



MAGYAR TUDOMÁNYOS AKADÉMIA  
KÖZGAZDASÁGTUDOMÁNYI INTÉZET

---

**BUDAPESTI MUNKAGAZDASÁGTANI FÜZETEK**  
**BWP – 2010/2**

**MINTAVÉTEL ÉS ELEMZÉSI MÓDSZEREK AZ OKTATÁSI  
INTEGRÁCIÓS PROGRAM HATÁSVIZSGÁLATÁBAN,  
ÉS A HATÁSVIZSGÁLATBÓL LEVONHATÓ  
KÖVETKEZTETÉSEK**

**KÉZDI GÁBOR – SURÁNYI ÉVA**

Budapest Working Papers On The Labour Market  
Budapest Munkagazdaságtani Füzetek  
BWP – 2010/2

Mintavétel és elemzési módszerek az oktatási integrációs program hatásvizsgálatában, és a hatásvizsgálatból levonható következtetések

Magyar Tudományos Akadémia Közgazdaságtudományi Intézet  
Budapesti Corvinus Egyetem, Emberi Erőforrások Tanszék

Szerzők:

Kézdi Gábor  
associate professor  
Közép-európai Egyetem  
tudományos főmunkatárs  
Magyar Tudományos Akadémia Közgazdaságtudományi Intézet  
E-mail: kezdig@ceu.hu

Surányi Éva  
kutatási igazgató  
MATT Humán Tanácsadó Kft.  
E-mail: suranyi@matt-od.com

2010. április

ISBN 978 963 9796 98 0  
ISSN 1785 3788

Kiadja

a Magyar Tudományos Akadémia Közgazdaságtudományi Intézete

# **Mintavétel és elemzési módszerek az oktatási integrációs program hatásvizsgálatában, és a hatásvizsgálatból levonható következtetések**

Kézdi Gábor – Surányi Éva

## **Összefoglaló**

Az Országos Oktatási Integrációs Hálózat 2003 őszén egy olyan, általános iskolás gyerekekre irányuló programot indított útjára először 45 iskolában, melynek célja a hátrányos helyzetű gyerekek lemaradásainak ellensúlyozása és továbbtanulási esélyeinek növelése volt integrált és magas minőségű oktatási/nevelési környezet biztosításával. A program hatását vizsgáló kutatásunk módszereit és eredményeit a Magyar Tudomány 2010. márciusi számában súlyos bírálatok érték. Ez a tanulmány a hatásvizsgálat mintavételét és elemzési módszereit ért kritikákra válaszol. A nem-kognitív készségeket mérő eszközök kidolgozására irányuló bírálatokra külön cikkben válaszolunk, amely cikk önmagában véve érdekes módszertani tanulmány, függetlenül a bírálatokra adott válaszoktól. Tanulmányainkban részletesen kimutatjuk, hogy a bírálatok megalapozatlanok. Ebben a tanulmányban összefoglaljuk a hatásvizsgálat eredményeiből levonható következtetéseket is.

**Tárgyszavak:** integrált oktatás, program hatásvizsgálat, módszertan

**JEL:** I20, I38, C81

Válaszcikk Kabai Péternek a Magyar Tudomány című folyóirat 2010. márciusi számában megjelent „Egy sikeres iskolai integrációs program elemzése“ /Analysis of a Study on a Successful School Integration Program/ című – a szerzők általános iskolai integráció hatását elemző kutatását bíráló – cikkére.

# **Sampling and Methodology in the Evaluation of the National Education Integration Program, and the Conclusions of the Evaluation**

Gábor Kézdi – Éva Surányi

## **Summary**

The National Educational Integration Network (OOIH) launched a program in 2003 that focused on the integrated education of primary school children (grades 1 through 8) in 45 schools in Hungary. The goal of the program was to compensate for the educational disadvantages of children from poor and/or minority families by providing quality education in an integrated environment. The report on the impact evaluation of the program is available at Kézdi and Surányi (2009). The results and methods of the impact evaluation research were criticized in a paper published (in Hungarian) in *Magyar Tudomány*, March 2010. This study answers the criticism on sampling and methodology. A separate study addresses the issues of our measurement of non-cognitive skills. It is an interesting methodological study on its own right, regardless of the debate. In both paper we show that the critiques are unfounded. In this study we also summarize the conclusions from the impact evaluation.

**Keywords:** integrated education, program evaluation, methodology

**JEL:** I20, I38, C81

## BEVEZETŐ

Kabai Péter a Magyar Tudomány című folyóirat 2010. márciusi számában megjelent cikkében súlyos kritikákat fogalmazott meg egy az általános iskolai integráció hatását elemző kutatásunk vizsgálati módszereivel és eredményeivel kapcsolatban. Ez a tanulmány foglalkozik a hatásvizsgálat mintavételi eljárását és statisztikai módszereit ért kritikákkal. A nem-kognitív készségeket mérő eszközök kidolgozására irányuló bírálatokra a Budapesti Munkagazdaságtani Füzetek 2010/1. számában megjelent cikkünk válaszol.<sup>1</sup>

Az Országos Oktatási Integrációs Hálózat 2003 őszén egy olyan, általános iskolás gyerekekre irányuló programot indított útjára először 45 iskolában, melynek célja a hátrányos helyzetű gyerekek lemaradásainak ellensúlyozása és továbbtanulási esélyeinek növelése volt integrált és magas minőségű oktatási/nevelési környezet biztosításával. A program az egyes iskolákra fókuszált, természeténél fogva érintetlenül hagyta az iskolák közötti szegregáció problémáját. A program előírta a különböző családi háttérű tanulók integrált oktatását iskolán belül, lényegi eleme pedig az ilyen környezetben folyó hatékony és előítéletektől mentes oktatás elősegítése volt. Ennek érdekében a pedagógusok olyan képzéseken vettek részt, melyeken módjuk volt elsajátítani a modern, kompetenciaorientált, gyermekközpontú és kooperációt segítő oktatási módszerek használatát. Az iskolák emellett egy koordinátori hálózat működtetésén keresztül segítséget kaptak az iskolai élet hatékonyabb szervezéséhez, valamint az iskola és környezete közötti kapcsolat fejlesztéséhez. Kutatásunk e program hatását vizsgálta a tanulók fejlődésére és iskolai eredményességére. A kutatás 2005 tavaszán indult, amikor a program már másfél éve zajlott. 4000 tanulót követtünk nyomon három éven át, akik 2005-ben második, illetve hatodik osztályosak voltak, és eredményeiket 2007-ben mértük, negyedik, illetve a nyolcadik osztályos korukban. A program hatását vizsgáló kutatásunk beszámolóját 2008-ban tettük közzé (angol nyelven 2009-ben), lásd Kézdi-Surányi (2008a, 2008b, 2009).

