

TÉNYEK ÉS ÉRVEK

MIT MÉR A MŰSZER?

A TANULÓI TELJESÍTMÉNYMÉRÉSEK
ALKALMAZHATÓSÁGÁRÓL

SZERKESZTETTE:

SZÉLL KRISZTIÁN

OKTATÁSKUTATÓ
ÉS FEJLESZTŐ
INTÉZET



MIT MÉR A MŰSZER?

A TANULÓI TELJESÍTMÉNYMÉRÉSEK
ALKALMAZHATÓSÁGÁRÓL

SZERKESZTETTE:
SZÉLL KRISZTIÁN

OKTATÁSKUTATÓ
ÉS FEJLESZTŐ
INTÉZET



Oktatókutató és Fejlesztő Intézet
Budapest, 2015

Szerkesztette:
Széll Krisztián

Lektorálta:
Fehérvári Anikó

Szerzők:
Horn Dániel, Nahalka István, Ostorics László, Széll Krisztián, Szemerszki Marianna

Olvasószerkesztő:
Gyimesné Szekeres Ágnes

Tördelőszerkesztő:
Pattantyus Gergely

Tördelés:
WOW Stúdió Kft. (1138 Budapest, Népfürdő u. 3/B.)

© Oktatókutatató és Fejlesztő Intézet, 2015

© Horn Dániel, Nahalka István, Ostorics László, Széll Krisztián, Szemerszki Marianna, 2015

ISSN: 1785-1432

ISBN: 978-963-682-799-1

Oktatókutatató és Fejlesztő Intézet
1143 Budapest, Szobránc utca 6–8.
www.ofi.hu
Felelős kiadó: Kaposi József

Tartalom

Előszó	5
Szemerszki Marianna: A tanulói teljesítménymérések szerepe a tényekre alapozott oktatáspolitikában.....	9
Nahalka István: Tanulói teljesítménymérések alkalmazhatósága a neveléstudományban	23
Ostorics László: A tanulói teljesítménymérések jellemzői, jövőbeni irányvonalai, kritikái.....	37
Horn Dániel: Az iskolai hozzáadott érték mérése.....	63
Szell Krisztián – Szemerszki Marianna: A tanulói teljesítménymérések megítélése és intézményi hasznosulása.....	91

Előszó

Az Oktatási Indikátor Munkacsoport¹ az Oktatókutató és Fejlesztő Intézettel² és a Tempus Közalapítvánnyal közösen 2014 végén műhelykonferenciát szervezett a tanulói teljesítménymérések eredményeinek értelmezési lehetőségeiről, azok esetleges korlátairól, előnyeiről és hátrányairól. A műhelykonferencia célja volt, hogy az oktatás eredményességének javítása érdekében bemutassa és ütköztesse a mérés-értékeléssel kapcsolatos eltérő szempontokat és aspektusokat, hozzájárulva a tanulói teljesítménymérések hazai értelmezéséhez, feltárva azok alkalmazhatóságának feltételeit. A műhelykonferencia első részében felkért előadók tartottak vitaindító előadásokat, majd ezt követően lehetőség nyílt szakmai konzultációra, a felvetett kérdések megvitatására is.

A műhelykonferencia tapasztalatai alapján számos észrevétel és javaslat fogalmazódott meg, melyek közül az alábbiakat emeljük ki:

- A minőségi oktatás érdekében a hazai oktatási rendszer fejlesztéseket támogató, felelős és elszámoltatható rendszerré alakítandó, melynek az egyik, de nem az egyetlen eszköze a teljesítménymérések eredményeinek nyomon követése.
- A teljesítménymérések jelentős potenciált rejtenek magukban az oktatáspolitikai és fejlesztési beavatkozások tervezését illetően. A szakpolitika a lehetőségek kiaknázását egyrészt célzott kutatásfinanszírozással, másrésztől értékelésekkel egybekötött, a 2014–2020-as uniós fejlesztési időszak során megvalósítható 3–5 éves szakpolitikai pilot programok támogatásával mozgathatja elő.

¹ A 2005-ben létrehozott Oktatási Indikátor Munkacsoport olyan konzultatív szakértői testület, melynek alapvető feladata a meglévő és a fejlesztés alatt álló adatforrások oktatáspolitikai célú alkalmazási lehetőségeinek feltárása, továbbá javaslatok kidolgozása és szakmai konszenzus kialakítása a hazai és a nemzetközi adat- és indikátorfejlesztés prioritásaira vonatkozóan. A Munkacsoport ügyvivői feladatait az Oktatókutató és Fejlesztő Intézet Kutatási, Elemzési és Értékelési Központja látja el.

² A TÁMOP-3.1.1-11/1-2012-0001 azonosítójú kiemelt projekt támogatásával.

- Nemzetközi tapasztalatok alapján a gyenge és az erős ösztönzők, szankciók, valamint kizárólag a tanulói teljesítménymérések eredményei alapján történő értékelések negatív hatásokat eredményezhetnek. Ebből következően szakpolitikai hiba volna a finanszírozási beavatkozások tervezését és megvalósítását teljes mértékben a teljesítménymérési eredményekre alapozni. Alapvető kérdés, hogy a teljesítménymérések eredményeihez milyen szinten milyen ösztönzők rendelkeznek. A megfelelő ösztönzők megtalálása érdekében az oktatás valamennyi szereplőjét érdekeltté kell tenni a teljesítménymérések eredményeinek fejlesztést és értékelést támogató felhasználásában.
- Javasolt olyan, különböző szakterületekről érkező szakértőkből álló munkacsoport felállítása, amely (1) elemzi a teljesítménymérések megbízhatóságát és a mérést lehetségesen torzító tényezőket; (2) sorra veszi és kiválasztja a kompetenciamérésekhez kapcsolható lehetséges ösztönzőket és azok alkalmazhatósági kritériumait; (3) ajánlásokat fogalmaz meg a kompetenciamérések fejlesztési és értékelési folyamatot támogató szerepéről, valamint az adaptálási és cselekvési folyamatok megvalósíthatóságáról és fenntarthatóságáról.
- Javasolt a kompetenciamérési eredmények felhasználásának intézményi szintű diagnózis irányába történő elmozdítása, valamint az iskolákban zajló nevelési-oktatási folyamatok fejlesztési és értékelési kereteinek alapos, széles körű szakmai konszenzuson alapuló kidolgozása.
- Segíteni kell az iskolákat, iskolavezetőket és pedagógusokat a kompetenciaméréseken alapuló olyan cselekvési (fejlesztési, beavatkozási) tervek kidolgozásában, amelyek mögött valódi cselekvés húzódik meg. Ennek fontos eleme a kapott eredmények és alkalmazott gyakorlatok széles körű disszeminációja és implementálása, az iskolák egymás közötti tanulása, tudásmegosztása.
- A nemzetközi tanulói teljesítménymérések alapvető jelentőséggel bírnak, általuk nemzetközi szinten is összehasonlíthatóvá válnak a hazai folyamatok, eredmények. Különös hasznót jelent, hogy adatainkat más országok hasonló adataival is összevethetjük, ez sok esetben segíti a saját hazai helyzetkép értelmezését is. Ugyanakkor ebben az esetben is figyelembe kell venni a teljesítménymérések korlátait, valamint a nemzeti sajátosságokat is.

A szakmai közéletben a tekintetben nincs különösebb vita, hogy a hazai és nemzetközi tanulói teljesítménymérések releváns és értékes információt nyújtanak az oktatáspolitikai számára. Ugyanakkor a mérési eredmények értelmezése, oktatáspolitikai alkalmazhatósága és felhasználása már olyan terület, ahol nagy eséllyel ütköznek az érvek és ellenérvek. Jelen tanulmánykötet ehhez a szakmai vitához kíván hozzájárulni.

A tanulmányok témái szorosan kötődnek a műhelykonferencia tematikájához, több tanulmány a műhelykonferencián elhangzott vitaindító előadás bővített, továbbgondolt változata. A kötet elején *Szemerszki Marianna* tanulmánya történeti aspektusba ágyazottan tárgyalja a tényekre alapozott oktatáspolitikai előtérbe kerülését, az intézményértékelés elveit és gyakorlatát, továbbá a tanulói teljesítménymérések és az oktatáseredményességi kutatások fejlődését, és azok intézményértékelésben betöltött szerepét. Ezt követően *Nahalka István* méréselméleti alapokon nyugvó, izgalmas vonalvezetésű tanulmányában értekezik arról, milyen szakmai problémákkal kell szembesülnie a mérési eredményeket felhasználó nevelési és oktatási szakembereknek, a neveléstudományi kutatások számára miben és mennyiben jelenthetnek alapot az országos és a nemzetközi tanulói teljesítménymérések eredményei. Ehhez kapcsolódóan *Ostorics László* írása áttekinti a hazai és nemzetközi tanulói teljesítménymérések legfontosabb jellemzőit és tendenciáit, különösen fókuszálva a magyar tanulók teljesítményére, továbbá ismerteti a mérések jövőben várható fejlődésének irányvonalait és az érvényességükkel kapcsolatosan megfogalmazott releváns kritikákat, valamint az arra adható szakmai válaszokat. Mindezek után *Horn Dániel* tanulmánya bemutatja az elméletileg lehetséges hozzáadott érték típusú modellek alaptípusait, majd igazi kuriózumként, a hazai kompetenciamérési adatokon alapulva tesz kísérletet a különböző, iskolai szintű hozzáadott érték típusú eredményességi mutatók összehasonlítására. Végül, de nem utolsósorban *Széll Krisztián és Szemerszki Marianna* egy közelmúltban megvalósult online kérdőíves adatfelvétel adataira támaszkodva elemzi a tanulói teljesítménymérésekkel kapcsolatos igazgatói és pedagógusi véleményeket, attitűdöket, kitérve a mérések iskolai munka értékelésében betöltött szerepére és hasznosíthatóságára is.

Kötetünket a tanulói teljesítménymérések és az oktatáseredményességi kutatások iránt érdeklődő kutatóknak, fejlesztőknek, szakpolitikusoknak, intézményvezetőknek és pedagógusoknak egyaránt ajánljuk. Reményeink szerint a tanulmányok mindegyike hasznos információkkal szolgál a széles körű szakmai közösség számára, és további alapot nyújt a teljesítménymérésekkel kapcsolatos szakmai párbeszéd mélyítéséhez.

Budapest, 2015. május 18.

A szerkesztő

Szemerszki Marianna

A tanulói teljesítménymérések szerepe a tényekre alapozott oktatáspolitikában

A tényekre alapozott oktatáspolitikai (*evidence based education policy*) az utóbbi évtizedek egyik legfontosabb trendjeként körvonalazható az oktatás területén. A folyamat összhangban van azokkal a változásokkal, amelyek az oktatást napjainkban egyre jobban áthatják, így többek között a piaci mechanizmusok oktatásban való megjelenésével, a minőségközpontú szemlélet térhódításával, benne az értékelések szerepének növekedésével, a standardizálásra való törekvésekkel. A megfelelő módon kivitelezett mérés-értékelés elemei szorosan illeszkednek ebbe a fogalmi körbe, támogatva az eredményes és minőségi oktatás különböző aspektusait. A jól működő oktatási rendszerekben a mérés-értékelés különböző elemei (tanulói értékelés, pedagógusértékelés, vezetői értékelés, iskolaértékelés) egymást erősítve alkotnak koherens egységet, a természetes módon adódó szinergiákat kihasználva, a felesleges redundanciákat elkerülve támogatják az oktatáspolitikai döntéseket, valamint az osztálytermi gyakorlatot és az iskolafejlesztéseket is.

A tényekre alapozott oktatáspolitikai térnyerése és az eredményességkutatások

A tényekre alapozott oktatáspolitikai megközelítések létjogosultsága nehezen kérdőjelezhető meg, hiszen a tudományosan igazolható megfontolások minden bizonnyal jobb alapot jelentenek a különféle döntésekhez, mint például a pusztán hiedelmeken, vagy ideológiákon alapuló. Amint arra *Halász* (2009) rámutat, a

Szemerszki Marianna (2015): A tanulói teljesítménymérések szerepe a tényekre alapozott oktatáspolitikában. In: Széll Krisztián (szerk.): Mit mér a műszer? A tanulói teljesítménymérések alkalmazhatóságáról. Budapest: Oktatókutatás és Fejlesztő Intézet, pp. 9–22.

kifejezés és maga a gyakorlat az orvostudomány területéről terjedt át az oktatásra, s először a 2000-es évek elején az Egyesült Államokban került a figyelem középpontjába a *No Child Left Behind (NCLB)* oktatásfejlesztési program kapcsán. E nagy volumenű beavatkozás egyik központi elve az oktatás tudományos megalapozottságával kapcsolatos kívánalmak rögzítése volt, így a programhoz kapcsolódó törvény olyan standardokat rögzített, amelyeknek a tanulás/tanítás eredményességéről gyűjtött adatoknak is meg kellett felelniük, s csak az ily módon alátámasztott eredmények alapján voltak jogosultak a szereplők különféle támogatásokra. A tényekre alapozott oktatáspolitikai paradigmája ugyanakkor nem nélkülözi a kritikát sem. A 2000-es évek kezdete óta sokan – többek között éppen az NCLB program egyik kidolgozója, az amerikai oktatáspolitikai formálásában is szerepet vállaló *Diane Ravitch* – felhívták a figyelmet a program nem szándékolt hatásaira, továbbá arra, hogy mennyire fontos az elővigyázatosság a mérhető eredmények értelmezése és oktatáspolitikai hasznosítása, valamint a nyilvánosság tájékoztatása során (*Tomasz, 2011*).

Az oktatás eredményességének és minőségének javítását célzó modern oktatáspolitikának szüksége van olyan információkra, amelyek alapján el tudja dönteni, hogy bizonyos beavatkozások várhatóan milyen következményekkel járnak, illetve, hogy az oktatás területén elért bármilyen eredmény mely tényezőknek milyen mértékben tulajdonítható. Az oktatás azonban olyan terület, ahol – különböző technikai és elvi/morális okok miatt – a kísérletező jellegű beavatkozások és azok eredményeinek mérése kevésbé evidens, mint az egészségügyben, ugyanakkor több tényező együttes hatására már a 20. század utolsó évtizedeiben is egyre több ország ismerte fel az adatokkal, tényekkel alátámasztott oktatáspolitikai, illetve oktatáspolitikai fejlesztések szükségességét. Hozzájárultak ehhez azok az iskolaeredményességi vizsgálatok, amelyek a 60-as évek második felétől kezdve elsőként az Egyesült Államokban, majd a 80-as évektől Angliában, Hollandiában és számos más országban azt próbálták meg feltárni, hogy milyen szerepük van az iskoláknak a tanulói teljesítmények alakulásában. Az első – kezdetben nem túlságosan ígéretesnek tűnő – eredmények után a kutatók, egyre kifinomultabb mérési eszközök és egyre szofisztikáltabb megközelítések birtokában, mind több bizonyítékot találtak az iskolai hatásokra, jóllehet a tanulói teljesítményben a családi háttér szerepe továbbra is meghatározónak mutatkozik.

Az eredményességkutatások kezdetben leginkább egyszerű bemeneti (input) és kimeneti (output) adatokra támaszkodtak, később azonban már egyes minőségi tényezők vizsgálata is előtérbe került, s a kutatások sok esetben kvalitatív módszereket is alkalmaztak. A számítástechnikai háttér fejlődésével egyre inkább előtérbe kerültek azok a vizsgálatok, amelyek nagyobb adatbázisokon, többváltozós elemzé-

sekkel, többszintű modellezést is lehetővé tévő megközelítésekkel dolgoztak. Erre kiváló lehetőséget nyújtottak a különböző teszt alapú tanulói teljesítménymérések, amelyek az oktatás eredményességéről korábban gyűjtött információkat részben kiegészítették, részben új megközelítésbe helyezték (lásd még e kötetben *Nahalka István* írását).

A bemeneti változók, a folyamatok és a kimeneti tényezők közötti összefüggés vizsgálata sokkal kifinomultabbá vált, amikor megjelentek a hozzáadott érték (value added) típusú modellek (fejlődésükről lásd *Saunders* 1999), amelyek tanulói szinten voltak képesek figyelembe venni a családi és az iskolai háttértényezőket, továbbá a megelőző iskolai eredmények hatását (*Sammons és mtsai.*, 1997) (lásd még e kötetben *Horn Dániel* írását). A tesztek és felmérések terjedésével nemcsak az alapvető képességek (írás, olvasás, matematikai feladatmegoldás) jobb mérésére nyílt lehetőség, hanem a magasabb szintű kognitív, sőt az affektív és a szociális eredmények tanulmányozására is (*Creemers és mtsai*, 2010).

A tanulói szintű méréseket is egyre inkább megpróbálják kiterjeszteni a tantárgyakhoz nem köthető általános kompetenciákra (*cross-curricular competencies*). A PISA-vizsgálatokban is szerepel ilyen kérdésblokk; 2012-ben a matematikai területhez kapcsolódóan a problémamegoldást vizsgálták, amelyen a magyar tanulók az átlagnál rosszabb eredményeket értek el (*OECD*, 2014). Mindez azért érdemel figyelmet, mert a XXI. században a munkaerőpiac által is nagyra értékelt készségek közé többek között olyanok tartoznak, mint a problémamegoldás, a kritikus gondolkodás, a kreativitás, a kommunikáció, az IKT-használat, valamint a szociális és állampolgári kompetenciák. Ehhez képest különösen aggasztónak tűnik a számítógéppel asszisztált problémamegoldás során elért eredmények alacsony szintje, hiszen, amint arra az egyik tanulmány rámutat, a magyar diákok 35%-ának a feladatmegoldása teljesen értékelhetetlen volt, a kettes teljesítményszintet sem érte el (*Csapó és mtsai.*, 2014).¹

A tanulói szintű, iskolai szinten aggregált teszteredmények oktatáspolitikai hasznosíthatósága egyes országokban felvetette az elszámoltathatóság kérdését is. Az elszámoltathatósági célú használat leginkább az Egyesült Államokban és Angliában került előtérbe, s fő céljai közé tartozik a nyilvánosság tájékoztatása és a minőségi munka támogatása azáltal, hogy visszacsatolást ad az elért eredményekről,

¹ Az eredmények értelmezésénél azonban felhívjuk a figyelmet arra, hogy bármilyen nemzetközi összehasonlításnál érdemes figyelembe venni az egyes országok közötti kulturális különbségeket is, jelen esetben például azt, hogy milyen a digitális eszközellátottság, vagy a korosztály diákjai elsősorban mire, milyen célból, mennyire kreatív feladatok megoldására használják a számítógépet.

illetve ösztönzésül szolgál a további fejlődéshez, fejlesztéshez. Az elszámoltathatóság különböző szinteken történhet, a kutatások az oktatás eredményessége szempontjából a legcélravezetőbbnek az iskolai szintet tekintik, mert ez az a szint, amely leginkább képes befolyásolni az oktató-nevelő munkát, annak gyakorlati megvalósulását (Horn, 2011). Emellett egyes országokban a pedagógusok elszámoltathatóságának irányában is tettek lépéseket, az Egyesült Államok néhány tagállamában például az egyes pedagógusok értékelése során a tanulói teljesítményeket is figyelembe veszik (Linn, 2005). Az egyes tanárok tanulói teljesítményekhez való hozzájárulásának mértéke, illetve annak mérhetősége ugyanakkor erőteljesen vitatott, hiszen a tanulók teljesítményének változása még a hozzáadott érték-modelleket figyelembe véve is nehezen köthető egy-egy konkrét pedagógus teljesítményéhez.

A mérések elszámoltathatósági célú használatával kapcsolatban egyaránt találhatunk érveket és ellenérveket, amelyek egyidejű gondos mérlegelésére van szükség, különösen akkor, ha a döntések komoly következményekkel járnak. Az elszámoltathatósági rendszerekkel kapcsolatban részben ugyanazok a dilemmák merülnek fel, mint amelyek magukkal a teljesítménymérésekkel kapcsolatban is, hiszen az elszámoltathatósági rendszerek alapja leggyakrabban valamilyen tanulói teljesítménymérés. A teszteredményekkel kapcsolatban felmerülhet a szándékos torzítás lehetősége is (azaz például a csalás vagy a teszteredmények hamisítása), de sokkal gyakoribbak és rejtettebbek a nem szándékolt hatások. Ilyen lehet a „*tesztre tanítás*” esete, vagy az, amikor az iskola kizárólag az adott mérésre és az abból eredő mutatóra koncentrál, s ezáltal éppen a minőség és a teljesítménynövekedés ellenében tesz (Fitz-Gibbon, 1996). A teszteken és vizsgaeredményeken alapuló elszámoltathatóság befolyást gyakorolhat a tanulási-tanítási folyamatra is, például azáltal, hogy háttérbe szorítja az innovatív tanítási gyakorlatok alkalmazását, vagy azáltal, hogy az oktatás csak bizonyos, a mérésekben is szereplő készségek fejlesztésére koncentrál, miközben a mérésben nem érintett készségek fejlesztése háttérbe szorulhat (Faubert, 2009).

Az elszámoltathatósági célú használat további fontos mérlegelendő tényezője, hogy a teljesítménymérések mennyire adnak valós képet egy adott intézmény működéséről. Az oktatásban a kimenet sokféle lehet, és egyes eredmények nem is tulajdoníthatók kizárólag az iskolának, illetve azon belül is nehezen különíthető el az egyes tényezők hatása. Egy iskola nyilvánvalóan akkor eredményes, ha valamilyen teljesítményt, hozzáadott értéket hoz létre, azonban az oktatás meglehetősen komplex rendszer ahhoz, hogy ezt a teljesítményt egyszerűen mérni tudjuk. Az egyik probléma, hogy az eredményesség mérésére használt különféle teljesítménymutatók nem képesek teljes képet nyújtani még akkor sem, ha jó és megbízható az adatok minősége. A tanulói teljesítménymérések és a vizsgaeredmények ugyanis

jó, de nem kizárólagos mutatói az oktatási rendszer, illetve az egyes intézmények teljesítményének. További problémát jelent az oktatási rendszer komplexitása, az iskolák összetettebb céljai, melyek egy része kevésbé mérhető és számszerűsíthető. Ráadásul az egyes mérőszámok egymáshoz viszonyított súlya is ismeretlen, ezért önmagukban a mérőszámok sokszor nem elegendőek az oktatás minőségének és eredményességének a méréséhez (*Hanushek és Lockheed, 1994*).

Az elszámoltathatóság, illetve az eredményesség kapcsán az is sokszor felmerül kérdésként, hogy vajon mennyire konzisztensek a kialakított indikátorok. A mutatók közötti kapcsolatot több szempontból is mérni lehet, például fontos lehet megvizsgálni azok időbeni konzisztenciáit, vagy az iskolán belüli évfolyamok közötti, illetve az egyes tantárgyak közötti egyezést. Az adatok azt mutatják, hogy az iskolai teljesítményt, eredményt tekintve az időbeli stabilitás általában nagy, míg az egyes tantárgyakra vonatkozó eredményességmutatók közötti összefüggés iskolai szinten már jóval kisebb (*Scheerens és mtsai., 2000*).

Az intézményértékelések elvei és gyakorlata

Az intézményértékeléseket a szakemberek olyan lehetőségnek és egyben szükség-szerűségnek is tartják, amelyek célja a tanulási folyamat állandó fejlesztése általt, hogy az iskolák visszajelzést kapnak teljesítményükről, ami segíti őket tanítási gyakorlatuk javításában, s felelőssé teszi őket saját teljesítményükért, eredményükért. Az intézményértékelések tehát fejlesztési célokat és elszámoltathatósági célokat egyaránt szolgálnak, s országonként is változó, hogy hol, melyik dimenzió kerül jobban előtérbe, kik végzik az értékeléseket (külső vagy belső értékelés, vagy ezek kombinációja zajlik), és az értékelés eredményei milyen nyilvánosságot kapnak. Ahol a külső értékelés erőteljesebben jelen van, ott az elszámoltathatósági célú használat is általában erőteljesebb, de számos országban ilyen körülmények közepette is a fejlesztési célú használat kerül előtérbe, nem csupán a közvetlen visszajelzés, ajánlások révén, hanem azáltal is, hogy a külső értékelési eredmények ismeretében lehetőség nyílik az iskola erősségeinek és gyengeségeinek a megvitatására, ami a belső önértékelési munkára is ösztönzőleg hat. A kizárólag intézményi önértékelésre építő oktatási rendszerekben érthető módon a fejlesztő funkció áll a középpontban. Az intézményeket különböző mutatók mentén összehasonlító értékelési rendszerek a legtöbb esetben mindkét célt szolgálják általt, hogy visszajelzéseket adnak az egyes intézmények számára, amelyeket azok fejlesztési célokra használhatnak, ugyanakkor a nyilvánosság tájékoztatása révén egyfajta elszámoltatható-

sági funkciót is betöltenek, lehetővé téve az intézmények közötti összehasonlítást, például az iskolaválasztás során (OECD, 2013).

Az, hogy melyik ország milyen gyakorlatot alakít ki az oktatási intézmények értékelése tekintetében, sok körülménytől függ. A külső intézményértékeléseket is folytató országokban gyakran építenek az önértékelések eredményeire, de arra is találhatunk példát, hogy szándékosan elkülönítik a kettőt egymástól, azt a megfontolást szem előtt tartva, hogy az önértékelések külső értékelésekhez való használata gyengíti azok intézményfejlesztő funkcióját. Több kutatási eredmény alapján arra a következtetésre juthatunk, hogy a külső értékelések fejlesztő, tanácsadó jellegének erősítése – a tisztán ellenőrző funkció helyett – az objektív és pontos jelentések, továbbá az összehasonlító elemzések révén maga is hozzájárulhat az intézményfejlesztésekhez (Matthews és Sammons, 2004; Dederig és Müller, 2010).

Akár önértékelésről, akár külső értékelésről legyen szó, az intézményértékelések területei között a tanulói előrehaladást a legtöbb országban nagy súllyal veszik figyelembe, legyen szó a vizsgaeredményekről, a teszteredményekről, vagy a továbbhaladási, illetve lemorzsolódási mutatókról. A tanulói teljesítményadatok mind az intézményi önértékelések, mind a külső értékelések számára hasznos adatforrásként szolgálnak, a standardizált mérések használata nem csupán elszámoltathatósági és iskolafejlesztési célokat szolgál, hanem lehetővé teszi az egyes intézmények közötti összehasonlítást is, ami a nyilvánosság nagyobb kiszolgálását is biztosítja (OECD, 2013).

Többen felismerték azonban, hogy a „nyers”, abszolút tanulói teszteredményeken vagy a vizsgaeredményeken alapuló teljesítményadatok publikálása sokkal inkább képet ad egy adott iskola bemeneti jellemzőiről, mint magáról az oktatás eredményességéről. Az eredmények és a szocioökonómiai státusz a legtöbb országban szorosan összefüggnek, emiatt az egyszerű átlageredmények nem kis részben inkább az intézmény szocioökonómiai jellemzőit, a tanulók társadalmi összetételét mutatják, mintsem az iskolák hozzájárulását tanulóik teljesítményéhez. Egy másik megközelítést jelent, amikor minden egyes iskola teljesítményét a saját korábban elért teljesítményéhez viszonyítják, mint ahogyan az Egyesült Államok bizonyos tagállamaiban teszik. Ennek a módszernek hátránya viszont, hogy nem tudja figyelembe venni a tanulói populáció évről-évre eltérő összetételét, sajátosságait. Ezzel szemben a hozzáadott érték típusú megközelítések középpontjában az egyes diákok teljesítményének a változása áll, ami sokkal igazságosabb megközelítés lehet, még akkor is, ha tudjuk, hogy a standardizált tesztek és a vizsgaeredmények csak korlátozott dimenzióit képesek mérni az iskolák teljesítményének (Faubert, 2009). A korai kutatási eredmények felhívták a figyelmet arra, hogy a család szociokulturális háttere a magasabb iskolai szinteken is meghatározó, s hogy az ún. intézményi

hozzáadott érték vizsgálata során mind ezt, mind pedig a korábbi iskolai végzettséget, teljesítményt figyelembe kell venni, különben azok az iskolák hátrányba kerülnek, amelyek kedvezőtlenebb háttérű tanulókat iskoláznak be (*Sammons és mtsai.*, 1994).

A hozzáadott érték típusú megközelítésekkel szemben is megfogalmazódnak azonban kritikák, amelyek részben módszertani jellegűek, de oktatáspolitikai szempontból is figyelemre méltóak. A kutatások arra hívják fel a figyelmet, hogy a tanulók tanulmányi teljesítménynövekedése alapján becsült iskolai/tanári hatás, vagyis a pedagógiai hozzáadott érték számításának eredménye – még ha ugyanazzal a módszerrel állítják is elő –, különbségeket mutathat az olyan modellspecifikációk szerint, mint a választott függvényforma vagy a bevont változók köre (*Steedle*, 2011; *OECD*, 2008). A kontextuális hozzáadott érték-modellekre vonatkozóan további megfontolást jelent az adatok nehéz értelmezhetősége, hiszen a kapott értékek bonyolult számítások eredményei, s éppen azt nem tudják megmutatni, hogy miért magasabb egy iskola teljesítménye, mint egy másiké, ezért iskolafejlesztési célokra kevésbé alkalmasak (*Masters*, 2012). Ráadásul az ilyen komplex modellek használata lehetővé teszi az eredmények „kozmetikázását” is, továbbá teret engedhet az alacsony teljesítménnyel szembeni megengedőbb hozzáállásnak, s akár intenzívesítheti is az alacsony elvárásokat (*OECD*, 2013).

Az eredmények mérése egyes országokban nem korlátozódik pusztán a kognitív dimenziókra, s azon belül az egyes tantárgyi teljesítményekre, hiszen az iskoláktól nemcsak azt várják el, hogy ezeken a területeken fejlesszék tanulóikat, hanem azt is, hogy például a szociális, állampolgári kompetenciáik, vagy az olyan összetett készségeik is fejlődjenek, mint a csoportos munkavégzés, kreativitás, problémamegoldás. Az iskolák értékelése tehát e dimenziók mentén éppúgy releváns lehet, mint a méltányosság, vagy az olyan dimenziók figyelembe vétele kapcsán, amikor például arra fókuszálnak, hogy a legrosszabbul teljesítő és a legjobban teljesítő tanulói csoportok közötti szakadék minél kisebb legyen. Ez utóbbi már a korábban említett NCLB program esetében is követelményként fogalmazódott meg, ugyanakkor kiderült, hogy a különböző tanulói összetételű iskolák eltérő mértékben tudnak ennek a célnak megfelelni (*Linn*, 2005).

Összességében tehát az egyes országok tapasztalatai azt mutatják, hogy az iskolai teljesítménymérések használata nagyon sok előnnyel járhat, akár az eredmények önértékelésre és iskolafejlesztési célokra való használatát tekintjük, akár a külső értékelések eredményességének növelését és az elszámoltathatóságot tartjuk szem előtt. Az egyes iskolák, illetve magának az oktatási rendszer teljesítményének az értékelése azonban elválaszthatatlan az esélyegyenlőség és a méltányosság kérdéskörétől. Az összehasonlító nemzetközi tanulói teljesítménymérések és az azok kapcsán

készülő elemzések alapján sokan jutottak arra a következtetésre, hogy a méltányos oktatási rendszerben tanuló diákok összességében magasabb szintű teljesítményt nyújtanak, mint a szelektívebb iskolarendszerekben. Az oktatás terén sikeres országok példái azt mutatják, hogy azok az iskolarendszerek teljesítenek jobban, amelyek képesek valamennyi diákjukat sikeresen fejleszteni. A PISA-méréseken jól szereplő országok egyik közös jellemzője, hogy valamennyi diák számára igyekeznek megfelelő feltételeket nyújtani a fejlődéshez (McKinsey, 2007).

A tanulói teljesítménymérések fejlődése

A nemzetközi tanulói teljesítménymérések kezdetei az IEA² megalakulásához és első méréseihez köthetők. Az IEA-vizsgálatok három nagy terület, a szövegértés, a matematikai és a természettudományos tudás vizsgálatára fókuszálnak. Magyarország 1970-től kezdve vesz részt a társaság munkájában és az általa koordinált nemzetközi mérésekben. Az utóbbi évtizedben rendszeres időközönként sorra kerülő TIMSS³ és PIRLS⁴ vizsgálatok a kezdeti mérések folytatásának tekinthetők, s egyre több ország bevonásával történnek. A TIMSS mérések a matematikai és természettudományi területek oktatásáról a 4. és 8. évfolyamokban adnak képet 5 évente, s a mérések során a tantervekben szereplő tananyagra koncentrálnak, ami lehetővé teszi a tanulói tudás és a tantervi követelmények összevetését, a legeredményesebb oktatási gyakorlatok megismerését. A PIRLS mérések célja a 4. évfolyamon tanulók szövegértésének vizsgálata, amely mérésre 4 évenként kerül sor; utoljára 2011-ben történt mérés. Mindkét mérés eredményeit elsősorban oktatáskutatási és fejlesztési célokra használják, céljuk, hogy nemzetközi szinten összehasonlítható adatokat szolgáltatassanak a tanulási és oktatási folyamatok eredményességéről és az oktatási rendszerek egyéb jellemzőiről, ezért a mérésekhez tanulói, tanári és intézményvezetői háttérkérdőív is tartozik (Balácsi és Horváth, 2011). A mérési eredmények azt mutatják, hogy a távol keleti országok – Szingapúr, Dél-Korea, Tajvan, Japán és Hongkong – oktatási rendszerei a legsikeresebbek, Európán belül pedig Finnország és Oroszország eredményei figyelemre méltóak. A 4. évfolyamos magyar diákok eredményei mindhárom mérési területen a legjobbak közé tartoznak Európán belül, 8. évfolyamon pedig matematikából a nemzetközi átlaggal egyenértékű, természettudományból a TIMSS-országok átlagánál jobb tudással rendelkeznek

² International Association for the Evaluation of Educational Achievement

³ Trends in Mathematics and Science Study

⁴ Progress in Reading Literacy Study

(Balázsi és mtsai., 2012a; Balázsi és mtsai., 2012b) (lásd még e kötetben *Ostorics László* írását).

Míg az előbbi tesztek elsősorban az iskolai tudást mérik, addig az OECD PISA-mérése⁵ inkább a megszerzett tudás gyakorlati alkalmazhatóságára, a mindennapi élet során fontos kompetenciák, készségek vizsgálatára koncentrál. A 2000-ben indított, háromévente ismétlődő felméréssorozat a 15 évesek szövegértési, matematikai és természettudományos képességeinek vizsgálatára koncentrál, ezáltal tehát olyan információkat, tudáselemeket gyűjt, amelyek a későbbi életrészekben, így például a munkavállalás során is hasznosíthatóak. Egy-egy képességterület részletes vizsgálatára 9 évente kerül sor, legutóbb, 2012-ben a matematikai kompetenciák és az azok alakulását befolyásoló iskolai, társadalmi tényezők feltárása volt fókuszban. A legutóbbi PISA-eredmények szövegértés területen kismértékű romlást, matematika és természettudományok területén az OECD átlagánál lényegesen gyengébb eredményt mutattak, amely ráadásul az előző méréshez képest szignifikáns csökkenést is jelent (Balázsi és mtsai., 2013) (lásd még e kötetben *Ostorics László* írását).

Mindhárom nemzetközi felmérés megerősíti a családi háttér jelentős hatását a tanulók teljesítményére. A magyar tanulók mért képességeit ez a tényező határozza meg a leginkább, aminek hatása erősebb a többi országban tapasztalhatónál. A szülők végzettségének eltérései ráadásul számottevő szociális és gazdasági különbségekkel is jár, ami az iskolaválasztáson keresztül az iskolarendszerre is kivetül, és az iskolák minőségi egyenlenségéhez vezet. A PISA-mérések szerint hazánkban az alulteljesítők aránya a szövegértés és a természettudomány területen az OECD átlagnak megfelelő, matematikából azonban magasabb annál, míg a kiemelkedő eredményt elérőké minden területen az átlagosnál kedvezőtlenebb. Fontos sajátossága a hazai adatoknak, hogy bemutatja a hazai iskolarendszer nemzetközi összehasonlításban is erőteljesen szelektív voltát. Az iskolák közötti különbségek Magyarországon jóval nagyobbak, mint sok más országban, az Országos kompetenciamérés eredményei pedig arra is rámutatnak, hogy az idősebb korosztályoknál nagyobbak, mint a fiatalabbaknál, ami azt jelzi, hogy az iskolarendszer nem kiegyenlíti, hanem tovább növeli a különbségeket (Balázsi és mtsai., 2014).

Feltehetően a nemzetközi teljesítményméréseknek is szerepük van abban, hogy ma már egyre több OECD-országban találkozhatunk saját mérési rendszerekkel, amelyek ugyanakkor változatos jellemzőkkel rendelkeznek. Az OECD összehasonlító nemzetközi tematikus áttekintése szerint az alapfokú iskolázásban a standar-

⁵ Program for International Student Assessment

dizált mérések nagyon gyakoriak, míg a középfokon tanulók esetében kevésbé elterjedtek (lásd 1. táblázat). Bár vannak országok, amelyek csupán mintavétel-alapú méréseket végeznek, a legtöbb országban az adott évfolyamon tanulók teljes körére kiterjednek a mérések (OECD, 2013).

1. táblázat

A tanulók számára következményekkel nem járó központi tanulói teljesítménymérések a különböző képzési szinteken (2012)*

Nincs ilyen	Csak az általános programokra
Csak minta alapú	Minden képzési programra

Ország	Alapfok	Alsó középfok	Felső középfok
Ausztrália			
Ausztria			
Belgium (Fl.)			
Belgium (Fr.)			
Kanada			
Chile			
Csehország			
Dánia			
Észtország			
Finnország			
Franciaország			
Németország			
Görögország			
Magyarország			
Izland			
Izrael			
Olaszország			
Japán			
Dél-Korea			
Luxemburg			
Mexikó			
Hollandia			
Új-Zéland			
Norvégia			
Engyelország			
Portugália			
Szlovákia			
Szlovénia			
Spanyolország			
Svédország			
Törökország			
Anglia			
Észak-franciaország			
Skócia			
Egyesült Államok			

Forrás: OECD, 2013 (Figure 4.5.)

*Az adatok jelzésértékűek csupán, nem alkalmasak a szigorúan vett összehasonlításra.

A nemzetközi tanulói teljesítménymérések egyfajta katalizátorként hatottak a 2001-ben elindított Országos kompetenciamérésre is, a mérési rendszer létrehozását azonban számos más körülmény is erőteljesen befolyásolta. Ezek között voltak a nemzetközi folyamatokba illeszkedők éppúgy, mint a hazai előzményekre és sajátosságokra reagálók (Berényi, 2010). A kompetenciamérés az indulását követő néhány évben jelentős változásokon ment keresztül, amelyek egy része valamilyen oktatáspolitikai kihívásra adott reakció volt, míg mások inkább technikai vagy módszertani jellegűek voltak (lásd még e kötetben *Ostorics László* írását). A változások sok esetben a pedagógusok és az igazgatók tanulói teljesítménymérésekkel kapcsolatos hozzáállására is hatást gyakoroltak (lásd még e kötetben *Széll Krisztián és Szemerszki Marianna* írását).

