

Lannert Judit: A PISA adatok használata és értelmezése a módszertani kritikák tükrében¹

1. PISA módszertan.....	1
2. A PISA fogadtatása	4
3. PISA kritika.....	6
A PISA filozófiája érzéketlen az országok specifikumaira.....	6
A PISA globális nagyhatalommá vált	6
A PISA nem jól mér, módszertani problémák	8
Kultúrafüggetlen teszt, fordítási problémák.....	8
Mintavétel, válaszadási arány, hiányzó adatok	9
Különböző tesztfüzetek és itemek.....	10
IR modell, Rasch módszer	11
A PISA leegyszerűsít	14
4. Mire jó a PISA?.....	15
Az oktatási mítoszok leleplezése.....	16
Az adatok másodelemzése és egyéb adatokkal való összevetése a hazai sajátosságok felderítése céljából.....	17
5. Az elemzés korlátai és a PISA tudás társadalmasítása.....	18
6. Összegzés	20
7. Irodalom:	21

2000 óta az OECD egyik legnagyobb hatású nemzetközi programja a PISA néven ismert nemzetközi tanulói teljesítménymérés (International Student Assessment) szinte forradalmasította az oktatásról való tudásunkat. Ez a vizsgálat a mai napig egyedülálló módon teszi lehetővé, hogy időben és térben is összehasonlító elemzéseket tudjanak végezni a világ oktatási rendszereiről.² Napjainkra valódi nagyiparrá vált ez a terület. Már a kezdetekben is 32 ország több mint 250 ezer tanulójának teljesítményét mérték három (matematika, szövegértés és természettudomány) területen, napjainkra pedig már 510ezer tanuló tölti ki a teszteket a világ 65 országában és már nem csak a matematika, szövegértés és természettudomány területeken, de a digitális szövegértés, problémamegoldás és pénzügyi ismeretek területeken is. Egy ekkora vállalkozásnak csodálói, irigyei és ellenlábasai egyaránt akadnak. Már a kezdeti szakaszban is sok kritika érte a felmérést és a belőle kinyert eredményeket és ez mára sem csitult el. Tanulmányomban azt foglalom össze, hogy milyen kritikák érik ma is a PISA felmérést, és ezek hogyan árnyalják a mai PISA-ról vallott felfogásunkat és a PISA adatok felhasználásának célját és módjait.

1. PISA módszertan

Érdeemes kicsit részletesebben kitérni a vizsgálat módszertani sajátosságaira, mivel nagyon sok kritika ezen a téren éri a PISA-t. A PISA felmérés bevallottan nem az adott országokban tanított tananyag ismeretére kérdez rá, hanem arra, hogy a tanulók mennyire képesek

¹ Köszönettel tartozom Balázi Ildikónak és Kádárné Fülöp Juditnak a tanulmányhoz fűzött értő megjegyzéseikért.

² Az egyéni kompetenciák nagymintás felmérése nem előzmény nélküli. A PISA módszertanában sokat merített egyrészt az IEA (Evaluation of Educational Achievement) által folytatott TIMSS és PIRLS vizsgálatokból, másrészt az OECD és a kanadai statisztikai hivatal által celebrált felnőttírásbeliség (IALS) vizsgálatokból.

alkalmazni tudásukat valós élethelyzeteket szimuláló problémák megoldásában. Ezáltal nem pusztán a kompetenciákat méri fel, de a tanulóiban rejlő, az életben át tartó tanuláshoz szükséges potenciált is. A kérdőívek nem csak a tudásukat méri fel, de rákérdeznek benne a tanulók motivációira és tanulási stratégiáira is.³

Egy ekkora vállalkozásnak sokféle szempontot kell figyelembe vennie. A fő kihívások, amire a PISA szakértők szerint is válaszolni kell:

- egy olyan minta kialakítása, ami egyaránt megfelel az iskolai és a tanulói szintű mintavétel követelményeinek,
- a többnyelvű és kultúrájú közeg miatt kiemelt figyelmet kell fordítani arra, hogy a tesztek az eltérő kulturális kontextusban is működjenek,
- a tesztek kidolgozása, az adatok feldolgozása terén olyan minőségbiztosítás kell, ami garantálja, hogy az országok közti összehasonlításra alkalmas adatbázisok jönnék létre,
- megfelelő módszertan alkalmazása: a skálák létrehozása és az adatelemzés esetén a PISA jelentések az IRT (Item Response Theory) módszert alkalmazzák.

Lebonyolítók

2000 és 2012 között a vizsgálat megtervezése és lebonyolítása egy, az Ausztrál Oktatáskutató Tanács (ACER) által vezetett konzorcium feladata volt⁴, amelyben két belga, két német, két amerikai és egy-egy japán, norvég és luxemburgi cég vett részt. A 2012-es matematika kompetencia felmérés kereteit az amerikai Achieve cég dolgozta ki. A PISA felvétel lebonyolítását a PISA irányító testülete felügyeli, ahol minden résztvevő ország magas politikai szinten képviselteti magát. A kompetenciaterületekre szakértői bizottságok létesülnek, a helyi lebonyolítást a nemzeti PISA koordinátorok kontrollálják.

Kérdőív

Papír alapú kérdőíveket használnak általában, amit két órán át töltenek ki a tanulók, de némely országban ezen felül egy negyven perces számítógép alapú felmérést is végeznek a matematika, szövegértés és problémamegoldás területeken. A kérdések egy részénél többféle felajánlott megoldásból választhatnak tanulók, másik része pedig nyílt végű, ahol maguk alkotják meg válaszaikat. A kompetenciaterületeken kívül a felmérés háttéradatakra is kíváncsi, a tanulókat családjukról, iskolájukról és magukról is megkérdezi. Az iskolaigazgatók is kitöltenek egy kérdőívet az iskolarendszerről és az iskolai tanulási környezetről. Opcionálisan szülői kérdőíveket is felvesznek némely országban a gyermeküknek nyújtott támogatásról és a továbbtanulásával kapcsolatos elképzeléseikről.⁵

Minta

15 éves tanulókat mér fel a vizsgálat, pontosabban 15 éves és három hónapostól 16 éves és két hónaposig, alkalmazkodva a két földteke közötti évszakok és ezáltal tanév eltéréséhez. Minimum 150 iskolának szerepelnie kell a mintában, minden iskolában véletlenszerűen mintegy 35 tanulót választanak ki. Összességében országonként 4500 tanuló a mintanagyság. A válaszadási ráta iránti elvárás első körben 85%, amennyiben 65 és 85 százalék között marad az arány, lehetőség van új iskolák kiválasztására egy, a mintavételi eljárásban előre meghatározott listáról. Ha a kiválasztott iskolákban sok tanuló hiányzik, mód van

³ A PISA jelentések, adatbázisok és egyéb információk a www.pisa.org honlapon érhetők el.

⁴ 2009-ben a CITO-val megosztva volt az ACER a konzorcium vezetője.

⁵ 2015-től már kizárólag számítógépes teszteket fognak alkalmazni.

pótmérésekre. Ugyanakkor, ha mindezek ellenére sem jön össze a megfelelő válaszadási arány, akkor az ország eredménye nem kerül be a jelentésbe.⁶ A mintavétel a PPS módszert követi (probability propotional to size) ahol az iskolai minta kialakításában egyrészt az elsődleges (Magyarország esetén pl. az iskola típusa) és másodlagos országonként változó rétegzési szempontokat, valamint az iskolaméretet is figyelembe veszik. A kisiskolák problémájára is külön módszert találtak ki. A mintavétel a PISA felvétel egyik legkidolgozottabb része, ezt részletesen leírja a technikai jelentés. A koncepció kialakítása, a nyelvi és kulturális eltérések kezelése, a mintavétel megalapozottsága, az adatfelvétel biztonságos lebonyolítása és az adatok kezelésének és elemzésének módszertanilag megalapozott módszerei mind helyet kapnak a technikai jelentésben.