---

<sup>1</sup> Kabai Péter egyik kritikái megjegyzése a referált folyóiratok és a "peer review" fontosságáról szól, és felrója nekünk, hogy kutatásunk eredményeit nem ilyen módon publikáltuk. Elveivel teljes mértékben egyetérünk, és törekszünk arra, hogy legfontosabb kutatási eredményeinket színvonalas referált nemzetközi folyóiratokban jelentessük meg. A társadalmi programokon elvégzett hatásvizsgálatok eredményeinek publikálása azonban sajátos utat követ: a kutatási beszámolók nem folyóiratokban jelennek meg, referált cikkek csak később születnek. A talán legismertebb hatásvizsgálat, a mexikói PROGRESA eredményeit is először kutatási beszámolók formájában publikálták (Gertler, 2000; vagy a számos műhelytanulmány eredményeire épülő összefoglalót, Skoufias, 2005), és csak később jelentek meg eredmények tudományos folyóiratokban (például de Janvry et al, 2006). Az okok között a terjedelmi szempontok mellett a társadalomtudományokban – különösen a közgazdaságtudományban – rendkívül lassú publikálási folyamat is szerepet játszik. Hatásvizsgálatunkkal mi is a nemzetközileg bevett utat tervezzük járni.

Általánosságban a társadalmpolitikai intervenciók hatásvizsgálatainak legfontosabb kérdése az, hogy milyen – társadalmi szempontból is jelentős – előnyökkel vagy esetleg hátrányokkal jár az adott kormányzati beavatkozás a programban résztvevők számára. A gyermekekre irányuló oktatáspolitikai beavatkozások esetén a leggyakrabban vizsgált eredmény a gyermekek fejlődése, későbbi iskolai sikeressége, néhány esetben családjuk helyzetének alakulása a program hatására.

Technikai értelemben tekintve a hatásvizsgálatoknak azt kell megbecsülniük, hogy a gyerekek hogyan fejlődtek, összehasonlítva azzal, ahogyan akkor fejlődhetek volna, ha nem vesznek részt a programban. A hatásvizsgálatok egyik legnagyobb módszertani nehézségét így az jelenti, hogy a „mi történt volna a gyerekekkel a program hiányában” eset nem mérhető közvetlenül. A kutatások ez a problémát úgy oldják meg, hogy a gyerekek fejlődésében bekövetkező változásokat alkalmasan választott kontrollcsoport teljesítményéhez viszonyítva értékelik, ami bizonyos feltevések mellett jól reprodukálja azt, hogy hogyan fejlődtek volna ugyanezek a gyermekek a nevelési környezet változatlanul hagyása esetén. Mindebből következően a hatásvizsgálatok megtervezésének módszertani értelemben két sarokpontja van: az első a mérési dimenziók és a hozzájuk tartozó mérőeszközök meghatározása (vagyis annak a kérdésnek a megválaszolása, hogy a gyerekek komplex fejlődési mintázatában mely összetevők azok, melyeknek mérése lehetővé teszi a program társadalmi hatásainak megfelelő nyomon követését); a második a kontrollcsoport megfelelő kiválasztása, ami egyben alapvetően meghatározza az alkalmazható becslési eljárásokat és az adatokból levonható következtetések hatókörét is. Ebben a cikkben ez utóbbi témakörrel foglalkozunk részletesen, elsősorban a Kabai Péter által bírált szempontokra koncentrálna. A mérőeszközök kifejlesztésének részleteit és Kabai Péter ezekre vonatkozó bírálataira adott válaszainkat egy másik - a Budapesti Munkagazdaságtani Füzetek sorozatban megjelent, illetve a Magyar Tudomány online kiadásában is elérhetővé tett - tanulmányban találhatják meg a hatásvizsgálat módszertana és a szakmai vita iránt érdeklődő olvasók.

## A KONTROLL CSOPORT KIVÁLASZTÁSA ÉS A PROGRAM HATÁS MÉRÉSE

A kontroll csoport kiválasztásának klasszikus természettudományos módszere a randomizált kísérlet, amelyben a potenciális résztvevőket véletlenszerűen osztják kezelt és kontroll csoportokba. A kiválasztás véletlenszerűsége miatt a két csoport összetétele gyakorlatilag azonos minden program előtti változóban, ezért a program után mért bármilyen különbség a program eredményének tekinthető. A társadalomtudományokban azonban a randomizált kísérletek a ritka, bár annál értékesebb kivételeket jelentik<sup>2</sup>. Nem randomizált hatásvizsgálatoknál általában adott az, hogy kik vesznek részt a programban, a kutatóknak erre nincs hatása (tipikus esetben a valamilyen szempontból jobbnak ítélt jelentkezőket választják ki a program lebonyolítói). A kontrollcsoport kiválasztásának és az elemzésnek a módszertana ilyenkor a randomizált kísérleteknél jóval bonyolultabb, és az eredmények interpretálása is problémásabb lehet. Mindez azonban nem jelenti azt, hogy a nem randomizált hatásvizsgálatok szükségszerűen alkalmatlanok a programok hatásának mérésére és a társadalompolitika általános informálására. Azt, hogy egy nem randomizált hatásvizsgálat eredményei milyen mértékben hihetőek, a változók megfelelő mérése mellett a kontroll csoport kiválasztása és az elemzés statisztikai-ökonometriai módszerei határozzák meg (Imbens és Wooldridge, 2009).