Amint azt korábban is láttuk, az intézményértékelések kapcsán – beleértve a tanulói teljesítménymérések eredményeit is – az egyes országokban jellemzően kettős cél körvonalazódik: a fejlesztési és az elszámoltathatósági célú hasznosíthatóság. A

két funkció természetesen járhat együtt is, gyakoribb azonban valamelyik előtérbe kerülése. Ugyancsak megjelenik a nyilvánosság vagy a szakmai támogatás céljára történő adathasználat, ami részben az előző kérdéskörrel is összefügg. Az intézményközi összehasonlításra alkalmas eredmények nyilvánosságra hozatala terén eltérőek az egyes országok megfontolásai. Bár a nyilvánossá tétel ösztönözheti az egyes intézményeket illetve pedagógusokat eredményeik javítására, továbbá segítheti a szülők iskolaválasztási döntéseit, ugyanakkor – kutatási eredmények szerint – a hatások vegyesek, hiszen ez az ösztönző hatás nem mindig az elvárt módon érvényesül. Az eredmények szakmai célú visszacsatolása az intézmények számára akkor igazán hasznos, ha azonnali, s lehetőséget ad az egyes tanulók tanulási szükségleteire való gyors reagálásra. Kutatások ugyanakkor azt találták, hogy amennyiben csak az intézményvezető kap visszajelzést az eredményekről, az nem mindig jut el a pedagógusokhoz, vagy ha el is jut, esetenként kapacitáskorlátok vagy értelmezési nehézségek miatt nem hasznosul eléggé. A számítógépekkel támogatott tesztek mindenesetre nagyfokú előrelépést jelenthetnek e téren (OECD, 2013).

A tanulói teszteredmények fejlesztési célú használatára példa lehet Dánia esete, ahol 2010-ben olyan számítógép alapú adaptív mérési rendszert vezettek be, amely gyors visszajelzést képes adni az iskoláknak illetve a pedagógusoknak a tanulók különböző területeken nyújtott teljesítményéről. Az egyes tanulók eredményei egy elemző szoftver segítségével összehasonlíthatók az országos átlagokkal, éppúgy, mint az adott iskola többi tanulójának, különböző tanulócsoportjainak átlageredményeivel. Mind a visszajelzés gyorsasága, mind az elemző funkciók rugalmassága hozzájárul a tanulói teljesítménymérések fejlesztési célú használatához, hiszen a pedagógusok olyan hathatós pedagógiai eszközt kapnak kézhez, amely lehetővé teszi számukra a tanulók fejlődésének követését, egyéni fejlesztési tervek kialakítását (Shewbridge és mtsai., 2011).

A tanulói teljesítménymérések elszámoltathatósági szempontú értékelésre való használatát és a nyilvánosság teljes körű kiszolgálását találjuk például Ausztráliában, ahol azonban a teljesítménymérési eredmények nem önmagukban, hanem az intézmény egyéb jellemzőivel együtt jelennek meg, így például egy olyan index kíséretében, ami a tanulói összetételt mutatja (ICSEA-index)⁶. A nyilvánosságra került adatok így árnyaltabb képet adnak egy-egy intézmény munkájáról, hiszen az eredmények adott iskolai kontextusban értelmezhetők. Az index nem csupán a tanulók társadalmi összetételére vonatkozó információkat tartalmaz, hanem bizo-

⁶ Index of Community Socio-Educational Advantage. Az index összetevőiről és számítási módjáról többek között lásd a következő infografikát: letöltés: http://www.acara.edu.au/verve/_resources/About_icsea_2014.pdf. [(2015. 04. 30.)

nyos iskolai jellemzőket is, mint például az iskola földrajzi elhelyezkedésére, vagy a bevándorlók arányára vonatkozó mutatókat. A *My School* weboldalon⁷ lehetőség van az egyes iskolák eredményeinek az országos átlageredményekkel, illetve hasonló kiinduló helyzetben lévő iskolák eredményeivel való összehasonlítására is. (OECD, 2013).

Összegzés

Napjainkban a fejlett országok oktatáspolitikai döntései egyre inkább olyan adatokon, mérhető eredményeken alapulnak, amelyek objektív, összehasonlítható információkat jelentenek mind rendszerszinten, mind az egyes intézmények vonatkozásában. Ezekhez gyakran használják a mérés-értékelés különböző elemeit, így többek között a standardizált tanulói mérések, vizsgák eredményeit. Az értékelések különböző módszerekkel és különféle megfontolásokkal készülnek, mindig szem előtt tartva azokat az oktatáspolitikai célokat, amelyeket egy adott ország el kíván érni. Az intézményértékelések eredményei fejlesztési és elszámoltathatósági célokat egyaránt szolgálhatnak, amit a kivitelezés során is fontos szem előtt tartani.

A mérési eredmények alapján készülő pedagógiai hozzáadott érték számítások nagyon fontos jelzésekkel szolgálhatnak nem csupán elszámoltathatósági, hanem iskolafejlesztési szempontból is, továbbá információval láthatják el a politikai döntéshozókat, szakpolitikásokat, az oktatás különböző szereplőit (tanulókat, szülőket, pedagógusokat, igazgatókat) és a szélesebb nyilvánosságot egyaránt. A mérési eredmények a méltányossági dimenzió vizsgálatára is alkalmasak lehetnek, akár azáltal, hogy egyéni szintű követésre adnak lehetőséget, akár azáltal, hogy az adott iskolán belül a különböző összetételű tanulói csoportok eredményeit, vagy a legjobban és a legkevésbé jól teljesítők eredményeinek viszonyát is képesek megjeleníteni.

A teszteredmények fejlesztési célú használata akkor eredményes, ha gyors a visszacsatolás és a pedagógusok számára is értelmezhető, a pedagógiai gyakorlatban közvetlenül is hasznosítható információkat tartalmaz. Az elszámoltatási funkció és a nyilvánosság biztosítása ugyancsak fontos kívánalom, azonban ezzel kapcsolatban is sokféle megfontolást érdemes figyelembe venni. Amikor komoly következmények kapcsolódnak a teszteredményekhez, az iskolák viselkedése nem feltétlenül a kívánatos irányba változik. Félrevezető lehet, ha kizárólag a teszteredmények javulásán mérjük az oktatás eredményességét, hiszen az iskola rendszerint sokkal szélesebb spektrumban fejleszti tanulóit, mint amelyek a tanulói teljesítménymérések során

⁷ <http://www.myschool.edu.au/>

megjelenítést kapnak. Ráadásul kutatások azt mutatják, hogy a teszteredmények esetleg anélkül is javulhatnak, hogy az iskolákban az oktatás színvonala érdemben javult volna. Megoldást jelenthet, ha az intézmények értékelését többféle tanulói teljesítmény mérésére és több év mérési eredményeire alapozzák.

Hivatkozások

- Balázi Ildikó és Horváth Zsuzsanna (2011): A közoktatás minősége és eredményessége. In: Balázs Éva, Kocsis Mihály és Vágó Irén (szerk.): *Jelentés a magyar közoktatásról 2010.* (pp. 325–362.), Oktatókutató és Fejlesztő Intézet, Budapest.
- Balázi Ildikó és mtsai. (2012a): *PIRLS és TIMSS 2011. Összefoglaló jelentés a 4. évfolyamos tanulók eredményeiről.* Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/pirls/PIRLS_TIMSS_2011_osszefoglalo_jelentes_4evf_eredmenyeirol.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2012b): *TIMSS 2011. Összefoglaló jelentés a 8. évfolyamos tanulók eredményeiről.* Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/timss/TIMSS_2011_Osszefoglalo_jelentes_8evf_eredmenyeirol.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2013): PISA 2012 Összefoglaló jelentés. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/pisa/pisa2012_osszefoglalo_jelentes.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2014): Országos kompetenciamérés 2013. Országos jelentés. Oktatási Hivatal, Budapest. Elérhető: https://www.kir.hu/okmf/iles/OKM_2013_Orszagos_jelentes.pdf [Letöltés dátuma: 2015.04.30.]
- Berényi Eszter (2010): A mérési iskoláktól az iskolák megméréseig. *Educatio*, 19(4), 601–613.
- Creemers, B. P. M., Kyriakides, L. és Sammons, P. (eds.) (2010): *Methodological Advances in Educational Effectiveness Research.* Routledge/Taylor & Francis, London and New York. Elérhető: http://rufismada.files.wordpress.com/2012/02/methodological_advances_in_educational_effectiveness_research_quantitative_methodology_series_.pdf [Letöltés dátuma: 2015.04.30.]
- Csapó Benő és mtsai. (2014): Az iskolai teljesítmények alakulása Magyarországon nemzetközi összehasonlításban. In: Kolosi Tamás és Tóth István György (szerk.): *Társadalmi Riport 2014.* TÁRKI, Budapest. 110–136. Elérhető: <http://www.tarki.hu/adatbank-h/kutjel/pdf/b327.pdf> [Letöltés dátuma: 2015.04.30.]
- Dederling, K. és Müller, S. (2010): School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change*, 12(3), Springer, 301–322.
- Faubert, V. (2009): *School Evaluation: Current Practices in OECD Countries and a Literature Review.* OECD Education Working Paper No. 42. Elérhető: <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP%282009%2921&docLanguage=En> [Letöltés dátuma: 2015.04.30.]
- Fitz-Gibbon, C. T. (1996): *Monitoring education. Indicators, quality and effectiveness.* Continuum, London.

- Hanushek, E.A. és Lockheed, M.E (1994): *Concepts of educational efficiency and effectiveness*. In: Human Resources Development and Operations Policy Working Papers http://www-wds.worldbank.org/servlet/WDSContentServer/IW3P/IB/1994/03/01/000009265_3961005235512/Rendered/PDF/multi_page.pdf [Letöltés dátuma: 2015.04.30.]
- Halász Gábor (2009): Tényekre alapozott oktatáspolitikai és oktatásfejlesztés. In: Pusztai Gabriella és Rébay Magdolna (szerk.): *Kié az oktatáskutatás? Tanulmányok Kozma Tamás 70. születésnapjára* (pp. 187–191.). Csokonai Könyvkiadó, Debrecen.
- Horn Dániel (2011): *Az oktatási elszámoltathatósági rendszerek elmélete*. MTA-KTI Műhelytanulmányok MT-DP – 2011/13. Elérhető: <http://econ.core.hu/file/download/mtdp/MTDP1113.pdf> [Letöltés dátuma: 2015.04.30.]
- Linn, R. L. (2005): *Issues in the Design of Accountability Systems*. University of California. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles. Elérhető: <https://www.cse.ucla.edu/products/reports/r650.pdf> [Letöltés dátuma: 2015.04.30.]
- Masters, G. N. (2012): *Measuring and Rewarding School Improvement*. A discussion paper commissioned by the Commonwealth Department of Education, Employment and Workplace Relations (DEEWR). Elérhető: http://research.acer.edu.au/cgi/viewcontent.cgi?article=1014&context=policy_analysis_misc [Letöltés dátuma: 2015.04.30.]
- Matthews, P. és Sammons, P. (2004): *Improvement through Inspection: An Evaluation of the Impact of Ofsted's Work*. Crown Copyright.
- McKinsey & Company (2007): *Mi áll a világ legsikeresebb iskolai rendszerei teljesítményének hátterében?* Elérhető: http://www.onfejlesztoiskolak.hu/2008/mckinsey_magyar.pdf [Letöltés dátuma: 2015.04.30.]
- OECD (2008): *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. OECD Publishing, Paris.
- OECD (2013): *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*. OECD Reviews of Evaluation and Assessment in Education.
- OECD (2014): *PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V)*. OECD Publishing, Paris. Elérhető: <http://dx.doi.org/10.1787/9789264208070-en> [Letöltés dátuma: 2015.04.30.]
- Sammons, P. és mtsai. (1994): *Continuity of School Effects: A Longitudinal Analysis of Primary and Secondary School Effects on GCSE Performance*. *School Effectiveness and School Improvement*, 6(4), 285–307.
- Sammons, P., Thomas, S. és Mortimore, P. (1997): *Forging links: effective schools and effective departments*. Paul Chapman, London.
- Saunders, L. (1999): A Brief History of Educational „Value Added”: how Did We Get To Where We Are? *School Effectiveness and School Improvement*, 10(2), 233–256.
- Scheerens, J., Bosker, R.J. és Creemers, B.P.M. (2000): Time for self-criticism: on the reliability of school effectiveness research. *School Effectiveness and School Improvement*, 12(1), 131–157.
- Shewbridge, C. és mtsai. (2011): *OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium*. OECD Publishing, Paris. Elérhető: <http://www.oecd.org/edu/school/49177679.pdf> [Letöltés dátuma: 2015.04.30.]
- Steedle, J. T. (2011): Selecting value-added models for postsecondary institutional assessment. *Assessment & Evaluation in Higher Education*, 37(6), 637–652.
- Tomasz Gábor (2011): Tesztek helyett oktatást! – avagy egy lakásfelújítás szellemi hozzáadéka. *Educatio*, 20(1), 123–127. Elérhető: <http://www.hier.iif.hu/hu/letoltes.php?fid=tartalomsor/2054> [Letöltés dátuma: 2015.04.30.]

Nahalka István

Tanulói teljesítménymérések alkalmazhatósága a neveléstudományban

Kérdésfeltevés

A tanulói teljesítménymérések és a neveléstudomány között természetesen rendkívül szoros kapcsolatok vannak. Maguk a mérések a tudomány eredményeire építenek, a vizsgálatok kivitelezése tudományos szakemberek feladata, a konkrét tevékenységet a tudomány által kidolgozott eszközrendszerek használata jellemzi. Ez a szoros kapcsolat azonban abban is megnyilvánul, hogy a tudomány szerves részeként jelennek meg olyan kutatások, amelyek támaszkodnak a tanulói teljesítmény vizsgálata során született eredményekre. E tanulmány arról szól, hogy milyen problémákkal kell számolni akkor, amikor a neveléstudomány felhasználja a tanulói teljesítménymérések eredményeit.

Ez a kérdés a mérhetőség kérdése. A neveléstudomány – majd hamarosan látjuk, hogy karöltve a matematikával – megteremti az elvi alapjait a pedagógiai méréseknek, megfogalmazza azokat az elvárásokat, amelyeket a méréseknek teljesíteniük kell. Ezek az elvárások vagy teljesülnek, vagy nem. A kapott adatokat a neveléstudomány – legalábbis ez lenne a „normális” helyzet – csak akkor használhatná fel, ha az adott méréssel minden rendben van. Vajon minden rendben van? A szakemberek többsége által is sokra tartott, a kvantitatív tudományos vizsgálódások értékrendje alapján a legmagasabb szintre helyezett empirikus kutatásokkal lehet bármi probléma? És ha vannak nehézségek – majd látjuk, hogy igen, rendkívül nehezen megválaszolható kérdések merülnek fel –, akkor az érinti a „nagy méréseket”, a PISA, TIMSS, IALS, PIRLS vizsgálódásokat, a hazai Országos kompetenciamérést (OKM) is? Netán ezekkel alapvető bajok vannak?

Méréselméleti megközelítés, történeti megjegyzések

Nincs tudomásom olyan, a neveléstudományi szakemberek nem elhanyagolható körében ismert, magyar nyelvű szakmai munkáról, amely a matematikai méréselmélet neveléstudományi konzekvenciáival foglalkozna (a neveléstudomány területén valószínűleg igen kevés ismert forrás: *Kehl*, 2011). A méréselmélet, mint matematikai részdiszciplína története *Hermann von Helmholtz* munkásságáig, 1887-ben megjelent, és e terület fejlődésében a nyitányt jelentő művéig vezethető vissza. Ettől kezdve a mai napig egy izgalmas, és éppen a humán tudományok (mindenekelőtt a pszichológia és a neveléstudomány) számára sokat mondó fejlődés tanúja lehetett több kutató generáció. Hazánkban magyar nyelven ennek a fejlődésnek nincs különösebben nyoma a szakirodalomban. A kutatásmetodológiai munkák sem szólnak erről a kérdésről, és az érintett konkrét kutatási programok (pl. a képességfejlettségmérések eredményeit használók) sem foglalkoznak méréselméleti megalapozással.

Eközben a nemzetközi szakirodalomban a kérdéskör nagyon is jelen van. A matematikai méréselmélet *Helmholtz* művének megjelenése óta több paradigmaváltáson ment át, és ma is fejlődik. (Csak példaképpen lásd *Trendler*, 2009; *Cervone* és *Caldwell*, 2008; *Rossi*, 2007; *Luce* és *Suppes*, 2002; *Narens*, 2002, 1985; *Maraun*, 1999; *Suppes*, 1998; *Díez*, 1997; *Luce*, 1996, 1992; *Cliff*, 1992; *Michell*, 1990; *Roberts*, 1979; *Krantz és mtsai.*, 1971.) E fejlődés egyik érdekessége, hogy bár a matematika tudományának belső szükségletei is generálták, mégis fontos szerephez jutottak éppen a mi tudományaink, a pedagógia és a pszichológia is, mivel a humán képességek, attitűdök, valamint a személyiség vizsgálata számos új matematikai problémát vetett fel.

A természettudományokban a mérés az újkor hajnalán lezajlott nagy átalakulások idején vált igazán fontossá, és a 20. század elejére a természettudományi mérések precíz, axiomatikus elmélete is létrejött (*Hölder*, 1901). *Klasszikus méréselméletnek*¹ is nevezik a kialakuló matematikai elméletet², amely azonban csak az összeadódó (additív) mennyiségek mérését alapozta meg. Az additív mennyiségek azok, amelyek esetében van egy „természetes összeadódás”, szakkifejezéssel élve *konkatenáció* a mérésben részt vevő objektumok között, aminek szinte tükörképe az objektumoknak megfelelően számok összeadása (*Narens*, 1981). Az ács azért adhatja össze két, egymás meghosszabbításaként a tetőn felerősített lécszárakat,

¹ Vigyázat, a klasszikus méréselmélet nem azonos a klasszikus tesztelmélettel (erről is lesz szó a tanulmányban). A méréselméletek kérdésfelvetése jóval tágabb, mint a tesztelméleteké. Igaz, mindkettő esetében létezik egy-egy, a „klasszikus” jelzővel ellátott változat.

² De közkeletű megnevezés az „extenzív struktúrák elmélete” is.

mert ez a bizonyos meghosszabbítás, a *konkatenáció* ugyanúgy viselkedik, mint a számok összeadása. Tehát vannak bizonyos összefüggések a „lécek világában”, amelyek jól tükröződnek a számok világában. A „dolgok” struktúráját adja vissza a számok struktúrája. Ez az alapgondolata valójában a modern, *reprezentációs méréselméletnek*.

Könnyű a természettudományoknak! Elméleti rendszereik bőséggel kínálnak összeadódó mennyiségeket. A hossz mellett a tömeg, az idő, az elektromos töltés mennyisége, az energia, és még számtalan más, a természettudományos elméletek megalapozásában döntő szerepet játszó mennyiség additív (a fizika ezeket extenzív mennyiségeknek nevezi). Bár a hőmérséklet, és a sűrűség kivételek. Jeannak az ismert viccben ura megparancsolja, hogy nyissa ki az ablakot, mert a 18 fokos szobába be kellene engedni a külső 5 fokot. De harmadikos gyerekek is rendszeresen írják a feladatlap megfelelő kérdésére, hogy a két pohárban lévő, egyaránt 30 fokos vizet egy edényben összeöntve 60 fokos lesz a víz (Nahalka, 2002). Ezek a „tévesztések” visszavezethetők arra, hogy az érintettek a hőmérsékletet is összeadódó mennyiségként kezelik. Hölder elméletébe még nem fértek bele ezek a mennyiségek, azonban a fizika számára a kiterjesztés természetes, és a teoretikus háttér is könnyen megteremthető volt. A nem összeadódó mennyiségek leszámaztathatók az additívakból (függvényszerű kapcsolatokban állnak azokkal), az összeadódó mennyiségek mérése megteremti a „kimaradók” mérésének lehetőségét is. Az additív és a leszámaztatott mennyiségek mérésére vonatkozó elmélet már megoldotta a fizika minden elemi mérési problémáját (a mérés ettől függetlenül nagyon izgalmas kérdés maradt a természettudományokban, de már egészen más okokból, amelyek – ma még – a humán tudományokban irrelevánsak).

A pszichológia – és annak nyomában később a neveléstudomány – problémái azonban éppen itt kezdődnek. E tudományokban lényegében nincsenek összeadódó, additív mennyiségek. Ellenpélda ugyan a reakcióidő, a „darabszám”, de ezekre és a még néhány előforduló kivételre komoly pszichológiai és pedagógiai elméleteket nem lehet építeni. Ettől függetlenül a pszichológusok már a 19. században is mérnek. Így alapozzák meg az intelligencia fogalmát, és elkezdik mérni az intelligenciahányadost (IQ), amit aztán az egyes képességek fejlettségi szintjének, még később az attitűdök erősségének mérése követ. Ma is igaz, hogy tesztek felvételén alapszik számos döntés (pl. álláshelyek betöltése, sajátos nevelési igényűvé nyilvánítás, iskolai osztályba való besorolás, egyetemi felvétel). Így nyernek adatokat számos professzionálisnak mondott kutatás számára is (pl. képesség- és attitűdvizsgálatok, a személyiség kutatása). És tesztek veszünk fel a PISA, TIMSS, IALS, PIRLS és még számos nagy nemzetközi vizsgálatban is, és ezek jelentik a kiindulópontját

az Országos kompetenciamérésnek (OKM) is (lásd még e kötetben *Ostorics László* írását).

A probléma meglehetősen korán világossá vált. 1933-ban a *British Association for the Advancement of Science* kiadta jelentését, amelyben a tekintélyes testület tagjai szinte diszkreditálták a pszichológiai méréseket. A verdikt kemény volt: mivel a pszichológiában nincsenek összeadódó mennyiségek, ezért a mérés ebben a tudományban lehetetlen. A pszichológia (és erős szövetségeseiként a pedagógia) harca azóta tart a mérések erős elméleti megalapozásáért.

Egy amerikai halláskutató, *Stanley Smith Stevens* tett sokat azért, hogy a pszichológiai mérések elfogadottá váljanak. *Stevens* nem volt matematikus, mégis a mérési skálákon alapuló, nem teljes precizitással kifejtett rendszere sokat jelentett a későbbi kutatások számára. (*Stevens*, 1946). A halláskutató alapgondolatára érdemes rávilágítani: a klasszikus méréselméletben (az összeadódó mennyiségek elméletében) lefektetett alapok túl szűkek. Nemcsak a konkatenálhatóság adhatja a mért „dolgok” közötti viszonyokat, hanem itt figyelembe vehetők más matematikai struktúrák is. Szűkkeblűek voltak a természettudósok, amikor „kijelölték” a mérhetőség határait.

Reprezentációs méréselmélet

Mindig *valamiket* mérünk. Van a mért „dolgoknak” egy halmaza, amelynek jól elkülöníthető elemei vannak. A mérés során az elemekhez bizonyos szabályok szerint számokat rendelünk hozzá. Hogy milyen szabályok érvényesülnek, az jelentős mértékben befolyásolja, hogy egyáltalán mit tehetünk a számokkal, milyen számításokat, például milyen statisztikai elemzéseket hajthatunk végre.

A megértést talán jól segítheti egy negatív példa. Ma sok olyan kutatás zajlik³, amelyben tesztekkel használnak a szakemberek. Például vizsgálják valamely képesség fejlettségét. Egy megfelelően kiválasztott mintába tartozó gyerekekkel a kutató megoldatja a feladatokat. Minden gyerek esetén kiszámítja a helyesen megoldott feladatok százalékos arányát, és máris megtörtént a számok hozzárendelése. Innen-től már – legalábbis a nem túl felkészült kutató ezt gondolja – csak szoftverkezelési (pl. SPSS), és statisztikai számítási kompetenciáján múlik, hogy mit tud kihozni a sok esetben temérdek számból. Vannak felkészültebbek, akik ismerik a tesztfelkészítés klasszikus tesztelmélet szerinti kötelező kúrjeit. Elővizsgálatot folytatnak, itemanalízist végeznek, Cronbach-alfát számolnak, az itemeket szelekciónak vetik

³ E helyen szándékosan nem hivatkozom konkrét forrásokra, mivel nem kipellengérezni akarok, nem a kritika az elsődleges célom.

alá, mindaddig, amíg a tesztnek nincsenek klasszikus tesztelméleti értelemben vett, megfelelő jóságmutatói. A még ennél is felkészültebbek nem a klasszikus, hanem a modern tesztelmélet modelljeit alkalmazzák, és addig cserélgetik benne a feladatokat megfelelőbbekkel, amíg nem teljesülnek mondjuk valamelyik Rasch-modell⁴ feltételei. Ekkor már nincs akadálya a feladatsor alkalmazásának, a pontszámok szoftveres támogatást igénylő meghatározásának, és akár a legfejlettebb statisztikai számítási eszközök bevetésének. Mindjárt megmutatom, hogy az első, a tesztelméleti tudást nem használó kutatók eljárása dilettáns, a klasszikus tesztelmélettel a feladatot megoldók egy bizonyíthatóan hibás metódust alkalmaznak. A modern tesztelméleti eszközöket bevetők módszereiben nincs ugyan hiba, de nagyon sokszor olyan képességekre (attitűdökre, személyiségvonásokra) alkalmazzák a (hibátlan) módszereket, amelyekre azok nem használhatók.

Mik akkor a lehetséges szabályok? Mi a baj ezekben a példákban? A reprezentációs méréselmélet a 20. század közepén alakult ki (Scott és Suppes, 1958; Suppes, 1951), és a Stevensi gondolatot öntötte matematikai formába. Az a lényege, hogy a mérés során úgy rendelünk számokat a mért „objektumokhoz”, hogy a számok valamilyen struktúrája pontosan megfeleljen a mért „objektumokon” kimutatható struktúrájának. Mint ahogy a lécek egymás folytatásaként történő, egy egyenesbe eső elhelyezése (e mennyiség esetén a konkatenáció) pontosan megfelel a számok összeadásának. Ha két lécet egymás folytatásaként helyezünk el, akkor kvalitatív értelemben ugyanolyan hosszú egyesítést kapunk, mint amilyen egy harmadik léce. Keresd a mért dolgok között a struktúrát, és ha megtalálod, remélheted, hogy a számok hozzárendelésével e szerkezetet reprezentálni tudod a számok világában, s remélheted, hogy ez a struktúra még komoly statisztikai számításokra is feljogosít.

Lehet, hogy a számok csak a kategóriákat különböztetik meg egymástól. A szabály az, hogy az ugyanabba a kategóriába sorolt egyedekhez mindig ugyanazt a számot rendeld hozzá, és ha két egyed tulajdonságai különböznek, akkor semmiképpen se kaphassák a hozzárendelés során ugyanazt a számot. Egyáltalán nincs szerepe a számok nagyságának, nem érdekes a sorrendjük sem. Csak az a fontos, hogy a mért dolgokhoz rendelt számok az azonosságukkal és a különbözőségükkel tükrözzék a mért „dolgok” közti azonosságot (ugyanolyan tulajdonságot) és az elté-

⁴ Georg Rasch dán matematikus írta le először a modern tesztelmélet legegyszerűbb modelljét. Ebben a képességfejlettségek becslése, az adott képességhez tartozó feladatok nehézségparamétere, illetve az ezekkel függvénykapcsolatban álló, az adott személyre vonatkozóan az adott feladat jó megoldásának valószínűsége játsszák az alapvető szerepeket (Rasch, 1960; Molnár, 2013). A Rasch-modell szerinti képességfejlettség-, feladatnehézség számításokhoz ma már szoftverek nyújtanak segítséget.

rést. Az ilyen adatokkal is sokat lehet kezdeni, az így kialakuló skálát (a skála maga a hozzárendelés) *nominálisnak* nevezzük.

Lehet, hogy a struktúra a *sorba rendezhetőség*. Valamilyen értelemben az egyik mért egyed (az adott tulajdonságát tekintve, például képességfejlettségében) előrébb van, mint egy másik. És ezt bármely két párról meg tudjuk állapítani. A matematikában egy-két jól formalizálható szabály érvényesül. Az egyik például azt mondja, hogy ha A megelőzi B -t, és B megelőzi C -t, akkor A megelőzi C -t is. Még mielőtt bármit mondanánk a számok hozzárendeléséről, egy ilyen struktúra felfedezhető a mért „dolgokon”. Ilyen például a gyerekek apjának iskolai végzettsége. Sok vizsgálat van, amelyben (1) a maximum 8 általános iskolai osztály, (2) a szakképzettség, (3) az érettségi és (4) a felsőoktatásban szerzett diploma azok az értékek, amelyekkel a mért „dolgok” jellemezhetők. Ezeknek az értékeknek van egy határozott sorrendje. A számokat úgy kell hozzárendelni az egyedekhez, hogy azok sorrendje tükrözze az apák iskolai végzettségének sorrendjét. Pl. az 1, 2, 3, 4 számok erre éppen megfelelnek, de nyugodtan választhatnánk az 1, 13, 15, 27 számokat is. Csak a sorrend a fontos. Így jönnek létre az *ordinális skálák*. Az ordinális skálák is többféle statisztikai vizsgálatra adnak lehetőséget. Fontos azonban, hogy a nagyon sokszor használt statisztikai számítások, mint az átlag, a szórás meghatározása; ebből következően számos statisztikai próba (t-próba, szóráselmélet); és számos többváltozós elemzés (faktoranalízis, klaszterelemzés, regresszióelemzés) ilyen adatokkal nem végezhető el. Ugyanakkor jó tudni, hogy hasonlóak léteznek az ordinális változókra is, tehát egyáltalán nem vagyunk eszköztelenek (Csíkos, 1999).

A „dolgokon” megnyilvánuló struktúra lehet még összetettebb. Lehet, hogy pontosan meg tudjuk mondani, hogy párok közötti kvalitatív értelemben vett „eltérés”, „különbség”, „távolság” mikor *nagyobb*, ha azt összehasonlítjuk egy másik párt jellemző „eltéréssel”, stb. Ha ugyanannyi vizet két különböző folyamatban melegítünk, akkor a melegítés ideje meghatározza, hogy melyik folyamatban fog jobban emelkedni a víz hőmérséklete. Úgy kell a számokat hozzárendelni az állapotokhoz, hogy ezt a viszonyt a számok is tükrözzék. Nem véletlen, hogy nem írtam pedagógiai példát. Amit jó szívvel írhatnék, az nagyon bonyolult lenne (modern tesztelméleti interpretációt adva valamilyen nagyon egyszerű képesség, mondjuk az egyjegyű, pozitív egész számok összeadása fejlettségével kapcsolatban). Más, nagyon sokak által elvégzett kutatásokban szereplő példát viszont nem szívesen írnék, mert azokat nem tartom korrekteknek.

Az így létrejövő skálák az *intervallumskálák*, és ennek az egész elemzésnek a középpontjában állnak. Ugyanis ha sikerül egy kutatásban ilyen skálán elhelyezkedő adatokat kapni, akkor nyert ügyünk van, intervallumskálán elhelyezkedő adatokkal már szinte minden olyan statisztikai vizsgálat elvégezhető, amelynek

számunkra nagy a jelentősége. Azok is, amelyeket az ordinális skála esetén, mint „tiltottakat” mutattam be.

A szokásos leírásokban itt következik még egy fontos skála bemutatása, ez az *arányskála*, amely ugyanaz, mint amit a klasszikus méréselmélet tartalmaz. Láttuk már, hogy értelmezése során a kiindulópont a konkatenálhatóság⁵, ami az intervallumskálákhoz viszonyítva még több lehetőséget teremt a statisztikai vizsgálatok számára. Érdekes módon az arányskála tulajdonságaira a legtöbb esetben a pszichológiai és pedagógiai vizsgálatok esetén már nincs szükség. Nem véletlen, hogy az SPSS statisztikai program, amelyet a humán kutatások esetén szívesen használnak a szakemberek, nem is ajánlja fel a változók jellegének meghatározására az arányskála lehetőségét.

Vajon miért pazaroltam a mindig szűkös karakterszámot a sokak számára unalomig ismert négy, itt bemutatott skálára? Azért, mert nem látom, hogy ezekből az ismeretekből levontuk volna a megfelelő következtetéseket. És most vegyük elő ismét a negatív példákat! A „tesztet” szinte vakon alkalmazó kutató egyáltalán nem foglalkozik azzal a kérdéssel, hogy milyen skálát alakít ki. Örül annak, hogy végre vannak számai, és szinte fejjel megy a falnak, amikor ezeket begépezi az SPSS-be, és már végzi is mindazokat a számításokat, amelyek csakis akkor lennének véggezhetőek, ha intervallumskálája lenne. Ezt azonban nem biztosítja semmi. A kutató adatai, eredményei így teljesen érdektelenek.

A klasszikus tesztelméletet alkalmazó szakember már egy lépéssel tovább lép. Nem teljes mértékben reménytelen vállalkozás, hogy egy teszt segítségével, a jól megoldott itemek arányának kiszámítása után kapott adatok intervallumskálán helyezkedjenek el. Ennek szükséges feltétele az lenne, hogy az adott képesség fejlettségének méréséhez minden egyes teszt alapján kialakított pontszámok egymással ún. pozitív lineáris kapcsolatban álljanak. Ez azt jelenti, hogy ha vesszük bármely két lehetséges teszttel elvégzett vizsgálat két számsorát, akkor az egyikből a másikat úgy kell kiszámolni, hogy minden egyes tanulónál az egyiknek a pontszámát meg kell szorozni egy minden pár esetére ugyanazon számmal, és minden esetben ugyanazt a számot kell hozzáadni. Ahogyan a Celsius és a Fahrenheit hőmérsékleti skálákon mért értékeket is át lehet számolni egymásba. Sokan ismerik ennek módját:

$$F = \frac{9}{5}C + 32,$$

⁵ A matematikai struktúra kifejtésében a konkatenációs reláció mellett szerepet kap még a mért „dolgozók” közötti, a megfelelő matematikai tulajdonságokkal rendelkező rendezési reláció is.

ahol F jelöli a hőmérséklet Fahrenheit skálán mért értékét, és C jelöli a hőmérséklet Celsius skálán mért értékét. Bármilyen Celsius fokokban meghatározott hőmérséklet Fahrenheit fokokban mérhető értékét kiszámolhatjuk, ha az előbbit megszorozzuk $9/5$ -del ($1,8$ -del), és a kapott értékhez hozzáadunk 32 -t. Ugyanez kellene, hogy érvényesüljön bármely ugyanazon képesség fejlettségének meghatározására szánt két teszt esetén is. Hogy ez így van, mondjuk a problémamegoldás, az induktív gondolkodás, a szöveges matematika feladat-megoldás, az arányszámítási képesség fejlettségének mérésekor (hogy létező példákat említsek), azt soha senki nem bizonyította be. Meggyőződése, hogy határozottan nem így van. Ezt azonban – a modern tesztelméleti modellek alkalmazása során tapasztalható problémákkal együtt – szeretném megmagyarázni.

Sokan azt gondolják, hogy a modern tesztelmélet modelljei megoldják a problémát. Valóban, a modern tesztelmélet matematikai értelemben teljes, mert szemben a klasszikus tesztelmélettel, világosan bebizonyítja, hogy ha bizonyos feltételek fennállnak, akkor igenis intervallumskálán helyezkednek el az adatok. Az is kétségtelen, hogy vannak bizonyos egyszerű képességek, amelyeknek fejlettsége e modellek alkalmazásával mérhető. A gondot az okozza, hogy számos modern tesztelméleti modell alkalmazásának az a feltétele, hogy *maga a vizsgált jelenség egydimenziós legyen*. Ez a tulajdonság többféleképpen is megmagyarázható, de egy példával hadd érzékeltessem a jelentést. A problémamegoldás képessége akkor lenne egydimenziós, ha legalább az igaz lenne (valójában ennél is többre van szükség), hogy ha egy tanuló egy feladatot nagyobb valószínűséggel tud megoldani, mint egy másikat, akkor ezzel a populációba tartozó minden tanuló így van, és bármelyik feladatpár esetén érvényesül ez a szabály. És fordítva, ha egy adott feladatot az egyik tanuló nagyobb valószínűséggel tud megoldani, mint egy másik tanuló, akkor minden feladattal, és bármelyik két tanuló esetén ez a helyzet. Természetesen nagyon kevés kivétel a gyakorlatban lehetséges, de ezek mennyiségének valóban elhanyagolhatónak kell lenni. Ezek a követelmények azonban a problémamegoldás esetén nem érvényesülhetnek. Nehézség számtalan területen megfogalmazható, és az is nyilvánvaló, hogy szinte minden tanuló esetén érvényes, hogy van a feladatoknak egy olyan köre, amelyek esetén ő jó teljesítményre képes, másokban viszont gyengébb eredményt ér el. Míg egy másik tanuló ugyanabban a populációban éppen ellentétes módon lehet jellemezhető. Ez, az összetett, bonyolult képességekre jellemző sajátosság nyilvánvalóvá teszi, hogy ezek a tulajdonságok nem lehetnek egydimenziósak. Ha nem egydimenziósak, akkor viszont semmilyen skálán nem tudjuk kifejezni a fejlettségi szintjüket.

Szinte hallom sok hozzáértő olvasóm ellenvetését, hogy a tesztek *reliabilitásának* becslése éppen azért szükséges, hogy az egydimenziósságot mutassuk meg a se-

gítségével, azt lássuk, hogy a teszt határozottan egy tulajdonság mérésére alkalmas. Ez az érvelés azonban több okból nem állja meg a helyét. A reliabilitás az a jóságmutatója a tesztnek, amely jelzi, hogy milyen mértékben vannak „közel” a ténylegesen mért értékek (a tesztpontszámok) a háttérben meghúzódó, a tényleges képességfejlettséget kifejező, valódi pontértékhez⁶. Ez utóbbi nem más, mint a teszttel kapható pontszámok várható értéke (az egyén esetében), ismeretlen, pontosan nem megmérhető, látens érték. Ha a ténylegesen, a vizsgálatban kapható tesztpontszámok ezen értékek közelében vannak, akkor magas a teszt reliabilitása, számértékét tekintve 1 közelében van, míg ha nem ez a helyzet, akkor kicsi a reliabilitás. A valódi érték nem határozható meg, a reliabilitás is elméleti érték. Becslésére a leginkább elfogadott, és legtöbbször használt mérték a Cronbach-alfa, amelyet az SPSS is meg tud határozni, de akár „kézzel” is kiszámolható. A kutatók örülnek, ha a Cronbach-alfa, a reliabilitás becslése mondjuk 0,9, mert ez azt jelenti – szerintük –, hogy a teszt jól mér, egydimenziós a jelenség, a teszteredményeket bátran tekinthetjük intervallumskálán elhelyezkedőknek. Csakhogy kimutatható, hogy a Cronbach-alfa nem jelzi jól az egydimenziósságot. Olyan adatsorokat lehet viszonylag könnyen előállítani, amelyekben mondjuk 2, vagy 3, vagy 4 dimenzió is jelen van, a reliabilitás becslése mégis nagyon magas értéket mutat (Sijtsma, 2009; Shevlin és mtsai., 2000). De jelentős problémát jelent az itemanalízis és a tesztfejlesztés maga is. Ugyanis a kutatók nagy valószínűséggel az itemek szelekciója és beválogatása által állítanak elő jó tulajdonságokkal (megfelelő jóságmutatókkal) rendelkező teszteket, miközben az eljárás értelme alapvetően megkérdőjelezhető. A problémamegoldás bármely feladatait tesztbe szervezve, a tesztnek megfelelőnek kellene lenni, csak ekkor beszélhetnénk arról, hogy képesek vagyunk mérni e képességet. „Műtesztekkel”, „gondos” feladatszelekció eredményeként előálló feladatsorokkal kaphatunk számadatokat, de azok nem mérhetik a problémamegoldás fejlettségét, mert ez a „valami” egész egyszerűen nem létezik. Az lehet, hogy valamit mérnek (mindegyik külön-külön valamit), de teljességgel indokolatlan azt a problémamegoldás fejlettsége mértékének tekinteni.