Testfejlesztés

A matematika, szövegértés és természettudomány standard itemjeit 2012-ben 13 klaszterbe rendezték (hét matematikai, három szövegértési és három természettudományi), mindegyik klaszter kitöltése 30 percet igényelt. Az itemeket 13 tesztfüzetbe rendezték, ahol mindegyik füzet négy klaszterből állt. A tanulókat minden országban véletlenszerűen rendelték hozzá egy-egy tesztfüzethez, akik két órán át töltötték ki a teszteket egy ötperces szünet beiktatásával. 2012-ben 2009-hez hasonlóan néhány, az OECD átlagtól jelentősen elmaradó ország könnyebb tesztfüzeteket használhatott.

A tesztek fejlesztési folyamatát előkészítette egy, a jól teljesítő országokban a matematikai standardokat felmérő survey. Az Archive ezek emellett konzultált szakemberekkel, valamint a résztvevő országokból 80 matematikus szakember reflexióját is begyűjtötték. Ezek alapján a korábbi keretrendszer felülvizsgálták és kialakították az újat. (Balácsi Ildikó et al., 2013).

A teszteket különböző tesztcentrumok készítették (ACER, Liege-i egyetem, három német központ, egy-egy japán, norvég, amerikai és luxemburgi cég). Ezen felül a résztvevő országok is küldhettek be feladatokat. A 110 kognitív matematikai item közt 36 volt olyan, amit 2003-ban is használtak (anchor item) és amelyek kapcsolatot teremtenek a vonatkozó évekre. 74 itemet újonnan fejlesztettek 2012-re. Az éles felmérést pilot tesztek és terepróba is megelőzi, a tesztek fordítása pedig több körben és ellenőrzéssel történik.

Skálák készítésének módszertana

A PISA azon a feltevésen alapszik, hogy létezik egyfajta logikus hierarchia az adatokban (Rasch modell). A válaszok alapján a tanulókat és az itemeket is besorolják képességszintekre illetve nehézségi fokozatokba. Azok az itemek, amelyeket többen oldanak meg, könnyebbnek minősülnek, mint amikre kevesebb jó válasz érkezik. A cél egy képességskála teljes lefedése, ami csak megfelelően nagy minta esetén hozható létre. Az adatok logaritmusos transzformációjával úgy határozzák meg egy személy és item képesség-, illetve nehézségparaméterét, hogy a közös logit skálán azokat az itemeket és személyeket teszik azonos szintre, ahol az adott személy képességparamétere alapján 50 százalékos valószínűséggel oldja meg helyesen az adott itemet (Molnár, 2005).

A Rasch modellt is magába foglaló ún. Item Response Theory (IRT) azzal az erős feltevésével él, hogy az itemek függetlenek az olyan lokális kontextusoktól, mint a tanuló érdeklődése vagy a nemzeti tananyag sajátosságai, vagyis egy adott kompetenciaterületen a tanulók inherens fix képességekkel, az itemek pedig országtól független inherens nehézségi fokozatokkal rendelkeznek. Így egy adott képesség esetén létezik egy globális skála, amire

⁶ Ilyen alapon esett ki egyszer az Egyesült Királyság is a PISA jelentésből.

mindenkit egyértelműen el lehet helyezni (Baird et al, 2011). Éppen ezért az itemeket elvileg ilyen szempontból is megvizsgálják, és amelyek nem illeszkednek ehhez a modellhez, azokat kiveszik. A nem jól működő kérdések csökkentése részben előzetes kiszűréssel, illetve item-splittinggel történik. Ez utóbbi esetben a megfelelően illeszkedő modell kialakítása érdekében országonként kiszámolták az infit és diszkriminációs koefficienseket és az itemek paraméter becsléseit összevetették 30 ország 15 ezer tanulójának a mintáján.

A klasszikus tesztelmélet szerint a mérés pontossága növelhető az itemek számának növelésével, így általában a teljesítménytesztek legalább 70 itemet tartalmaznak. Ugyanakkor nem mindig van lehetőség ilyen sok item lekérdezésére, a PISA tesztek során is nem minden kérdést válaszol meg minden tanuló minden országban. Így a tanulókra ún. valószínűsíthető (plausible) képességértékeket számítanak a válaszok, a háttérváltozók és a modell paramétereinek feltételes eloszlása alapján.⁷ Ezek az értékek nem tesztpontszámok a hagyományos értelemben, közvetítőként szolgálnak arra, hogy megfelelően tudják becsülni a populáció jellemzőit. Amennyiben a modell korrekt, akkor ezek a plauzibilis értékek megfelelően becslik a népesség jellemzőit (Yamamoto, 2002).

A PISA négylépcsős eljárásban alakítja ki az ország rangsorokat. Első lépésként kiszámolja az országonkénti Rasch paramétereket, majd 30 ország 15 ezer tanulójának véletlen almintáján ezeket a paramétereket az egész mintára is kiszámolja. A harmadik lépcsőben létrehozzák a plauzibilis értékeket, amelyek a mérni kívánt látens dimenzió poszterior eloszlása alapján létrehozott random számok. A negyedik lépcsőben varianciaanalízissel hasonlítják össze az országokat (Kreiner, 2012).

2. A PISA fogadtatása

A legtöbb kritika azért éri a PISA-t, mert – ahogy a bírálók fogalmazzák - az országokat egy dimenzió mentén sorba állítva leegyszerűsített képet ad egy ennél bonyolultabb világról, figyelmen kívül hagyva a nemzeti sajátosságokat. Sajátos paradoxon, hogy a PISA épp azért vált ennyire nagy hatásúvá, mert a média és a politika számára is egyszerű és hatásos üzenetet tudott közvetíteni. A rangsor könnyen értelmezhető és politikai célra jól hasznosítható termék. Ugyanakkor nem véletlen, hogy habár módszertani kritikák a kezdetektől fogva érték a PISA-t, ezek a kritikai hangok a 2012-es eredmények után erősödtek fel, amikor az ázsiai országok előretörésével Európa és az angolszász országok is hátrébb sorlódtak a PISA rangsorban. A legutolsó PISA felvétel eredményei alapján az első 10 ország közé nem került be angolszász ország, az első hét ország ázsiai (távol-keleti) és az Európai Unióból csak Hollandia fért be Svájc és Lichtenstein után. A finnek hátrébb kerültek és sokat rontottak a korábban jobb eredményt elérők, mint Új-Zéland, vagy Svédország.

A PISA eredmények fogadtatása persze sohasem volt konfliktusmentes. Alapvetően három tényezőtől függ, hogyan fogadja egy ország az eredményt. Egyrészt függ magától az eredménytől (jó eredmény esetén még nem hangzott el olyan, hogy esetleg mintavételi probléma okozhatta a kiugrást, míg romlás esetén elég gyakran hivatkoznak különböző okokra és vetnek fel kritikai megjegyzéseket). Másrészt függ attól, hogy politikai választások elé néz-e egy ország vagy nem. Előbbi esetben mindig kiélezettebb a fogadtatás és az adott érdekeknek megfelelő az eredmények értelmezése vagy éppen elhallgatása. Harmadrészt pedig függ a fogadtatás attól is, hogy vannak-e olyan szakemberek, akik rendelkeznek azokkal a kompetenciákkal, amik a PISA adatok elemzéséhez és értelmezéséhez szükségesek.

⁷ Nevezik ezt imputálásnak is, ami a hiányzó (missing) adatok helyettesítése egy becsült értékkel.