Az általunk vizsgált oktatási integrációs program esetében a programban résztvevő iskolákat a pályázatra jelentkezők közül válogatták ki, ahol a kiválasztás szempontja a várható sikeresség volt; hatásvizsgálatunk így szükségszerűen nem randomizált kiválasztásra épült. Tovább korlátozta a mozgásterünket az, hogy vizsgálatunk másfél évvel a program indulása után kezdődött, a programban való részvételben a hatásvizsgálat szempontjai így természetesen nem játszhattak szerepet. A résztvevő iskolák meghatározásának nem-random módját Kabai Péter is hangsúlyozza, sőt, a programba történő önszelekció miatt (ami a legtöbb hatásvizsgálat természetes velejárója) a mintavételt „torzítottnak”, az adatokat további elemzésre alkalmatlannak nyilvánítja, miközben a kontroll csoport kiválasztásának kérdését teljes mértékben figyelmen kívül hagyja. A szelekciós problémát pedig a hatásvizsgálatok a kontroll csoport kiválasztásának módszertani szigorúságával törekszenek kompenzálni, így ez a nem randomizált hatásvizsgálatokban az adatok megfelelő értelmezésének kulcskérdésévé válik (ezt kutatási

---

<sup>2</sup> A legismertebb példák a PROGRESA mellett a J-PAL program terep-kísérletei az M.I.T. egyetemen: <http://www.povertyactionlab.org/about-j-pal>. Praktikus okok miatt azonban többen vitatják a randomizált kísérletek felsőbbrendűségét a társadalompolitika megalapozásában – elismerve nyilvánvaló tudományos tisztaságukat ideális körülmények között –, lásd például Angus Deaton előadását (Deaton, 2009).

beszámolóink is részletesen ismerteti). Mi a kurrens módszertani ajánlások (Imbens és Wooldridge, 2009) szerinti optimális megoldások közül azt választottuk ki, amely esetünkre alkalmazható volt: propensity score alapján párosítottunk (Rosenbaum és Rubin, 1983). A párosítási eljárás lényege, hogy a vizsgált programiskolák mindegyikéhez olyan kontrolliskolát kerestünk, amely a rendelkezésre álló, program előtti állapotokat tükröző *valamennyi fontos, és egyben elérhető információ* szempontjából a lehető leginkább hasonlított az adott programiskolára. A modellben szereplő változók között a tanulók összetétele mellett az iskolák és a települések mérete, valamint kompetenciamérés eredményei szerepeltek. Magukat a változókat, azok átlagait a program- és a kontrolliskolákban és az országos átlagokat kutatási összefoglalóink 2.1. táblázata mutatja (Kézdi és Surányi, 2008a).

A kiválasztás arra a nagyon is figyelemreméltó eredményre vezetett, hogy a programiskolák és a kontrolliskolák tanulóinak összetétele szinte teljes mértékben megegyezett minden vizsgált ismérvben – ahogyan ezt Kabai Péter is elismeri. Az összetétel nemcsak a propensity score egyetlenben használt változók esetében volt gyakorlatilag azonos, ami a program előtti állapotokat tükrözte: a vizsgálatunkban szereplő *tanulók családi háttere és a programot megelőző iskolai eredményessége szerinti összetétele* is gyakorlatilag azonos volt valamennyi vizsgált ismérv esetében (Kézdi és Surányi, 2008a, 2.2. és 2.3. táblázatok). Nem egyszerűen arról van tehát szó, hogy a kontroll csoport kiválasztása megfelelő módszertant követett és megfelelő módon használta fel a rendelkezésre álló adatokat. A kiválasztás sikeres is volt, amit a tanulói összetétel utólagos összehasonlítása látványosan igazolt. A kiválasztás így olyan program-kontroll összehasonlításokra adott lehetőséget, amelyek alapján megvizsgálhattuk, hogy azonos háttérű tanulók mennyiben érnek el más eredményeket, ha a programiskolákba járnak. *A programban résztvevő és kontroll iskolák átlaga közötti „nyers” különbség a mintavétel sajátosságai miatt így a program-iskolákban tapasztalható eltérő gondozói környezet hatását méri.*

A tanulói összetétel hasonlóságából ugyanakkor természetesen nem következik az, hogy maguk az iskolák is hasonlóak lettek volna a program előtt. Bár a mintavétel során hasonló kompetenciaeredményű iskolákat párosítottunk össze, az iskolák sok más tekintetben eltérőek lehettek. A program kiválasztási algoritmusa miatt valószínű, hogy az integrált oktatáshoz, valamint a program tartalmi elemeihez pozitívan viszonyuló iskolák inkább jelentkeztek és nyertek. Ez az ún. önszelekciós probléma nagyon megnehezíti azt, hogy az eltérő oktatási környezet gyermekek fejlődésére gyakorolt hatásából elkülönítsük a kizárólag a program hatására bekövetkező változások mértékét: elképzelhető ugyanis, hogy az általunk mért pozitív különbségek nem feltétlenül magának a programnak köszönhetőek, lehet, hogy a programiskolák a program nélkül is jobb eredményeket értek volna el.