Már az ordinális skála tulajdonságai sem érvényesülnek. Mikor tekintjük az egyik gyerek problémamegoldó képességét fejlettebbnek a másik gyerekénél? Erre

⁶ Pontosabban: az elméleti reliabilitás nem más, mint a valódi értékek és a mérhető értékek, mint valószínűségi változók lineáris korrelációs együtthatója a populációban. Ha ismernénk egy mintába bekerült résztvevők valódi értékeit, és elvégeznénk a mérést (megoldatnánk a teszteket), akkor ennek az elméleti reliabilitásnak egy becslését közvetlenül ki tudnánk számolni. A valódi pontértékeket azonban nem ismerhetjük, ezért a reliabilitás becslésére (pl. Cronbach-alfa) más megoldást kellett keresni.

sosem született, mert nem is létezhet jó meghatározás. A kérdésnek végtelen sok összetevője van. Mint amikor egy egyenesen a pontok határozottan egy sorba rendezhető halmazt alkotnak, viszont ha a síkon „szórunk el” néhányat, azok esetében sorrend már nem alakítható ki. A matematika persze kiderítette, hogy valójában minden halmaz rendezhető, azonban ennek az elméleti lehetőségnek nincs pedagógiai relevanciája. Tegyük egy derékszögű koordináta-rendszert a síkra, rendezzük úgy a síkon lévő pontokat, hogy ezt a rendezést az origótól való távolságok határozzák meg! Csakhogy más és más módon elhelyezve a koordináta-rendszert, a pontok sorrendje más és más lesz. Ezt a reprezentációs méréselmélet nem engedi meg, ha ordinális skáláról van szó, akkor a sorrendnek bármely legitim skála esetén ugyanannak kell lenni. De furcsa módon még nominális skálákat sem alkotnak a modern tesztelméleti eszközökkel kapott eredmények. Senki nem tudja ugyanis megmondani, hogy két tanuló problémamegoldó képességét mikor tekinthetjük azonosnak. Ilyen meghatározás soha nem született, mert nem is születet.

A következtetés?

Nem tudhatom, hogy tisztelt olvasóim számára a fenti érvelés végeredménye mennyire meglepő. Ha nem látható át azonnal az eredmény furcsa volta, akkor annak érzékeltetésére álljon itt egy-két következmény:

1. Számos pszichológiai és pedagógiai kutatás eredménye megkérdőjelezhető. Olyan kutatásokról van szó, amelyek teszteket használtak képességek fejlettségének, attitűdök, személyiségjegyek erősségének mérésére. Ezek egy része egyáltalán nem használt tesztelméleti megfontolásokat, az ilyen kutatások semmilyen tudományosnak tekinthető alappal nem rendelkeznek. Egy részük klasszikus tesztelméleti alapokat használt, azonban méréselméleti megalapozás ezeket sem jellemezte, a skálák intervallum jellege nem bizonyított tény, mint láttuk, a magas reliabilitás-bebecslés sem menti meg a helyzetet. A modern tesztelméleti modellek alkalmazása matematikai értelemben korrekt lehet, azonban csak a képességek egy szűk körére lehet érvényes. A pedagógiát éppen érdeklő összetett képességek esetén (és hasonló a helyzet a bonyolultabb attitűdökkel, a komplex személyiségjegyekkel) szintén az egydimenziósság nem érvényesülése akadályozza meg a mérést.
2. Vannak pedagógusok, akik komoly tanulási folyamat után, viszonylag rendszeresen alkalmaznak tanulóik körében teszteket a biztonságosabb pedagógiai értékelés érdekében. Az ő tevékenységük eredményei is megkérdőjelezhetők, ha itt valóban arról van szó, hogy valamely összetettebb

képesség fejlettségét szeretnék mérni. Összetettebbnek számít már egy matematikai résztema feladataihoz köthető képesség is; egy példa: nagy valószínűséggel ilyen a másodfokú egyenletek megoldására szolgáló képleten alapuló feladatok rendszere is.

3. És az ebben a tanulmányban bennünket a leginkább érdeklő kérdés: mi a helyzet a „nagy mérésekkel” (PISA, OKM stb.)? Ez olyan fontos kérdés, hogy külön fejezetben taglalom.

A „nagy mérések” esete

A PISA-vizsgálatban a tanulók szövegértését, matematikai felkészültségét, és természettudományos feladatok megoldásában mutatott eredményességét vizsgálják a szakemberek. Kétségtelen, hogy mindhárom rendkívül összetett „képesség” (ha képességnek tekinthető egyáltalán). Az előző fejezetben éppen azt írtam le, hogy egy bizonyos, a reprezentációs méréselmélet alapján megfogalmazható érvelés szerint miért nem mérhetők az olyan összetett tulajdonságok, képességek, amilyenekről éppen ezekben az esetekben is szó van. Vagyis első megközelítésben a megelőző fejezetben azt mutattam ki, hogy mondjuk a PISA-vizsgálat, vagy az OKM teljesen labilis talajon állnak, sőt, valójában semmit nem mérnek. Azok a konstrukciók, amelyek az érvelésekben szerepelnek, mint a mérés tárgyai (matematikai eszköztudás stb.) valójában nem mérhetők.

És valóban arról van szó, hogy ha mi például a szövegértéssel kapcsolatosan vizsgálódunk, és kifejezetten a szövegértés képességének fejlettségét szeretnénk egyetlen számmal jellemezni, akkor az a fenti érvelés értelmében hiábavaló vállalkozás. Azt állítom azonban, hogy a „nagy mérések” esetén nem erről van szó. Ára van annak, hogy az ilyen vizsgálatok szakmailag korrektek legyenek, és ezt az árat mi megfizettük. Csak az a kérdés, hogy akik foglalkoznak az eredmények értékelésével, vajon tisztában vannak-e ezzel.

Az ár az volt, hogy valójában nem a szövegértés, a matematikai és a természettudományos eszköztudás fejlettségi szintjét mérjük (hiszen ezek nem léteznek), hanem valami olyan konstrukciót, amelyek értelmezése ugyan nem ismert, viszont nagyon jelentősek az oktatási rendszerek és alrendszerük eredményességének megítélése szempontjából (lásd még e kötetben *Szemerszki Marianna* írását). Megmagyarázom.

A szóban forgó mérések, vizsgálatok egyaránt modern tesztelméleti alapon zajlanak. Amennyiben sikerül biztosítani a tesztek egydimenziósságát, akkor a modern tesztelmélet modelljei garantálják, hogy intervallumskálán helyezkednek el

a mért adatok. Ez azt jelenti, hogy szabad mindazokat a statisztikai vizsgálatokat végezni, amelyeket szokásosan elvégeznek a kutatók. De hogyan lehetnek a tesztek egydimenziósak, amikor éppen arról van szó, hogy például a szövegértés „kétségbe-ejtően” sokdimenziós emberi tulajdonság? Ez az a pont, ahol lemondunk a szövegértés komplex képességének vizsgálatáról (ez az ár). Azt mondjuk, hogy a kutatók kezdetben kialakítottak egy tesztet, amely jó tulajdonságokkal rendelkezett, vagyis elsősorban egydimenziós volt, és olyan feladatok kaptak szerepet benne, amelyeknek a megoldása fontos a társadalom jövője szempontjából. Permanens módon úgy értékeljük (már akik úgy értékelik, magam igen) a tesztekben előkerülő feladatokat, hogy azok megoldásának színvonala sokat mond arról, hogy milyen eredményességgel működnek a jövőre való felkészülés kérdésében az iskolarendszerek és részrendszereik. A feladatok a szövegértés területéről valók, de a teszt maga nem méri a szövegértés komplex képességének fejlettségét. Van a szövegértéshez tartozó feladatoknak egy részhalmaza, amely feladatok egydimenziós sokaságot alkotnak. Fontos kérdés számunkra, hogy e feladatok megoldásában hogyan teljesítenek tanulóink. Ez a magyarázat! Teljes mértékben korrekt a mérés, ha szervezői nagyon vigyáznak arra, hogy mindig csak az eredeti egy dimenzióhoz tartozó feladatokat válasszanak be a tesztbe. Hogy mi ez a „mesterséges” dimenzió, azt nem tudjuk, és ez kissé aggodalomra ad okot. Ám a mérések adaptivitása, az eredmények fontossága oktatáspolitikai értelemben tagadhatatlan.

De mit jelent mindez a neveléstudományi kutatások számára? Azt, hogy a „nagy mérések” adatait nem lehet bizonyos képességek fejlettségének vizsgálatára alkalmazni. Rossz hír a szövegértés képességének kutatói számára, hogy a PISA vagy az OKM adatai semmit nem mondanak arról, hogyan alakul ez a komplex képesség a 15 évesek, vagy a 6., 8. és 10. osztályosok körében. A jó hír az, hogy az ugyan nem ismert, de az iskolai eredményesség szempontjából jelentősnek tartott teljesítménydimenziók mérése korrekt, megfelel azoknak a matematikai elvárásoknak, amelyek mai ismereteink szerint alapvető jelentőségűek a mérésekkel kapcsolatban. A reprezentációs méréselmélet, valamint a modern tesztelmélet „házasságából” igencsak sokat ígérő, sőt, ma már sokat jelentő mérési rendszer jött létre.

A „végeredmény” egyben arra is figyelmeztet, hogy szükség van az iskolarendszerek tanulás terén nyújtott teljesítményének más, sokféle mutató alapján történő vizsgálatára. Hiszen a „nagy mérések” csak egy-egy nem is egészen ismert, ám fontosnak tartott dimenzió tekintetében mutatják az eredményességet. Ahhoz, hogy komplexebb kép kialakulhasson, még több, újabb és újabb dimenziókat kell az értékelésbe bevonnunk.

Hivatkozások

- British Association for the Advancement of Science (1933): Interim Report of the Committee Appointed to Consider and Report upon the Possibility of Quantitative Estimates of Sensory Events. *Report of the Annual Meeting*, 277–334.
- Cervone, D. és Caldwell, T. L. (2008): From Measurement Theory to Psychological Theory, in Reverse. *Measurement*, 6, 84–88.
- Cliff, N. (1992): Abstract measurement theory and the revolution that never happened. *Psychological Science*, 3, 186–190.
- Csikos Csaba (1999): Nem-paraméteres statisztikai módszerek alkalmazási lehetőségei a pedagógiai kutatásban. *Iskolakultúra*, 9(2), 113–119.
- Díez, J.A. (1997): A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887-1990, Part I: The Formation Period. Two Lines of Research: Axiomatics and Real Morphisms, Csales and Invariance. *Studies in History and Philosophy of Science*, 28 (1), 167–185. Part II: Suppes and the Mature Theory. Representation and Uniqueness. *Studies in History and Philosophy of Science*, 28(2), 237–265.
- Helmholtz, H. V. (1887): Zahlen und Messen, erkenntnistheoretisch betrachtet. In: Helmholtz, *Schriften zur Erkenntnistheorie*. 70–108. (Angol fordítás: Numbering and Measuring from an Epistemological Viewpoint. In: Helmholtz, *Epistemological Writings*, 72–114.)
- Hölder, O. (1901): Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physische Klasse*, 53, 1–64.
- Kehl Dániel (2011): Skálák és statisztikák: a méréselméletéről és történetéről. *Statisztikai Szemle*, 1989(10-11), 1057–1080.
- Krantz, D. és Mészáros, I. (1971): *Foundations of Measurement*. Vol. 1. Academic Press, New York.
- Luce, R.D. (1992): A Path Taken: Aspects of Modern Measurement theory. In: Healy, A.F., Kosslyn, S. és Shiffrin, R. (Szerk.): *From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes, Vol. 1*. Erlbaum, Hillsdale. 45–64.
- Luce, R.D. (1996): The Ongoing Dialog between Empirical Science and Measurement Theory. *Journal of Mathematical Psychology*, 40, 78–98.
- Luce, R.D. és Suppes, P. (2002): Representational Measurement Theory. In: Pashler, H. és Wixted, J. (Szerk.): *Stevens' Handbook of Experimental Psychology, 3rd Edition, Vol 4*. Wiley, New York. 1–41.
- Maraun, M.D. (1999): Measurement as a Normative Practice: Implications of Wittgenstein's Philosophy for Measurement in Psychology. *Theory & Psychology*, 8(4), 435–461.
- Michell, J. (1990): *An Introduction to the Logic of Psychological Measurement*. Lawrence Erlbaum Associates, Hillsdale.
- Molnár Gyöngyvér (2013): *A Rash-modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat, Budapest.
- Nahalka István (2002): *Hogyan alakul ki a tudás a gyerekekben? Konstruktivizmus és pedagógia*. Nemzeti Tankönyvkiadó, Budapest.
- Narens, L. (1981): A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, 13, 1–70.
- Narens, L. (1985): *Abstract Measurement Theory*. MIT Press, Massachusetts.
- Narens, L. (2002): A Meaningful Justification for the Representational Theory of Measurement. *Journal of Mathematical Psychology*, 46, 746–768.

- Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests*. (Danish Institute for Educational Research, Copenhagen). The University of Chicago Press, Chicago.
- Roberts, F. S. (1979): *Measurement Theory with Applications to Decisionmaking, Utility and the Social Sciences*. Addison-Wesley, Massachusetts.
- Rossi, G. B. (2007): Measurability. *Measurement*, 40, 545–562.
- Scott, D. és Suppes, P. (1958): Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 23(2), 113–128.
- Shevlin, M. és mtsai. (2000): Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences* 28, 229–237.
- Sijtsma, K. (2009): On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107–120.
- Stevens, S. S. (1946): On the theory of scales of measurement. *Science*, 103, 667–680.
- Suppes, P. (1951): A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica*, 10, 163–172.
- Suppes, P. (1998): Theory of Measurement. In: Craig, E. (Ed.) *Routledge Encyclopedia of Philosophy*, (pp. 243–249), Routledge, London.
- Trendler, G. (2009): Measurement Theory, Psychology and the Revolution That Cannot Happen. *Theory and Psychology*, 19(5), 579–599.

Ostorics László

A tanulói teljesítménymérések jellemzői, jövőbeni irányvonalai, kritikái

A nagy nemzetközi és hazai tanulói teljesítménymérések napjainkra olyan jelenséggé váltak, amelyek nemcsak a közvetlen érintettek, azaz a közoktatásban tanuló, dolgozó, a szülők és az oktatáspolitikai döntéshozók életében, hanem a médiafogyasztók mindennapjaiban is elkerülhetetlenül megjelennek. Ezzel együtt a befogadók a méréseknek az eredményeivel, eljárásrendjével vagy éppen oktatáspolitikai hatásával kapcsolatban szeretnek állást is foglalni vagy szívesen értelmezik azokat. Tanulmányunkban olyan kérdéseket járunk körül, amelyeket szerintünk az eredmények megítélése során mindenképpen érdemes tekintetbe venni. Nem foglalkozunk azonban a pedagógiai mérés és értékelés bevett fogalmainak általános felsorolásával és definíciójával.

A következőkben áttekintjük azokat a vonásokat, amelyek a nagy tanulói teljesítményméréseket jellemzően meghatározzák. Bemutatjuk ebből a szempontból a három nagy nemzetközi példát, a PIRLS, a TIMSS, és a PISA vizsgálatokat, melyekben Magyarország indulásuk óta minden alkalommal részt vett, valamint ezek hazai rokonát, az Országos kompetenciamérést (OKM). Áttekintjük, hogy ezeknek a méréseknek az eredményei jeleznek-e kimutatható változásokat a magyar tanulók teljesítményében, majd felvázoljuk a mérések jövőben várható fejlődésének irányvonalait. Végül összefoglaljuk a nagy tanulói teljesítménymérések érvényességével kapcsolatban megfogalmazott kritikákat és az ezekre reagáló szakirodalom releváns elemeit.

Mi jellemez egy mérést?

A tanulmány tárgyául szolgáló méréseket jól megkülönböztetik egymástól a következő vonásaik: a vizsgálni kívánt célpopulációk; a mérés által vizsgált műveltség-területek, az, hogy mintán alapulnak vagy teljes körűek, illetve ebből következően az, hogy a közoktatási rendszerek mely szintjére tudnak visszajelzést biztosítani (lásd 1. táblázat).

1. táblázat

A nagy nemzetközi és hazai tanulói teljesítménymérések jellemző vonásai

	OKM	PISA	PIRLS	TIMSS
Megrendelő	A közoktatásért felelős miniszter	OECD	IEA	IEA
Célcsoport	6., 8., 10. évfolyamos tanulók	15 éves tanulók	4. évfolyamos tanulók	4. és 8. évfolyamos tanulók
Fő mérési területek	Szövegértés, matematikai eszköztudás	Szövegértés, alkalmazott matematikai műveltség, alkalmazott természettudományi műveltség	Szövegértés	Természettudomány, matematika
A tesztek szerkezete	Egy adott populáció tagjai azonos tartalmú tesztekkel találkoznak	Egy adott populáció tagjai különböző tartalmú tesztekkel találkoznak	Egy adott populáció tagjai különböző tartalmú tesztekkel találkoznak	Egy adott populáció tagjai különböző tartalmú tesztekkel találkoznak
Minta	A célcsoportok évfolyamain bizonyos kivételekkel teljes körű mérés*	150 iskola, iskolánként 35 tanuló, ~ 4 800 tanuló	150 iskola, iskolánként 1 osztály, ~ 5 400 tanuló	150 iskola, iskolánként 2 osztály, ~ 5 400 tanuló
Gyakoriság	évente	2000 óta 3 évente	2001 óta 5 évente	1995 óta 4 évente
Visszajelzés szintjei	<ul style="list-style-type: none"> • országos • régió • megye • fenntartó/tankerület • intézmény • telephely • osztály • tanuló 	<ul style="list-style-type: none"> • a résztvevő oktatási rendszerek közössége • országos • régiók • képzési programok 	<ul style="list-style-type: none"> • a résztvevő oktatási rendszerek közössége • országos • régiók 	<ul style="list-style-type: none"> • a résztvevő oktatási rendszerek közössége • országos • régiók

*Ez 2014-ben összesen 257 080 tanulót jelentett. Az adat forrása az Országos kompetenciamérés 2014. évi tisztított adatbázisa.

Triviális állításnak tűnik, de fontos leszögezni, hogy a nagy nemzetközi és hazai tanulói teljesítményméréseket meghatározó tényezők között a megrendelő alapvető fontosságú. A megrendelő, a felmérésért felelős szervezet jellege a mérni kívánt tartalomtól kezdve a célpopuláción keresztül a visszajelzések szintjéig egy mérés minden lényeges jellemzőjére kihat.

A fenti táblázatban szereplő megrendelők közül a legnagyobb múltra az IEA¹ tekinthet vissza. Az 1958-ban alapított tudományos társaság elsőként fogalmazta meg nemzetközileg intézményesült formában azt a kérdést, hogy az iskolai oktatás eredményessége hogyan mérhető, és hogyan vizsgálhatók a különféle tanulói teljesítmények mögött meghúzódó háttér folyamatok (lásd még e kötetben *Szemerszki Marianna* írását). Magyarország 1968-ban csatlakozott a szervezethez, hosszú ideig a vasfüggöny közötti országok közül egyedüliként. Az IEA által megrendelt mérések a közoktatás eredményességének egy olyan elsődleges felfogását képviselik, amely az iskolarendszereken *belüli* sikeresség vizsgálatából indul ki: a PIRLS és a TIMSS a szövegértési, a matematikai és a természettudományos ismeretek kapcsán a *tantervi tartalmak* elsajátíttóságának fokát kívánja mérni. Ebből az adottságból következik, hogy a mérni kívánt célpopulációkat ezek a vizsgálatok iskolaszakaszokat lezáró *évfolyamok* alapján határozzák meg, valamint az is, hogy a felmérésekben a kiválasztott iskolákból egész *osztályok* vesznek részt (*Balázsi és mtsai.*, 2012a).

Más felfogást képvisel a tanulói teljesítménymérések világát sok tekintetben radikálisan és rendszeresen megújító PISA megrendelője, az OECD. Míg az IEA kutatói tömörülés, addig az OECD elsősorban pénzügyi irányultságú társulás, a világ gazdaságilag legfejlettebb államainak elit klubja (az OECD és az oktatási rendszerek vizsgálatának kapcsolatáról lásd *Széll*, 2014). Ez a különbség jól tükröződik a PISA és a PIRLS&TIMSS közötti hangsúlyeltolódásokban is. Szemben az utóbbiakkal, a PISA fő kérdése nem a tanulók közoktatási rendszereken belüli sikerességére vonatkozik – azaz nem a tantervi tartalmak elsajátíttóságára összpontosít –, hanem arra, hogy az iskolában elsajátított tudást a tanulók hogyan tudják a gazdasági szférában, a munkaerőpiacon sikeresen kamatoztatni, vagyis a PISA *alkalmazott műveltséget* mér (lásd még e kötetben *Nahalka István* írását). Emiatt a vizsgált célpopulációt nem évfolyam, hanem *életkor* alapján határozza meg: a program kezdetekor, a XX. század utolsó éveiben az akkori tagországok legtöbbszörben a 15 éves volt az a korcsoport, amelyik a legközelebb állt a tanköteles kor végéhez, emellett pedig beiskolázási arányuk megközelítette a 100 százalékot. Mivel a mérés

¹ International Association for the Evaluation of Educational Assessment – Tanulói Teljesítmények Vizsgálatának Nemzetközi Társasága.

életkor alapú, a kiválasztott iskolákból nem osztályok vesznek részt, hanem 35, a korcsoport definíciójának megfelelő tanuló (lásd *Balácsi és mtsai.*, 2010a).

Az Országos kompetenciamérés évfolyam alapú, teljes körű, alkalmazott műveltséget vizsgál, azaz a magyar mérés mindkét fenti példából sokat merített, ugyanakkor fejlődése során a megrendelői oldalról jelentkező igények úgy alakították, hogy a felmért területek tekintetében tapasztalható hasonlóságoktól eltekintve különbözik mind az OECD, mind az IEA programjaitól.²

Az egyes vizsgálatok közötti legszembetűnőbb különbség abban mutatkozik meg, hogy a négy mérés a közoktatási rendszerek mely szintjén képes visszajelzést nyújtani. A nagy nemzetközi mérések arra töreksenek, hogy reprezentatív minta alapján a lehető legrészletesebb, statisztikailag megbízható információt szolgáltatassák a résztvevő országok oktatási rendszereinek teljesítményéről. Ezekben a vizsgálatokban gyakorlati okokból korlátozott számú tanuló vesz részt (mindhárom nemzetközi mérés esetében országonként mintegy 5 000 fő az alapértelmezés), továbbá a tesztfüzetek terjedelme is korlátozott, hiszen az egyes tanulókat nem szabad túlterhelni, emellett azonban arra van szükség, hogy minden egyes tesztfeladatra összességében viszonylag nagyszámú válasz szülessen. Ezt a három kényszert úgy kezeli a PISA és a PIRLS&TIMSS rendszere, hogy a tesztkérdéseket blokkokba rendezik, és az egymástól részben különböző tartalmú füzeteket a kérdésblokkok bizonyos sorrend szerinti ismétlése által hozzák létre. Így, noha minden részt vevő tanuló nem találkozik minden feladattal, egy adott kérdésre a blokkok ismétlődésének köszönhetően megfelelő számú válasz keletkezik. Ez költséghatékony és statisztikailag kezelhető eljárás, ugyanakkor azzal jár, hogy a PISA, a PIRLS és a TIMSS eredményei „csak” az oktatási rendszerek egészéről nyújtanak hiteles, nemzetközileg összehasonlítható információt, a tanulók, az osztályok és az iskolák szintjén viszont nem tudnak megbízható visszajelzést adni az oktatási rendszer teljesítményéről. Ezzel szemben az Országos kompetenciamérésnek pont ez a feladata, amelyre többek között teljes körű mivolta miatt alkalmas.

Tehát a Magyarországon rendszeresen lebonyolított nagy tanulói teljesítmény-mérések sok tekintetben rendkívüli sokszínűséget mutatnak: megtalálható közöttük évfolyam és életkor alapú, tantervi tartalmakat és alkalmazott műveltséget mérő, teljes körű és reprezentatív mintán alapuló, illetve az oktatási rendszerek és a tanulók szintjéig visszajelzést adó mérés is. Felmerül a kérdés, hogy ezen vizsgálatok eredményei alapján mennyire egyenemű a magyar diákok teljesítményéről rajzolható kép. A következőkben erre a kérdésre keressük a választ.

² Az Országos kompetenciamérés változásaiával kapcsolatban lásd *Berényi*, 2010; valamint *Balkányi és Ostorics*, 2011.

A magyar tanulók teljesítményének változásai

A magyar diákok eredményével kapcsolatban két rendkívül érdekes kérdés vetődő fel:

1. „Jeleznek-e változást a magyar gyermekek teljesítményében az idők során keletkező adatok?”, illetve:
2. „Milyen mértékben vethetőek össze egymással az egyes felmérések eredményei?”

A mérések eredményeiről részletes jelentések érhetőek el (*Balázsi és mtsai.*, 2010b, 2012b, 2012c, 2013), ezért a következőkben nem mutatjuk be részletesen a rendelkezésünkre álló adatokat. Inkább arra törekszünk, hogy általános képet adjunk a magyar tanulók eredményeiben megragadható változásokról, példákat adjunk arról, hányféle módon tudja leírni a gyermekek képességeit egy nemzetközi mérés, emellett pedig rámutassunk azokra a tényezőkre, amelyek kérdésessé teszik a mérési eredmények közvetlen összehasonlíthatóságát.

A PIRLS és a TIMSS mérések eredményei³

Először a 4. és 8. évfolyamos magyar tanulók természettudományos, matematikai és szövegértés-teljesítményének változásait vesszük szemügyre a TIMSS és a PIRLS 1995, illetve 2001 óta lezajlott adatfelvételei alapján. A PIRLS és a TIMSS eredményeinek értelmezésekor figyelembe kell venni a mérések skálájának keletkezési körülményeit és tulajdonságait. A két mérésben a tanulókat tesztelméleti módszerekkel számított 0-tól 1000-ig terjedő képességskálán helyezték el, amelyet a PIRLS2001, illetve a TIMSS1995 vizsgálatban részt vett tanulók eredménye alapján alakítottak ki úgy, hogy a nemzetközi átlag 500, a szórás 100 pont volt. Ennek alapján elmondható, hogy a magyar gyerekek eredménye minden esetben *meghaladta* ennek a skálának az átlagát. A részt vevő országok köre azonban az egyes ciklusok során nagymértékben módosult, mind átlag alatti, mind átlag feletti teljesítményű országok csatlakoztak ezekhez a mérésekhez. Közben az IEA úgy döntött, hogy az egyes adatfelvételek alkalmával nem számolják újra a résztvevők átlagát, hanem az eredményeket konvencióképpen az 500 pontos, úgynevezett PIRLS-skálaátlaghoz,

³ Alapvető források: *Mullis és mtsai.*, 2007, 2012a, 2012b.

illetve TIMSS-skálaátlaghoz viszonyítják, amely különbözik az országok adott évi eredményei alapján számítható nemzetközi átlagtól.

A rendelkezésre álló elsődleges adatok alapján azt meg tudjuk mondani, hogy a magyar tanulók eredménye a skálaátlaghoz képest elmozdult-e, azt azonban csak további (bár egyszerű) elemzések segítségével tudnánk leírni, hogy ez az eredmény az egyes felmérések alkalmával a résztvevő oktatási rendszerek átlagához képest hol helyezkedett el. E jelenség miatt az IEA ezekkel a mérésekkel kapcsolatban *nem közöl* helyezési tartományokat és viszonylagos sorrendeket a részt vevő oktatási rendszerek eredményei alapján. Mindezt figyelembe véve a magyar negyedikesek teljesítményében matematikából a 2011-es adatokban a 2007-es és 1995-ös eredményekhez képest nem volt statisztikailag értékelhető változás, a 2003-as mérésben részt vevőkhöz képest azonban gyengébb teljesítményt nyújtottak. A természettudomány tekintetében Magyarország átlageredménye a TIMSS2003 óta nem változott szignifikáns mértékben, a legelső, 1995-ös vizsgálathoz képest viszont nagy arányban, 27 képességponttal javult (lásd 2. táblázatot).

2. táblázat

A 4. évfolyamos magyar tanulók TIMSS-méréseken elért eredményeinek változása matematikából és természettudományból^{*}

Év	A magyar diákok átlageredménye	Különbség 2007	Különbség 2003	Különbség 1995
Matematika				
2011	515	6	-13	-6
2007	510	-	-19	-12
2003	529	-	-	7
1995	521	-	-	-
Természettudomány				
2011	534	-2	5	27
2007	536	-	6	28
2003	530	-	-	22
1995	508	-	-	-

Forrás: Balázs és mtsai., 2012b: p. 61.

^{*}A szignifikáns különbségeket a félkövér szedés jelzi. A táblázat azért nem tartalmaz az 1999-es évre vonatkozó eredményeket, mert abban az évben csak 8. évfolyamosok vettek részt a TIMSS felmérésben.

Ugyanezen két műveltségi terület 8. évfolyamon mért eredményei némileg más képet mutatnak. A TIMSS összes adatfelvételét tekintetbe véve az állapítható meg, hogy a magyar tanulók matematikából az első három, 1995-ös, 1999-es és 2003-as mérésben viszonylag stabil teljesítményt nyújtottak. A 2007-es mérésben azonban 12 pontos, majd 2011-ben újabb hasonló nagyságrendű csökkenést tapasztalhatunk. Visszatekintve a negyedik évfolyamos tanulók 2003 és 2007 közötti eredménycsökkenésére, feltételezhető, hogy ez előrevetítette a nyolcadikosok körében tapasztalt 2007 és 2011 közötti eredmények csökkenését. Reményt keltő lehet ebből a szempontból, hogy a 2007 és 2011 közötti negyedikes korosztály eredménye nem romlott, igaz, nem is javult jelentős mértékben. A TIMSS 2015-ös adatfelvétele talán nyújt valami fogódzót ebben a kérdésben. Természettudomány szempontjából azonban jóval baljósabb a helyzet: a magyar gyerekek legutóbbi eredménye minden addigi teljesítményüknél szignifikánsan alacsonyabb. Így a TIMSS eredményei alapján egy populáció esetében mutatható ki trend: az általános iskolát befejező évfolyam természettudomány-eredménye, noha még mindig meghaladja az 1995-ben beállított skálaátlagot, egy évtizede folyamatosan romlik (lásd 3. táblázat).

3. táblázat

A 8. évfolyamos magyar tanulók TIMSS-méréseken elért eredményeinek változása matematikából és természettudományból*

Év	A magyar diákok átlageredménye	Különbség 2007	Különbség 2003	Különbség 1999	Különbség 1995
Matematika					
2011	505	-12	-24	-27	-22
2007	517	-	-12	-15	-10
2003	529	-	-	-2	3
1999	532	-	-	-	5
1995	527	-	-	-	-
Természettudomány					
2011	522	-17	-20	-30	-14
2007	539	-	-4	-13	2
2003	543	-	-	-10	6
1999	552	-	-	-	16
1995	537	-	-	-	-

Forrás: Balácsi és mtsai., 2012c: p. 42., p. 61.

*A szignifikáns különbségeket a félkövér szedés jelzi.

A 4. évfolyamosok szövegértés-teljesítményével kapcsolatos PIRLS-adatok alapján tekintsük át, hogyan írhatóak le részletesebben egy mérés segítségével a diákok eredményeiben tetten érhető változások. A három PIRLS-mérés eredményeiről azt mondhatjuk, hogy a magyar tanulók minden esetben a skálaátlag felett teljesítettek, azonban a 2006-os teljesítmény szignifikánsan meghaladja mind a 2001-es, mind pedig a 2011-es eredményüket. Itt trendről nem beszélhetünk, az eredmény előbb mérhetően javult, majd romlott (lásd 4. táblázat).

4. táblázat

A 4. évfolyamos magyar tanulók PIRLS-méréseken elért eredményeinek változása szövegértésből¹

Év	A magyar diákok átlageredménye	Különbség 2006	Különbség 2001
2011	539	-12	-4
2006	551	-	8
2001	543	-	-

Forrás: Balázsi és mtsai., 2012b: p. 51.

¹A szignifikáns különbségeket a félkövér szedés jelzi.

Vajon a tanulók képességeloszlásának változásaiban tükröződött-e ez? A PIRLS a tanulók tudásának jellemzésére az 500-as átlagú képességskálán négy osztópontot jelölt ki a kiváló, a magas, az átlagos és az alacsony szint alsó határait 625, 550, 475, 400 pontban állapítva meg. Érdeemes megvizsgálni, hogy az egyes adatfelvételek során a tanulók milyen arányban oszlottak meg e négy képességszint között – figyelembe véve, hogy aki az egy adott szinthez rendelt műveleteket végre tudja hajtani, az az alsóbb szintekhez tartozó feladatokat is meg tudja oldani. A képességeloszlások szempontjából korántsem olyan egyértelmű a helyzet, mintha csak az átlag változásait vennénk figyelembe. A kiváló képességszinten a 2011-es arányok szignifikánsan magasabbak, mint 2001-ben voltak, és nem alacsonyabbak, mint 2006-ban, vagyis az átlageredmény változásai nem a kiválóan teljesítő tanulók hányadának módosulásaival jártak együtt. Ugyanakkor a magas képességszinten már mérhető a csökkenés az 2006-os adathoz képest, a 2001-es méréshez viszonyítva ugyanannyi tanuló érte el ezt a képességszintet – ez követi az átlag ingadozását. Az átlagos és az alacsony szint esetében azonban 2011-ben 2006-hoz és 2001-hez képest egyaránt szignifikánsan alacsonyabb a tanulók aránya. A képességeloszlások arányainak változásai alapján tehát azt látjuk, hogy míg az átlageredmény a három adatfelvétel során mintegy kilendült, majd visszaállt az egyébként is átlag feletti

kiindulási állapotba, addig folyamatosan növekszik azon tanulók aránya, akik a legalacsonyabb szintet sem érik el (lásd 5. táblázat).

5. táblázat

A 4. évfolyamos magyar tanulók képességeloszlásának változásai szövegértésből (PIRLS-mérés)

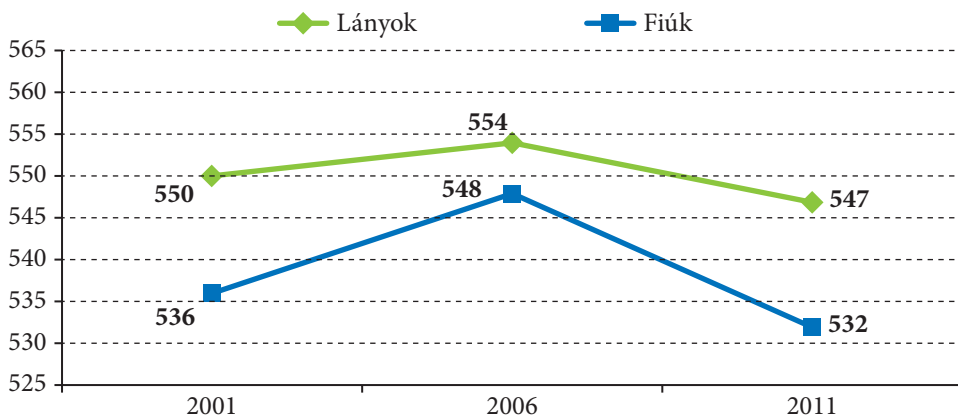
Képességszint	2011	2006	2001
Kiváló szintet elérők	12%	14%	10%
Magas szintet elérők	48%	53%	49%
Átlagos szintet elérők	81%	86%	85%
Alacsony szintet elérők	95%	97%	98%

Forrás: Balácsi és mtsai., 2012b.

Közelebb kerülhetünk a probléma pontosabb leírásához, ha tekintetbe vesszük az eredmények nemek szerinti megoszlásának változásait. Látható, hogy a lányok teljesítménye gyakorlatilag nem változott a PIRLS három ciklusa során: 2001-es (550 pont) és 2006-os (554 pont) eredményük között nincs szignifikáns különbség, 2011-es átlaguk (547 pont) gyengébb, mint a második ciklusban mutatott, azonban statisztikailag nem alacsonyabb, mint a 2001-es. Ezzel szemben a fiúk egyedül 2006-ban (548 pont) közelítették meg a lányok teljesítményét. Elmondható, hogy az átlag fentebb említett javulása és romlása szinte kizárólag a fiúk eredményeiben tapasztalható változásokkal járt együtt (lásd 1. ábra).

1. ábra

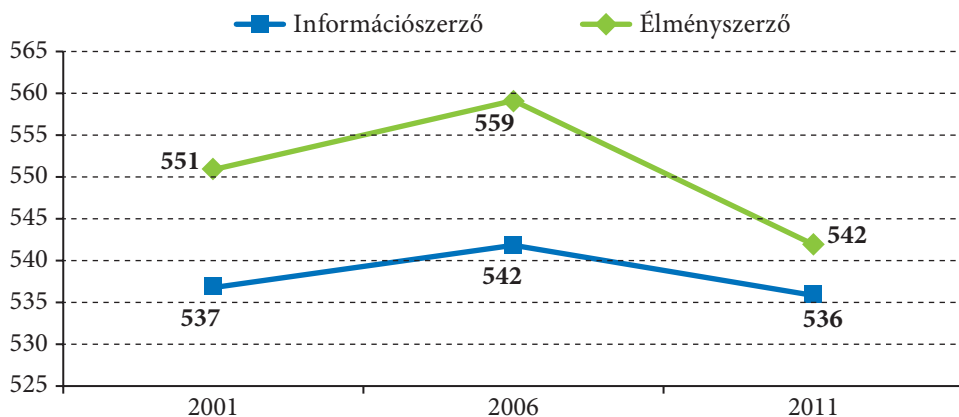
Az eredmények nemek szerinti megoszlásának változásai a PIRLS2001, 2006 és 2011 során



Forrás: Balácsi és mtsai., 2012b: p. 47.

A nemek szerinti megoszláshoz hasonló képet mutat, hogy miként módosultak a magyar 4. évfolyamosok egyes szövegtípusokhoz rendelhető eredményei.⁴ Az élményszerző szövegek esetében a 2001-es pontszámhoz képest 10, a 2006-os eredményekhez képest pedig 17 pontot romlott 2011-re az eredményünk, ami szignifikáns változás. Ugyanakkor az információszerző szövegekben a magyar diákok kiegyensúlyozottabban teljesítettek, szignifikáns változás a három ciklus alatt nem volt. Látható tehát, hogy a fiúk eredményének romlása mellett főleg az élményszerző szövegeken elért pontszámok csökkentek (lásd 2. ábra).

2. ábra
Az eredmények szövegtípusok szerinti megoszlásának változásai
a PIRLS2001, 2006 és 2011 során



Forrás: Balázi és mtsai., 2012b: p. 48.