A PISA eredményekre való reagálás is sokféle, sok országot sokkhatásként ért egy adott rosszabb eredmény. Jellemző módon a rangsor végén nagy a mozgás, nagyon sok ország kiszáll a számára nem előnyös „versenyből”. India esete a legszembeötlőbb, akik két tagállammal vettek részt 2009-ben és általános megrökönyödésükre az utolsó előtti helyen végeztek, 2012-ben pedig már nem voltak a résztvevő országok között.⁸

Franciaország a felnőtt írásbeliség vizsgálaton elért rossz eredmény miatt félve vett részt a PISA első körében és nem is nagyon adtak az eredményeknek nagy nyilvánosságot (Baird et al., 2011). Csak a második felvétel után ébredt nagyobb érdeklődés az eredmények iránt és 2006-tól vált a politika számára is érdekessé. Ugyanakkor a szakma kevésbé volt felkészülve.⁹ Pszichometriai, klasszikus statisztikai kompetencia és szaktárgyi didaktikai tudás terén meglehetősen egyenetlen képet mutatott a francia oktatáskutatók köre és ez a tény nem kedvezett annak, hogy kialakuljon egy, a PISA módszertant is birtokló kutatói kör. Éppen ezért a franciák PISA szkepticizmusának kritikai hatásosságát csökkentette az elfogultság és a szakmai hozzáértés hiánya. Így a PISA eredményeket általában az éppen aktuális aktorok, döntéshozók saját érdekeiknek megfelelően használták, illetve értelmezték (Mons et al., 2009).

A norvég oktatási miniszter a 2000-es eredményeket e szavakkal fogadta: „Norvégia egy iskolai lúzer, ez most már dokumentálva is van. Olyan ez, mintha a téli olimpiáról aranyérem nélkül jönnének haza. Ráadásul még csak azt se mondhatjuk, hogy a finnek doppingoltak.” (Sjoberg, 2012). A 2008-as Norvégiáról szóló OECD jelentés azt tanácsolta Norvégiának, hogy csukja be a kisiskolákat, növelje az osztályméretet, mérje az iskolák teljesítményét és az eredményeket tegye nyilvánossá.

A PISA eredményekre – nemcsak Norvégiában – gyakran építettek oktatási reformokat. Az, hogy milyen üzenetet vettek le az adott országok, nagyrészt az adott politikai és szakmai érdekek és kompetenciák konstellációjának függvényében alakult. A PISA eredmények ilyen hatását a nemzeti oktatáspolitikákra sokan nehezményezik. A mérési technikákban járatlanabb, vagy a kutatási diszciplinájuk visszaszorulásától féltő kutatók illetve az autonómiájukat féltő tanárok egyaránt megtalálhatók e körben. Utóbbiak legnagyobb félelme, hogy a neoliberális globalizáció és a piaci ideológia térnyerése csökkenti az oktatásra szánt költségvetési pénzeket és a külső elszámoltathatóság irányába mozdítja a rendszert (Froese-German, 2010).

Mindazonáltal a PISA-nak az oktatáspolitikára vonatkozó hatását elemző kutatók kiemelik, hogy a PISA megkerülhetlenné vált, hiszen olyan tényeket nyújt, amelyek – amennyiben jól értelmezik őket – nagyban segíti a jövőorientált oktatáspolitikák kialakítását. A rangsorok okozta „naming, blaming, shaming” hatáson túl a PISA normává vált és ereje elsősorban nem a bünbakképzésben rejlik, hanem abban, hogy a reform, a pozitív változás lehetőségét villantja fel. A PISA a szakmapolitikai problémák és megoldások igazolásához egy vélt univerzális, független szakmai tudás komfortját biztosítja. Az univerzalizmus és tudományosság együttese az, amely megkerülhetlenné teszi – vélik az elemzők (Carvalho-Costa, 2014).

⁸ <https://www.tes.co.uk/article.aspx?storycode=6376876> letöltve: 2015 április 10. és <http://ajayshahblog.blogspot.hu/2012/01/first-pisa-results-for-india-end-of.html> letöltve: 2015 április 10.

⁹ <http://knowandpol.eu/IMG/pdf/o31.pisa.france.pdf>

Ugyanakkor a PISA rangsor átalakulása okozta szorongás napjainkra felerősítette a PISA módszertant évek óta kritizálók táborának hangját is. A továbbiakban összefoglaljuk a PISA módszertant ért legfontosabb kritikákat az elmúlt évtizedben.

3. PISA kritika

A PISA-val kapcsolatos ellenvetéseket három – bár egymástól nem független - fő csoportra oszthatjuk:

A PISA filozófiája érzéketlen az országok specifikumaira

A bírálatok e csoportjának a legnagyobb ellenvetése az, hogy a PISA filozófiája különbözik attól, amit egy ország oktatása követ és az ország nevelési specifikumait figyelmen kívül hagyva elsősorban a munkaerőpiaci igényekre reflektál. A tesztek nem kultúra függetlenek, ha egy tanuló soha nem látott ATM automatát, akkor – volt az indiaiak érve - az erre vonatkozó kérdést sem értheti. Nem mérnek minden kompetenciát, holott ennél sokkal többről szól az iskola és tanulás, így a nemzeti és tantervi sajátosságok figyelembevételének a hiánya értelmezési problémákat is felvet. A norvég tanulók például nem találkoznak az olyan számukra fontos témákkal, mint az olajkitermelés északon, a vízgazdálkodás és haltenyésztés, vízi erőművek, természetmegőrzés, északi fény, az északi félteke klímaváltozása, stb. (Sjoberg, 2012). A frankofón kritikusok pedig a kutatási design, a tanácsadók és a tesztek terén érvényesülő angolszász hatást és annak torzító hatását vetik-vetették fel (Goldstein, 2004). Fontos megjegyezni, hogy a PISA vállaltan nem méri a tantárgyi kompetenciákat, ugyanakkor az eredeti kompetenciaterületeket mára már kiegészítette a problémamegoldó készség, pénzügyi tudatosság, digitális szövegértés területeivel és jelenleg tervezik a kreativitás mérésének bevezetését.

A PISA globális nagyhatalommá vált

Gyakori, az összeesküvéselméleteket idéző kritika az is, hogy a PISA globális nagyhatalommá vált és negatív hatással van az országok oktatási rendszerére (Sjoberg, 2012). A PISA olyan monopóliummá vált, amelynek hatalma túl nagy és ellenőrizhetetlen, az oktatásra pedig káros hatással van (Hopman-Brinek, 2007). Az a tény, hogy a 2015-ös PISA vizsgálat tartalmi keretének fejlesztését a Pearson konszern fogja végezni¹⁰, csak olaj volt a tűzre. A Pearson tanácsadó testületének tagjait sokan a legveszélyesebb oktatáspolitikusok közt tartják számon.¹¹ Így köztük van Michael Barber, korábban a McKinsey tanácsadó cég fő tanácsadója, Eric Hanoushek, ismert oktatással foglalkozó amerikai közgazdász és Andreas Schleicher az OECD PISA-ért felelős vezetője.¹²

2014 tavaszán több mint 100 akadémikus nyílt levelet intézett Andreas Schleicherhez, a PISA negatív hatásaira figyelmeztetve. Az aláírók többsége amerikai vagy német, ezen felül több ír, angol, új-zélandi, osztrák és svájci található, tehát egy angolszász-német csoport, ahol nem igazán találunk ázsiait (kivéve egy japán professzort), közép-európaiakat vagy latin nyelvű

¹⁰ <http://www.pearsoned.com/news/pearson-to-develop-frameworks-for-oecd-pisa-student-assessment-for-2015/#.U3ME4q1dWFw>

¹¹ <http://dianeravitch.net/2014/12/02/pearson-created-the-international-pisa-test-for-2015/>

¹² Fontos személy még a PISA szempontjából Irwin Kisch, a Globális Értékelési Központ (Centre for Global Assessment, ETS) vezetője, aki a PISA módszertan egyik fő kialakítója.

ország képviselőjét. Az angolszász országok esetén (elsősorban az Egyesült Királyság és az USA) a tesztek iránti ellenszenvet a PISA teszteken romló teljesítményük mellett nagy valószínűséggel a nemzeti oktatáspolitikáik erősen piaci és tesztbarát jellege okozhatja, míg a németek esetén régi hagyománya van a standardizált tesztek iránti antipátiának, lévén a német neveléstudományi tradíciók sokkal inkább az elméleti-filozófiai irányt részesítik előnyben. Mindenesetre figyelemreméltó, hogy megnyerték maguknak Noam Chomsky-t is, az MIT nyelvész-filozófus professzorát.