Valószínűtlen azonban, hogy a programnak ne lett volna hatása a programiskolák működésére. Ezt alátámasztja az a tény, hogy sok dimenzióban mások, mint a kontrolliskolák, és ezek mind olyan dimenziók, amik a program fókuszában álltak. A programiskolákban, ha vannak párhuzamos osztályok, azok között átlagosan kisebb az eltérés például a roma tanulók arányában, mint a kontrolliskolákban (Kézdi és Surányi, 2008a, 3. fejezet). Az alkalmazott pedagógiai módszerek és a tanár-diák kapcsolatok nagymértékben különböznek. Kabai Péter nem tesz említést róla, de kutatásunk fontos része volt mintegy 400 tanóra megfigyelése standardizált kérdőívek és alaposan betanított osztálymegfigyelők segítségével (Kézdi és Surányi, 2008a, 4. fejezet). Az eredmények azt mutatták, hogy a programiskolák inkább alkalmazzák a modern, kompetenciaorientált, gyermekközpontú és kooperációt segítő oktatási módszereket, mint a kontrolliskolák. Bár az egyes módszerek hatásának elkülönítése gyakorlatilag lehetetlen (Kézdi és Surányi, 2008a, 8. fejezet), eredményeink azt jelzik, hogy a személyközpontú és tanulói autonómiát erősítő pedagógia jelentősen hozzájárul a jobb tanulói eredményekhez, függetlenül a tanulók családi háttérétől.<sup>3</sup>

Annak érdekében emellett, hogy az eltérő oktatási-nevelési környezet tanulói eredményességre gyakorolt hatásában elkülönítsük magának a program bevezetésének a hatását, két módon is kísérletet tettünk (mindkettőt részletesen is ismerteti zárótanulmányunk). Az első módszer az eredményváltozók átlagos eltérése mellett megvizsgálta azok különbségét néhány további változó kiszűrése után is, sokváltozós regressziók segítségével<sup>4</sup>. A kontroll változók közül a legfontosabb a felsős tanulók esetében az „induktív gondolkodás,” az alsósok esetében a matematika teszteredmény volt; mindkettőt két évvel korábban mértük<sup>5</sup>. Ez az eljárás kiszűri mindazon nem megfigyelhető tanulói tulajdonságok hatását, amelyek ezeket a teszteredményeket befolyásolják. Fontos azonban látni, hogy emiatt olyan hatásokat is kiszűr, amiket nem szeretnénk: a kiszűrt teszteredményekben ugyanis nemcsak a tanulói háttér, hanem az iskola is szerepet játszott, sőt maga a program is hatással lehetett rá, hiszen azokat másfél évvel a program

---

<sup>3</sup> Kabai Péter nem kívánt megjegyzést fűzni ehhez a fejezethez sem, mondván „itt csak leíró statisztikát” közlünk. Valószínűleg elkerülte a figyelmét, hogy az itt bemutatott eredmények mögött diff-in-diffs jellegű modellek állnak, amelyek azt vizsgálják, hogy a program-kontroll különbségek (a program becsült hatása) mennyiben függ össze a programiskolák közötti különbségekkel egyes köztes oktatási változóknál (pl. tanulói autonómia). Talán az explicit teszteket hiányolta: ebben a fejezetben csak becslési eredmények vannak, grafikus formában is ábrázolva, szignifikanciaszintek nélkül. Az ok egyszerű: a teszteket a kettőnél több kategóriát összehasonlító modellekben a referenciacsoport meghatározásával lehet elvégezni, ami tovább bonyolította volna az eredmények közlését. A nagymértékű különbségek – mint például a tanulói autonómia vagy a személyközpontú pedagógia esetében – természetesen statisztikailag mindig szignifikánsak 5%-on.

<sup>4</sup> Ezt kontroll függvény módszernek is hívják a hatásvizsgálat irodalomban. A kontrollváltozók nemcsak additívan, de a program dummyval interaktálva is szerepelnek a jól specifikált kontroll függvényben, lásd például Imbens és Wooldridge (2009) – a mieinkben is így szerepeltek.

<sup>5</sup> Ezeket a teszteket, a később mért olvasáskészség tesztekkel együtt, Csapó Benő szegedi kutatócsoportja dolgozta ki és bocsátotta rendelkezésünkre.

indulása után mértük. A megmaradó program – kontroll különbségek emiatt abszolút értékben lefelé torznak, vagyis valószínűleg kisebb különbségeket mutatnak a valós hatásnál.

A második módszer során a tanulói eredményváltozók vizsgálatában egy mintaszűkítés segítségével megkíséreltük szétválasztani a program hatását az iskolák program előtti gyakorlatától (Kézdi-Surányi 2008a, 2.6. fejezet). Valamennyi program-kontroll összehasonlítást elvégeztünk a vizsgált iskolák egy szűkebb halmazán is. E szűkebb halmazba azok a programiskolák tartoztak – kontroll párjaikkal –, amelyek a program előtt nem oktattak integrált szemléletben. Ezekben az iskolákban a többi programiskolához képest az integráció és a program által támogatott pedagógiai módszerek kevésbé voltak jellemzőek a program előtt. Ha a programnak nem, vagy csak alig volt hatása a tanulói eredményváltozókra (és csak az számít, hogy az iskolák milyenek voltak alapvetően, a programtól függetlenül), akkor a szűkített mintában a teljeshez képest jóval gyengébb program – kontroll különbségeket kellett volna kimutatnunk. A vizsgált eredményváltozók többségénél azonban a szűkített mintában a teljes mintához nagyon hasonló program – kontroll különbségeket mértünk. Ezek alapján azt a következtetést vontuk le, hogy *a különbségek nagy valószínűséggel valóban a program hatását mutatják.*

Ennek fényében figyelemreméltó az az eredmény, hogy a vizsgált eredményváltozók többségénél a program – kontroll eredmény nem vagy alig csökken a kontrollváltozók bevonása után. Kabai Péter kiragadja a továbbtanulási eredményt, ahol a csökkenés a legnagyobb mértékű: a mért különbség felére csökken, és elveszti statisztikai szignifikanciáját. Nem válik azonban negatívvá egyik csoportban sem (ennek jelentőségéről lásd később). Kabai Péter állítása szerint a továbbtanulási eredmények értékelésekor „(...) nagyon egyszerű lett volna (vagy lenne akár most is) korrigálni a továbbtanulási mutatókat a program előtti adatok alapján. Ha például a roma tanulók továbbtanulási mutatói 10%-kal jobbak voltak a bázisiskolákban a program bevezetése előtt, mint a kontrollként kiválasztott intézményekben, akkor a programnak nincs hatása a továbbtanulási mutatókra.”