Az IEA tantervi alapú mérési eredményeinek áttekintése során azt láttuk, hogy a 4. és 8. évfolyamos magyar tanulók eredményei változásokon mennek keresztül: a nyolcadikosok természettudományi teljesítménye gyengülő tendenciát mutat, míg a matematika és a szövegértés eredményekben zajlottak ugyan szignifikáns változások, azonban ezek esetében trendről nem beszélhetünk. Láttuk, hogy az IEA-mérések elsődleges eredményei alkalmasak arra, hogy egy oktatási rendszer eredményességének változásait megbízhatóan leírják, illetve rávilágítsanak a változások mögött a tanulók teljesítményének képességszintek, nemek vagy egyéb változók (például szövegtípusok) szerinti eloszlásában bekövetkező módosulásokra, ezzel árnyalva a képet. A következő bekezdésekben azt vizsgáljuk meg, hogy tapasztalható-e hasonló tendencia az alkalmazott műveltséget mérő programok eredményeiben.

⁴ A PIRLS tartalmi kerete élményszerző és információközlő szövegtípusokat különböztet meg. Bővebben lásd Balázi és mtsai., 2012b.

A PISA-felmérés eredményei⁵

A PISA eddigi öt ciklust élt meg, amely mérés során három nagy mérési területének skáláit a PIRLS és a TIMSS esetében látott jellemzőkhöz hasonlóan alakították ki, ezeknek a skáláknak az átlaga is 500, szórásuk pedig 100. Minden egyes skálát akkor rögzítettek, amikor az adott műveltségi terület először került a mérés fókuszába: ez a szövegértés esetében 2000, az alkalmazott matematikai műveltségnél 2003, az alkalmazott természettudományi műveltségnél pedig a 2006-os év volt – ezért nem szerepelnek a korábbi adatfelvételek eredményei sem a táblázatban, sem pedig az ábrán. A résztvevők körének átlaga azonban, szemben az IEA-mérésekben tapasztaltakkal, állandó: a PISA eredményeit mindig az OECD-országok átlagához viszonyítjuk, ami a mérés eddig lezajlott öt ciklusa során a minden adatfelvételben részt vett OECD-tagokra számítva szignifikánsan nem változott.⁶

A PISA 12 éve alatt azonban történtek szignifikáns változások a magyar tanuló eredményeiben. A táblázat tanúsága szerint a 2009-es volt az egyetlen ciklus, amelyben a magyar 15 évesek minden fő területen elérték az OECD-országok átlagát. Egy olyan adatfeltétel volt ezen kívül, amelynek során egy területen átlagos teljesítményt mutattak fel tanulóink, ez a 2006. évi 504 pontos természettudomány átlagunk. A legutóbbi, 2012-es mérésben számszerűleg a magyar fiatalok minden eredménye romlott. Emellett azonban elhamarkodott kijelentés lenne azt állítani, hogy ez bizonyosan romló tendenciára is utal (lásd 6. táblázat).

6. táblázat

A 15 éves magyar tanulók eredményeinek változása a PISA-felmérésben*

Képességterület	2000	2003	2006	2009	2012
Szövegértés	480	482	482	494	488
Matematika	–	490	491	490	477
Természettudomány	–	–	504	503	494

Forrás: Balázsi és mtsai., 2013.

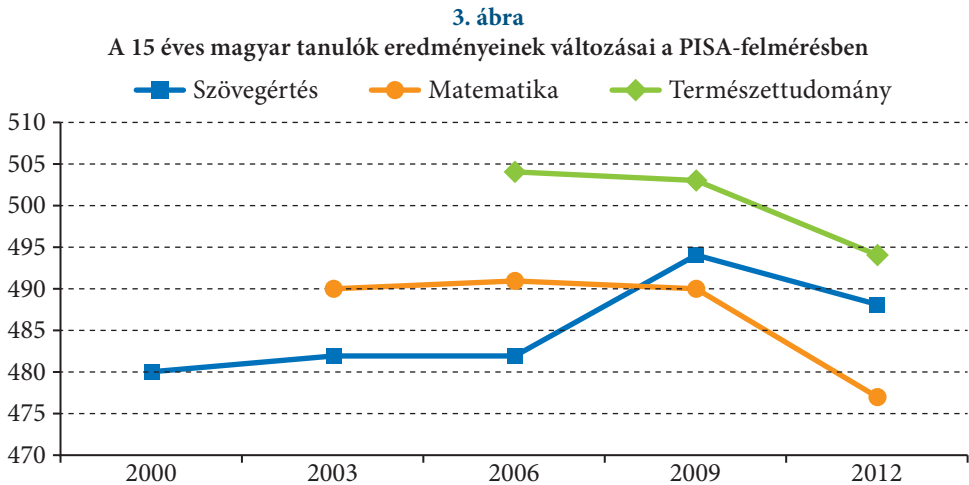
*Az OECD-átlag feletti szignifikáns eltéréseket félkövér szedés jelzi.

Jó példát jelentenek az eredmények változásának többféle módon való értelmezhetőségére a magyar 15 évesek szövegértés-pontszámának ingadozásai. Ha az

⁵ Alapvető források: OECD, 2013a, 2013b, 2013c, 2013d, 2013e.

⁶ A eredmények részletes bemutatásával és elemzésével kapcsolatban lásd Balázsi és mtsai., 2010b, 2013.

évenkénti eredményeket egymással összehasonlítjuk, azt fogalmazhatjuk meg, hogy eredményünk 2000 és 2006 között gyakorlatilag nem változott, megmaradt az OECD-átlag alatt, majd 2009-re szignifikánsan javult, és elérte az OECD-átlagot. 2009 és 2012 között az eredmény számszerűleg csökkent, ez a változás azonban nem volt szignifikáns, azaz a két eredmény között nincs statisztikai különbség, emellett azonban a romlás mértéke elég volt ahhoz, hogy eredményünk már átlag alattinak számítsion. Ezzel szemben, ha az egyes évek eredményeit a 2012-es eredményekkel külön-külön vetjük össze, akkor teljesen más képet kapunk, ugyanis ez a pontszámunk szignifikánsan egyetlen megelőző szövegértés-teljesítményünktől sem különbözik, azaz az eredmények ilyen szempontú leírása inkább azt az értelmezést erősítené, hogy a magyar 15 évesek szövegértésének eredményessége stagnál (lásd 3. ábra).



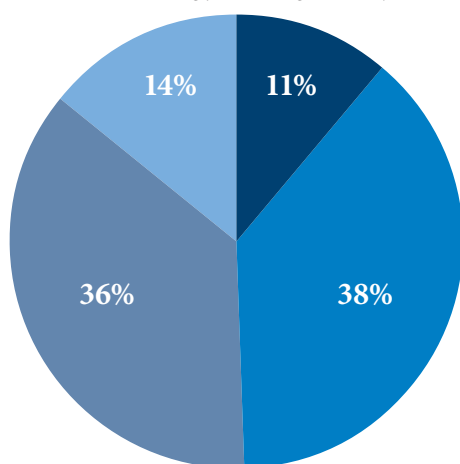
Forrás: Balázsi és mtsai., 2013.

Korábban láttuk, hogy az IEA mérései évfolyam alapúak, és csak az általános iskolát érintő programokra irányulnak, ezért nem releváns feltenni velük kapcsolatban az iskolatípusok eredményeinek különbségeivel kapcsolatos kérdést. Ezzel szemben a PISA a 15 éves korosztályt vizsgálja, és az országos mintában négy képzési forma (általános iskola, gimnázium, szakközépiskola, szakiskola) és négy évfolyam (7., 8., 9., 10.) között oszlanak meg a tanulók. Így értelmesen megfogalmazható az a kérdés, hogy a 2009 és 2012 között tapasztalható általános gyengülés vajon mely képzési típusok és évfolyamok tanulóinál jelentkezett elsősorban. A mérésben részt vevő diákok legnagyobb hányada szakközépiskolában és gimnáziumban tanul, és a célcsoport mintegy harmada oszlik meg az általános és a szakiskola kö-

zött. Emellett a PISA-populáció kétharmada 9. évfolyamra jár, ötöde 10., valamivel több, mint a tizede pedig a 7. és a 8. évfolyamon részesül képzésben. A PISA országosan reprezentatív mintájában az egyes képzési típusokhoz és évfolyamokhoz rendelhető tanulók aránya a legutóbbi két adatfelvétel között szignifikánsan nem változott (Balázsi és Ostorics, 2014) – azaz a változást nem tulajdoníthatjuk egyes gyengébben teljesítő csoportok aránynövekedésének (lásd 4. és 5. ábra).

4. ábra

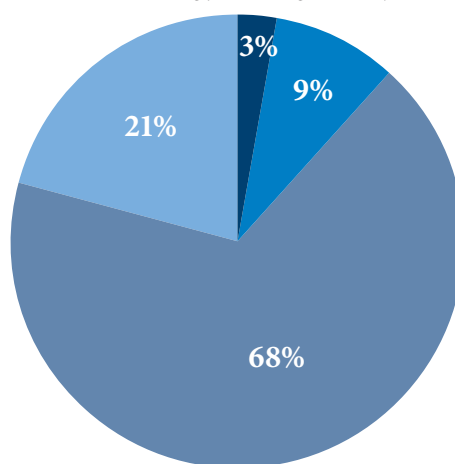
Az egyes képzési típusokba járó tanulók százalékos aránya a PISA2012 magyarországi mintájában



■ általános iskola ■ gimnázium
 ■ szakközépiskola ■ szakiskola

5. ábra

Az egyes évfolyamokra járó tanulók százalékos aránya a PISA2012 magyarországi mintájában



■ 7. évfolyam ■ 8. évfolyam
 ■ 9. évfolyam ■ 10. évfolyam

Forrás: Balázsi Ildikó számításai a PISA 2012 adatbázisa alapján.

Forrás: Balázsi és Ostorics, 2014.

Most nem szentelünk figyelmet annak a jelenségnek, hogy mennyire elválík egymástól a négy képzési típus teljesítménye – annyit jegyzünk meg, hogy ezt a teljesítménykülönbséget a hazai mérési rendszer is regisztrálja, ahogy azt is, hogy a legkisebb különbség az általános iskolások és a szakiskolások teljesítménye között jelentkezik (Balázsi és mtsai., 2014: p. 18). Ezúttal arra fordítjuk a figyelmet, hogy a két adatfelvétel között mely oktatási típusok eredményei változtak. Az vehető észre, hogy az általános iskolába járó 15 évesek teljesítménye ugyan nem szignifikánsan, de *minden területen javult* 2009 és 2012 között. A gimnáziumi tanulók teljesítménye némileg ingadozik, de egyik irányú elmozdulás sem szignifikáns, azaz az ő

eredményük tartósan és magasán az OECD átlag feletti. Ezzel szemben szignifikáns (13–20 pontos) *romlás* következett be a hazai PISA-minta felét képviselő szakközépiskolások és a szakiskolások matematika- és természettudomány-eredményeiben (lásd 7. táblázat).

7. táblázat

A PISA-felmérésben részt vevő magyar tanulók teljesítményének változásai képzési típus szerint*

Képzési típus	2009			2012		
	Szövegértés	Matematika	Természettudomány	Szövegértés	Matematika	Természettudomány
Általános iskola	378	378	390	388	386	413
Gimnázium	559	552	559	554	542	554
Szakközépiskola	491	487	504	487	470	490
Szakiskola	410	410	429	395	391	410

Forrás: Balázi Ildikó számításai a PISA2009 és 2012 adatbázisa alapján.

*A félkövérrel szedett értékek a mérési időszakok közötti szignifikáns különbséget jeleznek.

8. táblázat

A PISA-felmérésben részt vevő magyar tanulók teljesítményének változásai évfolyam szerint

Évfolyam	2009			2012		
	Szövegértés	Matematika	Természettudomány	Szövegértés	Matematika	Természettudomány
7. évfolyam	343	351	362	349	349	375
8. évfolyam	391	388	400	404	402	429
9. évfolyam	499	496	508	493	480	497
10. évfolyam	538	524	534	531	517	528

Forrás: Balázi Ildikó számításai a PISA2009 és 2012 adatbázisa alapján.

*A félkövérrel szedett értékek a mérési időszakok közötti szignifikáns különbséget jeleznek.

Szintén a természettudomány és a matematika területein figyelhetők meg szignifikáns változások akkor is, ha évfolyamok szerinti bontásban vizsgáljuk a tanulók teljesítményeit. Összhangban azzal, amit az általános iskolás 15 évesek eredményeinek viselkedésével kapcsolatban fentebb láttunk, azt tapasztaljuk, hogy a 8.

évfolyamon a tanulók természettudományi eredménye emelkedett. Ne felejtjük el, hogy ez mindössze a minta 8,7 százalékát jelenti, ami jelen esetben alig több, mint 400 tanuló; ez a változás, noha szignifikáns, nem érezhető hatását a teljes minta eredményén. A résztvevők derékhadát, mintegy 3 200 főt kitevő 9. évfolyamosok eredménye ugyanis mind matematikából, mind természettudományból szignifikánsan romlott (lásd 8. táblázat).

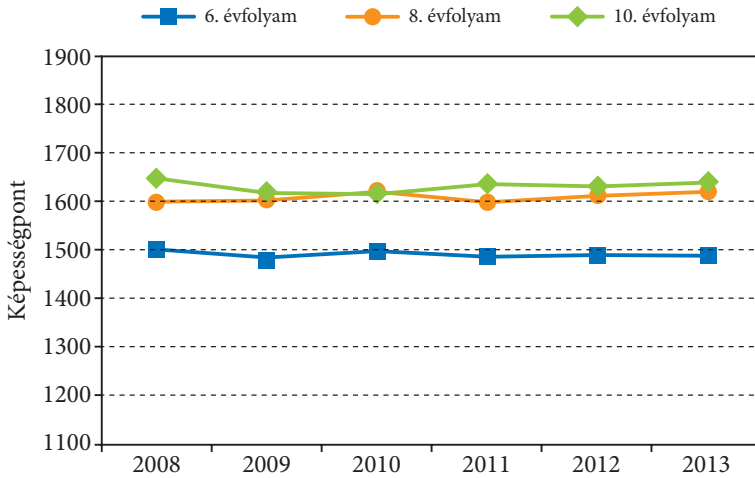
Láthatjuk, hogy a PISA 2009 és 2012 közötti eredményváltozásainak egyszerű tovább bontása (az évfolyamok és képzési típusok szerinti vizsgálat) segít abban, hogy releváns kérdéseket tegyünk fel a méréssel kapcsolatban. Azt mondhatjuk, hogy a „*Romlott-e a magyar 15 évesek teljesítménye?*” kérdésre az elsődleges eredmények alapján is csak több szempontú, árnyalt válasz adható, a „*Miért romlott a tanulók eredménye matematikából és természettudományból?*” kérdés pedig túl általános. A képzési típusokban három év alatt bekövetkező változásokra vonatkozó kérdések jobban megragadható kutatói felvetések felé vezetnek bennünket, és a közoktatási rendszerben jelentkező valódi gondokra deríthetnek fényt. A nemzetközi mérési eredményeink alapján rendkívül összetett kép rajzolódik ki a különböző tanulói nemzedékek eltérő területeken mutatott eredményességéről. Joggal merül fel a kérdés, hogy a magyar országos mérés eredményei mit mutatnak ezzel kapcsolatban.

Az OKM eredményei

A 2008. és a 2013. évi mérés között a 6. 8. és 10. évfolyamra járó tanulók szövegértésből és matematikából mutatott teljesítményeit vizsgálva az mondható el, hogy a különbségek értéke akár pozitív, akár negatív irányban a skála értékeihez képest oly kicsi, hogy azokkal sem a fejlődésre, sem a teljesítményromlásra vonatkozó következtetéseket nem lehet érvényesen alátámasztani. Egyedül a 6. évfolyamos szövegértés esetében látható, hogy 2008-hoz képest 2011-re 35 ponttal csökkent az átlageredmény, azonban mivel az ezt követő két évben az átlageredmény kismértékben nőtt, ez a különbség 2013-ra 3 pontra mérséklődött, azaz statisztikailag megszűnt. Összességében tehát a magyar tanulók egészének az eredményei az Országos kompetenciamérés adatai alapján egyáltalán nem változnak (lásd 6. és 7. ábra).

6. ábra

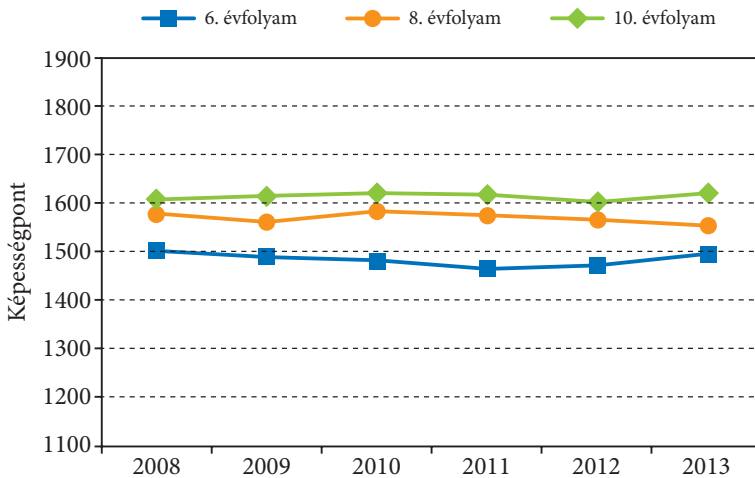
A 2013-as és a korábbi országos kompetenciamérések eredményei matematikából évfolyamok szerint



Forrás: Balázsi és mtsai., 2014: p. 9.

7. ábra

A 2013-as és a korábbi országos kompetenciamérések eredményei szövegértésből évfolyamok szerint



Forrás: Balázsi és mtsai., 2014: p. 9.

A tanulmány elején a nagy nemzetközi és hazai tanulói teljesítménymérések jellemző vonásait összefoglaló 1. táblázatra visszautalva szeretnénk felhívni a figyelmet arra, hogy a nemzetközi és a magyar mérések eredményei nem „monda-

nak ellent” egymásnak. Ezek az eredmények éppen a mérések alapvető vonásaiban lévő különbségek miatt közvetlenül nem hasonlíthatóak össze egymással. Nem mondhatjuk, hogy a PISA és az OKM adatai „cáfolják” egymást, még akkor sem, ha fentebb tudtunk érvényes kijelentéseket tenni a 10. évfolyamos tanulók PISA-eredményeinek változásáról, mert az egyik kizárólag a 15 évesek, míg a másik a 10. évfolyam egészének teljesítményéről szolgáltat információt. Ugyanígy, elgondolkodtató, de nem ellentmondás, hogy a TIMSS-ben a 8. évfolyamos matematikaeredmények 2007 és 2011 közötti változása miért nem tükröződik az Országos kompetenciamérés 2008 és 2011 közötti adataiban. A két mérés ugyanis nem ugyanazokat a képességeket méri, az egyik kizárólag tantervi alapú, a másik pedig matematikai eszköztudást vizsgáló felmérés.

A mérési eredmények sokféleségének jelenségét valószínűleg akkor értelmezzük helyesen, ha kijelentjük, hogy a nemzetközi programok számos változást jeleznek *egyed magyar tanulói populációk* bizonyos teljesítményeinek szintjében, ugyanakkor az Országos kompetenciamérésnek a 6., 8. és 10. évfolyamra vonatkozó idősoros eredményei arra utalnak, hogy 2008–2013 között nem történtek olyan oktatáspolitikai változások, amelyek *egész tanulói nemzedékek* teljesítményére lettek volna hatással.

A nagy teljesítménymérésekkel kapcsolatban áttekintettük, milyen változásokat mutattak a tanulók eredményei. A következőkben magukkal a mérésekkel foglalkozunk.

Változások a mérésekben

A nagy tanulói teljesítményméréseknek egyik fő céljuk az adatok időben való összehasonlíthatóságának biztosítása, így minden mérésnek vannak állandó területei. Emellett azonban mind az IEA, mind az OECD, mind pedig az Országos kompetenciamérés fejlesztését és lebonyolítását irányító Oktatási Hivatal törekszik arra, hogy új területekre és médiumokra fejlesszen mérőeszközöket.

Új mérési területek bevonásában a leginnovatívabb a fentiek közül az OECD; a PISA-vizsgálatban némiképp változtak az állandó, kötelező és választható mérési területek 2000 és 2015 között. Látható, hogy noha a három nagy mérési terület, a szövegértés, az alkalmazott matematikai műveltség és az alkalmazott természettudományi műveltség állandó eleme a mérésnek, az adatfelvétel médiuma 2015-re már kizárólag digitális. Emellett 2015-ben új mérési terület a digitális kollaboratív problémamegoldás, a 2018-as ciklussal kapcsolatban pedig az úgynevezett „globális kompetencia” vizsgálata merült fel (lásd 9. táblázat).

9. táblázat
A PISA mérési területeinek változásai

Mérési terület	2000	2003	2006	2009	2012	2015
Szövegértés (nyomtatott)	Fő terület	Kötelező	Kötelező	Fő terület	Kötelező	Megszűnik
Matematika (nyomtatott)	Kötelező	Fő terület	Kötelező	Kötelező	Fő terület	Megszűnik
Természettudomány (nyomtatott)	Kötelező	Kötelező	Fő terület	Kötelező	Kötelező	Megszűnik
Problémamegoldás (nyomtatott)	-	Kötelező	-	-	Megszűnik	-
Digitális természettudomány	-	-	Választható (hazánk nem)	-	-	Fő terület
Digitális szövegértés	-	-	-	Választható (hazánk igen)	Választható (hazánk igen)	Kötelező
Digitális problémamegoldás	-	-	-	-	Kötelező	-
Digitális matematika	-	-	-	-	Választható (hazánk igen)	Kötelező elem
Pénzügyi műveltség	-	-	-	-	Választható (hazánk nem)	Választható (hazánk nem)
Olvadási részképességek	-	-	-	-	Választható (hazánk nem)	Választható (hazánk nem)
Digitális kollaboratív problémamegoldás	-	-	-	-	-	Kötelező elem

Forrás: Ostorics, 2014.

A digitális médium bevonásának tendenciája jelentkezik a tárgyalt IEA-méréseknél is: az elektronikus információszerző szövegértést már a következő adatfelvételben, 2016-ban vizsgálni fogja az ePIRLS (lásd ePIRLS Brochure⁷), és a TIMSS is rendelkezni fog digitális komponenssel. Az Országos kompetenciaméréssel kapcsolatban ez a lehetőség gyakorlati okokból egyelőre nagyon távoli, ugyanakkor hamarosan napvilágot lát az a hatástanulmány, amely a digitális médiumra való áttérés lehetőségeit vizsgálja meg az Országos kompetenciaméréssel kapcsolatban.

⁷ Elérhető: http://timssandpirls.bc.edu/pirls2016/downloads/P2016_ePIRLS_Brochure.pdf
[Letöltés dátuma: 2015.04.30.]

Emellett a TÁMOP 3.1.8 program keretei között lezajlott az Országos kompetenciamérés természettudományi próbamérésének adatfelvétele – a mérési anyagok feldolgozása jelenleg is folyamatban van.

Az innovációra való folyamatos törekvés – különösen a PISA esetében – gyakran párosul nagy médianyilvánossággal, a figyelem pedig törvényszerűen kritikát is generál. A tanulmány végül röviden kitér a 2014. év mérésekkel kapcsolatos két legérdekesebb médiaeseményére, és az eddig leírtak fényében értelmezi ezeket.

Médianyilvánosság és méréskritika

A 2014. évben Magyarországon két jelentősebb hír foglalkozott mérésekkel, annak ellenére, hogy ebben az évben hozták nyilvánosságra a PISA2012 digitális problémamegoldás eredményeit. Az egyik az úgynevezett „PISA-levél” volt, a másik pedig a *GEMS Solutions* PISA-adatokra alapuló elemzése, amely úgymond, a világ leghatékonyabb oktatási rendszerei közé helyezte Magyarország oktatási rendszerét.

A PISA feltételezett káros hatásait előszámláló nyílt levél 2014. május 6-án jelent meg a *Guardian*ben⁸ – *Heinz-Dieter Meyer* német kutató címezte *Andreas Schleichernek* –, és száznál több, elsősorban egyetemi kutató és szülői szervezetekbe tömörülő aktivista írta alá. A levél részletes elemzése nem tárgya ennek a tanulmánynak, két főbb állítást vizsgálunk csak meg (részletesebben lásd *Meyer, 2014; Ostorics, 2014* és *Schleicher, 2014*):

1. Az OECD markánsan gazdasági irányultságának köszönhetően a PISA az oktatást a gazdasági előrejutás eszközévé redukálja, és elvonja a figyelmet olyan nem mérhető, de fontosabb célokról, mint a személyes fejlődésre, a demokratikus társadalomban való részvételre és az erkölcsös döntéshozatalra való felkészítés.
2. A PISA-adatok nagyon erősen befolyásolják a közoktatás gyakorlatát a résztvevő országokban. Az eredmények öncélú javításának érdekében az országok rövid távú előnyöket keresve reformálják meg oktatási rendszereiket.

A PISA-nak (és más alkalmazott műveltséget mérő programoknak, tehát az Országos kompetenciamérésnek is) természetesen központi eleme a boldogulni képes állampolgár víziója. Így a mérés alapidokumentumaiban, a műveltségi területek tartalmi kereteiben, mindegyik fő terület definíciójának is szerves része (lásd 10.

⁸ Elérhető: <http://www.theguardian.com/education/2014/may/06/oced-pisa-tests-damaging-education-academics> [Letöltés dátuma: 2015.04.30.]

táblázat). Anélkül, hogy hitvitázó hévvel támadnánk Meyer elgondolásait, ki kell jelentenünk, hogy a definíciós szövegek alapján a PISA tartalmi keretei alapjául szolgáló emberképnek a lényege a demokratikus elkötelezettség, az önfejlesztésre való képesség, a megfontolt és etikus döntéshozatal, valamint ezek elválaszthatatlansága az olvasás, a matematika és a természettudományok mindennapokban való alkalmazásától.

10. táblázat

A fő mérési területek definíciói a PISA-felmérésben

Matematika	Szövegértés	Természettudomány
„Az egyénnek az a képessége, hogy különböző kontextusokban megjelenő problémákat matematikailag megfogalmaz, matematikai ismereteit alkalmazva megold, és matematikailag értelmez. Idetartozik a matematikai gondolkodás, valamint a matematikai fogalmak, eljárások, tények és eszközök használata jelensége leírásához, magyarázatához, előrevetítéséhez. Segítségével az egyén felismeri a matematika szerepét a világban, és konstruktív, elkötelezett, megfontolt állampolgárként megalapozott ítéleteket és döntéseket hoz.”	„Az egyénnek az a képessége, hogy megértse, használja, felhasználja az írott szövegeket, illetve reflektáljon azokra, annak érdekében, hogy céljait elérje, fejlessze meglévő tudását, és részt vegyen a társadalom működtetésében.”	„Az egyénnek az a képessége, hogy a természettudományi ismeretek és azok alkalmazása segítségével képes kérdéseket feltenni, új ismereteket elsajátítani, meg tud magyarázni természettudományi jelenségeket, és megfogalmaz természettudományi problémákkal kapcsolatos, bizonyítékokkal alátámasztott következtetéseket. Az egyén megérti az emberi tudásként és emberi felfedezőmunkaként is értelmezhető természettudományok jellemző tulajdonságait, valamint azt, hogy a természettudományok és a technika hogyan alakítja fizikai, szellemi és kulturális környezetünket. Megfontolt állampolgárként hajlandó magát elkötelezni természettudományi vonatkozású problémák és elméletek mellett.”

Az oktatáspolitikai hatással kapcsolatban nem ennyire egyértelmű a helyzet: tény, hogy a résztvevők köreit, a vállalkozás nagyságát és a közölt adatok felhasználási lehetőségeit nézve a PISA mindenképpen a világ legtekintélyesebb mérésének szerepét követeli magának, és igényt is tart arra a befolyásra, melyet a levél tulajdonít neki (erről a kérdéskorról lásd *Halász*, 2014). Ugyanakkor a PISA sokszor em-

legetett „bezzeg-országi”, Finnország és Lengyelország pont arra kiváló példák, hogy a méréstől *függetlenül* meginduló oktatáspolitikai fejlesztések hatásai hogyan tükröződnek az adatokban, hiszen mindkét állam a 2000-es évek előtt, a még PISA nélküli világban vágott bele közoktatási rendszerük átalakításába. Összességében elmondható, hogy a PISA-eredményekre adott oktatáspolitikai reakciók államonként igen különbözőek: a cselekvés teljes hiányától a saját reformok igazolásán keresztül a rendkívül aprólékosan megfogalmazott célokig terjedhetnek, azaz mind nagyságrendjükben, mind szándékolt hatásukban, mind pedig lezajlásuk gyorsaságának tekintetében rendkívül sokfélék (Ostorics, 2014).

Hasonlóan vitatható üzeneteket hordozott a magyar sajtóban elsőként a HVG-ben⁹, szeptember 15-én hivatkozott elemzés, melyet a *GEMS Solutions* készített a PISA2012 adatainak felhasználásával. Ennek a médiafigyelem számára legfontosabb következtetése az volt, hogy a magyar közoktatás a kutatásban vizsgált 30 ország közül az előkelő 4. helyezésre sorolható. Anélkül, hogy a kutatás módszertanával és eredményeinek értelmezésével részletesen foglalkoznánk, csak azt tekintjük át, hogyan viszonyul egymáshoz a GEMS rangsorának és az eredeti PISA-adatoknak az üzenete.

A GEMS rangsora százalékos arányban adja meg egy oktatási rendszer „hatékonyságát” (a számítás alapja többek között a pedagógusok bére, az osztályméret és a matematikaeredmény volt), amellet, hogy közli az adott ország PISA2012-es matematikai eredménye alapján született helyezést. Meglepő, hogy a „PISA2012 matematikahelyezés” oszlopban található sorszámok mennyire nem tükrözik az OECD mérésének filozófiáját, hiszen az oly sokat kárhoztatott rangsorokban a PISA mindig csak lehetséges *helyezési tartományt* közöl. Az eredmények alapján csak bizonyos eséllyel lehet megállapítani a részt vevő országok egymáshoz képesti sorrendjét. Mindemellet a rangsort kommentáló cikkek a „meglepően jól teljesít a magyar közoktatás” következtetést olvasták ki ebből a felsorolásból, amely valóban ezt is sugallja (lásd 11. táblázat).

⁹ Elérhető: http://hvg.hu/plazs/20140915_Vilagelso_a_magyar_oktatas_hatekonysaga/
[Letöltés dátuma: 2015.04.30.]

11. táblázat

Hatékony közoktatási rendszerek a GEMS Solutions besorolása alapján

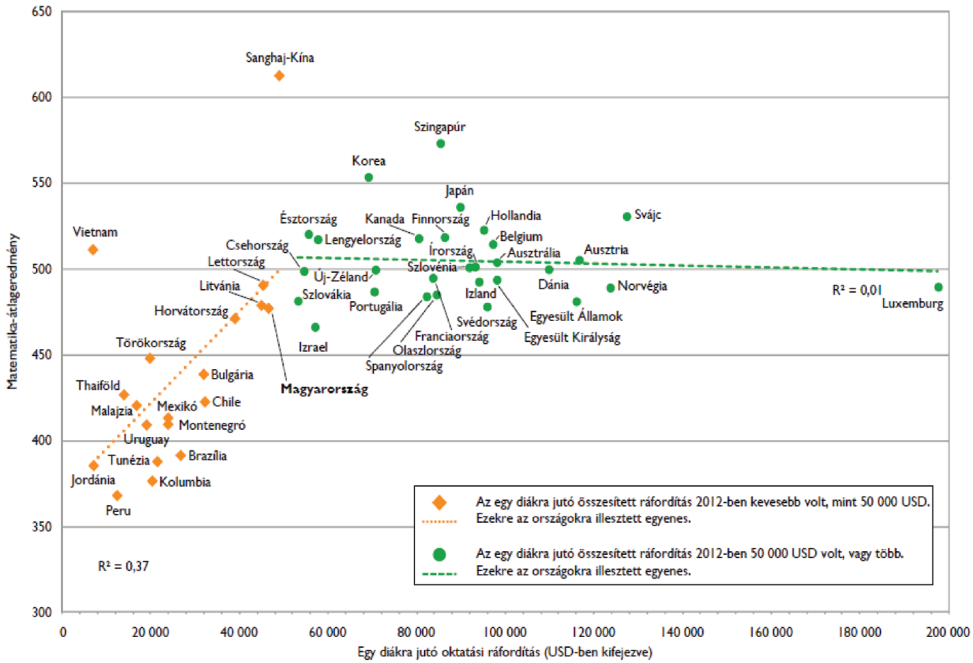
Helyezés	Ország	Hatékonyság-index	PISA2012 matematikahelyezés
1.	Finnország	87,81%	5.
2.	Korea	86,66%	1.
3.	Csehország	84,38%	14.
4.	Magyarország	84,08%	24.
5.	Japán	83,88%	2.
6.	Új-Zéland	83,30%	12.
7.	Szlovénia	83,28%	10.
8.	Ausztrália	81,23%	9.
9.	Svédország	80,58%	23.
10.	Izland	79,39%	17.

Forrás: GEMS Education Solutions, 2014

Az egy főre jutó oktatási ráfordítás és a matematikaeredmény közötti kapcsolatot – azaz azt az információt, amit a PISA az oktatási rendszerek pénzügyi értelemben vett hatékonyságával kapcsolatban közölni kíván – szemlélve megállapítható, hogy azoknál az országoknál és oktatási rendszereknél, ahol az egy diákra jutó kumulatív oktatási ráfordítás 50 000 USD alatt van (mint például Magyarországon), a magasabb oktatási ráfordítás magasabb PISA-eredményt jelez előre. Azaz a 10 000 USD-vel magasabb ráfordítás átlagosan 27 ponttal magasabb átlageredménnyel jár együtt (lásd 8. ábra). A mérés eredeti üzenete tehát nem csupán az, hogy „a magyar oktatás meglepően jól teljesít”, hanem az, hogy a ráfordított, nemzetközi összehasonlításban csekélynek számító összegeket olyan hatékonyan használja fel, hogy *bőségesen megtérülne további ráfordításokat befektetni.*

8. ábra

Az oktatási ráfordítás és a matematika-eredmények a PISA2012 országaiban



Forrás: Balácsi és mtsai., 2013: p. 29. (OECD, PISA 2012 database, Tables IV.SL.2 and 1.2.3a.)

Záró gondolatok

Össességében elmondható, hogy az egyes felmérések eredményeinek értelmezésekor minden alkalommal érdemes figyelembe venni az olyan trivialisokat is, mint a felmérések alapjellemezői: a célpopulációkat leíró kategóriák (életkor vagy iskolai évfolyamok), a tartalmi keretek (a felmért területek meghatározásai közötti különbségek), a tesztek szerkezete és a visszajelzés szintjei. Mindez fokozottan érvényes az olyan kijelentések megfogalmazásakor, amelyek olyan sokféleképpen értelmezhető jelenségekkel foglalkoznak, mint egy közoktatási rendszer hatékonysága. Ha a fentieket tekintetbe véve képesek vagyunk a megfelelő kérdések megfogalmazására, akkor a mérési eredmények értelmezése rendkívül gyümölcsöző vállalkozás.

Hivatkozások

- Balázi Ildikó és mtsai. (2007): *PIRLS 2006 Összefoglaló jelentés a 10 éves tanulók szövegértési képességeiről*. Oktatási Hivatal, Budapest. Elérhető: https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pirls/pirls2006_jelentes.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2008): *TIMSS 2007 Összefoglaló jelentés a 4. és 8. évfolyamos tanulók képességeiről matematikából és természettudományból*. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/timss/timss_2007_osszefoglalo_jelentes.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2010a): *A PISA2009 tartalmi és technikai jellemzői*. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pisa/A_PISA2009_tartalmi_es_technikai_jellemzoi.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2010b): *PISA2009 Összefoglaló jelentés – Szövegértés tíz év távlatában*. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pisa/pisa_2009_osszfojl_jel_110111.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2012a): *A PIRLS és TIMSS 2011 Tartalmi és technikai jellemzői*. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pirls/PIRLS_TIMSS_2011_tartalmi_technikai_jellemzoi.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2012b): *PIRLS és TIMSS 2011 Összefoglaló jelentés a 4. évfolyamos tanulók eredményeiről*. Oktatási Hivatal, Budapest. Elérhető: https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pirls/PIRLS_TIMSS_2011_osszefoglalo_jelentes_4evf_eredmenyeirol.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2012c): *TIMSS 2011 Összefoglaló jelentés a 8. évfolyamos tanulók eredményeiről*. Oktatási Hivatal, Budapest. Elérhető: https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/timss/TIMSS_2011_Osszefoglalo_jelentes_8evf_eredmenyeirol.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2013): *PISA2012 Összefoglaló jelentés*. Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merekek/pisa/pisa2012_osszefoglalo_jelentes.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és Ostorics László (2014): *A PISA és az Országos kompetenciamérés kapcsolata, összehasonlítása*. Tartalmi és szerkezeti változások a köznevelésben. XBI: Országos Közoktatási Szakértői Konferencia, Hajdúszoboszló. Elérhető: http://suliszerviz.com/images/stories/Dok/2014/konferencia_eloadasok/okt_9_csutortok_delelott_plenaris/Ostorics%20L%C3%A1szl%C3%B3.pdf [Letöltés dátuma: 2015.04.30.]
- Balázi Ildikó és mtsai. (2014): *Országos kompetenciamérés 2013. Országos jelentés.*: Oktatási Hivatal, Budapest. Elérhető: http://www.oktatas.hu/pub_bin/dload/kozoktatas/merekek/orszmer2013/Orszagos_jelentes_2013_03.pdf [Letöltés dátuma: 2015.04.30.]
- Balkányi Péter és Ostorics László (2012): *A szövegértés tartalmi keretének változása. Educatio 21(3), pp. 465–471.* Elérhető: <http://www.hier.iif.hu/hu/letoltes.php?fid=tartalomsor/2176> [Letöltés dátuma: 2015.04.30.]
- Berényi Eszter (2010): *A mérési iskoláktól az iskolák megméréseig. Educatio 19(4), pp. 601–613.* Elérhető: www.hier.iif.hu/hu/letoltes.php?fid=tartalomsor/2024 [Letöltés dátuma: 2015.04.30.]

