008–2019 (kétváltozós standardizált regressziós együtthatók)



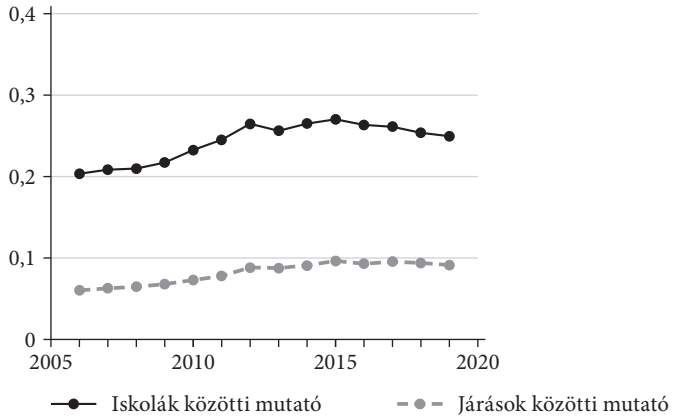
Megjegyzés: minden pont egy kétváltozós regressziós modell eredményét mutatja, ahol a függő változó a standardizált tesztpontszám, a jobb oldali változó pedig az évenként és évfolyamonként standardizált társadalmi háttérindex. Az ábra a becsült együtthatókat mutatja, 95 százalékos konfidenciaintervallumokkal (iskolaszinten klaszterezett standard hibák alapján).

B2. ábra
Az iskolák, illetve a járások közötti társadalmi elkülönülés országos mérőszámainak alakulása, 2006–2019 (8. évfolyamos tanulók)



A B2. ábra folytatása

C) Szegregációs index, felső ötöd

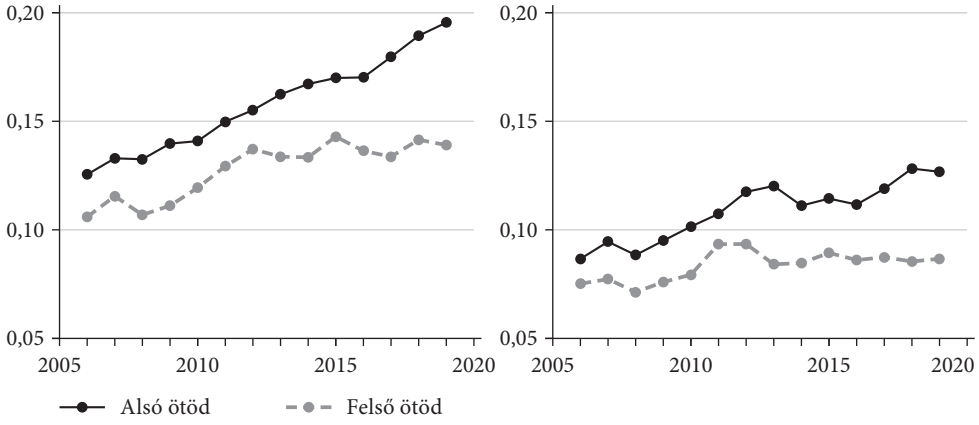


Megjegyzés: a számítási eljárásról lásd a 6. lábjegyzetet és a hozzá tartozó főszöveget.

B3. ábra
Az iskolák közötti társadalmi elkülönülést mérő szegregációs indexek alakulása az adott járáson, illetve városon *belüli relatív* társadalmi helyzet szerint, 2006–2019 (a járások, illetve városok szintjén mérve, 8. évfolyamos tanulók)

A) Járások

B) Városok



Megjegyzés: az alsó és felső ötöd a társadalmi háttérindex szerint az egyes járásokon, illetve városokon belüli ötödöket jelöli. Az egyes járásokra, illetve városokra becsült szegregációs mutatók az adott évre számolt súlyozatlan átlagok. Budapest egyetlen járásként szerepel ($N = 175$). Csak a teljes időszakban legalább két iskolával rendelkező városokra számoltunk szegregációs mutatókat ($N = 174$).