Mikor azonban a „nagyon egyszerű lett volna” állítást megfogalmazta, Kabai Péter nyilvánvalóan nem tájékozódott a lehetőségekről. A kontrollminta kiválasztásánál semmilyen továbbtanulási adatra nem támaszkodhattunk, mert ezek *nem állnak rendelkezésre Magyarország általános iskoláira.* (A propensity score egyenletben szerepeltek a 2003-as kompetenciamérés iskolai eredményei, mert azok nehezen ugyan, de hozzáférhetőek voltak.) Az elemzés során pedig akkor tudtuk volna kiszűrni a korábbi roma továbbtanulási arányok hatását, ha azt mérhető lett volna legalább a mintában szereplő iskolákban. Ilyen információt visszamenőleges adatok bekérésével tudtunk volna gyűjteni, hiszen kutatásunk másfél évvel a program kezdete után indult. Az adatokat akkor

lehetett volna begyűjteni, ha maguk az iskolák őriznék ilyeneket. Természetesen azonban az iskolák etnikai bontásban semmilyen adatot nem őriznek, így a roma tanulók továbbtanulási arányát sem. Az egyetlen ok, amiért hatásvizsgálatunkban a későbbiekben elemezni tudjuk a programban érintett roma tanulók továbbtanulási arányát az az, hogy mi magunk mértük az etnikai hovatartozást. A mérés teljes mértékben megfelelt a hatályos jogszabályoknak és az ombudsmani ajánlásoknak, és rendkívül sikeres volt (a tanári becslésekkel azonos arányban találtunk roma tanulókat mindkét mintában), amint arról kutatási jelentésünk részletesen beszámol (2.5. fejezet).

S még abban az esetben is, ha valakit mindezek a tények nem győznek meg teljes mértékben arról, hogy a program – kontroll különbségek magának a programnak a hatását mérik, érvényes az a következtetés, hogy *a programiskolák ugyanolyan összetételű tanulókkal jobb eredményeket értek el az integrált oktatás keretében*. Az eredmény tehát a megfelelő pedagógiával támogatott integrált oktatás hatását méri, még abban az esetben is, ha mindezt az iskolák nem kizárólag a program hatására teszik. Téves tehát Kabai Péter sommás megállapítása, mely szerint „a minta torzított és ezért semmilyen következtetés levonásra nem ad lehetőséget”.

## **TESZTELÉSI ELJÁRÁSOK, SZIGNIFIKANCIASZINTEK**

Az elemzés módszereit ért bírálatok második csoportja a tesztek és a statisztikai szignifikancia témáját érintik. Kabai Péter állítása szerint 10%-os szignifikanciaszintet fogadunk el, amit túl gyengének vél. Felrója, hogy nagyszámú eredményváltozóra külön-külön végezzük el a program – kontroll összehasonlítást Bonferroni-típusú korrekció alkalmazása nélkül. Kifogásolja emellett azt is, hogy a tanulók összességére és különböző csoportjaikra is megvizsgáljuk a program – kontroll különbségeket, s ezek során a csoportképző ismérvek nem függetlenek egymástól.

Az ezekre a kritikákra vonatkozó rövid válaszuk az, hogy módszereink teljes mértékben követték a társadalompolitikai programok hatásvizsgálatának nemzetközi irodalmában szokásos elemzési eljárásokat. Nem választottunk magunk számára szignifikanciaszintet, hanem megjelöltük, mely eltérések szignifikánsak 10, 5, illetve 1%-on, s az olvasóra bíztuk, hogy a 10%-on szignifikáns eredményeket elfogadja-e. Ugyanígy bevett gyakorlat az is az irodalomban, hogy a hatást nemcsak összességében, hanem különböző csoportokra külön is megvizsgálják, s ezek a csoportok sokszor átfedik egymást. Ugyancsak standard a hatást eredményváltozónként külön is tesztelni, Bonferroni-típusú korrekciók nélkül<sup>6</sup>. Kabai Péter felvetése alapján úgy tűnik tehát, ezek a standardok eltérnek a természettudományos kutatások gyakorlatától (bár azokat teljes mértékben

---

<sup>6</sup> Lásd például Blundell et al. (2004), Skoufias (2005), National Evaluation of Sure Start (2008).

nem ismerjük, hiszen nem vagyunk természettudósok, és nem is véleményezzük természettudományos kutatási eredményeit).

Ebben a tekintetben Kabai Péter bírálata érdemének tudható be, hogy e szakmai vita keretében rávilágít egy olyan, tágabb kutatási kontextust is érintő kérdésre, mely túlmutatva kutatásunk megfelelőségének boncolgatásán, távolabbra is kitekint, és felveti a kérdést, hogy mi állhat az eltérő tudományterületek eltérő gyakorlatának hátterében. Véleményünk szerint a magyarázat a tudományterületek eltérő kérdésfeltevésében és eredményeik eltérő felhasználási területében rejlik. A hatásvizsgálatok, amint korábban már részletesen szóltunk róla, általában nem a laboratóriumi kísérletek precizitásával megtervezett adatokkal operálnak, hanem sok esetben külső tényező által meghatározott információhalmazra épülnek. A hatásvizsgálatok ugyanakkor a politikai döntéshozatal közvetlen támogatói, s a következtetéseket egy-egy kormányzati intervenció sikerességére, társadalmi megtérülésére, fenntartható működtetésére vonatkozóan kell levonniuk. Ebben a tekintetben tehát eszközrendszerük és azok alkalmazásának gyakorlata is jelentősen eltér a természettudományok – és sok társadalomtudományi alapkutatás - eszközrendszerétől is.