- Breakspear, S. (2012): The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. *OECD Education Working Papers*, No. 71, OECD Publishing. Elérhető: <http://dx.doi.org/10.1787/5k9fdqfqr28-en> [Letöltés dátuma: 2015.04.30.]
- GEMS Education Solutions (2014): *The Efficiency Index. Which Education Systems Deliver The Best Value For Money?* Elérhető: <http://www.edefficiencyindex.com/book/> [Letöltés dátuma: 2015.04.30.]
- Halász Gábor (2014): Az OECD és az oktatás. In: Széll Krisztián (szerk.): *Az OECD az oktatásról – adatok, elemzések, értelmezések* (pp. 7–15.), Oktatókutató és Fejlesztő Intézet, Budapest. Elérhető: https://www.ofi.hu/sites/default/files/attachments/az_oecd_az_oktatasrol_ofi_2014.pdf [Letöltés dátuma: 2015.04.30.]
- Martin, M. O. és mtsai. (2012a): *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S. és mtsai. (2007): *PIRLS 2006 International Report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S. és mtsai. (2012a): *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S. és mtsai. (2012b): *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Meyer, H-D. (2014): OECD and PISA-tests are damaging academics worldwide. Elérhető: <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics> [Letöltés dátuma: 2015.04.30.]
- OECD (2012): *PISA 2009 Technical Report*. OECD Publishing. Elérhető: <http://dx.doi.org/10.1787/9789264167872-en> [Letöltés dátuma: 2015.04.30.]
- OECD (2013a): *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. Elérhető: <http://dx.doi.org/10.1787/9789264190511-en> [Letöltés dátuma: 2015.04.30.]
- OECD (2013b): *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science (Volume I)*. OECD Publishing.
- OECD (2013c): *PISA 2012 Results: Excellence through Equity: Giving Every Student the Chance to Succeed (Volume II)*. OECD Publishing.
- OECD (2013d): *PISA 2012 Results: Ready to Learn: Student Engagement, Attitudes and Motivation (Volume III)*. OECD Publishing.
- OECD (2013e): *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. OECD Publishing.
- Ostorics László (2014): *Valóban kártékony a PISA?* Raabe KLETT, Budapest (előkészületben).
- Schleicher, A. (2014): *Response To Points Raised in Heinz-Dieter Meyer 'Open Letter'*. OECD Publishing. Elérhető: <http://www.oecd.org/pisa/aboutpisa/OECD-response-to-Heinz-Dieter-Meyer-Open-Letter.pdf> [Letöltés dátuma: 2015.04.30.]
- Széll Krisztián (2014) (szerk.): *Az OECD az oktatásról – adatok, elemzések, értelmezések*. Oktatókutató és Fejlesztő Intézet, Budapest. Elérhető: http://www.ofi.hu/sites/default/files/attachments/az_oecd_az_oktatasrol_ofi_2014.pdf [Letöltés dátuma: 2015.04.30.]

Horn Dániel

Az iskolai hozzáadott érték mérése

Bevezetés

Az elmúlt években, évtizedekben egyre nagyobb hangsúlyt kap az iskolák elszámoltathatósága. A 2001-es amerikai *No Child Left Behind* törvényt¹ már azzal a feltevéssel fogadták el, hogy iskolai, illetve egyéni standardok felállítása és ezek elérésének rendszeres mérése, illetve ezekhez kapcsolódó ösztönzők kialakítása növelni fogja a tanulói – és egyben tanári – teljesítményeket.²

A 2008-as évtől kezdve az Országos kompetenciamérés is kihasználja azt a lehetőséget, hogy minden közoktatásban tanuló diák rendelkezik oktatási azonosítóval. Ez a kódszám, illetve az ebből számított mérési azonosító lehetőséget teremt a tanulók kompetenciamérés-eredményeinek évek közötti egyéni szintű összekötésére, ami, többek között, lehetőséget teremt a hozzáadott érték típusú eredményességi mutatók számítására. Jelenleg az iskolák főleg keresztmetszeti szintű adatokon becsült „iskolahatásokat”³ kapnak kézhez. Ez különösen azért probléma, mert ezen iskolahatások nem veszik figyelembe a diákok előző teljesítményét, vagyis jobbnak

¹ Elérhető: <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf> [Letöltés dátuma: 2015.04.30.]

² Feltehetőleg hasonló premissán alapul az EMMI 2014/35-ös rendelete, illetve különösen ennek 9§-nak 10. bekezdése is, amely az Országos kompetenciamérések alapján készült jelentések figyelembevételével rendeli el iskolai intézkedési tervek kialakítását és ezek végrehajtását.

³ Iskolahatás alatt az iskola valamilyen statisztikai módszerrel számított eredményességét értjük a tanulók valamilyen jellemzőire (pl. teszteredményeire). Vagyis a továbbiakban az iskolahatás az összes olyan mutatószám összefoglaló neve, beleértve a hozzáadott-érték típusú mutatókat is, amely az iskola tanulóira gyakorolt hatását próbálja számszerűsíteni.

A tanulmány részben támaszkodik az MTA-KTI A közoktatás teljesítményének mérése-értékelése, az iskolák elszámoltathatósága 2008–2011 között futó programjának ACC 1503. számú produktumára.

mutatják a képességek szerint szelektáló iskolákat.⁴ A telephelyi jelentések ugyan tartalmazznak egy „komplex modellt”, amely egy bizonyos típusú hozzáadott érték mérés eredményeit mutatja, ám ez csupán egy indikátor a több tucat között. Hozzáadott érték típusú indikátorból azonban nem csak egyféle létezik.

Jelen tanulmánynak nem célja e külső mérések és az ehhez kapcsolt ösztönzők jogosságának vagy eredményességének elemzése. A tanulmány bemutatja az elméletileg lehetséges hozzáadott érték típusú indikátorok alapverzióit és néhány nemzetközi példát is hoz ezekre. Továbbá kísérletet tesz arra, hogy magyar adatokon megvizsgálja, valóban más eredményeket kapnának-e az iskolák, ha nem keresztmetszeti, hanem több időpontban mért, tanulói szinten összekötött adatokon becsült, hozzáadott érték típusú eredményességi mutatókat kapnának kézhez. Ez a kutatás csupán egy *pilot*, amely nem alkalmas iskolai szintű következtetések levonására, de az oktatáspolitikai döntéshozóknak kulcsfontosságú következtetésekkel szolgálhat.

A tanulmány első fele az elméleti, míg a második az empirikus fejezeteket tartalmazza. Az elméleti rész érvel a hozzáadott érték típusú mutatók fontossága mellett, és kifejti néhány alaptípusát a hozzáadott érték típusú modelleknek. A hozzáadott érték típusú modellekből négy, a gyakorlatban is felhasznált modell is bemutatásra kerül. Az empirikus rész a jelenleg használt Országos kompetenciamérést mutatja, továbbá az elméletben bemutatott eredményességi mutatókat veti össze az iskolajelentések eredményeivel.

Hozzáadott érték-modellek az elméletben

Hogy mit tekintünk az iskola teljesítményének, egyaránt definíciós és módszertani kérdés (lásd még e kötetben *Szemerszki Marianna* írását). Az elszámoltathatósági rendszerek leginkább a tanulói teszteredményeket használják fel kimenetként. Természetesen az e tesztek által mért tanulói teljesítmények, illetve az ezekből számított iskolai teljesítmény csupán kiragadott mutatói az iskola valódi teljesítményének, ám feltehetőleg a rendelkezésre álló mérések közül még mindig a legpontosabbak, legmegbízhatóbbak. Ezen teszteredmények nagy előnye, hogy általában fontos alapképességeket mérnek (például szövegértés, matematika, természettudományok), ráadásul folytonos mutatók (szemben például a diszkrét to-

⁴ A hozzáadott érték típusú modellek előnyeiről és hátrányairól lásd például *Ladd és Walsh* (2002), vagy *Kane és Staiger* (2002), a felhasznált termelési függvényről pedig *Todd és Wolpin* (2003), vagy *Dolton* (2002) tanulmányait.

vább tanul/nem tanul tovább, vagy a kategorikus dolgozik/munkanélküli/inaktív mutatókkal), ami módszertanilag megkönnyíti az indikátorok számítását (lásd még e kötetben *Nahalka István* írását).

A tanulmány e teljesítménymutatókból számított iskolai eredményesség indikátorainak sokféleségét hangsúlyozza, és nem foglalozik a különböző eredményességfelfogásokból adódó eltérésekkel. Egy adott tanulói teljesítménymutatóból is nagyon sok eredményességi indikátort lehetséges előállítani. A gyakorlatban is szinte feltérképezhetetlen variációkban fordulnak elő egyszerű keresztmetszeti adatokon, vagy több időpontban történő méréseken nyugvó hozzáadott érték típusú eredményességi mutatók, amelyek ugyanazon kimenetet használnak fel. A tanulmány e mutatók leggyakoribb és az oktatáspolitikai döntéshozók számára talán leginkább felhasználható variánsait mutatja be.

A *keresztmetszeti adatokon nyugvó*, iskolai eredményességet leíró indikátoroknak két nagyon fontos előnyük van. Az *első*, hogy szinte azonnal, a felmérést követően és általában sokkal nagyobb arányban rendelkezésre állnak, mellyel szemben a hozzáadott érték típusú mutatókhoz minimum két időpontban szükséges megfigyeléseket végezni egy adott iskolában, és így az ezeken alapuló visszajelzés is sokkal lassabb. Ráadásul a két időpontban végzett felmérések elkerülhetetlenül nagyobb válaszmegtagadással vagy mintakopással járnak, ami szintén gyengítheti a mérés megbízhatóságát. A *másik előnye* e mutatóknak az egyszerű értelmezhetőségük. A keresztmetszeti adatokon nyugvó nyers mutatók általában egyszerű iskolai átlagok, amik – egy jellemző példával szemléltetve – például azt mutatják meg, hogy egy adott iskolában a tanulók hány százaléka teljesít egy megadott minimumszint alatt. E mutatóknál nincs szükség bonyolult statisztikai vagy ökonometriai számításokra, s így az indikátorok visszacsatolási funkcióját is feltehetőleg könnyebben betöltik. Ráadásul adott indikátorok közpolitikailag is fontos üzenetet hordozhatnak: az előbbi példa esetében például azt, hogy függetlenül az iskola társadalmi összetételétől, programtípusától, fenntartójától vagy egyéb jellemzőitől minden tanulótól (illetve a tanulók megadott arányától) elvárt, hogy a minimumszint felett teljesítsen. Nyilvánvalóan némely iskolának egy ilyen indikátornak való megfelelés nem jelent kihívást, míg egy másiknak igen. Ugyanakkor éppen ez a keresztmetszeti adatokon nyugvó indikátorok egyik legnagyobb hátránya is: a tanulók jellemzői, illetve az iskolától független egyéb jellemzők nagymértékben meghatározzák, hogy egy adott iskola miként teljesít. Egy átlagosan jobb képességű tanulókat oktató iskolában várhatóan akkor is sokkal nagyobb arányban fognak a tanulók a megadott teljesítménymínimum felett teljesíteni, ha az iskola semmivel nem jobb (a szó általános értelemben), mint egy másik iskola, ahol a tanulók kevésbé jó ké-

pességűek. Épp ennek a torzításnak a csökkentése miatt fontos a hozzáadott érték típusú mutatók számítása.

Ahhoz, hogy a hozzáadott érték típusú mutatók előnyét be lehessen mutatni, be kell vezetni az oktatási témákkal foglalkozó közgazdászok (oktatásgazdászok) által használt oktatási termelési függvény fogalmát.⁵ A termelési függvény azt feltételezi, hogy az iskola egy meghatározott kimenetét (például tanulói tesztpontszámok) különféle, az iskolától független, illetve az iskolától függő változók befolyásolják. *Todd és Wolpin* (2003) részletesen összefoglalja a különböző termelési függvények becslésének módszereit, feltevéseit és minimális adatigényüket. Cikkük alapján *Dolton* (2002) a következő egyszerűsített modellt mutatja be.

Feltéve, hogy a tanuló iskolai teljesítménye az egyéni veleszületett képességek és tulajdonságok, az iskolai és a családi erőforrások függvénye, a következő általános termelési függvényt írhatjuk fel:

- (1) $Y_{ijkt} = g(X_{it}, \Sigma S_{ijkt}, \Sigma F_{it}, \mu_i)$, ahol
- Y_{ijkt} a t -edik évben, j -edik iskolába, k -edik osztályba járó, i -edik tanuló teljesítménye;
 - X_{it} a t -edik évben az i -edik tanuló minden olyan személyes jellemzője, amely befolyásolhatja a teszteredményét;
 - S_{ijkt} a t -edik évben, j -edik iskolába, k -edik osztályba járó, i -edik tanuló rendelkezésére álló iskolai erőforrások;
 - F_{it} a t -edik évben az i -edik tanuló minden családi jellemzője, erőforrásai;
 - μ_i az i -edik tanuló veleszületett képességei.

Azaz az adott évben megfigyelt tanulói jellemzők, az adott és az azt megelőző években megfigyelt iskolai erőforrások és családi jellemzők, erőforrások, illetve a tanuló veleszületett képességei mind befolyásolják valamilyen módon az adott időpontban megfigyelt tanulói teljesítményt.

Így az első időszakra például a következő egyszerű függvényt becsülhetjük (feltéve, hogy az összefüggés a tanulói teljesítmény és az inputok között lineáris és additív):

$$(2) \quad A_{ijkt} = \alpha_1 X_{it} + \beta_1 \Sigma S_{ijkt} + \delta_1 \Sigma F_{it} + \gamma_1 \mu_i + \varepsilon_{it} + u_{jt}$$

⁵ *Horn* (2011) felhasználásával.

ahol α , β , γ , δ becslési paraméterek, míg ε_i és u_j egyéni és iskolai szintű nem megfigyelt változók hatását mutatja.

Az ilyen termelési függvényekkel végzett elszámoltathatósági vizsgálat célja általában az u_j , azaz az iskolai szinten nem megfigyelt változók hatásának becslése. Ez a hatás egy úgynevezett reziduum-, vagyis maradék-hatás. Az u_j minden olyan iskolaszintű hatást magában foglal, amit az egyéb, a modellbe bevont változók nem magyaráznak. Ez a szám mutathatja meg, hogy egy adott időszakban hogyan teljesített az iskola az átlaghoz képest, figyelembe véve az iskolába járó tanulók képességeit és tulajdonságait, illetve az iskola és a család erőforrásait. A becslés igencsak adatigényes, hiszen például ismertnek feltételezi a diák veleszületett képességeit, illetve az elmúlt időszakok iskolai és családi ráfordításait, amelynek mérése nem egyszerű feladat.

A közgazdászok e nagy adatmennyiség kikerülésére a következő módon becsült, úgynevezett „hozzáadott-érték” vagy „student-gain” (tanulói-nyereség) típusú becslést szokták elvégezni. Elméletileg hasonlóan az első időszakhoz, a második időszakra is meg lehet becsülni a termelési függvényt:

$$(3) \quad A_{ijk2} = \alpha_2 X_{i2} + \beta_2 \sum S_{ijk2} + \delta_2 \sum F_{i2} + \gamma_2 \mu_i + \varepsilon_{i2} + u_{j2}.$$

A kettő különbségét véve ezt kapjuk:

$$(4) \quad \Delta A = \alpha_2 X_{i2} - \alpha_1 X_{i1} + \beta_2 \sum S_{ijk2} - \beta_1 \sum S_{ijk1} + \delta_2 \sum F_{i2} - \delta_1 \sum F_{i1} + \gamma_2 \mu_i - \gamma_1 \mu_i + \varepsilon_{i2} - \varepsilon_{i1} + u_{j2} - u_{j1}.$$

Ami az alábbi feltevésekkel a következő formára egyszerűsíthető:

$$(5) \quad \Delta A = \alpha X_i + \beta S_{ijk} + \delta F_i + \varepsilon_i + u_j.$$

Feltételezések:

1. A tanulók személyes jellemzői nem változnak az idő múlásával ($X_{i1} = X_{i2}$ így $\alpha_2 X_{i2} - \alpha_1 X_{i1} = \alpha X_i$). Ez természetesen azt jelenti, hogy olyan jellemzők, mint a tanulói motiváció vagy erőfeszítés sem változik. Ez nyilvánvalóan nem felel meg a valóságnak, így eléggé erős feltételezés.
2. Az iskolai erőforrásokat egy adott időpontban megfelelően meg lehet figyelni, így $\beta_2 \sum S_{ijk2} - \beta_1 \sum S_{ijk1} = \beta S_{ijk}$, vagyis az előző ráfordítások hatása kiesik az egyenletből.

3. A családi ráfordításokat egy adott időpontban megfelelően meg lehet figyelni, így $\delta_2 \Sigma F_{i2} - \delta_1 \Sigma F_{i1} = \delta F_p$, vagyis az előző ráfordítások hatása kiesik az egyenletből.
4. A tanulói veleszületett képességek hatása az idővel nem változik, azaz $\gamma_2 = \gamma_1$, így a veleszületett képességek hatása kiesik az egyenletből. Ez a feltevés is erős, hiszen valószínű, hogy az életkor növekedésével a gyermek képességeinek hatása változik.
5. A család és az iskola általi ráfordítások nem függenek össze a tanuló(k) korábbi teljesítményével, azaz S_2 és F_2 nem korrelál Y_0 -al illetve μ_i -vel. Tekintve, hogy például várhatóan a szülők sokkal több energiát fordítanak gyermekük oktatására mint korábban, ha kiderül, hogy rosszul teljesít az iskolában, ez a feltevés sem felel meg teljes mértékben a valóságnak.

Bár a felvételek igencsak erősek, ha elfogadjuk ezeket, az iskolai hatás (u_j) becsléséhez már reális mennyiségű adat is elegendő. Ez esetben szükség van legalább két időpontban mért, tanulói szinten összekötött adatokra a tanuló iskolai teljesítményéről, illetve legalább egy megfigyelésre az iskola és a család ráfordításairól, és a tanuló tulajdonságairól. Ráadásul, ha az utóbbi változókat mindkét időpontban megfigyeljük, azaz részletesebb adatok állnak rendelkezésre, a feltételezések egy része feloldható.

Hozzáadott érték-modellek a gyakorlatban

A hozzáadott érték típusú modellek bemutatását a legegyszerűbb lineáris modellekkel kezdeni. Ezek mind felvételeikben, mind adatigényében viszonylag egyszerűek, és pont ezért könnyen átláthatóak és relatíve könnyen megérthetőek. A modellek néhány, gyakorlatban is alkalmazott, nem lineáris specifikációját a következő alfejezet tárgyalja.

A hozzáadott érték típusú modellek közös tulajdonsága, hogy minimum két időpontban megfigyelt, tanulói szinten, az időpontok között összekötött adatokat használnak fel. Egy jelentős leegyszerűsítéssel élve a modelleket két dimenzió alapján lehet csoportosítani:

1. az egyik dimenzió, hogy a múltbeli teljesítmény-pontszámokat függő vagy független változóként veszi-e figyelembe,
2. a másik pedig, hogy az iskola hatásának random vagy fix hatást becsül.

Az *első dimenzió* szerint 3 típust különböztethetünk meg: (1) a múltbeli teljesítményt, mint független változót felhasználó („kiigazított”) modellek, (2) a több

időpontban megfigyelt kimenetek különbségét felhasználó („*fejlődés*–”) modellek, illetve (3) ezek kombinációja („*kombinált*” modellek). A *másik dimenzió* szerint – hogy random vagy fix hatással becsülik-e az iskola hatását, vagyis a hozzáadott értéket – szintén *három eset* lehetséges: random hatás becsülésénél lehet (1) az egyéni hibák iskolai átlagát vagy (2) az iskolai random hatást vizsgálni, míg fix hatás esetében (3) az iskolai fix hatások felelnek meg a hozzáadott értéknek.

Az 1. táblázat összefoglalja a két dimenziót, és az általuk meghatározott hozzáadott érték mutatókat, illetve azt a jelölést, amelyet a továbbiakban a tanulmány alkalmaz.

1. táblázat

A tanulmányban számolt indikátorok és ezek rövidítései

		Keresztmetszeti adatokon alapuló iskolahatás	Hozzáadott érték típusú mutatók		
			„kiigazított”	„fejlődés”	„kombinált”
Random hatás	Egyéni hibák iskolai átlaga	IH_E	$HÉ_{E1}$	$HÉ_{E2}$	$HÉ_{E3}$
	Iskolai random hatás	IH_I	$HÉ_{I1}$	$HÉ_{I2}$	$HÉ_{I3}$
Fix hatás	Iskolai fix hatás	IH_F	$HÉ_{F1}$	$HÉ_{F2}$	$HÉ_{F3}$

Az alábbiakban először az első dimenzió logikája kerül bemutatásra a legegyszerűbb, egyéni hibák iskolai átlagát figyelembe vevő modell segítségével, majd az iskolai random és fix hatás különbsége kerül bővebb kifejtésre.

Random hatás-modellek – az egyéni hibák iskolai átlaga

A három felsorolt modelltípus közül ez a legegyszerűbb. Az iskolahatás a tényleges és a modell által becsült érték közötti különbség (az egyéni hiba) iskolai átlaga; vagyis ugyanaz a logika, mint az Országos kompetenciamérés telephelyi jelentéseiben (lásd alább). A modell feltételezi, hogy az egyéni hiba normális eloszlású, egyéni szinten várható értéke 0 és szórása konstans, illetve független minden modellbe illesztett kontrollváltozótól.

A) „Kiigazított” (covariate adjustment) modell

A modell legáltalánosabb formája:

$$(6) \quad y_{ijt} = a_0 + a_1 y_{ij(t-1)} + bX + e_{ij},$$

ahol a fentebb bemutatott jelölések mellett t az idő jele. A hozzáadott érték ($H\hat{E}_{E1}$) a modell által becsült és a valós érték közötti eltérés, azaz

$$(7) \quad H\hat{E}_{E1} = \text{átlag} (y_{ijt} - \hat{y}_{ijt}) = \hat{e}_{ij},$$

ahol \hat{y}_{ijt} az adott diákra becsült érték.

A modell tehát azt feltételezi, hogy a tanuló adott évi teljesítménye függ az előző teljesítményétől, és egyéb egyéni (családi) és iskolai jellemzőitől.⁶ Ezek segítségével lineáris becsléssel közelíti a jelenbeli teljesítményét. Amennyiben egy iskolában a tanulók átlagosan ennél, azaz az elvártnál jobban teljesítenek, az adott iskola hozzáadott értéke pozitív, ellenkező esetben negatív.

B) „Fejlődés” (gain) modell

Ennél valamelyest restriktívebb feltételezésekkel él a következő modell:

$$(8) \quad (y_{ijt} - y_{ij(t-1)}) = a_0 + bX + e_{ij}.$$

A hozzáadott érték ($H\hat{E}_{E2}$) itt is a hibatag iskolai átlaga:

$$(9) \quad H\hat{E}_{E2} = \hat{e}_{ij}.$$

A két modell közötti legfontosabb különbség, hogy míg $H\hat{E}_{E1}$ azt feltételezi, hogy a múltbeli teljesítmény és az egyéni, illetve iskolai jellemzők befolyásolják a jelenbeli teljesítmény szintjét, addig a $H\hat{E}_{E2}$ szerint az egyéni és iskolai jellemzők a teljesítménynövekedés ütemét befolyásolják. Azaz, míg a „kiigazított” modell megengedi, hogy az azonos egyéni vagy iskolai jellemzővel rendelkező diákok másképp fejlődjenek az adott periódus alatt, addig a „fejlődés”-modell azt feltételezi, hogy az azonos egyéni, illetve iskolai jellemzőkkel rendelkező diákok azonos ütemben fej-

⁶ Az, hogy a gyakorlatban a modell milyen egyéni és iskolai változókat vesz/vehet figyelembe, külön tanulmányt érdemelne, és az adott kérdésben az oktatáspolitikának a modellválasztás mellett egyik kulcskérdése kell, hogy legyen.

lódnek.⁷ Hasonlóan az előző modellhez, az egyéni és iskolai jellemzők alapján meg lehet becsülni a $t-1$ és a t időpont közötti fejlődés elvárható értékét, majd ennek a becsült értéknek és a megfigyelt valós értéknek a különbségét iskolánként átlagolva, kiszámolni az iskola hatását.

C) „Kombinált” modell

A két modell kombinációjához már nem csupán két, hanem három év tanulói szinten összekötött adataira van szükség:

$$(10) \quad (y_{ijt} - y_{ij(t-1)}) = a_0 + a_1 y_{ij(t-2)} + bX + e_{ij},$$

ahol HE_{E3} a hibatag iskolai átlaga:

$$(11) \quad HE_{E3} = \hat{e}_{ij}.$$

Ez a modell a fejlődés-modellhez képest azt is feltételezi, hogy a tanulói teljesítménynövekedés üteme nem csupán a családi háttértől (egyéni jellemzőktől) és az iskolai jellemzőktől függ, hanem a tanuló előző teljesítményétől is.

A három modell feltételezéseiben (egyéni szinten normál eloszlású hiba, amely független a bevont kontrollváltozóktól) megegyezik egymással. Értelmezésük is viszonylag egyszerű, adatigényük is hasonló egymáshoz.

Random hatás-modellek – iskolai random hatás (többszintű modellek)

Az egyszerű lineáris modellek továbbfejlesztett változatai az iskolai random hatást is magukban foglaló modellek. Ezek az úgynevezett többszintű vagy hierarchikus modellek két regressziós egyenletből állnak: egy egyéni és egy iskolai szintű becslésből. Ez annyiban tér el a fentebbi lineáris becsléstől, hogy a konstans tagot iskolák között engedi szóródni, és a hozzáadott érték (HE_{ij}) az iskolai szintű regresszió (12b) hibatagjának felel meg (13):

$$(12) \quad \begin{aligned} (a) : y_{ijt} &= a_{0j} + a_1 y_{ij(t-1)} + bX + e_{ij}, \\ \text{illetve} \\ (b) : a_{0j} &= A + \delta_{0j}, \end{aligned}$$

⁷ Amennyiben az előző modellben az $a_1 = 1$, a két modell azonos eredményre vezet.

$$(13) \quad H\hat{E}_{11} = \delta_{0j}$$

Az egyenletben becsült két hibatag (az e_{ij} egyéni és a δ_{0j} iskolai) függetlenek egymástól, és mindkettő 0 átlagú fix szórású random tag. Ezeket a modelleket általában „random hatás” modelleknek is hívják, mivel az iskola hatását egy random változó írja le, szemben a fix-hatás modellekkel. A modelleket a fentebb leírt módon lehet „fejlődési” ($H\hat{E}_{12}$) és „kombinált” ($H\hat{E}_{13}$) modellekként is becsülni.

Fix hatás-modellek – iskolai fix hatás

A random hatással szemben itt az iskola hatását nem egy random, hanem egy fix taggal becsüljük. Vagyis a lineáris regresszióba illesztett iskolai dummy változók koefficiensei lesznek az iskolahatások.

$$(14) \quad y_{ijt} = a_0 + a_1 y_{ij(t-1)} + bX + \Theta_j + e_j,$$

ahol

$$(15) \quad H\hat{E}_{F1} = \Theta_j,$$

vagyis az iskolai fix hatás a hozzáadott érték mutató. Hasonlóan a fentiekhez az iskolai fixhatás-modelleket is lehet „fejlődési” ($H\hat{E}_{F2}$) és „kombinált” ($H\hat{E}_{F3}$) modellekként is becsülni.

Az egyes modellek összevetése⁸

Gyakorlatban *mindegyik modell használható*, sőt – mint az a tanulmány empirikus fejezetében látható lesz – eredményeiben is hasonlóak egymáshoz, azonban a becslések lényegesen különbözhetnek az egyes iskolára vonatkoztatva. Bár teoretikusan a „kombinált” modellek feltétlenül előnyösebbnek tűnnek, mint a „kiigazított” vagy

⁸ A modellek kiválasztásához köthető statisztikai, módszertani problémákat – mint például a kihagyott változók, a hiányzó értékek problémája vagy a modell hatásossága, illetve torzítatlansága – a tanulmány nem tárgyalja. A témáról lásd részletesebben az OECD jelentését (OECD, 2008) vagy a Rand Corporation-nak készített szakértői jelentést (McCaffrey és mtsai., 2003).

a „*fejlődés-*”modellek, kivitelezhetőségük sokkal problematikusabb. Elsősorban három időpontban megfigyelt tanulói szinten összekötött adatbázist sokkal nehezebb (költésesebb és időigényesebb) létrehozni, mint a két időpontban megfigyelt adatokat tartalmazó adatbázisokat. Ám, még ha sikerül is az összekötés, az adatvesztés (évismétlők, lemorzsolódók, válaszmegtagadó stb.) által okozott problémák sokkal nagyobbak, mint a másik két esetben. Ez természetesen további módszertani problémákat is felvet. Ráadásul *Ladd és Walsh* (2002) úgy érvel, hogy ha rendelkezésre áll 3 év tanulói szinten összekötött adatbázisa, akkor a legelső év pontszámait sokkal célszerűbb a második év pontszámainak instrumentumaként felhasználni, és ezáltal szűrni ki a pontszámok – a szerzők eredményei szerint igen jelentős mértékű – mérési hibáját.

A „*kiigazított*” és a „*fejlődés-*”modellek között nincs ilyen egyértelmű átváltás. Ugyanazon az adatbázison meg lehet valósítani mindkettőt, ám a kapott eredmények mást jelentenek. Míg a „*kiigazított*” modellek esetén az adott évben az egyes iskolák tanulójának átlagos teljesítménye szintbeli különbségét vetjük össze (figyelembe véve, hogy előtte hogyan teljesítettek, illetve milyen egyéni és iskolai jellemzőik vannak), addig a „*fejlődés-*”modellek esetében a két időpont közötti fejlődésük ütemét vetjük össze, és azt feltételezzük, hogy ezt a fejlődési ütemet meghatározzák az egyenletbe bevont egyéni és iskolai változók. Vagyis kontrollváltozók bevonása nélkül a két becslés tökéletesen azonos eredményeket produkálna.

Az iskolai random és fix hatás-modellek nagyon hasonlóak egymáshoz. De míg a fix tag használata esetén a hozzáadott érték mutató pusztán a kontrollként bevont változóktól függ, azaz csak az iskola (és az oda járó diákok) saját jellemzőitől, random hatás használata esetén a modell az összes adatbázisba bevont iskola jellemzőit is figyelembe veszi, amelyet egy úgynevezett „összehúzás” (*shrinking*) segítségével ér el. Ez a legegyszerűbb iskolai random hatás-modellek esetén a következő:

$$(16) \quad \theta_j = \lambda_j (\hat{y}_j - \hat{y}),$$

ahol \hat{y}_j az iskola mért tanulói teljesítményének az átlagát jelöli, míg \hat{y} az országos átlag. A λ_j egy 1-nél kisebb súly, azaz az iskolai átlag és az országos átlag közötti hiba „össze van húzva” nulla felé. A fix hatás-becslés ezzel szemben pontosan a két átlag különbségével egyezik meg (lásd bővebben *McCaffrey és mtsai.*, 2003: p. 65).⁹

A *random*, illetve *fix hatás-modellek* között tehát hasonlóan nehéz a választás (*McCaffrey és mtsai.*, 2003; *OECD*, 2008). Mint az fentebb látható, az iskolai ran-

⁹ Míg a random hatás-becslés esetén a \hat{y} az iskolai átlagok súlyozott átlaga, azt feltételezve, hogy $\sum(\hat{y} - y) = 0$, addig a fix hatás-becslés esetén \hat{y} a valós átlag.

dom és fix hatás-modellek között az a különbség, hogy a random hatás szórása, az „összehúzás” miatt, valamelyest kisebb lesz, mint a fix hatásé, ugyanis a random hatás-modellek a nagyobb „hibával” rendelkező iskolákat az átlag felé húzzák. Továbbá minden hozzáadott érték-becslésnél fontos figyelembe venni, hogy a becslés megbízhatósága nagyban függ az iskola méretétől (*Kane és Staiger, 2002*). A kis iskolákra adott becslés mindig megbízhatóbb, mint a nagy iskolákra adott. Így a random hatás-modellekben az iskola mérete miatti eltérések kevésbé jelennek meg. Mivel a random hatás becslése a szélsőségesen jó vagy rossz iskolákat inkább az átlag felé húzza, így várhatóan épp a kisebb iskolák lesznek azok, amelyekre a random hatás-becslés más eredményeket ad. Vagyis míg országos közpolitikai szempontból a random hatás-modellek kisebb hibával becsülnek, az egyes iskolákra nézve a random hatás, az „összehúzás” miatt a becsült érték el fog térni a valós értéktől, azaz torzított lesz a becslés. A random hatás használata mellett szól továbbá, hogy ezeket a modelleket sokkal egyszerűbb úgy kibővíteni, hogy figyelembe vegyenek egyéb iskolai tényezőket is (például random meredekséget, vagy egyéb iskolai szintű jellemzőket), bár ezek fix hatás számításánál is megoldhatók, csak kevésbé egyértelműen.

Összességében a random hatás-becslés akkor lehet jobb, ha a cél közpolitikai jellegű, például ha az iskolák egyes tulajdonságainak az iskola eredményességére való hatását vizsgáljuk. Egyéb esetekben – például amikor az iskolák hatását elszámoltathatósági céllal mérjük – a választás nem egyértelmű, a kisiskolákkal kapcsolatos probléma ugyanis mindkét becslésnél fennáll (lásd például *Kane és Staiger, 2002*). Bár a fix hatás-becslés esetében a becslés és a valós érték közötti eltérés (a hiba) nincs mesterségesen összehúzva, így a becslés nem lesz torzított (várható értékben minden iskola pont olyan becsült értéket kap, amekkora a feltételezett iskolahatás), azonban a kisiskolákban és/vagy a nagyon heterogén iskolákban a becslés így is nagyon bizonytalan lesz, nagy szórással. Vagyis ezekben az esetekben két rossz közül kell választani:

1. vagy a hibát mesterségesen csökkentjük, tudva, hogy az extrém esetekben – a különösen rossz és a különösen jó iskolák esetében – a hozzáadott érték becslése az átlag felé fog torzítani,
2. vagy kockáztatjuk, hogy a becslés bizonytalansága miatt nagyon nagy lesz egy-egy iskola hozzáadott értékében az évek közötti eltérés, függetlenül az iskola valós teljesítményétől.

Az egyéni hibák iskolai átlaga alapján becsült iskolahatás is torzított, ha azt felételezzük, hogy az iskolai fix hatás a torzítatlan közelítés. Mert például, ha a jó tanulók az eredményesebb (azaz a nagyobb fix hatást mutató) iskolákban, míg a rossz

tanulók a kevésbé eredményes iskolákban csoportosulnak, akkor az egyéni hibák átlagán alapuló hozzáadott érték-becslés lefelé torzított lesz a jó, és felfelé torzított lesz a rossz iskolák esetében. Természetesen, ha a képességek szerinti csoportosulás ennek ellenkezője, akkor a torzítás iránya is megváltozik; vagyis a torzítás iránya empirikus kérdés.

A következő fejezet konkrét példákat sorol fel az egyes, itt bemutatott hozzáadott érték-modellek nemzetközi felhasználásáról, illetve olyan, a gyakorlatban is felhasznált modelleket mutat be, amelyek a fentebb felsorolt modelleknek továbbfejlesztett, sokkal adatigényesebb és bonyolultabb verziói. Nyilvánvalóan mindegyik, így az egyszerűbb, gyakorlatban is felhasznált modell is különbözik valamelyest az elméleti modellektől.

Hozzáadott érték típusú mérésen alapuló modellek a gyakorlatban

Az alábbi fejezet három, gyakorlatban is alkalmazott hozzáadott érték típusú modellt mutat be. A lengyel modell módszertanilag nem nyújt újdonságot, de érdekes látni, hogy Kelet-Európában is kísérleteznek hasonló modellekkel. A karolinai modellek a hozzáadott érték típusú iskolahatás-számítások egyik legkorábban kialakított típusa, amelyet számos tanulmányban elemeztek már. Végül a dallasi modell szintén elég korai, de a karolinai modellekkel szemben nem az egyszerűséget, hanem sokkal inkább a hiteles iskolahatás-becslést tekinti elsődleges célnak.

Lengyelország

Lengyelországban az elszámoltathatósági rendszer kialakítása még gyerekipőben jár, de *Jakubowski* (2008) leírása alapján már szépen felvázolható a rendszerhez felhasznált iskolai hozzáadott érték típusú eredményességmérés. A felhasznált modell egy az egyben a „kiigazított” egyéni hibák iskolai átlagán alapuló hozzáadott érték becslésére épül (*lásd (6)-os és (7)-es egyenleteket*). A kimeneti változó a diák által az alsó-középfokú (*gimnazjum*, azaz a 9. osztály vége) képzési szint befejeztével kötelezően megírt kimeneti teszt eredménye, míg a $t-1$ -edik időpontban mért teljesítmény a 6. osztály után szintén kötelezően megírt teszt. Míg a 9. évfolyam utáni teszt „éles”, azaz eredménye beszámít a középfokú továbbtanulásba, addig a 6. osztály utáni tesztek pusztán informatív jellegűek. Utóbbi eredményei csak az adott diák számára, illetve aggregált formában elérhetőek, és az eredményeknek inkább visszajelzés funkciója van. Bár a közeljövőben a döntéshozók szándékában

áll iskolai és egyéni változók kontrollálása is a modellben, jelenleg sem megfelelő részletezettségű iskolai jellemzők, sem egyéni szintű családi státuszjellemzők nem állnak a kutatók rendelkezésére. A *Jakubowski* (2008) tanulmányában bemutatott modell így csupán a nemet és egy diszlexiát mutató bináris változót tartalmaz egyéni szintű kontrollként.

E modell kiválasztását három érveléssel támasztja alá az erre kijelölt terület (*Jakubowski*, 2008: p. 5). A három általános érv, hogy a modell teoretikusan megállja a helyét: (1) egyszerű megérteni, (2) semleges, vagyis az iskola eredményessége független az oda járó tanulók eredeti pontszámaitól, képességeitől, és (3) végül egyszerű a kivitelezése. *Jakubowski* (2008) tanulmánya mindemellett azzal érvel, hogy a kiválasztott modell eredményeiben nem térne el jelentősen egy nem lineáris, vagy egy fix vagy random hatást felhasználó modelltől, viszont felhívja a figyelmet *Ladd és Walsh* (2002) fentebb említett kritikájára, miszerint a pontszámokat jelentős hibával mérik, így szükséges a korrekciójuk például egy instrumentális becslés segítségével.

Észak és Dél Karolina (Egyesült Államok)

Az iskolák elszámoltathatóságára Észak- és Dél-Karolinában is a legegyszerűbb, egyéni hibák iskolai átlagán alapuló „kiigazított” (Dél-Karolina), illetve „fejlődés” (Észak-Karolina) típusú elszámoltathatósági modelleket alkalmaznak (*Clotfelter és Ladd*, 1996; *Ladd és Walsh*, 2002; illetve magyarul *Muraközy és Horn*, 2005).

A) Dél-Karolinában a (6)-os modell egy változatát becslik:

$$(17) \quad y_{ijt} = a_0 + a_1 y_{ij(t-1)}^m + a_2 y_{ij(t-1)}^0 + a_3 (y_{ij(t-1)}^m)^2 + a_4 (y_{ij(t-1)}^0)^2 + a_5 (y_{ij(t-1)}^m * y_{ij(t-1)}^0) + e_{ij},$$

$$(18) \quad HE_{DK} = \hat{e}_{ij},$$

ahol y^m a matematika, míg y^0 az olvasás-szövegértés pontszám, a pedig a becsült koefficiensek.