Bírálatukban kiemelik, hogy a PISA – bár módszertanában hagy maga után kívánnivalót – mégis filozófiája annyira elterjedt, hogy világszerte túlzott mértékben dominánsá vált a standard tesztek használata és az erre alapuló döntéshozatal az oktatáspolitikai terén. A PISA hároméves ciklusa szerintük felerősíti a rövidtávú szemléletet, miközben az oktatásban zajló folyamatok jóval hosszabb időtávot igényelnek. A mérhető területekre való koncentrálása miatt így olyan oktatási célok kerülnek háttérbe, mint a fizikai, erkölcsi vagy művészeti nevelés. A legfőbb kritika az, hogy az OECD túlzottan is az iskolák és az oktatás gazdasági szerepét hangsúlyozza és hatást próbál gyakorolni az országok oktatáspolitikájára, holott – ellentétben az UNESCO-val vagy UNICEF-fel – az OECD-nek nincs ilyen felhatalmazása. Az OECD ezáltal olyan multinacionális cégekkel kerül szoros kapcsolatba, amelyek ebből profitot húznak és – elsősorban az észak-amerikai kontinensen – már így is túlzottan jelen vannak az iskolák életében. Végezetül a legfontosabb ellenvetésük, hogy a PISA rezsim a folyamatos globális teszteléssel árt a tanulóknak és elszegényíti az osztálytermi folyamatokat és csökkenti a pedagógus autonómiáját. Ezáltal növelik az iskolák stressz szintjét és veszélyeztetik a tanulók és tanáraik jóllétét.¹³

Ugyanakkor több javaslattal is előállnak, mint

- alternatív rangsorok létrehozása, amely jobban értelmezhető és kevésbé érzékeny, mivel nincs értelme összehasonlítani a fejlett országokat azokkal a fejlődő országokkal, ahol a gyermekmunka még mindennapi gyakorlat.
- a PISA által képviselt tudást társadalmisítani szükséges, be kell vonni az oktatás terén tevékenykedőket. Ma az oktatás terepén legnagyobb befolyással a pszichometriához, statisztikához és közgazdaságtanhoz értő szakemberek bírnak. A pedagógusokat és szülőket is be kell vonni – mondják - (Mortimer, 2009), ugyanúgy mint az olyan diszciplínák képviselőit, mint az antropológia, történelem, filozófia, nyelvtudomány, művészetek.
- a mérési standardok és módszerek kialakításába be kell vonni azokat a nemzetközi és nemzeti szervezeteket is (pl. ENSZ), amelyek nem gazdasági szempontokból közelítik meg a közoktatást, hanem a jóllét, boldogság, egészség szempontjait is érvényesítik.
- tegyék nyilvánossá a PISA költségeit, hogy az adófizető polgárok eldönthessék akarják-e a pénzüket tesztekre költeni.
- független nemzetközi monitoring teamekre lenne szükség, hogy a koncepcióalkotástól a tesztek kialakításán át a konkrét pontozásig az elfogultság és igazságtalan összehasonlítások vádjá ne merülhessen fel.
- részletes beszámolás/elszámolás szükséges a profitorientált cégek részvételéről.
- le kell lassítani a tesztgyárat. Ahhoz, hogy kellő mélységben lefolyhassanak a szükséges viták egy ciklust ki kell hagyni a PISA mérésekből. Ez elég időt adna arra,

¹³ <http://oecd-pisaletter.org/2014/04/28/open-letter-to-andreas-schleicher-oecd-paris/>

hogy a kollektív tanulási folyamat eredményét beépíthessék egy javított értékelési modellbe.

Andreas Schleicher válaszában kifejti, hogy a PISA éppen a rövidtávú szemlélet ellen dolgozik és új perspektívákat nyit a nemzetközi összehasonlítás segítségével, mint ahogy elősegíti olyan nemzetközi fórumok létrejöttét, ahol ezeket a kérdéseket meg lehet vitatni. S habár valóban az oktatás terén a változás nem mindig gyors, éppen a PISA eredmények igazolják, hogy rövid idő alatt is lehet látványos fejlődést produkálni, mint ahogy ezt Lengyelország vagy Németország példája mutatja.¹⁴

A PISA nem jól mér, módszertani problémák

A módszertani kritikák is elsősorban azt a kérdést feszegetik, lehetséges-e egy többnyelvű és kulturájú közegben kontextustól függetlenül mérni a tanulók és ezáltal az országok teljesítményét. A mintavétel, a tesztek fordítása, a Rasch módszer megfelelő alkalmazása, a rangsorok mind ösztűz alá kerültek. Egyre többen számolják újra az adatokat más modellek alapján és ennek eredménye általában az, hogy a rangsor eleje és vége ugyan nem változik, de a középmezőnyben nagyon nagy eltérések is tapasztalhatók attól függően, hogy milyen módszert használnak.¹⁵ Ezért azt állítják, hogy a PISA nem jól mér. Az országok nem összehasonlíthatók egy egy dimenziós skálán, a rangsorok félrevezetőek, attól függően milyen tesztkérdéseket választunk, eltérő rangsorokat kaphatunk. A mintavétel során is történnek hibák, a Rasch módszer nem megfelelő (Goldstein, 2004; Kreiner, 2010).

Az alábbiakban a módszertani kritikákat vesszük sorra.

Kultúrafüggetlen teszt, fordítási problémák

Habár a PISA tesztek fordításának komoly procedurális minőségbiztosítása van, mégis újra és újra felmerülnek fordítási, interpretálási problémák. A skandináv országok esetén a Dolly klónozásáról szóló történet finn, svéd és norvég fordítása alapvető eltéréseket tartalmazott (Sjöberg, 2014). Az ugyanolyan nyelvterületen lévő, de különböző országok esetén is meglepő eltéréseket lehet tapasztalni, így az osztrák és német tesztek összehasonlításánál is találhatóak különbségek (Wutke, 2007). Ezt annak tudják be, hogy dacára a komoly procedurális eljárásnak, a fordítók gyakran kifutnak az időből és a kapkodás eredménye a fordításbeli eltérések.

Az sem mindegy, hogy hogyan vannak megfogalmazva a kérdések. Egy francia tanulmány szerint annak a kérdésnek, hogy „Quels taux...?” egy direkter formában való feltevése („Énumérez tous les taux...”) 31%-kal megváltoztatta a helyes válaszok arányát (Blum–Guérin-Pace, 2000).

A 2000-es adatbázison az angol és francia tanulók adatait kétdimenziós többszintű bináris faktorelemzéssel újraszámolva Goldstein azt találta, hogy a francia diákok ugyanarra a kérdésre nyílt végű formában szisztematikusan jobban válaszoltak, mint az angol diákok, míg ez utóbbiak ugyanazon a területen a több választási lehetőséget kínáló (multiple choice)

¹⁴ <http://www.theguardian.com/education/2014/may/06/academics-international-school-league-tables-killing-joy-of-learning>

¹⁵ Leslie és David Rutkowski számításai alapján például Magyarország Észtország után a 8. helyre került és Németország mögöttünk van néhány helyvel. <https://ec.europa.eu/jrc/sites/default/files/2014-pisa-learning-iu.pdf>

itemeknél voltak jobbak a franciáknál. Ez arra utal, hogy a kérdezés módja sem kultúrafüggetlen (Goldstein, 2004).

A 2000-es PISA tesztben szerepelt az influenza elleni oltásról egy szöveg. Habár a válaszokat a szöveg alapján kellett megadni, mégis – az adott ország kultúrájának és tananyagának is betudhatóan – a francia tanulók a technológiát, a német tanulók a természet erejét pártolták (Wutke, 2007).