A Kabai Péter által javasolt Bonferroni korrekció lényege például az, hogy abban az esetben, ha több hipotézist tesztelünk egyenként, akkor az alkalmazandó szignifikancia szintet a hipotézisek számával arányosan szigorítsuk. A szignifikancia szintek szigorításával az ún. elsőfajú hiba bekövetkezési valószínűségét csökkentjük (vagyis annak valószínűségét, hogy úgy mondjuk ki a program hatását, hogy az a valóságban nem létezik). Ezzel együtt azonban a másodfajú hiba növeljük valószínűségét. Csökken tehát a teszt ereje, vagyis nő annak a valószínűsége, hogy bár az intervenció képes pozitív változások elindítására, a kutatás mégis azt a következtetést vonja le, hogy a programnak nincs a társadalom számára érezhető eredménye. A szignifikancia szint megválasztása így nem „módszertani precizitás” kérdése, hanem a kétféle hibázási lehetőség közötti mérlegelés eredménye, aminek státusa a program hatásvizsgálatokban jelentősen eltérhet a természettudományokban megszokottól. A Fisher-Neyman-Pearson-féle tradícióban az elsőfajú hiba elkerülése valóban elsőbbséget élvez: inkább fogadjuk el gyakran a nullhipotézist még ha az nem is igaz, minthogy 5 (vagy 1) százaléknál gyakrabban utasítsuk el a nullhipotézist, ha az igaz. Kormányzati beavatkozások hatásvizsgálataiban azonban ez azt jelentené, hogy inkább döntsünk úgy, hogy nincs hatás, miközben van, minthogy az elhanyagolhatóan nagyobb valószínűséggel tulajdonítsunk hatást egy olyan programnak, aminek nincsen hatása. Abban az esetben viszont, ha az intervenció nem valami teljesen új dolgot vezet be, hanem egy már alkalmazott gyakorlattal szemben kínál alternatívát, mint a mi esetünkben is, komoly érvek szólnak a kétféle hibázási lehetőség egyenrangúsága mellett, hiszen a tévesen megítélt „nincs hatás” komoly, társadalmi szempontból is jelentős törekvések leállítását eredményezheti. Úgy véljük, ezek az érvek

alapot szolgáltatnak a program hatásvizsgálatokban használt eljárásokhoz, de nyilvánvaló, hogy a diszciplínák közötti mély módszertani különbségek megértésére és elfogadására ennél alaposabb viták szükségesek.

Kabai Péter mindezek után megjegyzi, hogy a vizsgált program-kontroll különbségekben „a szignifikanciaszint negyvennyolc esetben nem érte el a 10%-ot”, amit aggasztónak ítélt. Ezzel szemben bármilyen részhalmazát nézzük is az eredményváltozóknak, az ilyen szinten szignifikáns különbségek száma mindig magasabb a nem szignifikáns különbségek számánál. Mindamellet Kabai Péter a nem szignifikáns kapcsolatokkal kapcsolatban is megjegyzi, hogy azok túlnyomó többsége pozitív, ami „nem lehet véletlen, a nem szignifikáns különbségek is tendenciára utalnak.” Ez az intuíció megegyezik azzal, ahogyan mi értékeljük az eredményeket. Az eredmények tartalmát tehát, úgy tűnik, végül is nem érintik a statisztikai módszerekről megfogalmazott kritikák.

## **AZ EREDMÉNYEK ÉS A BELŐLÜK LEVONHATÓ KÖVETKEZTETÉSEK**

Eredményeink alapján megállapítottuk, hogy a programiskolákban folyó integrált oktatás csökkenti a nem roma tanulók romákkal szembeni távolságtartását. A programiskolákban a tanulók olvasáskészsége és továbbtanulási esélyei is jobbak kismértékben, akár a roma, akár a nem roma, akár a hátrányos helyzetű, akár a nem hátrányos helyzetű tanulókat tekintjük. Mindezek mellett úgy tűnik, hogy a programnak jelentős pozitív hatása van valamennyi tanuló önértékelésére, a sorsirányítás képességébe vetett hitére, valamint a nehéz helyzetekkel való megküzdési készségére; ami azért kiemelkedő eredmény, mert ez utóbbi készségek jelentősége a munkaerőpiacon és az élet más területein is a kognitív készségekhez és kompetenciákhoz mérhető. A különböző családi háttérű tanulókra gyakorolt hatás mértéke és megnyilvánulási formája is sokszor különböző, de hangsúlyoznunk kell, hogy egyáltalán nem találtunk olyan fejlődési területet, ahol a megfelelő pedagógiai módszerekkel támogatott integrált nevelési környezet negatív hatással lett volna a gyerekek bármely csoportjára nézve is, sőt, a gyerekek minden (etnikai hovatartozás és családi háttér) szerinti csoportja vonatkozásában található olyan fejlődési terület, melyet a megváltozott oktatási/nevelési környezet pozitívan befolyásolt.

A mért hatások Kabai Péter szerint „nem tűnnek jelentősnek”, részben ezért nem tartja indokoltnak a kutatási jelentés címében szereplő „siker program” jelzős szerkezetet (Kézdi és Surányi, 2008a). A hatásnagyságok megítélése során azonban több olyan szempontot is érdemes figyelembe venni, amivel Kabai Péter úgy véljük, nem számolt. Az első az, hogy a társadalmi integrációt célzó oktatási programok sikerességét a probléma orvosolásának alternatív formáihoz érdemes viszonyítani. Ebből a szempontból

hasznos lehet az összehasonlítás azon programok hatásával, amelyek felnőttkorban próbálják a hosszútávú munkanélkülieket rendszeres munkához juttatni. Hudomiet és Kézdi (2008) az úgynevezett aktív munkaerőpiaci programokkal kapcsolatban (képzési programok, foglalkoztatástámogatási programok, közcélú foglalkoztatás, aktív segítség a munkakeresésben) bemutatja, hogy a „módszertani szempontból megbízható vizsgálatok az esetek nagy részében elhanyagolható vagy legfeljebb kismértékű hatást mutatnak”. Az általunk vizsgált oktatási program ezek viszonylatában a sikeres intervenciók körébe sorolható. A gyermekkorban történő beavatkozások nagyobb eredményessége egyébként nem meglepő eredmény az irodalomban. Számos nemzetközi kutatás bizonyítja mára azt, hogy a társadalmi integráció elősegítése és a hosszútávú életesélyek javítása céljaira az oktatáspolitikai beavatkozások hatásosabbak, mint a felnőttkorra irányuló programok (Heckman, 2006).