Az alapmodellből a dél-karolinai modell két fontos tulajdonságában tér el. Egyrészt közvetlenül nem veszi figyelembe sem az egyéni családi háttér jellemzőit, sem az iskola egyéb tulajdonságait, másrészt nem lineáris hatást feltételez a jelenbeli és a múltbeli teszteredmények között. Az egyéni hatás kontrollálása ellen az szól, hogy ha figyelembe veszik, akkor explicit módon is kisebb teljesítményt várnak el az alacsonyabb státuszú tanulóktól, mint a magasabb státuszúaktól. Így az egyéni

jellemzők kontrollálása (és különösen az etnikai hovatartozásé) politikailag igen kényes kérdés. A nem lineáris hatás használatát azzal indokolják a modellt alkalmazók, hogy a jobb tanulók empirikusan igazolhatóan gyorsabban is tanulnak, azaz a múltbeli iskolai teljesítmény hatása a jelenbeli teljesítményre nem lineáris. Másként fogalmazva, a korábbi képességnek nem feltétlenül ugyanaz a hatása a jelenbeli teljesítményre, vagyis azt teszik fel, hogy a jobban teljesítő tanulók még jobban, míg a rosszabbul teljesítők relatíve még rosszabbul fognak teljesíteni.

Mivel ez a becslés a jobb háttérű tanulókat oktató iskolákat előnyben részesíti, a dél-karolinai döntéshozók az iskolákat öt kvantilisre bontották a tanulók családi háttere alapján. Az iskolákat a kvantiliseken belül hasonlítják össze a hozzáadott értékük alapján. A torzítás logikája a következő: mivel az egyéni teljesítményeket hibával lehet csak becsülni, az a regresszió, amely a t időpontban becsült teljesítményeket a $t-1$ időpontban becsült teljesítményekkel közelíti, kétszeresen torzított ugyanabban az irányban, ami a konfidencia intervallumokat a nulla felé torzítja. Minél nagyobb a mérési hiba, annál nagyobb a torzítás, és annál inkább fog a becsült hozzáadott érték a független változóként felhasznált t időpontban mért teljesítményekkel korrelálni, így a becsült hozzáadott érték annál inkább fog korrelálni az egyéni családi háttérrel is (*Clotfelter és Ladd*, 1996).

Az iskola hozzáadott értékének a dél-karolinai modellben is e_j -t feleltetik meg, azzal a különbséggel, hogy nem az egyéni hibák átlagát, hanem annak iskolai mediánját használják. A hivatalos érvelés szerint a medián használata kiküszöböli azt a nem kívánatos ösztönzést, hogy a tanárok esetleg a jó tanulók eredményeinek további javításával emeljék az átlagot a rosszabb tanulók rovására. A kritikusok kiemelik, hogy a medián használata viszont a „széleket vágja le”, azaz arra ösztönzi a tanárokat, hogy a középben lévő tanulókkal foglalkozzon, s így erőforrásokat csoportosítson át a leszakadó vagy a kiemelkedő tanulóktól.

B) Az észak-karolinai modell a (8)-as alapmodell egy verziója:

$$(19) \quad (y_{ijt} - y_{ij(t-1)}) = a_0 + a_1 Y1_{ij(t-1)} + bY2_{ij(t-1)} + e_{ij},$$

$$(20) \quad HE_{EK} = \hat{e}_{ij}.$$

ahol $Y1$ a diák előző évi teljesítményének a szintje (*proficiency level*) és $Y2$ az adott diák tesztpontszámának az állami átlagtól való eltérése.

Az $Y1$ -re való kontrollálást azzal indokolják a modell megalkotói és a döntéshozók, hogy a magasabb teljesítményszinten lévő tanulók gyorsabban is fejlődnek,

míg az Y_2 az átlaghoz való konvergálást (*regression to the mean*)¹⁰ hivatott korrigálni. Hasonlóan a dél-karolinai modellhez, a családi háterre itt sem kontrollálnak közvetlenül a modellben (*Ladd és Walsh, 2002*).

Dallas (Texas, Egyesült Államok)

A dallasai modell talán az egyik legszofisztikáltabb az iskolák elszámoltathatóságára gyakorlatban is alkalmazott modellek közül (*Clotfelter és Ladd, 1996; Ladd és Walsh, 2002; OECD, 2008; Webster és Mendro, 1997*).

A dallasai rendszer alapja egy kétlépcsős, „kiigazított” random iskolahatás-becslés, ahol az első lépcsőben (21) a nyers pontszámokat számos egyéni és iskolai/körzet szintű jellemzőtől „megtisztítják”:

$$(21) \quad y_{ijt} = a_0 + b_1 X_{1ijt} + \dots + b_p X_{pijt} + e_j.$$

Az X itt a gyerek nemét, etnikai hovatartozását, anyagi helyzetét, ezek első és második tagú interakcióit – azaz a változók egymásra gyakorolt kereszt hatásait (szorzatait) – és néhány iskolakörzeti változót jelent.

A második lépcsőben az első lépcső becslőt hibáját standardizált formában használják fel (legyen \sim a standardizálás jele). A második lépcső egy többszintű, random hatás-becslés, ahol nem csupán a diákok múltbeli teljesítményére kontrollálnak (2 évre visszamenően), hanem további iskolai szintű jellemzőket is bevonnak az egyenletbe:

$$(a) : \tilde{e}_{ijt} = c_{0j} + c_{1j} y_{ij(t-1)} + c_{2j} y_{ij(t-2)} + \delta_{ij}$$

illetve

$$(22) \quad (b) : c_{0j} = G_{00} + \sum_{k=1}^m G_{0k} W_{0k} + u_{0j}$$

$$(c) : c_{1j} = G_{10} + \sum_{k=1}^m G_{1k} W_{1k}$$

$$(d) : c_{2j} = G_{20} + \sum_{k=1}^m G_{2k} W_{2k}$$

¹⁰ Átlaghoz való konvergálásnak azt az empirikus megfigyelést hívjuk, hogy az első időszakban jól teljesítők a második időszakban általában relatíve rosszabbul, míg a rosszul teljesítők relatíve jobban teljesítenek. Ennek számos oka lehet, a mérési hibától kezdve, az egyéni pszichikai („múltkor jól ment, most nem figyelek annyira”) okokig bezárólag.

ahol c , illetve G becslött koefficienseknek felelnek meg, míg W m darab olyan iskolai jellemző, ami az iskola intézkedéseitől független (*exogén*), mint például az iskola néhány demográfiai jellemzője, az iskolába járó diákok társadalmi státuszára vonatkozó változók vagy az iskolába járó diákok fluktuációja vagy az iskolai túlsúlyosság mutatója. A számított iskolahatás így az u_{0j} lesz:

$$(23) \quad H\hat{E}_{Dallas} = u_{0j}.$$

A dallasi modell felfogásában szinte gyökeres ellentéte a tervezett lengyel modellnek. Eredményeit nehéz még egy képzett szakember számára is megérteni, bonyolult és adatigényes a kivitelezhetősége, viszont feltehetőleg jelenleg a politikailag legkorrektebb modell, amely a legkevésbé enged teret a módszertani hibáknak, kontrollál a legtöbb ismert iskolák döntésein kívül eső tényezőre és még alkalmas a gyakorlati politikai következtetések levonására is.

Országos kompetenciamérés¹¹

Jelenleg Magyarországon minden iskola, minden telephely és iskolafenntartó évente kap egy rövid összefoglaló értékelést a tanulók összesített teljesítményéről. Ezek a jelentések mindig az Országos kompetenciamérés (OKM) előző évi adatain alapulnak. A kompetenciamérés több oktatáspolitikai célt is szolgál. Egyrészt a központi kormányzatot hivatott tájékoztatni az országos, regionális és helyi oktatás színvonaláról, illetve egyéb jellemzőiről. Másrészt az iskolák, telephelyek vezetőit és az iskolafenntartókat tájékoztatja az egyes iskolák teljesítményéről, az előző év adatai alapján. Harmadrészt adalékul szolgál a tanároknak a mérési-értékelési gyakorlatuk fejlesztéséhez, végül a szülők is tájékozódhatnak a gyermekük és az iskolák eredményéről.

Az Országos kompetenciamérést 2001-ben kezdte el megvalósítani a központi oktatási kormányzat, azóta összesen 6 alkalommal mérték fel teljes körűen a 6., 8. és 10. évfolyamos tanulók olvasási és matematikai műveltségét (lásd 2. táblázat).

¹¹ Az OKM felméréséről lásd bővebben *Hermann és Molnár (2008)*, valamint *Balázsi és munkatársai (2014)* tanulmányát, illetve e kötetben *Ostorics László* írását.

2. táblázat

Az Országos kompetenciamérés adatbázisainak néhány jellemzője

Év	6. évfolyam	8. évfolyam	10. évfolyam
2003	iskolánként 20 fő	–	iskolánként és képzési típusonként 20 fő
2004	iskolánként 20 fő	iskolánként 20 fő	iskolánként és képzési típusonként 20 fő
2006	195 országosan reprezentatív intézményben teljes körű	teljes körű	minden iskola minden telephelyéről képzési formánként 30 fő
2007	200 országosan reprezentatív intézményben teljes körű	teljes körű	minden iskola minden telephelyéről képzési formánként 30 fő
2008*	teljes körű	teljes körű	teljes körű
2009*	teljes körű	teljes körű	teljes körű
2010*	teljes körű	teljes körű	teljes körű
2011*	teljes körű	teljes körű	teljes körű
2012*	teljes körű	teljes körű	teljes körű
2013*	teljes körű	teljes körű	teljes körű
2014*	teljes körű	teljes körű	teljes körű

*Oktatási azonosítón alapuló egyéni mérési azonosítók alapján felvett mérések.

A kompetenciamérés, azon kívül, hogy egy kidolgozott, több órás kérdőívvel felméri a tanulók olvasási és matematikai műveltségét, tartalmaz egy tanulói háttérkérdőívet, illetve egy telephelyi és egy iskolai szintű kérdőívet is. A tanulói kérdőív a tanulók, illetve a szüleik legfontosabb szociokulturális és gazdasági jellemzőit méri fel, a telephelyi és az iskolai kérdőívből készült adatbázisok pedig az adott telephelyek és iskolák legfontosabb erőforrás-ellátottságáról tartalmaznak információkat.

Az iskolajelentések ezen adatokon alapulnak. Az úgynevezett FIT (Fenntartói, Intézményi, Telephelyi) jelentések 2006 óta iskolánként szabadon hozzáférhetőek minden érdeklődő számára.¹² A jelentések alapvetően 3 típusú információt tartalmaznak:

1. megmutatják az iskolák átlageredményét mindkét mért műveltségterületen (olvasás-szövegértés és matematika),
2. megmutatják az iskolák tanulóinak képességeloszlását (azaz, hogy a tanulók hány százaléka van az egyes képességszinteken), illetve

¹² <https://www.kir.hu/okmfit>

3. ezeket az adatokat összevetik az adott településtípus és a képzés jellege szerinti országos átlaggal.

Az iskolai jelentések ezeken az információkon túl megmutatják az iskola elvárható eredményét is. Az elvárható eredményt a következő egyenletből becslik:

$$(24) \quad y_{ij} = a_0 + bX + e_{ij},$$

ahol i a diákot, j az iskolát, y a diák mért teljesítményét jelöli, míg X az egyéni és iskolai jellemzők (a kontrollváltozók) vektora. Az X vektor tartalmazza a tanulók családi háttérindexét (CSHI), illetve közvetve az iskola típusát és településének típusát, amennyiben ezeket összeveti az adott képzési-, illetve településtípusú intézmények eredményeivel.¹³ Az a_0 és a_1 , illetve b vektor becsült koefficiensek, míg e az egyéni hiba jele.

Az elvárható eredményt (\hat{y}_{ij}) az iskolajelentés összeveti a tényleges eredménnyel (y_{ij}), és bár a jelentés közvetlenül nem számolja ki az elvárt eredménytől való eltérést, vagy iskolahatást, a tanulmány további részében ezt vetem össze a hozzáadott érték típusú mutatókkal:¹⁴

$$(25) \quad IH = \text{átlag}_j (y_{ij} - \hat{y}_{ij}) = e_j.$$

Hozzáadott érték típusú mutatók az OKM-adatokon

A tanulmányban felsorolt elméleti modelleket a magyar Országos kompetenciamérés adatain is ki lehet számolni. Az alábbiakban a legegyszerűbb alapmodelleket vetem össze egymással (lásd 3. táblázat).

¹³ Csak abban az esetben számol a jelentés elvárt eredményt, ha a tanulók legalább kétharmadának van érvényes családi háttérindexe. A telephelyi jelentés az átlageredmények, a képességeloszlás és az elvárt eredmények mellett az átlageredmények változását is megmutatja az előző évi eredményekhez viszonyítva.

¹⁴ A tanulmányban használt iskolahatás a tanulók tényleges és elvárt eredményének iskolai átlaga (vagyis az egyéni hibát felhasználó modell), míg a jelentésekben a tanulók tényleges eredményének iskolai átlaga és az elvárt eredményének iskolai átlaga van megadva; vagyis ez utóbbi nem egyezik meg pontosan a tanulmány által használt, keresztmetszeti adatokon becsült iskolahatással. Ennek ellenére úgy gondolom, hogy a tanulmány által használt mutató jobban összevethető a többi, az irodalomban megjelenő hozzáadott érték típusú mutatóval.

3. táblázat
A különböző modellek rövidítései

Rövidítés	Iskolahatás
Keresztmetszeti adatokon becült mutatók	
IHe*	Keresztmetszeti adatokon becült iskolahatás (tényleges – CSHI alapján elvárt eredmény), egyéni hiba
IHi	Keresztmetszeti adatokon becült iskolahatás (tényleges – CSHI alapján elvárt eredmény), iskolai random hatás
IHf	Keresztmetszeti adatokon becült iskolahatás (tényleges – CSHI alapján elvárt eredmény), iskolai fix hatás
Hozzáadott érték típusú mutatók	
HEe1	Egyéni hiba, „kiigazított”
HEi1	Iskolai random hatás, „kiigazított”
HEf1	Fix hatás, „kiigazított”
HEe2	Egyéni hiba, „fejlődés”
HEi2	Iskolai random hatás, „fejlődés”
HEf2	Fix hatás, „fejlődés”
HEe3	Egyéni hiba, „kombinált”
HEi3	Iskolai random hatás, „kombinált”
HEf3	Fix hatás, „kombinált”

*A FIT-jelentésekben számolt elvárható eredményhez leginkább hasonló mutató.

Jelen számítás csupán egy személtetése annak, hogy mennyiben befolyásolja az egyes iskola értékelését a módszertan. Fontos hangsúlyozni, hogy az elméleti modellek egyike sem jobb vagy rosszabb, mint a másik. Különbség köztük csupán az alapfeltevésekben van. A modellek az előző időszak(ok)ban mért tesztpontszámok mellett csupán a családi háttérindexre kontrollálnak¹⁵: nem veszik figyelembe a tanulók egyéb jellemzőit, az iskola típusát vagy egyéb iskolai jellemzőket, vagyis semmiképp nem tekinthetők végleges eredményességmutatóknak.

A 2012-es évre matematikából és szövegértésből számolt páronkénti korrelációs együtthatókat vizsgálva jól látható, hogy a keresztmetszeti adatokon alapuló elvárt eredmény és valós eredmény különbsége (IHe, IHi, IHf) leginkább a „kiigazított” becslésekkel (IHe1, IHi1, IHf1) korrelál. A korrelációs együttható minden esetben 0,7 felett van, de egyes mutatók között nagyon szoros (0,96) együtt járás is tapasztalható. Vagyis a FIT-jelentésekben megjelenő iskolahatás-mutatók (a családi háttérindexre kontrolláló modellek) nem térnek el alapjaiban a FIT-jelentések „komplex modelljeitől”. A FIT-jelentések komplex modelljei megfeleltethetők az itt bemutatott

¹⁵ A keresztmetszeti adatokon becült modellek (IHe, IHi, IHf) csak a családi háttérindexre kontrollálnak.

„kiigazított”, egyéni hibát használó modelleknek. Ez nem jelenti azt, hogy egy-egy iskolára nézve ne adna más eredményt az egyéni hibát használó iskolahatás (IHe) vagy a „kiigazított” egyéni hibát használó (HEe1) modell (lásd 4. táblázat).

4. táblázat

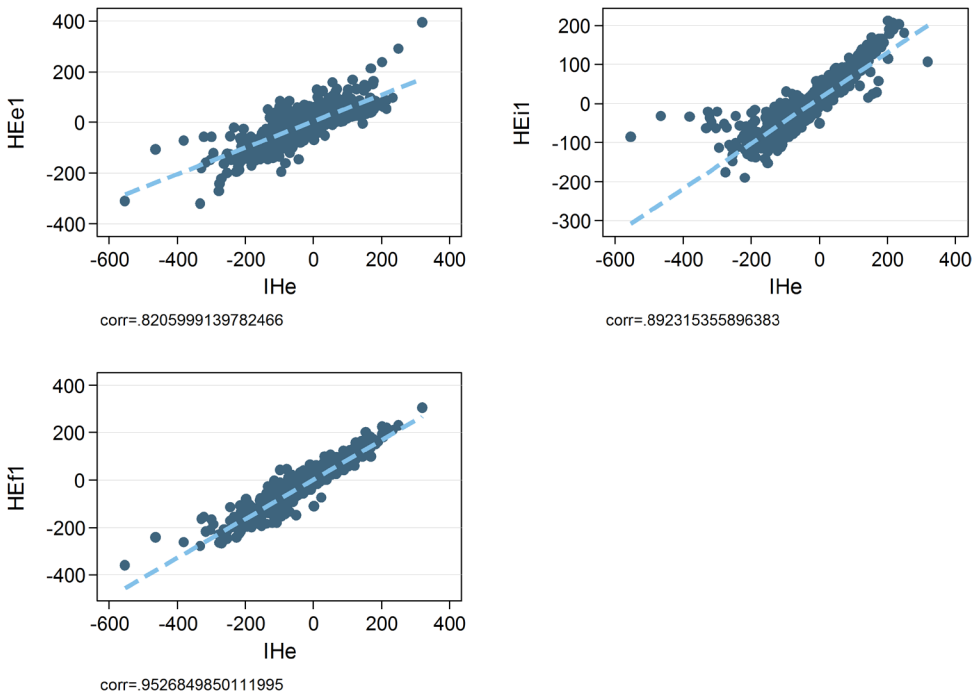
Páronkénti korrelációs együtthatók az iskolahatások egyes mutatói között (2012)

Matematika	IHe	IHi	IHf	HEe1	HEi1	HEf1	HEe2	HEi2	HEf2	HEe3	HEi3	HEf3
IHe	1,00	-	-	-	-	-	-	-	-	-	-	-
IHi	0,91	1,00	-	-	-	-	-	-	-	-	-	-
IHf	0,95	0,97	1,00	-	-	-	-	-	-	-	-	-
HEe1	0,82	0,70	0,75	1,00	-	-	-	-	-	-	-	-
HEi1	0,89	0,96	0,92	0,76	1,00	-	-	-	-	-	-	-
HEf1	0,95	0,93	0,96	0,84	0,95	1,00	-	-	-	-	-	-
HEe2	0,48	0,37	0,41	0,89	0,47	0,54	1,00	-	-	-	-	-
HEi2	0,49	0,45	0,45	0,84	0,56	0,56	0,91	1,00	-	-	-	-
HEf2	0,50	0,42	0,46	0,90	0,50	0,57	1,00	0,92	1,00	-	-	-
HEe3	0,50	0,39	0,44	0,90	0,48	0,56	0,98	0,90	0,98	1,00	-	-
HEi3	0,52	0,48	0,48	0,85	0,58	0,59	0,90	0,99	0,91	0,91	1,00	-
HEf3	0,54	0,45	0,50	0,91	0,53	0,60	0,98	0,91	0,98	1,00	0,92	1,00
Szövegértés	IHe	IHi	IHf	HEe1	HEi1	HEf1	HEe2	HEi2	HEf2	HEe3	HEi3	HEf3
IHe	1,00	-	-	-	-	-	-	-	-	-	-	-
IHi	0,91	1,00	-	-	-	-	-	-	-	-	-	-
IHf	0,95	0,96	1,00	-	-	-	-	-	-	-	-	-
HEe1	0,83	0,71	0,77	1,00	-	-	-	-	-	-	-	-
HEi1	0,87	0,96	0,92	0,74	1,00	-	-	-	-	-	-	-
HEf1	0,92	0,91	0,96	0,82	0,94	1,00	-	-	-	-	-	-
HEe2	0,47	0,36	0,41	0,88	0,43	0,51	1,00	-	-	-	-	-
HEi2	0,44	0,41	0,40	0,78	0,51	0,48	0,86	1,00	-	-	-	-
HEf2	0,49	0,40	0,45	0,89	0,47	0,55	1,00	0,86	1,00	-	-	-
HEe3	0,47	0,37	0,42	0,87	0,44	0,53	0,98	0,84	0,98	1,00	-	-
HEi3	0,37	0,32	0,32	0,74	0,43	0,41	0,86	0,98	0,86	0,86	1,00	-
HEf3	0,42	0,32	0,37	0,84	0,40	0,49	0,97	0,82	0,97	0,99	0,85	1,00

Az 1. ábra bal felső ábráján jól látható, hogy egyes iskolák jóval az azonosságot mutató szaggatott vonal alatt vagy felett helyezkednek el. Minél távolabb esik egy iskola a szaggatott vonaltól, annál jobban eltér a két mutatóból levonható következtetés az adott iskola esetén.

1. ábra

A „kiigazított” egyéni és iskolai random hatás, illetve a fix hatás-modellek összevetése az egyéni hibát felhasználó, keresztmetszeti adatokon becsült iskolahatás-mutatókkal (OKM 2012/10. évf.–2010/8. évf., matematika)

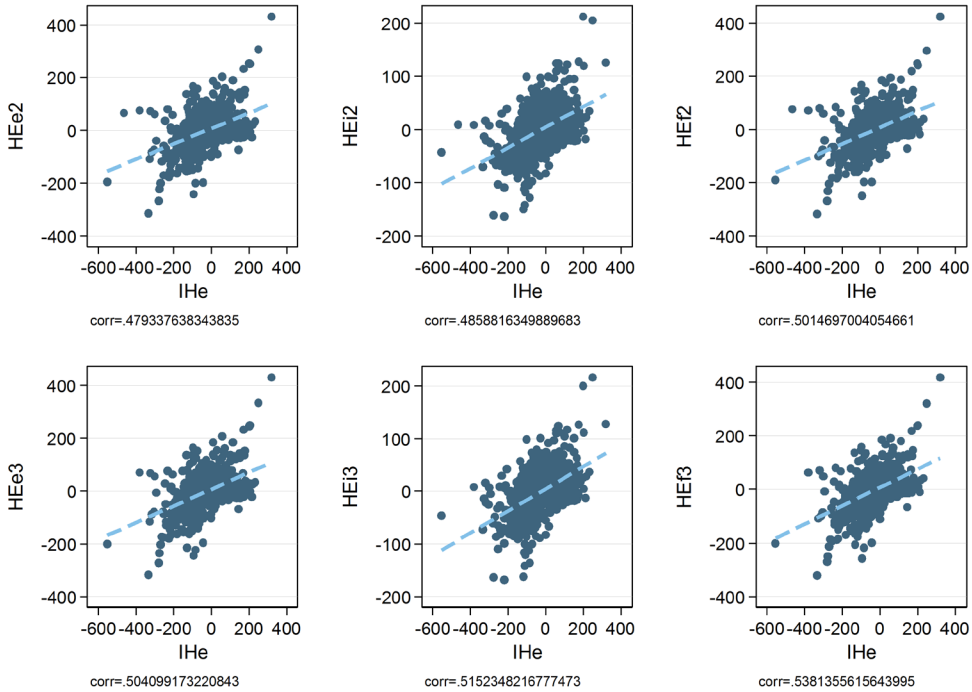


A 4. táblázatból az is leolvasható, hogy a FIT-jelentések által közölt elvárható értékhez leginkább hasonló egyéni hibát használó, keresztmetszeti adatokon becsült iskolahatás-mutatók együtt járása már korántsem olyan szoros, ha a „fejlődés” vagy a „kombinált” hozzáadott érték típusú modellekkel vetjük őket össze. A korrelációs együtthatók itt szinte sehol nem haladják meg a 0,5-ös értéket. Ráadásul, ha nem is az egyéni hibát használjuk a keresztmetszeti adatokon becsült iskolahatás kiszámításához, hanem iskolai random hatást (IHi), és ezt vetjük össze a szintén iskolai random hatás modellel kiszámolt „kombinált” modellel (HEi3), a korrelációs együttható 0,32-re csökken a szövegértés esetén. Tekintettel arra, hogy mindkét mutató azzal a céllal jött létre, hogy az iskolák eredményességéről adjon tájékozta-

tást, ez az együtt járás kifejezetten gyengének mondható. Jól látható, hogy számos iskola jóval a szaggatott vonal alatt vagy felett van (lásd 2. ábra). Ezen iskolákra nézve különösen nagy tétje lenne annak, ha az egyik vagy a másik számítási módszer szerint lennének értékelve.

2. ábra

Az „fejlődés”, illetve a „kombinált” egyéni és iskolai random hatás, illetve a fix hatás-modellek összevetése az egyéni hibát felhasználó, keresztmetszeti adatokon becsült iskolahatás-mutatókkal (OKM 2012/10. évf.–2010/8. évf., matematika)

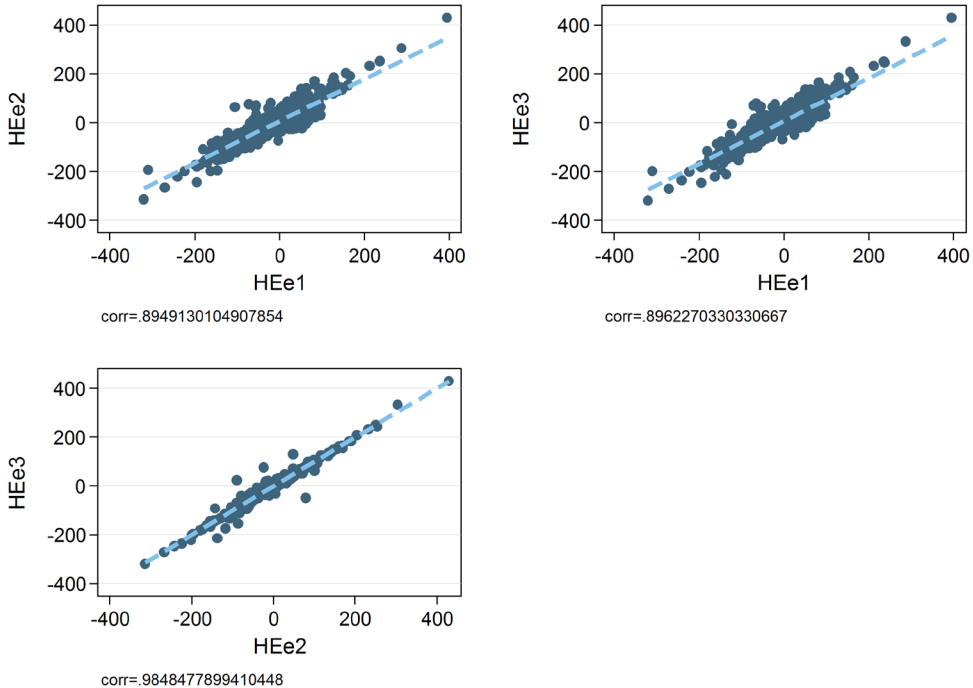


A 3. ábrán az egyes hozzáadott érték típusú – vagyis a korábbi tesztpontszámokra és a családi háttérindexre is kontrolláló – mutatók, a „kiigazított”, a „fejlődés” és a „kombinált” modellek egyéni hibát felhasználó eredményei vannak összevetve egymással. Jól látható, hogy önmagában az, hogy milyen modellt feltételezünk, nem határozza meg jelentősen az egyes iskolák helyzetét: a korrelációk együttthatók mindenütt 0,9 körüliek, vagy e feletti. Vagyis úgy tűnik, a másik dimenzió sokkal inkább meghatározó: az, hogy egyéni hiba, iskolai random hatás vagy iskolai fix hatás-modellekkel számolják az iskolák eredményességét, nagyobb mértékben meghatározza az egyes iskolák helyzetét, mint az, hogy a múltbeli

teszteredmény(ek)e)t független változóként („*kiigazított modell*”), függő változóként („*fejlődés modell*”) vagy mindkét helyen („*kombinált modell*”) figyelembe veszik.

3. ábra

A „kiigazított”, „fejlődés” és „komplex” hozzáadott érték modellek egyéni hibát felhasználó eredményein alapuló mutatók összevetése (OKM 2012/10. évf.–2010/8. évf., matematika)



Az eredmények azonban még egy dimenzióban óvatosságra intenek: az egyes módszerek által számított eredményességmutatók jelentősen eltérhetnek egymástól az évek között. Jól látható, hogy a 2012. és a 2011. évre ugyanolyan módon számolt mutatók korrelációs együtthatói 0,5–0,8 között mozognak (lásd 5. táblázat).

5. táblázat

Páronkénti korrelációs együtthatók a „kiigazított” hozzáadott érték modellek között 2011-ben és 2012-ben*

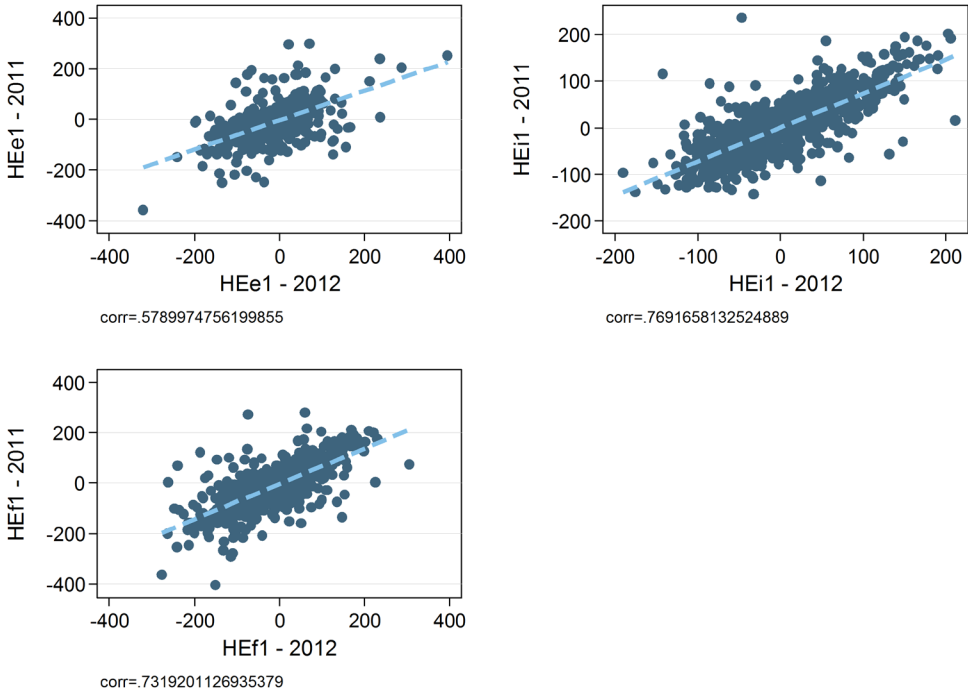
Matematika		2011			2012		
		HEe1	HEr1	HEf1	HEe1	HEr1	HEf1
2011	HEe1	1,00	–	–	–	–	–
	HEr1	0,79	1,00	–	–	–	–
	HEf1	0,87	0,95	1,00	–	–	–
2012	HEe1	0,58	0,54	0,57	1,00	–	–
	HEr1	0,53	0,77	0,71	0,76	1,00	–
	HEf1	0,58	0,74	0,73	0,84	0,95	1,00
Szövegértés		2011			2012		
		HEe1	HEr1	HEf1	HEe1	HEr1	HEf1
2011	HEe1	1,00	–	–	–	–	–
	HEr1	0,73	1,00	–	–	–	–
	HEf1	0,80	0,94	1,00	–	–	–
2012	HEe1	0,46	0,52	0,52	1,00	–	–
	HEr1	0,50	0,82	0,76	0,74	1,00	–
	HEf1	0,49	0,76	0,75	0,82	0,94	1,00

*A félkövérrel jelölt számok a különböző évek adatain, azonos módon számított mutatókat jelölik.

Mivel nem valószínű, hogy az egyes iskolák *valódi* eredményessége két egymást követő évben nagymértékben változna, így feltételezhető, hogy a hozzáadott érték típusú mutatók sem teljesen függetlenek olyan egyéb hatásoktól, mint amilyen például a kohorsz-hatás, hiszen az egyetlen fontos eltérés a két időpontra számolt mutatók között, hogy más tanulók szerepelnek benne. Az egyikben azok, akik 2012-ben voltak 10. évfolyamosok, míg a másikban azok, akik egy évvel korábban, 2011-ben. Ennek megfelelően az egyes évek között feltehetően éppen a kohorsz-hatás miatt jelentősen változik/változhat egy-egy iskola hozzáadott érték mutatóval mért eredményessége (lásd 4. ábra).

4. ábra

A különböző „kiigazított” hozzáadott érték típusú mutatók összevetése 2011 és 2012 között



Ezekből a megfigyelésekből azt a következtetést lehet levonni, hogy érdemes az iskolák eredményességét nem egy év, hanem több év összevont eredményessége alapján vizsgálni. Hogy ez az összevonás miként történik, feltehetően szintén jelentősen befolyásol(hat)ja az egyes iskolák hosszabb távú megítélését.

Összefoglalás

A tanulmány az elméletileg lehetséges hozzáadott típusú indikátorok alapverzióit mutatja be, és nemzetközi példákat is hoz ezekre. A tanulmány kísérletet tesz arra is, hogy magyar adatokon megvizsgálja, valóban más eredményeket kapnának-e az iskolák, ha nem keresztmetszeti, hanem több időpontban mért, tanulói szinten összekötött adatokon becsült, hozzáadott érték típusú eredményességi mutatókat kapnának kézhez. Fontos hangsúlyozni, hogy a kutatás egy *pilot*, amely nem alkalmas iskolai szintű következtetések levonására, de az oktatáspolitikai döntéshozók számára kulcsfontosságú információkkal szolgálhat.

A tanulmány két fontos dologra hívja fel a figyelmet: egyrészt arra, hogy a keresztmetszeti adatokon nyugvó statisztikák nagy valószínűséggel torzított képet fognak nyújtani egy adott iskola valós eredményességéről. E torzítás mértékének a csökkentésére lehetőség van hozzáadott érték típusú indikátorok figyelembevételével. A tanulmány másik fontos állítása azonban az, hogy a hozzáadott érték típusú mutatók sem tökéletesek. Egyrészt nem oldják meg, csak csökkentik a mérésből fakadó bizonytalanságot, másrészt nem létezik egy tökéletes „hozzáadott érték mutató”, hanem sokféle, különböző feltevéseken alapuló *hozzáadott érték típusú mutató* létezik, amelyek adott esetben más és más eredményeket mutatnak egy adott iskolára nézve.

A tanulmány fő konklúziója, hogy gyakorlatban mindegyik modell használható, sőt eredményeiben is hasonlóak egymáshoz, azonban a becslések lényegesen különbözhetnek az egyes iskolára vonatkoztatva. Még a legegyszerűbb hozzáadott érték típusú modellek is jelentősen megváltoztatnák az egyes iskolák eredményességi mutatóit, ha azonban kicsit bonyolultabb, ám az iskolai munka értékelése szempontjából korrektebb mutatókat alkalmaznánk, az eredmények több esetben drasztikusan eltérnének. Mindezek mellett – bár a hozzáadott érték típusú mutatók közti azonos éven belüli korreláció sem tökéletes –, ha két különböző évet hasonlítunk össze, akkor az eltérés már jelentős. Így érdemes egyrészt többféle módszerrel, másrészt több évre kiszámolni az iskolák eredményességét, s így talán egy lépéssel közelebb kerülünk az iskolák valós eredményességének becsléséhez.

Hivatkozások

- Balázi Ildikó és mtsai. (2014): *Az Országos kompetenciamérés tartalmi keretei – szövegértés, matematika, háttérkérdőívek*. Oktatási Hivatal, Budapest.
- Clotfelter, C. T. és H. F. Ladd (1996): Recognizing and Rewarding Success in Public Schools. In: H. F. Ladd (Ed.): *Holding Schools Accountable: Performance-based Reform in Education* (pp. 23–63), The Brookings Institution, Washington D. C.
- Dolton, P. J. (2002). Improving Educational Quality: How Best to Evaluate Our Schools? (Discussion). In Y. K. Kodryzcki (Ed.): *Education in the 21st Century* (pp. 225–235), Federal Reserve Bank of Boston.
- Hermann Zoltan és Molnár Tímea Laura (2008): *Országos kompetenciamérési adatbázis*. Elérhető: <http://adatbank.mtaki.hu/files/dokum/7.pdf> [Letöltés dátuma: 2015.04.30.]
- Horn Dániel (2011): Az oktatási elszámoltathatósági rendszerek elmélete. *MTA-KTI Műhelytanulmányok*, no. 13.
- Jakubowski, M. (2008): Implementing Value-Added Models of Schools Assessment. *EUI Working Papers RSCAS*, no. 6.
- Kane, T. J. és Douglas, O. S. (2002): The Promise and Pitfalls of Using Imprecise School Accountability Measures. *The Journal of Economic Perspectives* 16 (November), 91–114.

- Ladd, H. F. és Randall, P. W. (2002): Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right. *Economics of Education Review*, 21(1), 1–17.
- McCaffrey, D. F. és mtsai. (2003): *Evaluating Value-Added Models for Teacher Accountability. Monograph*. RAND Corporation.
- Muraközy Balázs és Horn Dániel (2005): Teljesítményalapú ösztönzőrendszerek a közoktatásban. In: Hermann Zoltán (szerk.): *Hatékonyági problémák a közoktatásban* (pp.165–185), Országos Közoktatási Intézet, Budapest.
- OECD (2008): *Measuring Improvements in Learning Outcomes – Best Practices to Assess the Value-Added of Schools*. OECD Publishing, Paris.
- Todd, P. E. és K. I., Wolpin (2003): On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, 113(485), 3–33.
- Webster, W. J. és R. R., Mendro (1997): *The Dallas Value-Added Accountability System*. Dallas Public Schools.

A tanulói teljesítménymérések megítélése és intézményi hasznosulása

A tanulói teljesítménymérések új nézőpontból világítanak rá egy-egy ország oktatási rendszerének eredményeire, s jelentősen hozzájárulnak az egyes országok közpolitikai gondolkodásának fejlődéséhez, a tényeken alapuló döntéshozatal fontosságának felismeréséhez (Szemerszki, 2014). Másik nagyon fontos hozadékuk, hogy lehetőséget teremtenek az esélyegyenlőség, az eltérő családi háttérből eredő hatások erősségének vizsgálatára, az iskolai eredmények és a szociális helyzet összefüggéseinek feltárására (Fehérvári és Széll, 2014). A tanulói teljesítménymérések tehát fontos információt biztosítanak az iskolák teljesítményéről, az oktatás minőségéről, eredményességéről a közvetlen oktatási szereplők (tanulók, szülők, pedagógusok, igazgatók), valamint a kutatók, a szakpolitikusok és a szélesebb nyilvánosság számára. A mérési eredmények ismerete az intézményi szintű irányítás egyik fontos eszközeként segítheti az intézményi szintű önértékelést és fejlesztést is.