Mintavétel, válaszadási arány, hiányzó adatok

A PISA a 15 éves tanulókat méri, de sok országban a 15 évesek egy része már nincs az iskolarendszerben. Mexikó, Törökország esetén a beiskolázás ebben a korban már 60 százalék alatt van (Wutke, 2007). Ilyen jellegű torzításnak tudható be valószínűleg az is, hogy az arab országok esetén a lányok teljesítménye átlagon felüli, de valószínűleg csak a legjobb helyzetben lévő családok lánygyermekei találhatók iskolában ezekben az országokban ebben az életkorban.

A PISA szabályai megengedik, hogy a népesség 5 százalékáig a tanulási vagy magatartási zavaros tanulókat kivegyék a mintából. Ez az arány ugyanakkor nagyon változó, az OECD országok közt 0,7%-7,3%-ig terjed. Kanadában, Dániában, Új-Zélandon, Spanyolországban és az USA-ban meghaladja az 5 százalékos megengedett limitet. Minél nagyobb arányú a kivétel, annál inkább emelkedik az átlagos pontszám. Van, ahol az ilyen tanulók egy részét nem kivették a mintából, hanem rövidebb tesztet írtak velük, amire a PISA lehetőséget adott. Így például Ausztria esetében 2003-ban az adott népesség 1,6%-át vették ki a mintából és 0,9% írt rövid tesztet. Magyarország esetén ezek az arányok 3,9% és 6,1% voltak. Ennek megfelelően a rövid tesztet írók esetében a magyar tanulók 200 ponttal többet produkáltak, mint az osztrák tanulók¹⁶(Wutke, 2007).

Előfordul, hogy bizonyos rétegek nincsenek kellőképpen reprezentálva a mintában. Ezt általában súlyozással korrigálják, de nem mindig veszik észre a hibát. 2000-ben az osztrák mintában például felülreprezentált volt a szakképzés, vagy Korea esetén feltűnően alacsony a lányok aránya, mintegy 40% (Wutke, 2007). Ráadásul a súlyozás nem tudja korrigálni azokat a láthatatlan torzításokat, amit például az okoz, ha szisztematikusan a legrosszabbul teljesítők hiányoznak a felmérés napján (Baird et al, 2011)

Van, amikor feltűnően hiányoznak a hiányzó adatok (missing). 2003-ban Lengyelország esetén nem volt egy tanuló sem, aki kevesebb, mint 25 kérdésre válaszolt volna és van hét olyan kérdés, amire mindenki válaszolt. A külső szemlélőben felmerül a gyanú, hogy a hiányos tesztfüzeteket nem értékelték (Wutke, 2007).

A kétórás teszt meglehetősen megterhelő, így nem meglepő, hogy gyakran nem tudják befejezni a tanulók és így sok hiányzó (missing) adat is van. Ugyanakkor ennek aránya nagyon változó országonként. 2003-ban a negyedik blokkban az el nem ért kérdések aránya a hollandok esetén 1%, a mexikói tanulók esetén több mint egynegyed. A meg nem válaszolt kérdések aránya a holland tanulók esetén az első blokkban 2,5%, a negyedik blokk esetén 4%,

¹⁶ 2003-ban a 15 évesen még általános iskolában tanulók rövid tesztet írtak Magyarországon. Később ezt megszüntették, mert az évvizsga önmagában nem indokolta a könnyített füzet írását. Ugyanakkor nem csoda, hogy a rövid füzetben a magyar diákok jobb eredményt értek el, mint az osztrákok. Ausztriában valószínűleg csak azok töltötték ki ezeket, akik valóban rászorultak.

ugyanaz az arány a görög tanulók esetén 11%-ról 24,4%-ra duzzadt. Az idő szűkét nagyon különbözőképpen menedzselik a különböző országok tanulói. A holland tanulók szinte minden kérdést megpróbálnak megválaszolni, de a kérdőív végére megnő a találgatás aránya, az osztrák és német tanulók sok kérdést átugranak, de így a többi kérdés megválaszolására elegendő idő jut. A görög tanulók éppen ellenkezőleg, meglepődnek amikor szembesülnek, hogy már kevés idejük van. Az első blokkban még jobb válaszokat produkálnak, mint a portugál tanulók, de az utolsó blokk esetén az el nem ért és hiányzó válaszok aránya eléri a 35 százalékot, ami miatt a PISA rangsorban le is csúsznak. Ez is egy olyan dimenzió, az elfáradás és időmenedzsment, aminek a hatását nem méri a PISA modell (Wutke, 2007).

A PISA mintáját legfrissebben ért kritika Shanghai kiemelkedő teljesítményére reflektál. Tom Loveless egy harvardi professzor megkérdőjelezte a kínai eredményt¹⁷, mondván, több mint gyanús, hogy bár 2009-ben 12 kínai tartomány is részt vett a PISA-ban, de csak a Shanghai eredmények lettek publikusak, ahol bár a 15 évesek hivatalos beiskolázási aránya 84 százalék, a 23 milliós városban csupán 108 ezer 15 évest tartanak számon (összehasonlításképpen a 10 milliós Magyarországon is hasonlóan 111 ezer 15 éves van). A furcsaság magyarázata, hogy a migráns tanulókat, gyerekeket nem számolják be a statisztikába, és ők a PISA felvételen sem vesznek részt. Andreas Schleicher ingerült válaszában visszautasítja azt a gyanút, hogy Kína csalna, és leszögezi, hogy Kína egy gyorsan fejlődő ország, ahol a rezidens migránsok éppen olyan mértékben vannak reprezentálva a mintában, mint más országok esetében is. Bár bevallja, hogy valójában a 2012-es minta a shanghai tanulók 79 százalékát öleli fel, de ez nem precedens nélküli, hiszen az USA esetében is csak 90 százalék ez az arány, ahol ráadásul Puerto Rico – amely valószínűleg nem felfelé húzná az átlagot (sic!) - nincs is benne a mintában.¹⁸

Különböző tesztfüzetek és itemek

Sokaknak talán meglepő, de a PISA felvétel során a tanulók nem egyforma feladatokat oldanak meg és nem is minden kompetenciaterületen töltenek ki tesztfüzetet. Többféle tesztfüzet készül, hogy elegendő számú item legyen a modell megalkotásához, ugyanakkor életszerűen egy tanulóra nem juthat életkorához képest megterhelő számú kérdés. Mintegy 165 itemből 60-nál több nem jut egy tanulóra. Ugyanakkor a tesztfüzetek nehézségét nem mindig sikerül egyformára kalibrálni, legalábbis erre utal, hogy sokszor ugyanazon területet mérő különböző tesztfüzetek kitöltése eltérő eredményeket produkál. Az egyik legnagyobb különbséget az USA-ban tapasztalták, ahol az egyik tesztfüzetet kitöltők átlagosan 444 pontot értek el, míg a másikon dolgozók 512 pontot. Ez felveti annak a gyanúját, hogy az itemek csoportosítása gyakran önkényes és nem eléggé homogén (Wutke, 2007)

A tesztfüzetek mindegyike tartalmazza az adott ciklus fő kompetenciaterületét (2012-ben ez a matematika volt), de a többi kompetenciaterület nem mindegyik tesztfüzetben szerepel. Ugyanakkor minden tanulóra megadják mindegyik kompetenciaterületen a plauzibilis értékeket, amit a tanulóknak a más kompetenciaterületeken adott válaszai és egyéb adatok alapján becsülnek. Ez a fajta eljárás, amennyiben korrekt, csökkenti az átlag standard hibáját. Ugyanakkor érdekes eltérések is elfordulnak, Görögország esetén 2003-ban a hivatalos természettudományi kompetencia pontszám 481 volt, ami 16 ponttal volt magasabb, mint

¹⁷ http://www.brookings.edu/research/papers/2013/12/11-shanghai-pisa-scores-wrong-loveless?rssid=brown&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%253A%2bBrookingsRSS%252Fcenters%252Fbrown%2b%28Brookings%2bCenters%2b-%2bBrow

¹⁸ <http://oecdeducationtoday.blogspot.hu/2013/12/are-chinese-cheating-in-pisa-or-are-we.html>

azon görög tanulóknak az átlagos teljesítménye, akik valóban kitöltöttek természettudományos tesztkérdéseket is (Wutke, 2007).