A másik figyelembe veendő szempont az, hogy az oktatási befektetések kumulatív hatása miatt a gyermekkorban történő kis elmozdulások is komoly, hosszú távon érvényesülő társadalmi hasznokkal járhatnak (Heckman és Masterov, 2007, valamint Surányi és Danis, 2010), különösen, ha egyszerre több dimenzióban történik az elmozdulás, amelyek egymást erősíthetik.

A legfontosabb érv azonban arra, hogy az eredményeket az integrált oktatás szempontjából sikeresnek – de legalábbis biztatónak – tekintsük, az a tény, hogy a vizsgált *társadalmi csoportok egyikében sem tapasztaltunk negatív eredményeket*, sőt, minden csoport átlagos eredményeire pozitív hatással volt a programiskolákban folyó integrált oktatás. Ez azért jelentős felismerés, mert alapvető intuíciónk mellett számos kutatási eredmény is azt sugallja, hogy az integrált oktatásnak lehetnek vesztesei (a nem hátrányos helyzetűek). Meghir és Palme (2005) például kimutatják, hogy a svéd alsó középiskolai rendszer egységesítése és ezáltal integráltabbá tétele az 50-es években jelentősen megnövelte az intergenerációs mobilitást oly módon, hogy segítette a hátrányos helyzetű tanulókat a nem hátrányos helyzetűek rovására. Hasonló eredményeket talál Pekkarinen et al (2009) Finnországra.

Ezen megfontolások miatt értékeljük jelentős eredményként azt, hogy a program iskolákban folyó integrált oktatásnak úgy tűnik, nem voltak vesztesei. Ez a különbség a programiskolákban nagyobb teret nyert pedagógiai módszerek szükségességét és egyben hatékonyságát mutatja. Az svéd és finn említett beavatkozások ugyanis nem jártak együtt az általunk vizsgált programban alkalmazott pedagógiai szemlélet azonnali elterjedésével. Mindez arra enged következtetni, hogy kutatásunkban nem egyszerűen „az integráció” hatásáról van szó, hanem mindazoknak a pedagógiai, oktatásszervezési és hozzáállásbeli különbségeknek a hatásáról, amelyek az integrált környezettel kölcsönhatásban a programiskolákat mássá tették.

Az eredmények pozitív értékelését árnyalja az a tény, hogy a programiskolák előnye több eredményváltozó esetében leginkább akkor mutatható ki, ha a roma tanulók aránya nem haladja meg a 20-30 (egyres esetekben 40) százalékot. E következtetés az adatok korlátai miatt igen spekulatív, ezért robusztusságát és lehetséges okait hatásvizsgálatunk nyitva hagyta további kutatások számára.<sup>7</sup>

## ZÁRÓ GONDOLATOK

Ebben a tanulmányban a hatásvizsgálat mintavételét és elemzési módszereit ért kritikákkal foglalkoztunk. Megmutattuk, hogy a kritikák megalapozatlanok, és ismételten hangsúlyoztuk, hogy eredményeink hogyan interpretálhatóak az OOIH program hatása szempontjából és az integrált és egyszersmind minőségi oktatás szempontjából. A nem-kognitív készségeket mérő eszközök kidolgozására irányuló bírálatokra a Budapesti Munkagazdaságtani Füzetek 2010/1. számában megjelent cikkünk válaszol – kimutatva azok megalapozatlanságát.

Tisztában vagyunk mindazzal a felelősséggel, amit a hatásvizsgálat módszertani követelményei és a levonható következtetések körültekintő mérlegelése ruház ránk. Úgy érezzük ugyanakkor, hogy Kabai Péter nem vetett számot azzal a felelősséggel, amely a sok esetben nem megfelelően megalapozott kritikáinak társadalmi hatásaival jár. Az alaptalan, nem megfelelő körültekintéssel megírt kritikák nemcsak a kutatásnak és magának a programnak a megtervezésében és kivitelezésében résztvevők számára sértőek, hanem negatív hatással vannak az akadémiai kutatások általános megítélésére éppúgy, mint esetünkben maguknak az etnikai integrációs törekvéseknek a társadalmi fogadtatására is.

---

<sup>7</sup> Az elemzés a kutatási jelentés (Kézdí és Surányi, 2008a) 9. fejezetében található. Hasonlóan a program hatásmechanizmusaiával foglalkozó 8. fejezethez, itt is hangsúlyozzuk az eredmények bizonytalanságát, és ezért itt sem közlünk standard hibákat. Az egyik fontos korlátja az elemzésnek, hogy kvadratikus függvényformát illeszt az osztály összetétele és az eredményváltozók kapcsolatára, ami nyilvánvalóan óvatosságra int. A 69. lábjegyzetben erre külön fel is hívjuk a figyelmet: "Minthogy évfolyamonként bázis/kontroll bontásban mindenütt 40-50 esetszámról van szó, a nemlinearitásokat nem lehet megbízhatóan becsülni, ezért messzemenő következtetéseket belőlük levonni nem szabad." Kabai Péter számos bíráló megjegyzést fűz ehhez az elemzéshez, bírálati azonban arra vezethetőek vissza, hogy messzemenő következtetéseket keresett az elemzésben.