Az oktatás eredményességének mérése, az oktatási rendszer és az iskolák teljesítményének értékelése ugyanakkor sokkal komplexebb annál, minthogy kizárólag a tanulói teljesítménymérések eredményeire támaszkodjunk. A mérési eredmények értelmezésekor, az abból levont következtetések elemzésekor érdemes megfontolni a tudományterülettel szemben felmerülő kritikákat és az alkalmazott mutatók korlátait is (Gyökös és Szemerszki, 2014). A tanulói teljesítménymérések nagyon fontos aspektust vizsgálnak, de nem feltétlenül adnak teljes képet az oktatás minőségéről, eredményességéről. Ugyanakkor meggyőződésünk, hogy a tanulói teljesítménymérések olyan rendszerszintű jelzőeszközként funkcionálhatnak, melyek segítségével a pedagógiai munka komplexitásából, az iskolák eredményességének,

értékelésének és elszámoltathatóságának összetett dimenzióiból képesek lehetünk számos fontos oktatáspolitikai kérdéskört alaposan körbejárni (lásd még e kötetben Szemerszki Marianna és Nahalka István írását).

A tanulói teljesítményvizsgálatok hazai és nemzetközi elterjedésével az oktatási intézmények vezetői és pedagógusai egyre gyakrabban találkoznak a különböző kompetenciamérésekkel és azok tanulókra és intézményekre vonatkozó eredményeivel. A magyar oktatási rendszerről, a hazai tanulási folyamatokról teljes körűen a nemzetközi viszonylatban is kiemelkedő Országos kompetenciamérés (OKM) gyűjt információkat immár több mint egy évtizede (lásd még e kötetben Ostorics László írását). A kompetenciamérés eredményei és az eredményekről szóló egyre komplexebb jelentések nyilvánosak, azonban fontos kérdés, hogy a mérési rendszer visszacsatolási és elszámoltathatósági funkciója mennyiben tekinthető kihasználatnak, illetve mennyire tudja betölteni ezen funkcióját (lásd még e kötetben Szemerszki Marianna és Horn Dániel írását).

Tanulmányunkban elsősorban arra keressük a választ, hogy a jelenlegi oktatási rendszerben milyen vélemények körvonalazódnak a tanulói teljesítménymérésekkel kapcsolatosan, miként viszonyulnak az egyes iskolák igazgatói, pedagógusai a mérésekhez, azok eredményeihez. Ennek érdekében összegezzük, hogy az egyes iskolák mennyire ismerik, illetve használják fel az egyes mérések eredményeit, hogy a közoktatási intézmények vezetői, illetve pedagógusai miként ítélik meg a tanulói teljesítményméréseket, miként látják hasznosíthatónak a mérési eredményeket, illetve mit gondolnak a mérések és az eredményesség kapcsolatáról, a méréseknek az iskolai munka értékelésében betöltött szerepéről.

Jelen elemzés egy online kérdőíves adatfelvétel adataira támaszkodik.¹ A vizsgálat járási alapú (50 járás), régió, megye és lakónépesség alapján rétegzett valószínűségi mintán gyűjtötte össze a pedagógusok és az intézményvezetők véleményeit az óvodától a középiskoláig, valamennyi oktatási szinten. A standardizált kérdőíves online adatfelvételre két időpontban is sor került: az első hullám 2013 őszén, a második pedig 2014 őszén valósult meg. Mindkét adatfelvétel külön intézményvezetői és pedagógusi kérdőívekkel történt, amelyekben az iskolai eredményességre, valamint a tanulói teljesítménymérésekre vonatkozó kérdéscsoportok is helyet kaptak.

¹ A kutatás a Társadalmi Megújulás Operatív Program XXI. századi közoktatás (fejlesztés, koordináció) II. szakasz kiemelt projekt (TÁMOP-3.1.1-11/1-2012-0001) keretein belül valósult meg.

Vizsgálatunkban a téma jellegénél fogva csak a közoktatási intézmények vezetőinek és pedagógusainak válaszait elemezzük.²

A tanulói teljesítménymérések eredményeinek intézményi szintű felhasználása

A tanulói teljesítménymérések elterjedése és hangsúlyosabbá válása következtében a mérési adatok felhasználása többé-kevésbé általánossá vált az iskolákban. Több korábbi felmérés is foglalkozott azzal, hogy az oktatási intézmények vezetői miként viszonyulnak a hazai szempontból legmeghatározóbb tanulói teljesítményméréshez, az Országos kompetenciaméréshez. Az Országos kompetenciamérés fogadtatása az oktatási intézmények vezetői körében már egy évtizede is viszonylag kedvező volt: az iskolaigazgatók több mint háromnegyede ítélte hasznosnak a kompetenciamérést mind az egyes tantárgyi területek szempontjából, mind pedig a mérési-értékelési kultúra meglapozásának szempontjából. Az iskolák többségében elemezték is az adatokat, igaz, leginkább szóban (Sinka, 2006; Horn és Sinka, 2006). Fontos azonban megjegyezni, hogy az Országos kompetenciamérés a 2001. évi indulása óta számos változáson ment át, amellet, hogy céljai is némileg változtak, illetve kiegészültek. Lényeges változás, hogy a 2007/2008. tanévtől kezdődően teljes körűvé vált a mérés, vagyis azóta lényegében valamennyi 6., 8. és 10. évfolyamos tanulóra kiterjed. Ugyanilyen fontos állomás, hogy 2010-től évfolyamfüggetlen skála áll rendelkezésre, amely lehetővé teszi a különböző évfolyamokon tanulók eredményeinek összehasonlítását, valamint, hogy a mérési azonosító bevezetésével egyéni szinten is követhetővé váltak a mérések eredményei. A mérés bevezetése óta elsődleges célként az iskolák önértékelésének és fejlesztési munkájának segítése jelenik meg, ugyanakkor az utóbbi években egyre inkább előtérbe kerül az intézményfenntartók munkájának támogatása, a külső intézményértékelés adatokkal történő ellátása, valamint információk biztosítása a szülők, a tanulók és a tényekre alapozott oktatáspolitikára számára is (lásd még e kötetben Ostorics László írását).

A 2008. évi kompetenciamérés óta az Oktatási Hivatal külön Fenntartói, Intézményi és Telephelyi jelentéseket, ún. FIT-jelentéseket készít, amelyek még inkább elősegítik az eredmények nyilvánosságát és azok idősoros elemezhetőségét. Ezzel kapcsolatosan friss adat nem áll rendelkezésre, de a 2008. és 2009. évi eredmények letöltési adatai azt mutatták, hogy az érdeklődés meglehetősen nagy volt a jelen-

² A 2013. évi adatfelvétel elemszámai: 496 igazgató, 6 943 pedagógus. A 2014. évi adatfelvétel elemszámai: 601 igazgató és 5 261 pedagógus.

tések iránt (*Balázs és Horváth, 2010*). Egy, az oktatási intézmények és tagintézmények vezetői körében 2009-ben készült felmérés szerint az intézményvezetők többsége használta a kompetenciamérés iskolai eredményeit, a FIT szoftver önálló használata azonban igencsak ritka volt, s még ennél is kevesebben készítettek teljes körű tanulói elemzést. A saját elemzések készítésében az általános iskolák és a gimnáziumok voltak a legaktívabbak, ugyanakkor a többi intézménytípusban gyakran hangzott el az a válasz, hogy a kérdés nem vonatkozik rájuk (pl. a speciális osztályok vagy a felnőttképzés miatt).³

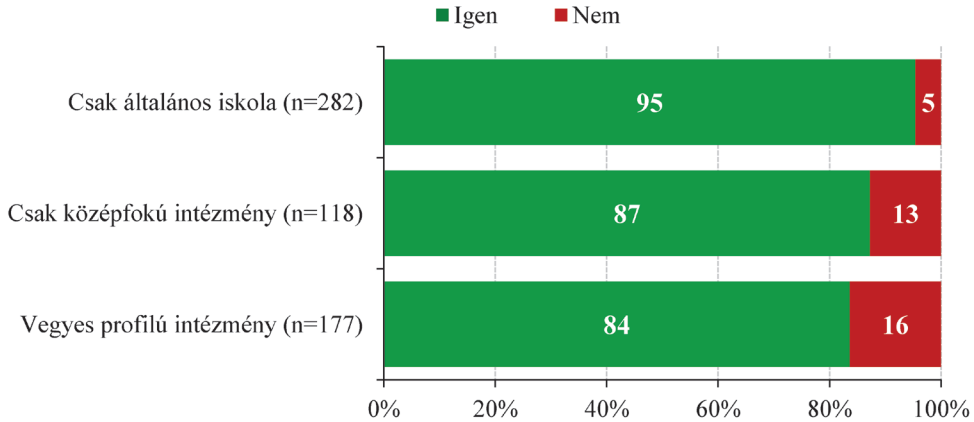
A 2014. évben felvett intézményvezetői kérdőív némiképp eltérően mérte fel a kérdéskört: egyrészt nem kizárólag az Országos kompetenciamérések, hanem általában a tanulói teljesítménymérések és vizsgák eredményeire fókuszált, másrészt arra tért ki, hogy az intézményben feldolgozzák, elemzik, értékelik-e valamilyen formában a központi mérések eredményeit, harmadrészt pedig arra, hogy a fenntartó felhasználja-e bármilyen módon az eredményeket. Az oktatási intézmények vezetőinek 90%-a válaszolt úgy, hogy a mérési eredményeket valamilyen formában a tantestületen belül közösen elemzik, megvitatják, ami nagyon hasonló érték a korábbi mérési eredményekhez. A legmagasabb arányban azokban az intézményekben élnek a leginkább ezzel a lehetőséggel, ahol csak általános iskolai képzés van, de a csak középfokú képzést nyújtó intézményekben is nagyobb arányban dolgozzák fel a mérési eredményeket, mint az olyan iskolákban, ahol általános iskolai és középiskolai szintű képzés egyaránt működik (ezeket az intézményeket a továbbiakban egyes profilú iskoláknak nevezzük) (lásd *1. ábra*). A KLIK-fenntartású, az önkormányzati és az egyéb állami intézményekben, valamint az egyházi iskolákban jóval elterjedtebb (91,8, illetve 89,1%) a mérési eredmények valamilyen szintű közös feldolgozása, mint az alapítványi, magán fenntartásban lévő iskolákban, ahol az igazgatók csupán hattizede számolt be erről.⁴ Az iskola elhelyezkedésének településtípusa szerint nem tapasztalható lényegi eltérés e tekintetben. Az eredmények értékelésével eddig különösebben nem foglalkozók, vagyis azon iskolák, akik nem értékelik, nem dolgozzák fel közösen a központi mérések eredményeit, nem is igazán érdeklődnek az ebben rejlő lehetőségek iránt, hiszen mindössze 44%-uk nyilatkozott úgy, hogy fontosnak tartaná az eredmények közös elemzését, értelmezését.

³ A felmérés 2009-ben készült az Oktatáskutató és Fejlesztő Intézet megbízásából. Az adatfelvétel során 962 iskolaigazgatót kérdeztek meg.

⁴ A 2014. évi mintában a KLIK, az önkormányzati és egyéb állami fenntartásban lévő iskolák aránya 85%, az egyházi iskoláké 11%, míg az alapítványi, magániskolák aránya 4%.

1. ábra

„Feldolgozzák, elemzik, értékelik-e közösen az Önök intézményében a központi mérések eredményeit?” – intézménytípus szerinti %-os arányok az intézményvezetők körében (2014)*



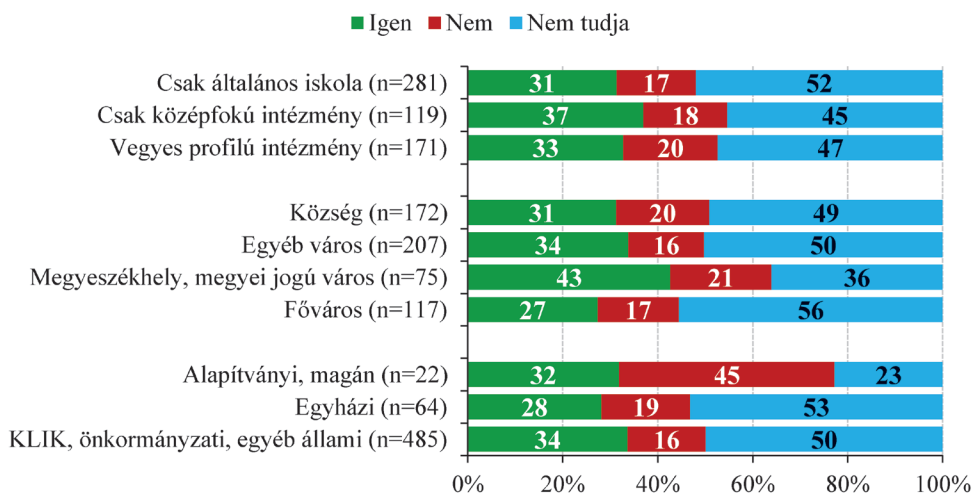
Forrás: Intézményvezetői adatfelvétel, 2014

*Szignifikáns különbség ($p < 0,001$)

2014-ben a megkérdezett közoktatási intézmények vezetőinek csaknem egyharmada volt biztos abban, hogy az elmúlt tanévben fenntartójuk használta valamilyen formában az intézményükre vonatkozó országos mérési eredményeket, közel egyötödük viszont biztos benne, hogy erre nem került sor. E tekintetben igen magas az információhiány, hiszen az intézményvezetők csaknem fele nem tudott választ adni erre a kérdésre. Az átlagosnál érezhetően gyakoribb fenntartói érdeklődést jeleztek a KLIK, önkormányzati és egyéb állami fenntartású iskolák, valamint a megyeszékhelyeken, megyei jogú városokban elhelyezkedő intézmények vezetői (lásd 2. ábra). Az információhiány egyértelműen nagyobb a csak általános iskolai képzést nyújtó, fővárosi, egyházi, illetve KLIK, önkormányzati és egyéb állami fenntartás alatt lévő intézményekben.

2. ábra

„Az elmúlt tanévben tudomása szerint az Önök iskolájának (akkori) fenntartója felhasználta-e valamilyen formában az Önök intézményének az Országos kompetenciamérésen elért eredményeit?” – intézménytípus, településtípus és fenntartó szerinti %-os arányok az intézményvezetők körében (2014)⁷



Forrás: Intézményvezetői adatfelvétel, 2014

⁷Fenntartó szerinti szignifikáns különbség ($p < 0,05$)

Az Országos kompetenciamérés adatai lehetőséget adnak arra, hogy stratégiai dokumentumok készítéséhez is használják őket, sőt, erre jelenleg is jogszabályi kötelezettségük is van az érintett intézményeknek.⁵ A kompetenciamérés eredményeit 2005-ben a legnagyobb arányban – az iskolák csaknem 40 százalékában – tantervfejlesztésre használták, de a tanítási gyakorlat gyakorlatorientáltabb feladatokkal történő gazdagítása, az eredmények tanulmányozása és megvitatása, valamint a fejlesztés és felzárkóztatás is 20% feletti arányokkal jelent meg az iskolaigazgatók válaszaiban, míg a más iskolákkal történő összehasonlítás, a szülők tájékoztatása, az iskola népszerűsítése még csekély mértékben került a felhasználás fókuszába (Sinka, 2006; Horn és Sinka, 2006). Az oktatási intézmények és tagintézmények vezetői körében 2009-ben készült felmérés szerint az intézményvezetők 79%-a számolt be a kompetenciamérési eredmények belső iskolai értelmezéséről, 75%-uk a mérések anyagainak tervezéshez, fejlesztési célok kijelöléséhez való alkalmazását, szintén 75%-uk a feladatok gyakoroltatását, míg 73%-uk a tanulók fejlesztéséhez, felzárkóztatásához történő használatot jelezte.

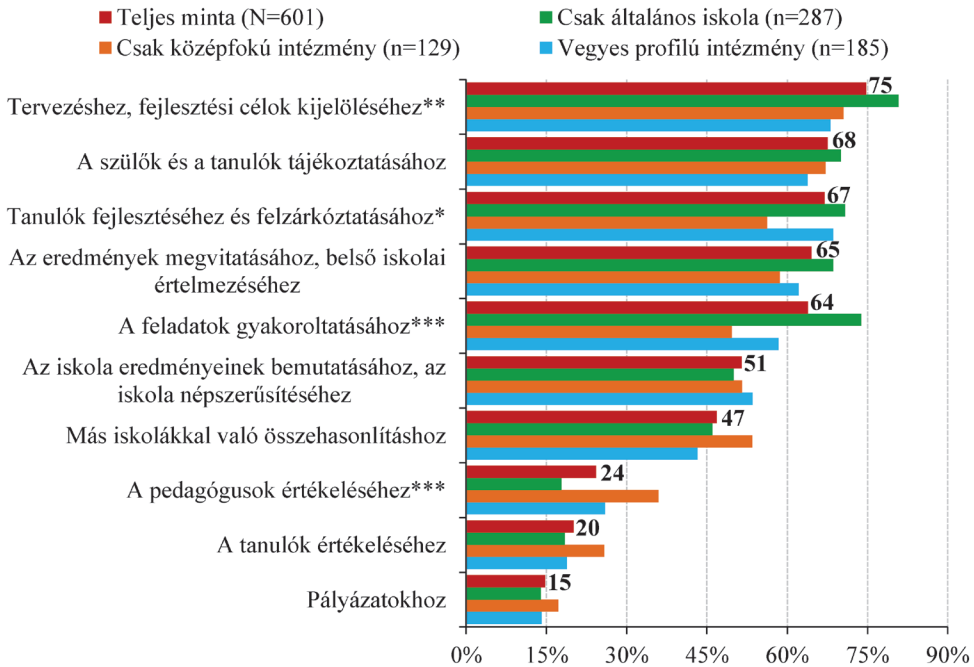
⁵ Lásd a 2014/2015. tanév rendjéről és az egyes oktatást szabályozó miniszteri rendeletek módosításáról szóló 35/2014. (IV. 30.) EMMI rendelet 9. §-ának 10. bekezdését.

Ugyanakkor ebben az évben már 71%-uk a szülők és a tanulók tájékoztatásához, 54%-uk pedig az iskola népszerűsítéséhez (is) felhasználta a mérési eredményeket. A vezetők 53%-a számára az is fontos szempont volt, hogy az eredmények ismerete más iskolákkal való összehasonlításra ad alkalmat. A tanulók és a pedagógusok értékelésére jóval kisebb arányban használták a méréseket.

A 2014. évi intézményvezetői adatfelvételben a kérdés nem csupán a kompetenciamérésre, hanem általában a mérések, vizsgák és tanulmányi versenyek eredményeinek hasznosítására vonatkozott. Az említési arányok így is igen hasonlóan alakultak a 2009. évi eredményekhez: a leggyakoribb a tervezési, fejlesztési és stratégiai jellegű használat; ezt olyan célok követik, mint a szülők és tanulók tájékoztatása, a diákok fejlesztése és felzárkóztatása, az intézményi adatok elemzésében rejlő lehetőségek hasznosítása, valamint a feladatok gyakoroltatása. Az iskola népszerűsítésére, illetve más iskolákkal történő összehasonlításra történő felhasználás már kevésbé jellemző, a legkevésbé pedig a tanárok, illetve a tanulók értékelése, valamint a különféle pályázatok kapcsán fordulnak a mérési eredményekhez. Megfigyelhető, hogy a kizárólag általános iskolai képzést nyújtó intézmények vezetői az átlagosnál jóval nagyobb arányban említették azokat a lehetőségeket, amelyek a fejlesztési célok meghatározásához, a tanulók egyéni szintű fejlesztéséhez, felzárkóztatásához, valamint a feladatok begyakoroltatásához kapcsolódnak. Középfokon viszont sokkal jobban megjelenik az értékelési szempont, és sokkal fontosabb az eredmények bemutatása, az intézmény népszerűsítése a mérési adatok – esetenként a saját eredmények más iskolákkal történő összevetése – révén (lásd 3. ábra).

3. ábra

„Milyen célokra használják a hazai/nemzetközi (standardizált) mérések, vizsgák, tanulmányi versenyek anyagait, eredményeit, adatait?” – említések %-os aránya az intézményvezetők körében (2014)



Forrás: Intézményvezetői adatfelvétel, 2014

Szignifikáns különbségek: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

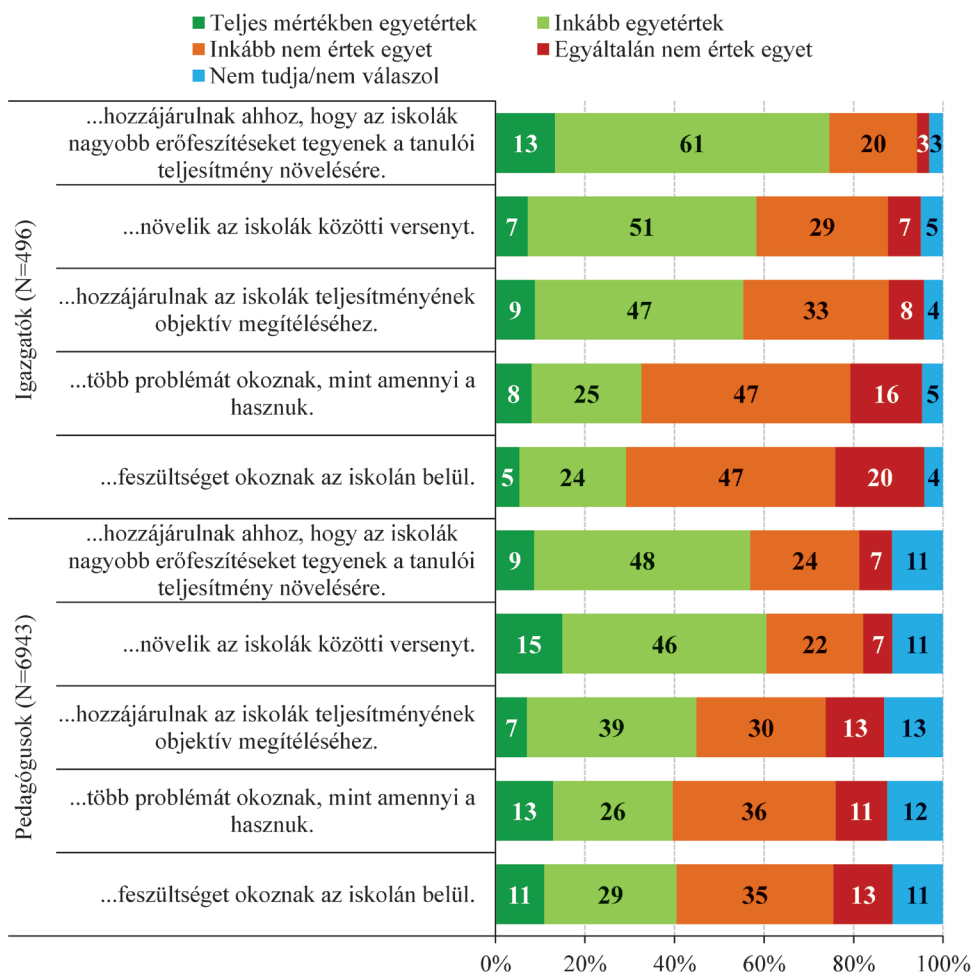
Az iskola településtípus szerinti elhelyezkedése is befolyásolja, hogy milyen célokra használják a mérések, vizsgák, tanulmányi versenyek eredményeit. Minél nagyobb városban található az intézmény, annál gyakoribb az iskola népszerűsítéséhez és a pedagógusok értékeléséhez történő felhasználás, míg a kisebb településeken a nagyobb városokhoz képest sokkal gyakrabban használják ezeket az eredményeket a tanulók fejlesztéséhez és felzárkóztatásához, valamint a feladatok gyakoroltatásához. Továbbá a KLIK, önkormányzati és egyéb állami fenntartásban lévő iskolák esetében az egyházi, illetve a magán, alapítványi fenntartású intézményekhez képest jóval gyakoribb a tervezés, a fejlesztési célok kijelölése, a tanulók fejlesztése és felzárkóztatása, valamint az eredmények megvitatása, belső iskolai értelmezése terén történő hasznosítás.

A központilag szervezett rendszeres tanulói teljesítménymérések megítélése

A központi tanulói teljesítménymérésekhez kötődő attitűdök fontos jelzéssel bírnak a tekintetben, hogy miként érdemes a mérések minél szélesebb körű felhasználását és elfogadottságát elősegíteni. Elmondható, hogy mind az intézményvezetők, mind pedig a pedagógusok inkább pozitív véleményeket társítanak a teljesítménymérésekhez, bár – elsősorban a pedagógusok közül – többen felhívják a figyelmet az esetleges kockázatokra is. Az intézményvezetők összességében kedvezőbben viszonyulnak a teljesítménymérésekhez, mint a pedagógusok, hiszen valamivel nagyobb arányban vélik úgy, hogy a rendszeres tanulói teljesítménymérések segítik az iskolai eredményesség objektív feltárását, valamint hozzájárulnak ahhoz, hogy az iskolák erőfeszítéseket tegyenek a tanulói teljesítmény növelésére. Ez álláspontunk szerint elsősorban a vezetői felelősség nagyobb arányú megnyilvánulásának, a vezetői szerepkörből adódó eltérő megközelítésnek és szemléletmódnak köszönhető. A teljesítménymérések hasznosságát illetően a pedagógusok összességében bizonytalanabbak, kevésbé érzékelik annak pozitívumait, fontosabbnak tartják viszont hangsúlyozni az esetleges veszélyeket, az iskolán belüli esetleges feszültségeket és az egyéb problémákat (lásd 4. ábra).

4. ábra

„A központilag szervezett rendszeres tanulói teljesítménymérések...” –
% -os arányok az intézményvezetők, illetve a pedagógusok körében (2013)



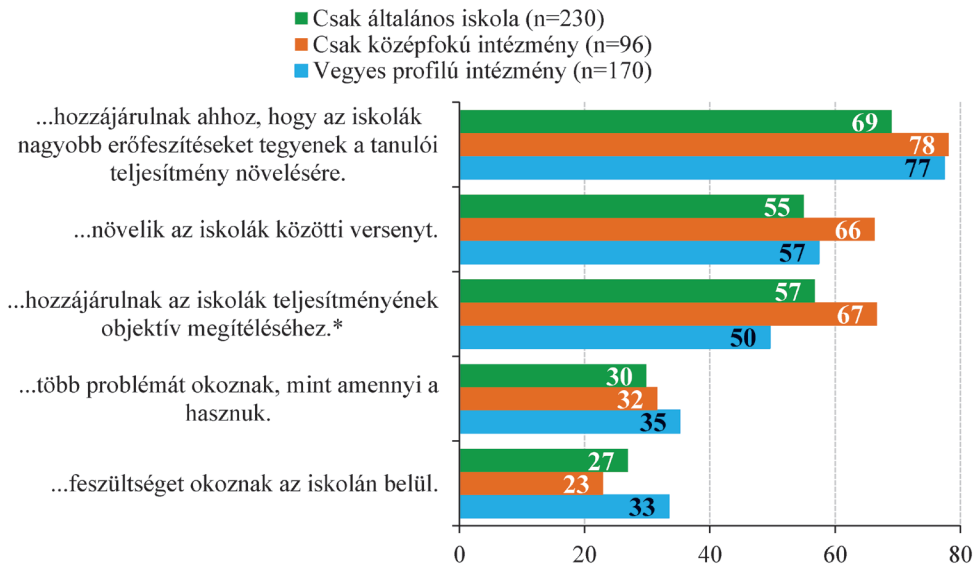
Forrás: Intézményvezetői és pedagógus-adatfelvétel, 2013

Az intézményvezetők kompetenciaméréssel kapcsolatos beállítódása leginkább az intézmény típusa és településtípus szerinti elhelyezkedése alapján tér el: úgy tűnik, hogy a középfokú intézményekben némiképp hangsúlyosabban jelennek meg a mérések hasznaira történő utalások, és gyakoribb az a meglátás, miszerint a rendszeres tanulói teljesítménymérések segítik az objektivitást. A csak általános iskolai képzést nyújtó iskolák, valamint a vegyes profilú intézmények igazgatói már kevésbé hangsúlyozzák a tanulói teljesítménymérések hasznait, és főként a vegyes típusú

iskolákban valamelyest jobban előtérbe kerülnek a mérések okozta hátrányok és problémák (lásd 5. ábra). Továbbá a községi iskolák vezetői kevésbé gondolják úgy, hogy a mérések fokozzák az iskolák közötti versenyt, illetve, hogy hozzájárulnak az iskolák teljesítményének objektív megítéléséhez.

5. ábra

„A központilag szervezett rendszeres tanulói teljesítménymérések...” – az inkább egyetért és a teljes mértékben egyetért válaszok intézménytípus szerinti %-os megoszlása az intézményvezetők körében (2013)



Forrás: Intézményvezetői adatfelvétel, 2013

*Szignifikáns különbség ($p < 0,01$)

Az öt állítást faktorelemzés segítségével megvizsgálva azt találtuk, hogy a pedagógusok körében a válaszok mögött két látens dimenzió húzódik meg: az egyik inkább a mérések negatívumait hangsúlyozó gondolkodásmód, a másik pedig inkább a mérések hasznosságát fontosabbnak tartó. Azon állítást, mely szerint a központi tanulói teljesítménymérések növelik az iskolák közötti versenyt, mindkét faktor tartalmazza, amely nem is annyira meglepő, hiszen a versenyhelyzetet lehet hasznosként, de lehet veszélyként is értelmezni (lásd 1. táblázat). A két faktor együttesen az öt állítás információtartalmának 54%-át őrizte meg.

1. táblázat

A központilag szervezett rendszeres tanulói teljesítménymérések megítélése a pedagógusok körében, 2013 (faktorsúlyok, N=5 652)*

	Faktorsúlyok	
	Hasznosság	Veszélyek
...hozzájárulnak ahhoz, hogy az iskolák nagyobb erőfeszítéseket tegyenek a tanulói teljesítmény növelésére.	0,763	-0,170
...hozzájárulnak az iskolák teljesítményének objektív megítéléséhez.	0,728	-0,274
...feszültséget okoznak az iskolán belül.	-0,060	0,763
...több problémát okoznak, mint amennyi a hasznuk.	-0,359	0,727
...növelik az iskolák közötti versenyt.	0,397	0,326

Forrás: Pedagógus-adatfelvétel, 2013

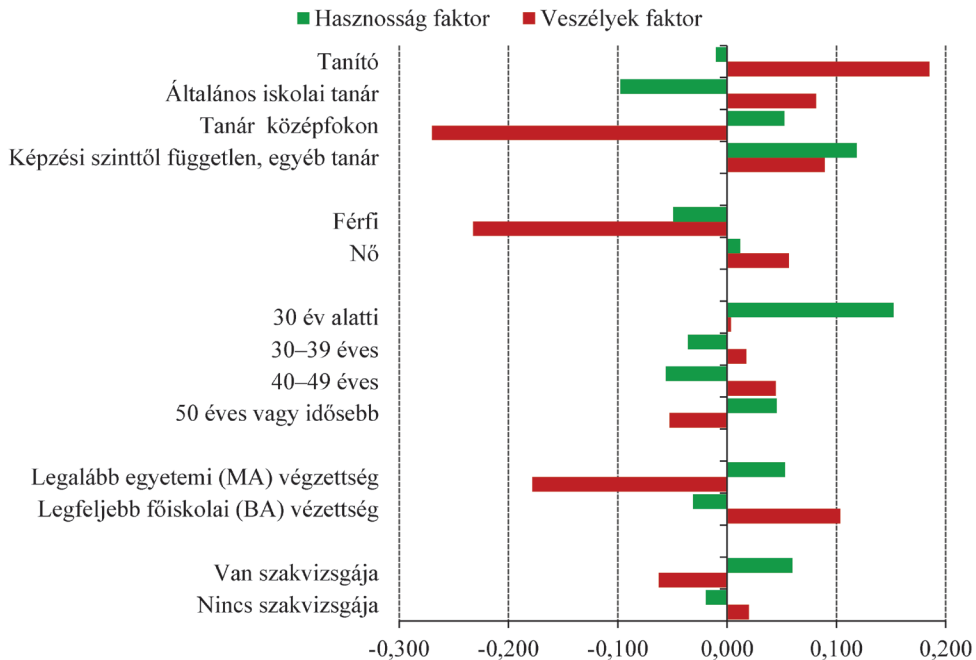
*Módszer: Maximum Likelihood, Varimax rotálás alkalmazásával. Goodness-of-fit teszt szign: 0,193, KMO értéke: 0,655, Bartlett teszt szign: 0,000. A faktorsúlyok azt mutatják meg, hogy a faktorelemzés során feltárt faktorokhoz milyen szorosan kapcsolódnak az egyes változók.

A mérések hasznosságának megítélése – mint ahogyan az igazgatói válaszok esetén is tapasztaltuk – egyértelműen kedvezőbb a középiskolai szinten, továbbá az egyetemi vagy annál magasabb iskolai végzettséggel, a szakvizsgával rendelkezők, valamint a legfiatalabb és a legidősebb korcsoportok esetében is. A fiatalabb pedagógusok körében feltehetőleg arról van szó, hogy ők már a képzésük során, a pedagógusképzési tananyag részeként is találkozhattak ezekkel a mérésekkel, ismerhetik azok előnyeit, míg az idősebbek inkább a folyamatos szakmai tapasztalás, nem ritkán a vezetői, szakértői tapasztalatok birtokában érzik át jobban ezek fontosságát, hasznosságát. A 60 év felettiek, illetve a 40 év feletti szakmai gyakorlattal rendelkezők körében a legmagasabb ugyanis azoknak az aránya, akik szerepelnek vagy a múltban szerepeltek valamely szaktanácsadói vagy szakértői névjegyzéken. Arányuk egyébként a magasabban kvalifikáltak körében is nagyobb, ezért elképzelhető, hogy a teljesítménymérésekkel kapcsolatban a körükben megnyilvánuló magasabb elfogadottság is részben az intenzívebb szakmai munkának és a nagyobb rálátásnak köszönhető. A mérésekben inkább veszélyeket látnak a tanító munkakörben dolgozók, az általános iskolai tanárok, a legfeljebb főiskolai végzettségűek, valamint a szakvizsgával, vagy azzal egyenértékű végzettséggel nem rendelkezők. Némileg összetettebb a képzési szinttől független, egyéb munkakörben dolgozó pedagógusok – úgy mint a fejlesztőpedagógusok, gyógypedagógusok, kollégiumi nevelők, könyvtáros tanárok stb. – álláspontja, mivel ők érzékelik a mérések hasz-

nait, de ugyanúgy veszélyeket is látnak bennük. A nők gyakrabban tartoznak azok közé, akik a mérések veszélyeire hívják fel a figyelmet, míg a férfiak nem tartanak a teljesítménymérések veszélyeitől, ugyanakkor azok hasznosságáról sincsenek meggyőződve (lásd 6. ábra).

6. ábra

Faktor-átlagértékek a pedagógusok különböző csoportjaiban, 2013 (N=5 652)*



Forrás: Pedagógus-adatfelvétel, 2013

*A 0 átlagú és 1 szórástú faktorok által megtestesített dimenzió jelentéstartalmának elfogadása a pozitív faktorértékek felé növekszik, míg a negatív értékek felé csökken. Szignifikáns különbségek mindkét faktor esetében valamennyi csoportképző ismérv alapján (min. $p < 0,05$).

A teljesítménymérések megítéléséhez, az eredmények megfelelő felhasználásához, hasznosításához a pedagógusoknak valamilyen mértékben birtokában kell lenniük a tanulók értékeléséhez, a kutatásokhoz, elemzésekhez kapcsolódó kompetenciáknak. A mintában szereplő pedagógusok (N=6 943) a pedagógiai folyamatok és a tanulók értékelése kapcsán a saját felkészültségüket összességében jónak tartják, csaknem kétharmaduk (63,5%) nyilatkozott úgy, hogy jó színvonalon rendelkezik ezzel a kompetenciával és egyötödük (20,8%) úgy, hogy tökéletesen birtokában van, egytizedük (9,2%) azonban érdemben nem válaszolt a kérdésre. Elmondható e tekintetben, hogy az általános iskolai tanítók és tanárok, a nők, az

idősebb korosztályokba tartozó, valamint a nagyobb szakmai gyakorlattal, magasabb iskolai végzettséggel és szakvizsgával rendelkező pedagógusok biztosabbak a tudásukban, és erősebbnek érzik ebbéli képességeiket. Az elemző, kutatói tudás megléte már jóval kevésbé jellemző, a pedagógusok alig háromtizede (29,9%) ítéli jó színvonalúnak, és csak 7%-a tekinti tökéletesnek ezen kompetenciáit, továbbá a bizonytalanok és a nem válaszolók aránya is magasabb, összességében több mint 16%. Elemző és kutatói kompetenciákkal az átlagosnál nagyobb mértékben rendelkeznek a középiskolákban tanító tanárok, a férfiak és a fiatalabb generáció tagjai, továbbá azok a pedagógusok, akik magasabb iskolai végzettséggel és szakvizsgával rendelkeznek.

A pedagógiai kompetenciák alakulásában az Országos kompetenciamérés eredményeinek felhasználása is szerepet játszhat. A megkérdezett pedagógusok csaknem egyötödénél (18,5%) egyáltalán nem, egynegyedénél (24,2%) kevés, háromtizedénél (29,6%) pedig közepes szerepet játszott a saját pedagógiai kompetenciáik fejlődésében a kompetenciamérési eredmények felhasználása. Ezzel szemben alig több mint egytizedük (11,5%) tulajdonít fontos szerepet a mérésnek e tekintetben, és mindössze 1% azok aránya, akik meghatározónak ítélik ezt a hatást. Igen magas, 15% azok aránya, akik nem válaszoltak a kérdésre, illetve nem tudták eldönteni a kérdést. Amennyiben csak az érdemben válaszolókat vizsgáljuk, azaz azokat, akik határozott véleményt adtak, akkor megállapítható, hogy az általános iskolai tanítók és tanárok, a nők, az idősebb és tapasztaltabb pedagógusok, valamint a legfeljebb főiskolai végzettséggel rendelkezők kompetenciáinak alakulásában játszott jelentősebb szerepet a hazai kompetenciamérés eredményeinek felhasználása. Ez összhangban van azzal, hogy épp ezen csoportok azok, amelyek az átlagosnál kisebb mértékben rendelkeznek elemzői és kutatói kompetenciákkal. Nagyobb mértékű a kompetenciamérések hatása a szakvizsgával rendelkező pedagógusokra is, esetükben arról lehet szó, hogy a szakvizsgához vezető úton többször is találkozhattak az Országos kompetenciamérés eredményeivel, azok elemzésével, amelyek gazdagították pedagógiai ismereteiket, kompetenciáikat (lásd 2. táblázat).