A tesztkérdések közül – a Rasch modell feltételeinek megfelelően – kiszűrik azokat, amelyek országoként vagy országcsoportonként másképpen működnek. Ez azt is eredményezi, hogy viszonylag kevés kérdés marad ahhoz, hogy megnyugtatóan tudjanak mérni a tesztelmélet szerint. 2009-ben 131 szövegértési, 35 matematikai és 53 természettudományos item szerepelt a tesztfüzetekben. Ezt sokan kevésnek tartják és a PISA egyik gyengeségének is egyben (Baird et al, 2011).

IR modell, Rasch módszer

Az IR modellt azért kritizálják, mert egydimenziós, éppen ezért csak akkor működik, ha minden kérdés esetében igaz, hogy a jobb képességű nagyobb valószínűséggel találja el a helyes választ, mint a gyengébb képességű. Goldstein (Goldstein, 2004) kritizálta ezt a fajta modell eljárást és többdimenziós modellek használatát javasolta, amikor azt feltételezzük, hogy több dimenzió húzódik meg egy egyén teljesítménye mögött és a teljesítménye ezen dimenziókban eltérhet. Az IRT model alapfeltevését, a feltételes függetlenséget számosan az adatelemzés legkritikusabb részének tartják. A feltételes függetlenség nem más jelent, mint hogy az azonos képességű válaszadók azonos valószínűséggel válaszolnak a kérdésekre, függetlenül egyéb, csoportjellegű tényezőtől (Yamamoto, 2002).

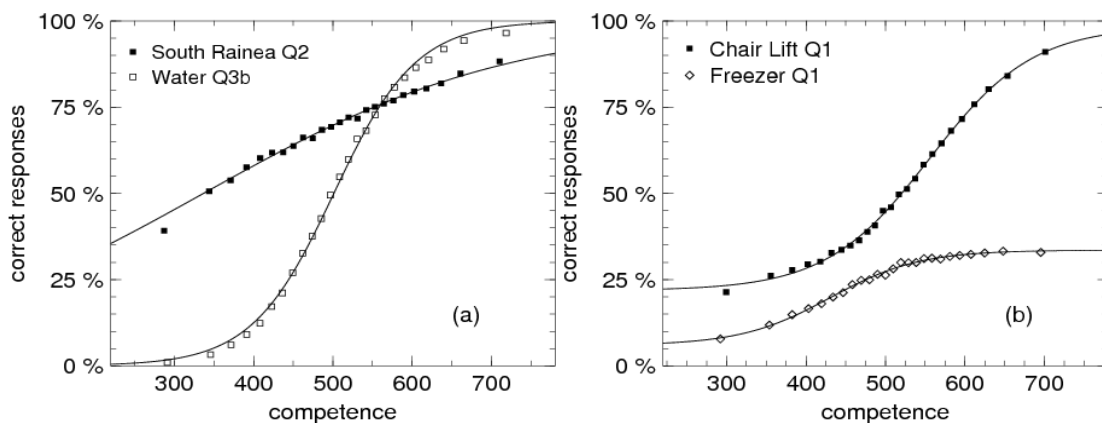
Többen is szóvá teszik, hogy a tanulókhöz rendelt értékek és az országátlagok számításai nem szerepelnek a technikai jelentésben. Az IR model alapján minden egyénre egy összesítő értéket hoznak létre, de ezek az értékek származtatott adatok és erről a folyamatról keveset tudunk és a hivatalos technikai jelentések is keveset árulnak el. (Brown, 2007, Kreiner, 2007). Goldstein szerint a többszintű elemzés módszerét kellene használni a PISA adatok esetén, vagyis figyelembe kell venni, hogy nem pusztán tanulói szintű, de iskolai szintű adatokkal is dolgoznak. Ezen felül javasolja a PISA adatok keresztmetszeti jellegét átalakítani oksági magyarázatokat is lehetővé tévő longitudinális panel vizsgálattá (Goldstein, 2004). Ez utóbbi minden bizonnyal szinte teljesíthetetlen, hiszen egy ilyen panelmintát egyrészt meglehetősen bonyolult és drága létrehozni, másrészt a minta lemorzsolódásának nagy a veszélye.

A PISA nagyméretű adatbázis, így az eredmények szignifikánsak. Ugyanakkor ez nem jelenti azt, hogy megbízhatók és érvényesek is. Hiába a nagy elemszám, ez nem csökkenti azt a sztohasztikus torzítást és bizonytalanságot, ami a PISA adatbázisban ott van (Wutke, 2007).

Többen is újraszámolták az adatokat az adatbázis alapján, hogy teszteljék valóban igaz-e, hogy – amint azt a PISA modellje feltételezi –, az egy kompetenciaterülethez tartozó itemeknek ugyanolyan alakzatot kell felvenniük, ahol az egyedüli eltérést horizontálisan csak az item nehézségi foka jelentheti. Általában ettől eltérő alakzatokat tapasztaltak kutatók több item esetén (Wutke, 2007; Kreiner, 2010).

Wutke négy – ugyanazt a kompetenciaterületet vizsgáló - itemet vizsgálva azt találta, hogy csak az egyik (Water q3) tudja valóban jól megkülönböztetni az adott területen a jobb és gyengébb képességű tanulókat (lásd 1. ábra). A Chair Lift Q1 item esetén azt láthatjuk, hogy alacsony kompetenciaszintnél szinte egyenes a görbe. Ez tipikus jele a találgatásnak. A Freezer Q1 item viszont már 35 százaléknál eléri a telítettséget. Ez arra utal szerinte, hogy a tanulók többsége nem találta ki, mire is kíváncsiak a tesztet készítőik. A South Rainea Q2 itemnek gyenge a képesség szerinti megkülönböztető (diszkriminációs) ereje, aminek több

magyarázata is lehet. A különböző alpopulációk esetén a nehézség foka különbözik, vagy a megoldási stratégiák különböznek alcsoportonként, jól sikerült találgatás is lehet mögötte vagy a mérendő látens képesség és a terület itemjeinek többsége között gyenge a korreláció (Wutke, 2007). Ez mind arra utal szerinte, hogy az utóbbi három item esetén további paraméterekre van szükség az értelmezésre. Az ábrán a folytonos vonalak többparaméteres modellek illeszkedését mutatja, ami még mindig tartalmaz lineáris mozgást, ami értelmezhető az item nehézségi fokaként. Ugyanakkor ezek a legjobban illeszkedő modellek +/- 30 pont eltérést mutatnak a hivatalos Rasch modell szerinti nehézségi fokokhoz képest. Amennyiben elfogadjuk a több paraméteres modellt, a tanulók rangsorba állítása máris önkényessé válik. Érdekes azt is megfigyelni, hogy a „South Rainea” és a „Chair Lift” kérdések esetén az első megfigyelt érték a görbe alatt található, vagyis a leggyengébb 4% még gyengébben teljesít, mint ahogy azt a modell feltételezi. Ezt magyarázhatja a kooperáció hiánya is, vagyis a tanulók nem akarják kitölteni a kérdőívet, vagy nem veszik komolyan. Ez is egy olyan dimenzió, amit a PISA egydimenziós Rasch modellje nem vesz figyelembe (Wutke, 2007)