## HIVATKOZÁSOK

- Angrist, J., Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114(2), 533-575.
- Blundell, R., Costa Dias, M., Meghir, C., Van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association* 2(4), 569-606.
- Deaton, A. (2009). *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*. The Keynes Lecture, British Academy, October 9th, 2008.  
[http://www.princeton.edu/~deaton/downloads/Instruments\\_of\\_Development.pdf](http://www.princeton.edu/~deaton/downloads/Instruments_of_Development.pdf)
- de Janvry, A., Finan, F., Sadoulet, E., Vakis, R. (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics* 79(2), 349-373.
- Gertler, P. (2000). *The impact of PROGRESA on health*. International Food Policy Research Institute (IFPRI) Research Report.  
<http://www.ifpri.org/publication/impact-progresa-health>
- Heckman, James J. (2006), "Skill Formation and the Economics of Investing in Disadvantaged Children" *Science*, 312 (June 30), pp. 1900-1902.
- Heckman, J. J., Masterov, D. V. (2007). The Productivity Argument for Investing in Young Children. NBER Working Paper No. 13016.
- Hudomiet, P., Kézdi, G. (2008). Az aktív munkaerő-piaci programok nemzetközi tapasztalatai. *Kormányzás* 3(1) 3-37. [http://www.kormanyzas.hu/081/01\\_Hudomiet-Kezdi.pdf](http://www.kormanyzas.hu/081/01_Hudomiet-Kezdi.pdf)
- Imbens, G. W., Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1) 5-86.
- Kézdi, G., Surányi, É. (2008a). *Egy sikeres iskolai integrációs program tapasztalatai*. Kutatási összefoglaló. Educatio Kht., Budapest  
<http://bin.sulinet.hu/tanar/szabadido/planetarium/8.1.1.pdf>
- Kézdi, G., Surányi, É. (2008b). Egy integrációs program hatása a tanulók fejlődésére. *Educatio*. 2008/IV. <http://www.hier.iif.hu/hu/letoltes.php?fid=tartalomso/1250>
- Meghir, C., Palme, J. (2005). Educational Reform, Ability, and Family Background. *The American Economic Review* 95(1), 414-424.
- National Evaluation of Sure Start Research Team (2008). *The Impact of Sure Start Local Programmes on Three Year Olds and Their Families*.  
<http://www.dcsf.gov.uk/everychildmatters/research/publications/surestartpublications/1974/>
- Németh, Sz., Papp Z. A. (2006). És mi adjuk az integráció vezérfonalát... Dokumentumelemzés és a kvalitatív vizsgálat eredményei című tanulmánya. In: Németh Szilvia (szerk.): *Integráció a gyakorlatban. A roma tanulók együttnevelésének iskolai modelljei*. Országos Közoktatási Intézet, Budapest.
- Pekkarinen, T., Uusitalo, R., Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics* (93), 965-973.
- Rosenbaum, P. R., Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41-55.



Skoufias, E. (2005). *PROGRESA and its impacts on the welfare of rural households in Mexico*. International Food Policy Research Institute (IFPRI) Research Report 139. <http://www.ifpri.org/publication/progresa-and-its-impacts-welfare-rural-households-mexico>

A sorozat korábban megjelent kötetei

2008

- Galasi Péter – Nagy Gyula: A munkanélküli-jövedelemtámogatások célzása Magyarországon 2004-ben. BWP 2008/01
- Szilvia Hámori: Adult education in the European Union – with a focus on Hungary. BWP 2008/02
- Galasi Péter: A felsőfokú végzettségű munkavállalók munkaerő-piaci helyzete és foglalkozásuk-iskolai végzettségük illeszkedése. BWP 2008/03
- Anna Lovász: Competition and the Gender Wage Gap: New Evidence from Linked Employer-Employee Data in Hungary, 1986-2003. BWP 2008/04
- Zsuzsa Blaskó: Does early maternal employment affect non-cognitive children outcomes? BWP 2008/05
- Galasi Péter: A túl- és az alulképzés bérhozama 25 országban. BWP 2008/06
- Galasi Péter - Nagy Gyula: Az aktív munkaerő-piaci programokba kerülés esélyei: képzés, bértámogatás, közhasznú munka. BWP 2008/07
- Galasi Péter: The effect of educational mismatch on wages for 25 countries. BWP 2008/08
- Gábor R. István: Életpálya-kereseti profilok és keresetingadozás. BWP 2008/09

2009

- Istvan Gábor R.: Experience-earnings profile and earnings fluctuation: a missing piece in some labour market puzzles? BWP 2009/01
- Anna Lovász – Mariann Rigó: Who Earns Their Keep? An Estimation of the Productivity-Wage Gap in Hungary 1986-2005. BWP 2009/02
- Köllő János: Miért nem keresnek állást a magyar munkanélküliek? Hipotézisek az Európai Munkaerőfelvétel adatai alapján. BWP 2009/03
- Bálint Mónika - Köllő János - Molnár György: Összefoglaló jelentés a KSH-ONYF adatfelvételről. BWP 2009/04
- Gábor R. István: "Minimálbér-paradoxon" - versenyzői munkaerőpiacon? Egy gondolkísérlet tanulságai. BWP 2009/05
- Kertesi Gábor - Kézdi Gábor: Általános iskolai szegregáció Magyarországon az ezredforduló után. BWP 2009/06
- Szilvia Hámori: Employment convergence of immigrants in the European Union. BWP 2009/07
- Gábor Kőrösi: Innovation and Rent Sharing in Corporate Wage Setting in Hungary. BWP 2009/08

---

**A Budapesti Munkagazdaságtani Füzetek** a Magyar Tudományos Akadémia Közgazdaságtudományi Intézetében működő Munkapiaci Kutatások valamint a Budapesti Corvinus Egyetem Emberi Erőforrások Tanszékének közös kiadványa. A kiadványsorozat angol nyelvű füzetei **“Budapest Working Papers on the Labour Market”** címmel jelennek meg. A kötetek letölthetők az MTA Közgazdaságtudományi Intézet honlapjáról: <http://www.econ.core.hu>