2. táblázat

„Véleménye szerint az Ön pedagógiai kompetenciáinak alakulásában milyen mértékben játszott szerepet az Országos kompetenciamérés eredményeinek felhasználása?” – ötfokozatú skála (1 = egyáltalán nem játszott szerepet, 5 = meghatározó szerepet játszott) átlagai az érdemben válaszoló pedagógusok különböző csoportjaiban (2013)

	N	Átlag	Szórás
Munkakör*			
Tanító	1 637	2,58	0,97
Általános iskolai tanár	1 699	2,56	1,02
Tanár középfokon	1 829	2,35	1,01
Képzési szinttől független, egyéb tanár	740	2,12	1,06
Nem*			
Férfi	1 133	2,30	1,01
Nő	4 772	2,48	1,02
Korcsoport*			
30 év alatti	365	2,19	1,09
30–39 éves	1 375	2,30	1,07
40–49 éves	1 998	2,40	0,98
50 éves vagy idősebb	2 167	2,62	0,98
Legmagasabb befejezett iskolai végzettség*			
Legfeljebb főiskolai vagy BA végzettség	3 700	2,50	1,01
Legalább egyetemi vagy MA végzettség	2 206	2,35	1,02
Szakvizsga megléte*			
Van szakvizsgája	1 455	2,54	1,00
Nincs szakvizsgája	4 451	2,41	1,02
Összesen	5 906	2,44	1,02

Forrás: Pedagógus-adatfelvétel, 2013

*Szignyifikáns különbség ($p < 0,001$)

A kutatás során arra is rákérdeztünk, vajon a pedagógiai munkájuk során milyen célok elérését tartják a legfontosabbnak a pedagógusok. A 2014. évi pedagógus-kérdőívben 15 különböző pedagógiai célt soroltunk fel, melyből a megkérdezettek kiválaszhatták a legfontosabbnak tartott 5 célkitűzést. A legtöbb választ

(76,6%) a jó tanórai légkör, a gyerekek érdeklődésének felkeltése kapta, de ugyanígy rendkívül fontosnak tartják a sikeres továbbhaladás, továbbtanulás elérését is (74,6%). Több mint felük (53,4%) jelölte meg a tanulási motiváció felkeltését és fenntartását, de a szilárd értékrend közvetítése és a képességek kibontakoztatásának egyénre szabott segítése is 40% feletti jelölést kapott. Sokkal kevésbé markáns célkitűzésként jelenik meg a különböző méréseken, vizsgákon való jó tanulói szereplés elérése, hiszen ezt a pedagógusok mindösszesen 15%-a jelölte célként, amely a 10–11. helyet jelenti a felsorolt 15 célkitűzés tekintetében. A jó mérési és vizsgaeredmény elérése az általános iskolai és a középfokon tanító tanárok, a férfiak, valamint a magasabb képesítéssel rendelkezők esetében az átlagosnál lényegesen magasabb arányban jelenik meg célként.

Hasonlóképpen megkérdeztük az intézményvezetőket is arról, hogy melyek az iskola legfontosabb célkitűzései. A 2014. évi intézményvezetői kérdőívben 16 célkitűzés közül választhatták ki a legfontosabb 5 intézményi célt az igazgatók. Az eredmények alapján látható, hogy a legtöbb intézmény számára (70,2%) a tanulói képességek kibontakoztatásának a segítése a legfontosabb, s fontosnak tekinthető a szilárd értékrend közvetítése (57,5%) és az intézmény jó légkörének a megteremtése is (57,3%). Az intézményvezetők csaknem fele (48,9%) a tanulók motivációjának felkeltését és fenntartását választotta, míg több mint egyharmaduknál megjelent a hátrányos helyzetű tanulók beilleszkedésének segítése is (34,3%). Akárcsak a pedagógusok esetén, az igazgatói válaszköznél is háttérbe szorult a célok között a kompetenciamérésen elért tanulói teljesítmények növelése, ezt az intézményvezetők kevesebb mint egynegyede (17,7%) jelölte, így ez a célkitűzés a fontossági sorrend alsó tartományába került. E tekintetben elsősorban fenntartó szerint tapasztalható jelentősebb különbség: a KLIK, önkormányzati, illetve egyéb állami fenntartású intézmények vezetői jóval magasabb arányban (19,9%) jelölték célként a mérési eredmények növelését, mint az alapítványi, magán (11,5%), vagy az egyházi iskolák igazgatói (4,4%), amely egyértelműen köthető a korábban már említett, főként az állami fenntartású intézményeket érintő szabályozási környezethez, jogszabályi kötelezettségekhez.

Tanulói teljesítménymérések mint az intézményben folyó szakmai munka minőségének mutatói

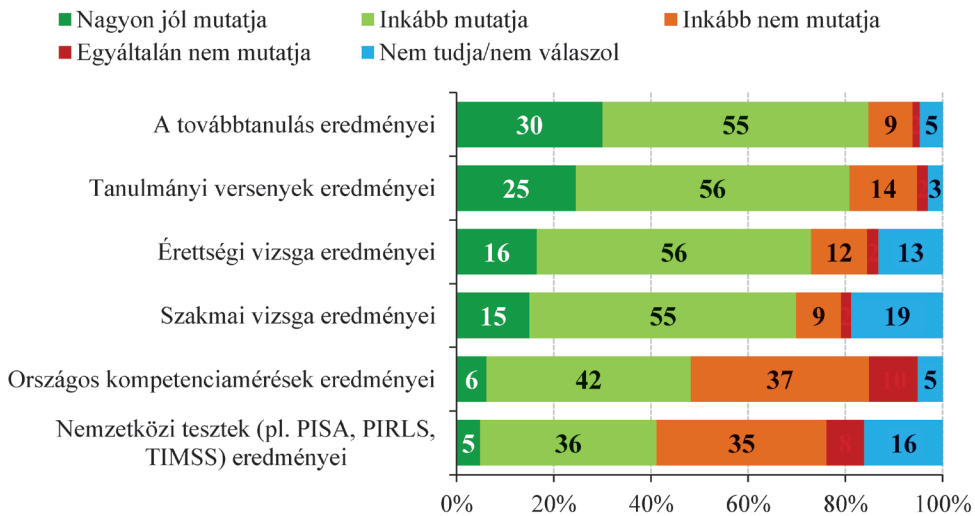
Fontos kérdés, hogy a pedagógusok miként vélekednek a különféle tanulói teljesítménymérések és az iskolában folyó szakmai munka eredményessége közötti kapcsolatról. A pedagógusok egy része nem tudott, illetve nem akart válaszolni arra

a kérdésre, hogy az általunk felsorolt, jellemzően a mérésekre, vizsgákra fókuszáló tényezők közül melyek mennyire mutatják az iskolában folyó szakmai munka minőségét, eredményességét. Kétharmaduk válaszolt valamennyi szempontra, 7%-uk viszont egyikre sem reflektált érdemben. Ez utóbbi csoportot kihagyva az elemzésből, megkapjuk azokat a pedagógus válaszadókat, akik a kérdés kapcsán legalább egy szempontot értékelték.

A legalább egy tényezőt értékelő pedagógusok válaszai azt mutatják, hogy leginkább a továbbtanulási eredményekre tekintenek úgy, hogy az jó mutatója lehet az intézményi munka minőségének, de a tanulmányi versenyeredményekről is sokan vélekednek így. Az általában középfeokon megjelenő, konkrét tantárgyakhoz kötődő szakmai, illetve érettségi vizsgaeredmények kapcsán is többen jelezték, hogy azokat elfogadhatónak tartják eredményességi mutatónak, a kompetenciamérésekkel és a nemzetközi tesztekkel kapcsolatban azonban már jóval kritikusabbak a pedagógusok (lásd 7. ábra).

7. ábra

„Ön szerint mennyire mutatják jól az alábbiak az iskolákban folyó szakmai munka minőségét, eredményességét?” (2013) – %-os arányok a legalább egy szempontot érdemben értékelő pedagógusok körében, N=6 461



Forrás: Pedagógus-adatfelvétel, 2013

A különböző vizsgaeredmények esetében tapasztalható magas bizonytalanság, illetve válaszhiány a közvetlen érintettség alacsonyabb fokára utal, vagyis az általános iskolákban tanító pedagógusok esetében jóval magasabb az érdemben nem

válaszolók aránya, mint a középfokon tanítók körében. Ezzel szemben az általános iskolai szinten és a középfokon is megjelenő nemzetközi teljesítménymérések esetén a nagyobb mértékű bizonytalanság inkább a nemzetközi tesztekre és azok eredményeire vonatkozó ismerethiányra utalhat. Például a PISA-vizsgálatok kapcsán egy 2012-es közvélemény-kutatás megállapítja, hogy a pedagógusok túlnyomó többsége ugyan már hallott a mérésekről, de a konkrét eredményekkel már jóval kevésbé vannak tisztában. A tájékozatlanság az eredmények alakulására és a tényleges pontszámokra egyaránt vonatkozik, hiszen a kutatás szerint a megkérdezett pedagógusok csupán 30 százaléka tudta, hogy 2009-re javultak a magyar szövegértési eredmények, és mindössze 16 százaléka volt tisztában azzal, hogy a szövegértés területén elért átlagos magyar teszteredmény 2000 és 2006 között 480 pont körül alakult. Mindkét kérdésre csak a megkérdezett pedagógusok 4 százaléka válaszolt jól (Lannert, 2012).

A tanulmányi versenyek szakmai és eredményességi mutatóként való kezelése kapcsán az általános iskolákban dolgozók, a nők, az alacsonyabb képesítéssel rendelkezők fejezték ki fokozottabb egyetértésüket, s ők azok, akik a továbbtanulási eredmények figyelembevételével mellett is erőteljesebben állnak ki. Az idősebb generáció tagjai a szakmai munka minőségét jelző tényezők közül valamelyest jobban kiemelik a tanulói továbbhaladás sikerességét, mint a tanulmányi versenyeredményeket, míg a fiatalabb korosztályban épp fordított viszonyulás tapasztalható. Az érettségi vizsgák és szakmai vizsgák eredményeinek figyelembe vétele érthető módon fontosabb azoknak, akik maguk is középfokon – mellyel összefüggésben általában magasabb végzettséggel – tanítanak. Az Országos kompetenciamérés és a nemzetközi mérések tekintetében a középfokon tanítók, a férfiak és a magasabban képzett pedagógusok jelezték nagyobb arányban elköteleződésüket, életkor és szakmai gyakorlat szerint pedig a két végpont közelében lévők, azaz a nagyon fiatalok és a leghosszabb gyakorlati idővel rendelkezők. Mint korábban is utaltunk rá, a fiatalok esetében arról lehet szó, hogy már a pedagógusképzés során is jobban megismerkedhettek ezekkel a mérésekkel és azok eredményeivel, míg az idősebb korosztály tagjainál a hosszabb gyakorlati idő, a folyamatos és intenzív szakmai tapasztalatok erősíthetik a pozitív hozzáállást (lásd 3. táblázat).

3. táblázat

„Ön szerint mennyire mutatják jól az alábbiak az iskolákban folyó szakmai munka minőségét, eredményességét?” (2013) – a „nagyon jól mutatja” és az „inkább jól mutatja” válaszok összevont %-os arányai a legalább egy szempontot érdemben értékelő pedagógusok körében, N=6 461

	Tovább- tanulás	Tanul- mányi versenyek	Érettségi vizsga	Szakmai vizsga	OKM	Nemzet- közi tesztek
	eredményei					
Munkakör*						
Tanító	87,2%	83,6%	68,0%	66,9%	43,2%	39,3%
Általános iskolai tanár	86,6%	80,9%	65,2%	63,3%	45,8%	36,3%
Tanár középfokon	81,0%	76,7%	85,0%	76,5%	54,2%	44,6%
Képzési szinttől független, egyéb tanár	85,0%	84,9%	70,6%	73,7%	48,9%	46,8%
Nem*						
Férfi	79,9%	74,6%	73,0%	69,9%	52,6%	42,6%
Nő	86,0%	82,3%	72,9%	69,8%	47,1%	40,8%
Korcsoport*						
30 év alatti	83,5%	85,4%	73,8%	74,2%	56,3%	49,9%
30–39 éves	82,9%	82,2%	74,6%	72,1%	45,9%	40,2%
40–49 éves	85,7%	80,4%	72,5%	68,6%	45,1%	38,2%
50 éves vagy idősebb	85,4%	79,2%	72,1%	68,6%	50,8%	42,6%
Legmagasabb befejezett iskolai végzettség*						
Legfeljebb főiskolai vagy BA végzettség	85,8%	83,3%	68,3%	68,7%	44,5%	39,0%
Legalább egyetemi vagy MA végzettség	83,1%	76,5%	80,9%	71,9%	54,5%	44,7%
Szakvizsga megléte*						
Van szakvizsgája	81,3%	77,9%	75,8%	72,6%	51,3%	43,5%
Nincs szakvizsgája	85,9%	81,7%	72,1%	69,0%	47,2%	40,4%

Forrás: Pedagógus-adatfelvétel, 2013

*Szignyifikáns különbség (min. $p < 0,05$)

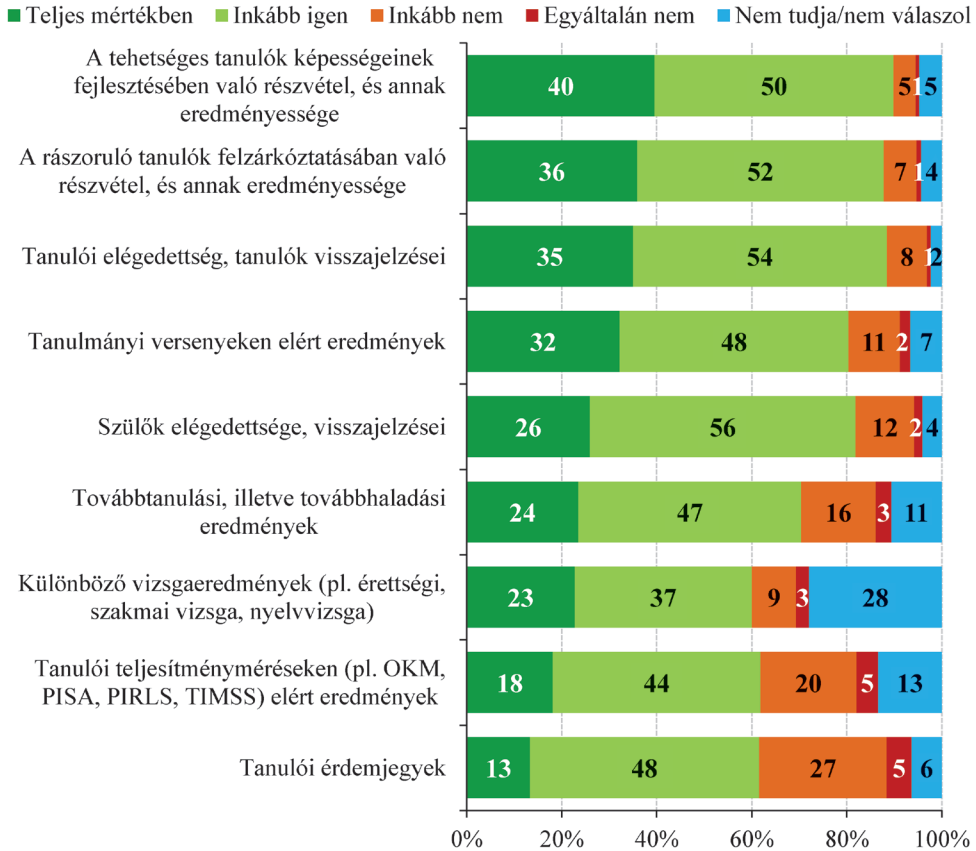
Tanulói teljesítménymérések, mint az iskolák és a pedagógusok értékelésének eszközei

A pedagógusokat és az intézményvezetőket a különböző mérési és vizsgaeredményekkel kapcsolatban arról is kérdeztük, hogy mennyire indokolt azokat a pedagógusok munkájának értékelése során figyelembe venni. A kérdés a különböző mérési és vizsgaeredmények túl olyan további fontos tényezőkre is kitért, mint a továbbtanulási eredmények, a tanulói és a szülői elégedettség, a felzárkóztatásban, illetve a tehetséggondozásban való eredményes részvétel.

A pedagógusok közül 7,5%-nyian egyik szempontra vonatkozóan sem válaszoltak érdemben a feltett kérdésre, őket kihagytuk az elemzésből, vagyis csak a legalább egy szempontot érdemben értékelő pedagógusok válaszait vizsgáltuk. A pedagógusmunka értékelésekor figyelembe veendő szempontok pedagógusok szerinti sorrendje nagymértékben hasonlít az eredményes szakmai munkát jelző szempontok sorrendjéhez, ugyanakkor a 2014. évi adatfelvétel során felsorolt plusz szempontokat a mérési és vizsgaeredményeknél rendre fontosabbnak értékelték a megkérdezettek. Az eredmények alapján a pedagógusok a saját munkájuk értékelésekor elsősorban a tehetséges tanulók képességeinek fejlesztésében, valamint a rászoruló tanulók felzárkóztatásában elért sikereket vennék figyelembe, de fontos szempontként jelennek meg a tanulók és a szülők visszajelzései, valamint a tanulmányi versenyeken elért eredmények is. A konkrét mérési és vizsgaeredményeket, valamint a tanulók érdemjegyeit viszont már kevésbé alkalmazzák a munkájuk megítéléséhez. Az adatok alapján észrevehető, hogy a kizárólag a középfokú intézményekben előforduló vizsgák (érettségi, szakmai vizsga), illetve a különböző, többnyire egy-két kompetenciaterületre irányuló tanulói teljesítménymérések kapcsán az átlagosnál többen nem nyilvánítottak véleményt, amely arra utal, hogy a kérdést a saját érintettséget is figyelembe véve válaszolták, illetve nem válaszolták meg a pedagógusok (lásd 8. ábra, vö. 7. ábra).

8. ábra

„Véleménye szerint az alábbi szempontokat mennyire kellene figyelembe venni akkor, amikor az Ön munkáját értékeli, vagy amikor a munkájára visszajelzéseket kap?” (2014) – %-os arányok a legalább egy szempontot érdemben értékelő pedagógusok körében, N=4 869



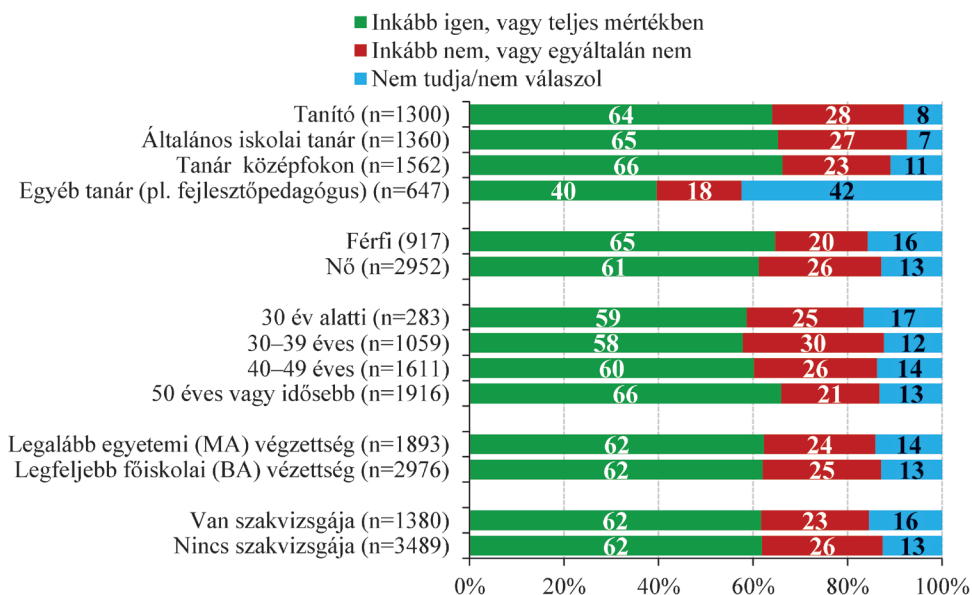
Forrás: Pedagógus-adatfelvétel, 2014

Általában elmondható, hogy a magasabb végzettséggel rendelkező, középfokon tanító tanárok és a férfi pedagógusok némiképp nagyobb arányban vennék valamilyen mértékben figyelembe a szakmai munkájuk értékelése során az olyan konkrét mérhető mutatókat, mint a tanulói érdemjegyek, a mérési- és vizsgaeredmények vagy a továbbtanulási eredmények. A „puhább” mutatók, mint a szülői elégedettség, a felzárkóztatásban, tehetséggondozásban való részvétel eredményessége inkább az alacsonyabb képzési szinteken tanítók és a nők esetében értékelődnek fel jobban. A tanulói visszajelzéseket a középfokon megjelenő, idősebb – és egyben érettebb – diákokat tanító tanárok vennék jobban figyelembe. Megfigyelhető továbbá, hogy valamennyi értékelési szempont a magasabb szakmai tapasztalattal

rendelkező, idősebb korosztály tagjainál jelenik meg hangsúlyosabban, ez alól a tanulói érdemjegyek és a vizsgaeredmények jelentenek némi kivételt. Amennyiben a tanulmányunk témáját adó tanulói teljesítményméréseken elért eredményeket vizsgáljuk részletesebben, akkor megállapítható, hogy a középfokon tanítók, az 50 év felettek és a férfi pedagógusok a szakmai munka értékelésekor az átlagosnál jobban támaszkodnának a mérési eredményekre (lásd 9. ábra).

9. ábra

„Véleménye szerint a tanulói teljesítményméréseken (pl. OKM, PISA, PIRLS, TIMSS) elért eredményeket mennyire kellene figyelembe venni akkor, amikor az Ön munkáját értékelik, vagy amikor a munkájára visszajelzéseket kap?” (2014) – %-os arányok a pedagógusok körében*



Forrás: Pedagógus-adatfelvétel, 2014

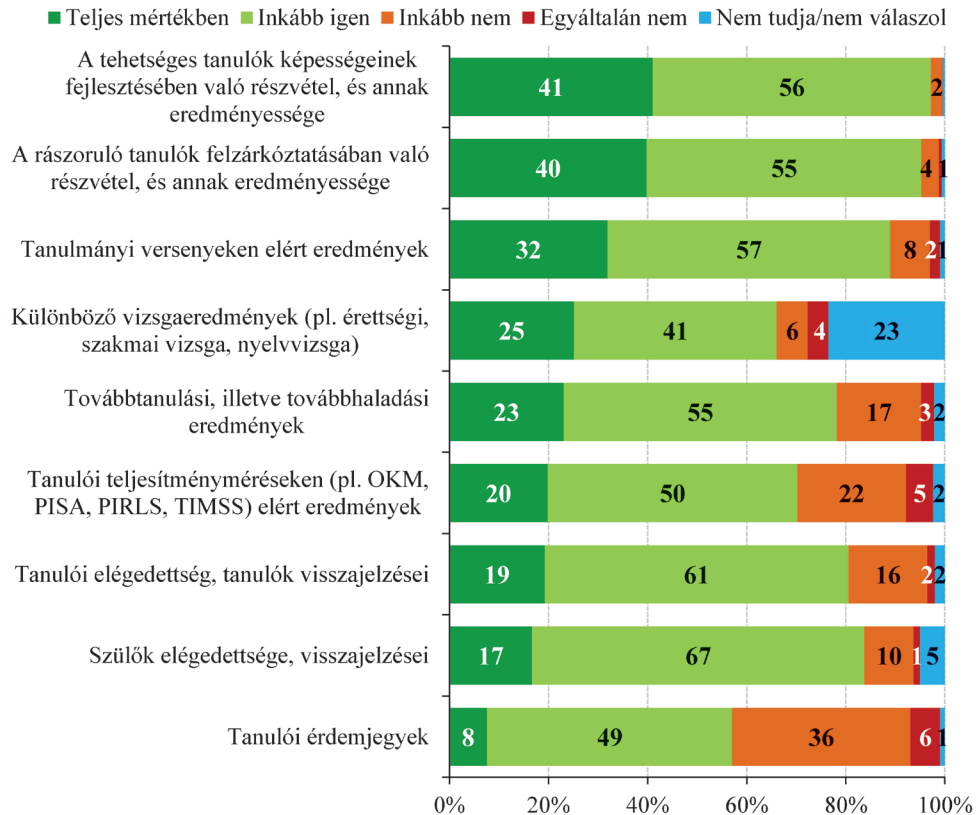
*Szignifikáns különbség a képzési szint, a nem és az életkor szerint ($p < 0,001$)

Az intézményvezetők körében az egyes értékelési szempontokat inkább vagy teljes mértékben indokoltnak tartók aránya a pedagógusok esetében tapasztaltnál némiképp magasabb. A pedagógusokhoz képest mind az egyes szempontokat, mind a szempontok összességét nézve kisebb az igazgatók körében tapasztalható válaszhiány – az intézményvezetők 5%-a nem adott érdemi választ egyik szempontra sem, őket nem vettük be az elemzésbe. Összességében az előzőekben felsorolt szempontok támogatottsága hasonlóan alakul az igazgatók esetében is. Ez alól a tanulói, illetve a szülői elégedettség, valamint a tanulói teljesítményméréseken elért eredmények értékelési szempontként történő figyelembevétele jelent némi

kivételt. A tanulók és a szülők elégedettségét kevésbé, míg a mérési eredmények figyelembevételét érezhetően indokoltabbnak tartják az intézményvezetők, mint a pedagógusok (lásd 10. ábra).

10. ábra

„Véleménye szerint mennyire indokolt az alábbiakat a pedagógusok értékelésekor figyelembe venni?” (2014) – %-os arányok a legalább egy szempontot érdemben értékelő intézményvezetők körében, N=571



Forrás: Intézményvezetői adatfelvétel, 2014

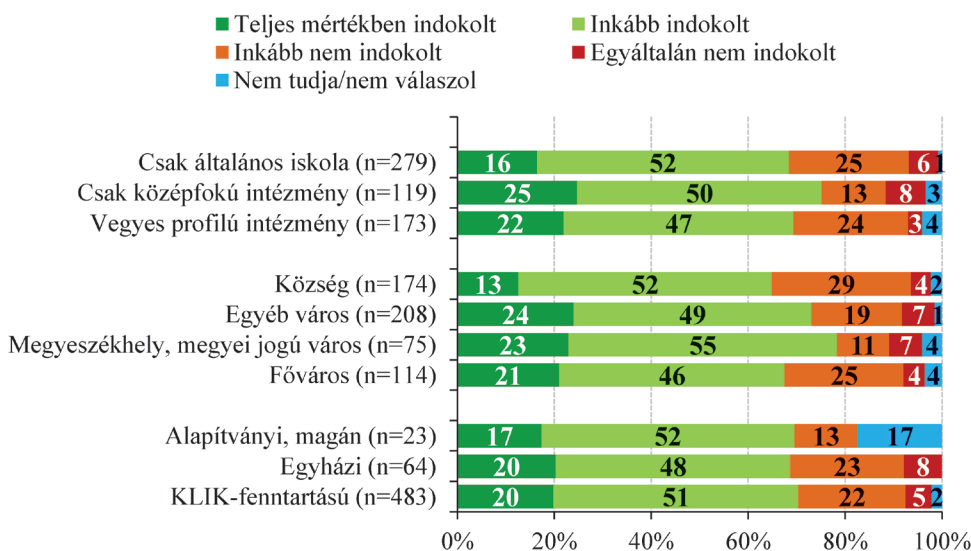
Az intézményvezetők körében a tanulói teljesítménymérések a pedagógusvéleményekhez viszonyítva elfogadottabbak a pedagógusok szakmai munkájának értékelési szempontjaként, amelyet jól mutat, hogy a válaszadó iskolavezetők héttizede inkább vagy teljes mértékben indokoltanak tartja a figyelembevételüket a pedagógusok értékelésekor. E tekintetben az elfogadottság szintje ugyanolyan mértékűnek tekinthető az utóbbi években, hiszen a 2008. évi TALIS nemzetközi tanárvizsgálatban a magyar iskolaigazgatók 70%-a számolt be arról, hogy a tanulók tesztekkel mért teljesítményeit fontos vagy nagyon fontos szempontként vették figyelembe a

tanári munka, a pedagógusok értékelésekor (OECD, 2009), amely arány – amennyiben az indokoltság felől közelítünk – a 2014. évi adatfelvételben sem változott, maradt a 70%-os szinten.⁶ Ebből következően a felsorolt, többé-kevésbé jól mérhető eredménymutatók közül a tanulói érdemjegyek kivételével a tanulói teljesítményméréseket az intézményvezetőknél is megelőzik a tanári munka értékelésének többi lehetséges mutatószámai, így a tanulmányi versenyeredmények, valamint a továbbtanulási, továbbhaladási eredmények.

Amennyiben a tanulói teljesítménymérések kapcsán adott igazgatói válaszokat képzéstípus, településtípus és iskolafenntartó szerint vizsgáljuk, megállapítható, hogy a középfokú intézményekben, valamint a (nagyobb) városok iskoláiban némiképp magasabb azon iskolavezetők aránya, akik a tanulói teljesítménymérések eredményeit jobban elfogadják értékelési szempontként (lásd 11. ábra).

11. ábra

„Véleménye szerint mennyire indokolt a tanulói teljesítményméréseken elért eredményeket a pedagógusok értékelésekor figyelembe venni?” (2014) – %-os arányok az intézményvezetők körében*



Forrás: Intézményvezetői adatfelvétel, 2014

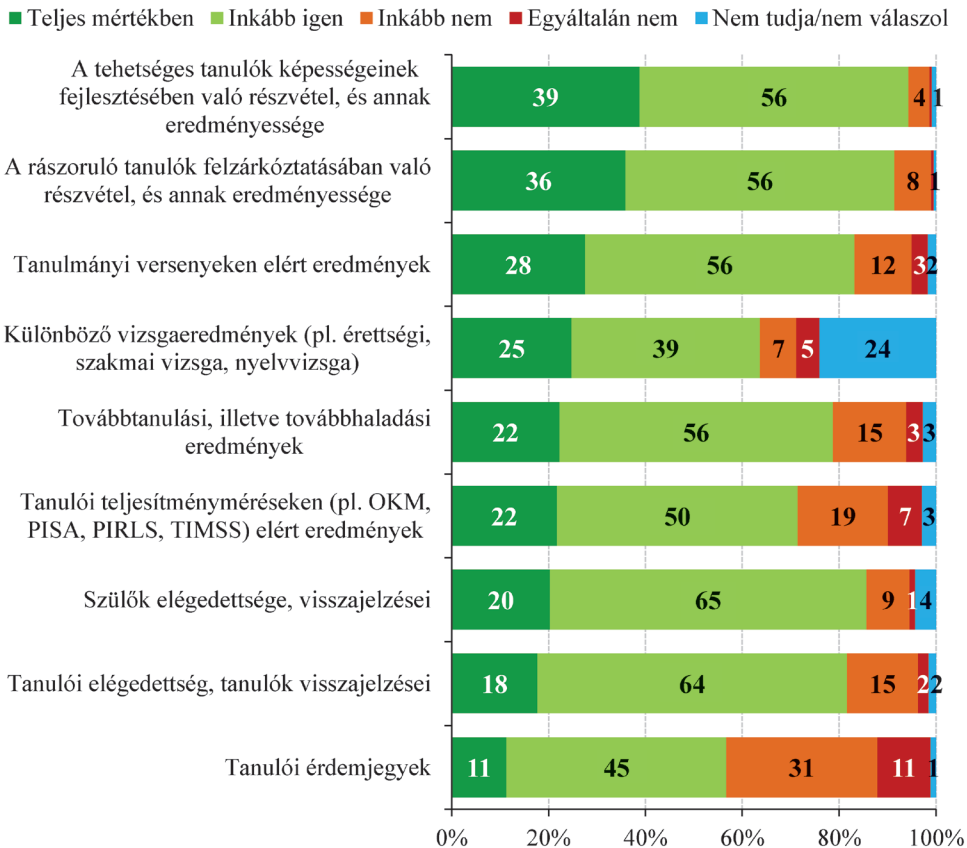
*Szignifikáns különbség (min: $p < 0,05$)

⁶ A teljesség kedvéért azonban megjegyezzük, hogy míg a TALIS-felmérés esetében a kérdés a tényleges gyakorlatra vonatkozott, addig a mi adatfelvételünkben a megfogalmazás az indokoltságot mérte. Továbbá a nemzetközi vizsgálat célcsoportja az ISCED2 szint (azaz a hazai felső tagozat) volt, esetünkben pedig a közoktatás valamennyi szintje.

Az intézményvezetőket arról is megkérdeztük, hogy mennyire indokolt a fenti-ekben tárgyalt értékelési szempontokat az iskolák értékelése során figyelembe venni. E tekintetben az érintett vezetők közül 4%-nyian nem válaszoltak érdemben a fel-tett kérdésre. Az intézményértékelés kapcsán az igazgatók hasonló válaszokat adtak, mint a pedagógusok szakmai értékeléséhez való használatot illetően, azzal az enyhe eltéréssel, hogy a tanulói teljesítménymérések esetében a „teljes mértékben indokolt” válaszok aránya egy kicsivel magasabb, összességében pedig az igazgatók 72%-a tartja inkább vagy teljes mértékben indokoltnak az iskolák értékelésekor figyelembe venni a különböző mérési eredményeket (lásd 12. ábra).

12. ábra

„Véleménye szerint mennyire indokolt az alábbiakat az iskolák értékelésekor figyelembe venni?” (2014) – %-os arányok a legalább egy szempontot érdemben értékelő intézményvezetők körében, N=577



Forrás: Intézményvezetői adatfelvétel, 2014

A tanulói teszteredményeknek az iskolák értékelésére való használata tehát valamivel elfogadottabb, aminek okát abban kereshetjük, hogy ezekben a nem diagnosztikus mérési eredményekben a pedagógusok munkáján és az iskolai tananyagon kívül nagyon sok más tényező is megjelenik, s az eredmény vagy eredménytelenség többnyire nem csupán egy pedagógus munkájához, hanem az iskola pedagógiai munkájához köthető. E tekintetben képzéstípus, településtípus és fenntartó szerint statisztikailag sem tapasztalható jelentős eltérés.

Összegzés

Tanulmányunkban elsősorban arra kerestük a választ, hogy az intézményvezetők és a pedagógusok miként vélekednek a tanulói teljesítménymérésekről, miként viszonyulnak a különféle mérésekhez és eredményességi mutatókhoz. Az eredmények azt mutatják, hogy a tanulói teljesítménymérések eredményeinek valamilyen mértékű felhasználása többé-kevésbé általánossá vált az iskolákban. Az iskolák kilenczetedében a tantestületen belül valamilyen formában közösen elemzik, megvitatják a mérési eredményeket. A leggyakoribb felhasználási területként a tervezési, fejlesztési és stratégiai jellegű használat, valamint a szülők és a tanulók tájékoztatása, a diákok fejlesztése és felzárkóztatása jelenik meg.

Elmondható, hogy mind az intézményvezetők, mind a pedagógusok inkább pozitív véleményeket társítanak a teljesítménymérésekhez, bár az intézményvezetők összességében kedvezőbben viszonyulnak a különböző mérésekhez. A pedagógusok bizonytalanabbak e tekintetben, kevésbé érzékelik annak pozitívumait, fontosabbnak tartják viszont hangsúlyozni az esetleges veszélyeket, az iskolán belüli esetleges feszültségeket és egyéb problémákat. A mérések hasznosságának megítélése egyértelműen kedvezőbb a középfokon működő iskolákban, továbbá azoknál a pedagógusoknál, akik magasabb képesítéssel rendelkeznek, illetve az e téren már a tanárképzés során tapasztalatokat szerző legfiatalabb és a hosszabb szakmai tapasztalatokkal rendelkező legidősebb korcsoportok esetében. A fiatalok vélhetően a pedagógusképzés során is jobban megismerkedhettek ezekkel a mérésekkel és azok eredményeivel, míg az idősebb korosztály tagjainál a hosszabb gyakorlati idő, a folyamatos és intenzív szakmai tapasztalatok erősíthetik a pozitív hozzáállást.

A megkérdezett intézményvezetők és pedagógusok szerint a pedagógiai munka és az iskolák értékelésekor a mérési és vizsgaeredményeknél rendre fontosabbnak az olyan „puhább” mutatók, mint a felzárkóztatásban, tehetségfejlesztésben való részvétel és annak eredményessége. A többé-kevésbé jól mérhető eredménymutatók közül leginkább a tanulmányi versenyeredményekre és a továbbtanulási

eredményekre tekintenek úgy, mint amelyek jó mutatói és értékelési szempontjai lehetnek az intézményi munka minőségének, de – főként a közvetlenül érintettek közül – a szakmai, illetve érettségi vizsgaeredményekről is sokan vélekednek így. A kompetenciamérésekkel és a nemzetközi tesztekkel kapcsolatban azonban már kritikusabbak az igazgatók és a pedagógusok is, ugyanakkor elmondható, hogy az intézményvezetők körében a tanulói teljesítménymérések elfogadottabbak a szakmai és iskolai munka értékelésének mutatójaként. Összességében mind az intézményvezetők, mind pedig a pedagógusok válaszaiból az körvonalazódott, hogy a tanulói teljesítménymérések, tesztek, de még a vizsga- és versenyeredmények is a tanulók és az intézmények teljesítményének csak egy részét képesek mérni, sok olyan fontos további tényező (pl. tehetséggondozás, felzárkóztatás, tanulói- és szülői elégedettség) van jelen az iskolák életében, amelyek szintén nagymértékben hatást gyakorolnak a pedagógusok és az iskolák munkájára, annak eredményességére. A tanulmányi teljesítményméréseken túl ezek mérése, nyomon követése szintén elengedhetetlen a szakmai munka minőségének értékelésekor.

Hivatkozások

- Balázi Ildikó és Horváth Zsuzsanna (2011): A közoktatás minősége és eredményessége. In: Balázs Éva, Kocsis Mihály és Vágó Irén (szerk.): *Jelentés a magyar közoktatásról 2010.* (pp. 325–362.), Oktatókutató és Fejlesztő Intézet, Budapest.
- Fehérvári Anikó és Széll Krisztián (2014): Méltányosság az oktatásban: tanulói eredmények, szülők, iskola. In: Széll Krisztián (szerk.): *Az OECD az oktatásról – adatok, elemzések, értelmezések.* (pp. 41–51.) Oktatókutató és Fejlesztő Intézet, Budapest. Elérhető: https://www.ofi.hu/sites/default/files/attachments/az_oecd_az_oktatasrol_ofi_2014.pdf [Letöltés dátuma: 2015.04.30.]
- Gyökös Eleonóra és Szemerszki Marianna (2014): Hol tart ma az oktatáseredményességi kutatás? *Új Pedagógiai Szemle*, 64(1-2), 43–64.
- Horn Dániel és Sinka Edit (2006): A közoktatás minősége és eredményessége. In: Halász Gábor és Lannert Judit (szerk.): *Jelentés a magyar közoktatásról 2006.* (pp. 341–375.) Országos Közoktatási Intézet, Budapest.
- Lannert Judit (2012): *TÁRKI-közvélemény-kutatás az oktatás területén a lakosság és a pedagógusok körében.* Kézirat.
- OECD (2009): *Creating Effective Teaching and Learning Environments: First Results from TALIS.* OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264068780-en> [Letöltés dátuma: 2015.04.30.]
- Sinka Edit (2006): A kompetenciamérés hasznosulása és fogadtatása az iskolákban. In: Lannert Judit és Nagy Mária (szerk.): *Eredményes iskola. Adatok és esetek.* (pp. 79–91.) Országos Közoktatási Intézet, Budapest.
- Szemerszki Marianna (2014): Mérés és értékelés az oktatásban. In: Széll Krisztián (szerk.): *Az OECD az oktatásról – adatok, elemzések, értelmezések.* (pp. 17–27.) Oktatókutató és Fejlesztő Intézet, Budapest. Elérhető: https://www.ofi.hu/sites/default/files/attachments/az_oecd_az_oktatasrol_ofi_2014.pdf [Letöltés dátuma: 2015.04.30.]

OKTATÁSKUTATÓ
ÉS FEJLESZTŐ
INTÉZET