1. ábra

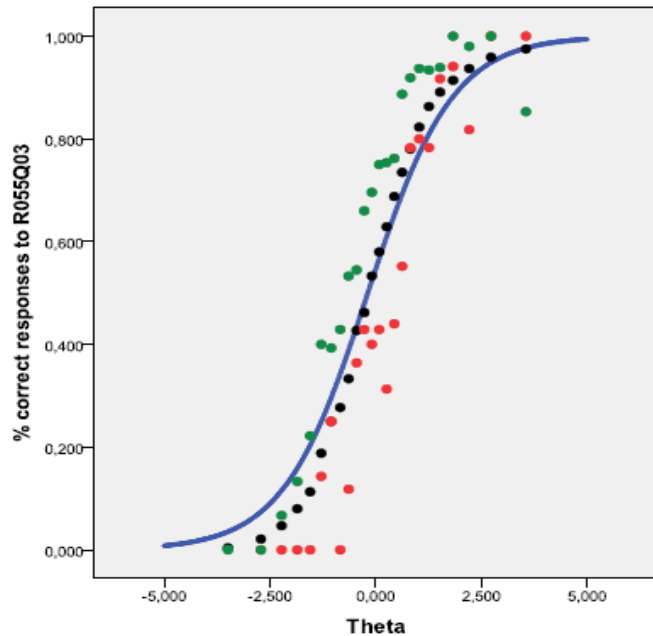
Néhány item jellemzője, ami a Rasch modelltől való eltérést mutatja. A folytonos vonal az (a) ábrán egy – a különbségeket magyarázó - két paraméteres modellre illeszkedik míg a négy-paraméteres (b) a találgatás és félreértést is engedő modellre illeszkedik.

Forrás: Wutke, 2007

A legnagyobb visszhangja egy Rasch tanítvány dán matematikusnak, Svend Kreiner kritikájának lett, aki a 2006-os PISA adatokat az angol és dán tanulók mintáján újraszámolta és azt találta, hogy dacára annak, hogy a PISA előzetesen megszüri a nem homogén módon működő itemeket, mégsem sikerült ezt a problémát megoldani. Az a tény, hogy a PISA által használt skálázási modellt alátámasztó tények és információk nincsenek publikálva és hogy az itemek egy része nem egyformán működik (DIF – differential item functioning), csökkenti a PISA rangsor hitelességét. Ez különösen akkor okoz zavart, ha viszonylag sok a DIF item és viszonylag kevés a rendszeresen feltett és megfelelően működő törzskérdés (anchor item). Az adatok robusztussága szintén függ a modell koncepciótól. Kismintás adatbázisok esetén a DIF itemeknek betudható szisztematikus torzítás/hiba elhanyagolható a nem szisztematikus véletlen hibákhoz képest. A nagymintás adatbázisoknál viszont kicsi a standard hiba, így egy kisebb számú nem megfelelően működő item is nagy hatást gyakorol a rangsorra.

Az eredményeket befolyásolja a hiányzó adatok nagysága és iránya is. Az angol és dán diákok válaszait újraelemelve Kreiner arra az eredményre jut, hogy az itemek különbözőképpen működnek és az adatok nem véletlenszerűen hiányoznak. A 2006-os PISA hatodik

tesztfüzetében lévő szövegértési itemjeinek vizsgálata azt mutatta, hogy az angol tanulóknak (zöld szín) szisztematikusan könnyebben ment a megoldás, mint a dán fiataloknak (piros), vagyis az itemek nem tekinthetők feltételesen függetlennek (lásd 2. ábra) Éppen ezért Kreiner a PISA rangsor módszertani megalapozottságát nem tartja megfelelőnek (Kreiner, 2012).



2. ábra

Becsült item tulajdonsággörbe az RN55Q03 esetén, ahol a valódi értékeket mutatják a színek, fekete=minden tanuló, zöld=UK, piros=Dánia

Kreiner eredményei a szövegértés területére érvényesek, a matematika és természettudományos területek működését nem vizsgálta. Fontos kiemelni, hogy bár a PISA rangsorokat nem tartja korrektnek, ugyanakkor a PISA adatok alapján megfelelő módszerekkel van mód összehasonlításra bizonyos országok, országcsoportok esetén. Ehhez meg kell találni a megfelelően működő itemeket és ezeken lehet – leginkább az egymáshoz kultúrában közel eső országok esetében – az összehasonlítást elvégezni (Kreiner, 2012).

Az OECD válasza Kreiner kritikáira

Az OECD nevében Ray Adams, az ACER egyik munkatársa válaszolt egy nyolc oldalas levélben (Adams, 2011). Ebben alapvetően – azon túl, hogy számtalan tárgyi tévedésre hívja fel a figyelmet – azzal utasítja el a kritikát, hogy Kreiner a 2006-os adatoknak csak egy részén végezte el a vizsgálatot és csak két ország esetén. A 28 itemből csak nyolcat vizsgált, így a szövegértési itemek négy kis alszettje túl kis méret ahhoz, hogy a következtetései módszertanilag megalapozottak legyenek. Egyben elutasítja az itemek tisztítására és csökkentésére vonatkozó javaslatot azzal, hogy ez bár módszertanilag lehet, hogy megbízhatóbb, de olyan kicsivé zsugorítaná a vizsgálat terepét, amitől a résztvevő országok és vezetőik elveszítenék érdeklődésüket iránta. Ha az adatok nem illeszkednek kellőképpen a modellhez, akkor el lehet hagyni a kevésbé működő adatokat, vagy egy jobban működő modellt lehet választani, vagy fel kell vállalni, hogy az eredmények akkor is hasznosak, ha az adatok és a modell illeszkedése nem tökéletes. Az OECD – lévén az első kettő kevésbé kivitelezhető - ezt a harmadik opciót választja.

Az alternatív modelleket ugyanis PISA kontextusba kell helyezni, vagyis

- teljes mértékben le kell fedni a konstrukciót
- olyan elemzési technikára van szükség, amely több mint 50 ország esetében működik
- a megfelelő mintavételi módszereket integrálni kell
- olyan adatbázist kell létrehozni, amelyet elérhetővé kell tenni a másodelemzések számára
- a leírt képességskálák konstrukcióját meg kell tudni támogatnia
- a teljesítmény minden szintjén lehetővé kell tenni az országok közötti összehasonlítást
- olyan skálákat kell létrehozni, amelyek az időben előrehaladva is megőrzik jelentésüket

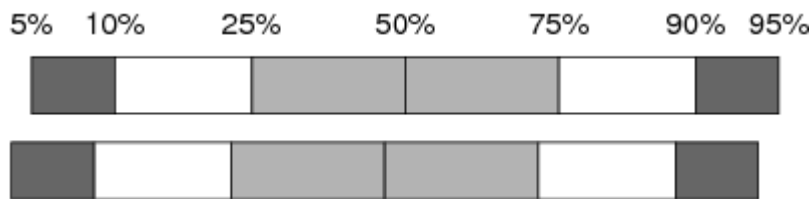
A PISA leegyszerűsít

Sokan azt vetik a PISA szemére, hogy túlzottan leegyszerűsít, de az adatoknak az értelmezése ettől nem lesz könnyebb. A teljesítmény pontoknak például nincs természetes mértékegysége. Az 500 pont átlagú és 100 pont szórással rendelkező skálát a felvételt szervezők választották, a pontszámok értelmezése pedig nem egyértelmű. A PISA jelentések, akár csak a TIMSS vagy a PIRLS, két alapvető értéket közölnek: 1.) azok arányát, akik egy adott képességszinten teljesítettek illetve 2.) az országonkénti teljesítményt az átlagpontszám szerint rangsorolva. Ugyanakkor a képességszintek besorolása nem dokumentált, azt a vizsgálat lebonyolítói állapítják meg, a PISA esetében a priori. Míg egy nemzetközi szegénység index esetén egy adott USA dollárban kifejezett szint mindenki által megfogható és értelmezhető, ez nem mondható el a PISA képesség pontszámairól (Brown et als, 2005).

Azáltal, hogy az országonként másképpen működő itemeket kiveszik, bár közelítenek a Rasch model által megkövetelt egydimenziós skálához, de éppen elvesznek a nemzeti sajátosságokat jobban tükröző szempontok (Goldstein, 2004).

A viszonylag bonyolult kérdezői repertoár és kognitív teszteredmények végül négy kompetenciaterületre aggregálódnak (matematika, szövegértés, természettudomány, problémamegoldás). A 2003-as PISA matematika alteszten az a tanuló, aki 500 pontot teljesített, a feladatok 46 százalékát oldotta meg. Az, aki 400 pontot ért el, 23%-át, aki 600 pontot, az a 71 %-át. Ez azt mutatja, hogy a kompetencia pontszám és a válaszadási arány között lineáris kapcsolat van, ahol 4 kompetenciapont a kérdések 1%-ának felel meg. Ez alapján a kompetenciaskála sokkal egyszerűbb jelentéssel bír (vagyis pusztán a jó válaszok arányát mutatja), mint amit neki tulajdonítanak (Wutke, 2007).

Miután mind a négy kompetenciaterületen a mérés egydimenziós, minden összehasonlítás egyben rangsorolást is jelent. A rangsor nem pusztán a média figyelmének felkeltését szolgálja, de mélyen benne gyökeredzik a PISA felmérés koncepciójában. A rangsorban elért helyek, különösen a rosszak, könnyen politikai üzenettké fordíthatók. Ezáltal a PISA eredményeket anélkül interpretálják, hogy valóban megértenék, mit is mértek. Nem mindegyik rangsorhelyek közötti különbség szignifikáns, ezt a PISA elemzők is jelzik. Ugyanakkor gyakran nagyon kis különbségek is szignifikánsnak bizonyulnak, akár 9 pontos eltérés is az lehet (lásd 3. ábra).



3. *ábra: Két Gauss eloszlás ahol az átlagértékek szórása 9 százalékkal tér el egymástól. Ez a kis különbség a két populáció között szignifikánsnak számít a PISA-ban.*
 Forrás: Wutke, 2007

4. Mire jó a PISA?

Fontos kihangsúlyozni, hogy a PISA módszertanát bírálók magát a PISA felmérés jelentőségét nem vonják kétségbe, kritikájuk elsősorban a PISA továbbfejlesztésére irányulnak. A legtöbb kritika magát a rangsort éri, mondván, hogy ez a kapott bonyolult információk túlzott leegyszerűsítésének veszélyét hordozza. Szofisztikáltabb módszereket és az egydimenzió helyett több dimenzió, paraméter bevonását javasolják a szakértők. Az országok összehasonlítását célzó kérdések és skála nem hasonlíthat egy intelligenciateszthez, ez egy komplex többdimenziós megközelítést igényel. Éppen ezért a PISA rangsort szinte mindegyik megszüntetni vagy átalakítani kívánja (Goldstein, 2004, Kreiner, 2010). Szerintük a nem tipikustól – vagyis ami valójában országspecifikus – nem szabad „megtisztítani” a vizsgálatot. Éppen ezért más megközelítésre van szükség a kérdések tartalmi és formai megalkotásánál és a válaszok elemzésénél. A PISA-t javaslatuk szerint társadalmisítani szükséges, az elemzési keret kialakításától a tanácsadók bevonásán át az eredmények és módszerek publikálásáig nagyobb nyilvánosságot és főleg transzparenciát kell biztosítani (Goldstein, 2004). A PISA – ahogy mondják - nagyszerű lehetőség arra, hogy még többet tudjunk meg arról, miben különbözünk, és jóval kevésbé célszerű versenyként felfogni.

A PISA-val szembeni kritikák, amelyek annak szűklátókörűségére, egydimenziós jellegére vagy korlátozottságára utalnak, nem feltétlenül jogosak. El kell tudni különíteni a PISA-t magát attól, hogy hogyan használják és értelmezik. A PISA-t létrehozók soha nem állították, hogy általános receptet nyújtanának. Ugyanakkor tény és való, hogy a pénzek felett diszponáló döntéshozókat és a médiát leginkább az egyszerű, de hatásos üzenetek érdeklik. Éppen ezért nem valószínű, hogy a rangsorok megszűnének. A szakmának ugyanakkor törekednie kell arra, hogy a mélyebb üzenetek is átjöjjenek, mert bár igaz, hogy a sorrendek azok, amik a legnagyobb figyelmet felkeltik, de éppen ezek a legkevésbé használható elemek. Ugyanakkor az adatbázis nyilvános, az eljárások nagy része transzparens, az eredmények jól dokumentáltak és továbbgondolhatóak. Felvállaltan nem mér „mindent”, de folyamatosan bővülnek a kompetenciaterületek. A PISA adatokat az OECD szakértői folyamatosan dolgozzák fel és értelmezik, az eredményeket egyre több kötetben teszik közzé. Sajnálatos módon ezeket nem csak az újságírók vagy politikusok nem olvassák, de a kutatók nagy része vagy a szakma képviselői, a pedagógusok sem. Pedig a térben és időben is kiterjedt elemzések egyik legizgalmasabb hozadéka, hogy a mai napig erőteljesen élő, de valójában nem működő mítoszokról rántja le a leplet.¹⁹

¹⁹ http://www.tarki-tudok.hu/files/mta_konferencia_csapobeno.pdf

Az oktatási mítoszok leleplezése

Az oktatási mítoszok leleplezésének jelentősége éppen abban áll, hogy a hagyományokra vagy „hüvelykujj” szabályokra alapozó oktatáspolitikai, amennyiben van erre készítés, eredményesebb üzemmódra állhat. A PISA eredményekből az évek során tartósan leszűrhető tanulság, hogy a méltányosság és eredményesség kéz a kézben járnak.²⁰ A hátrányos helyzetű tanulók nincsenek arra kárhóztatva, hogy gyengébben teljesítsenek, az ugyanolyan családi háttérű gyermekek között szintén nagyok lehetnek a különbségek, attól függően, hogy milyen iskolába járnak. Nincs összefüggés egy ország teljesítménye és a bevándorló gyermekek aránya között sem. Azok az országok teljesítenek jól, ahol nem a született tehetségben, hanem abban hisznek, hogy a szorgalom, befektetett energián múlik egy tanuló eredményessége.²¹ Ezen felül az is nyilvánvaló, hogy egy bizonyos szint felett már nem az oktatásra szánt erőforrások mennyisége, hanem azok minősége számít. Így önmagában az osztályméret és az országok teljesítménye között sincs összefüggés, hiszen a pedagógus felkészültsége az, ami számít. A digitális világ pedig más tananyagot és óraszervezést igényel. Az eredményesebb országok nem új tantárgyak bevezetésével és a tananyag növelésével érnek el eredményt, hanem éppen fordítva, kevesebbet, de mélyebben tanítanak.²²

Habár van, aki éppen a PISA mítoszromboló szerepét kérdőjelezi meg. Furcsának tartják, hogy a legjobban teljesítő országokban tapasztalni a legnegatívabb attitűdöket, hogy a PISA eredmények függetlenek a befektetett összegektől és az osztálymérettől és hogy az osztálytermi IKT használat és a tanulók digitális szövegértése között éppen negatív az összefüggés (Sjoberg, 2010). Különösen a skandináv országokat irritálja az az összefüggés, hogy a több pénz nem jelent feltétlenül jobb oktatást, ugyanis náluk viszonylag magas az oktatásra fordított erőforrások nagysága és így a PISA jó alkalmat teremthet a pénzek kivonására és a bérek befagyasztására (Sjoberg, 2012). Pedig csak rosszul fordították le az üzenetet, a pénz vagy a kis osztályméret és az eredményesség közötti kapcsolat hiánya nem azt jelenti, hogy le lehet csökkenteni a pénzt, hanem hogy önmagában a pénz kevés, azt rosszul is el lehet költeni, ha nem figyelnek rá, hogy mire és hogyan fordítják.

Ugyanígy azok, akik a PISA-t okolják azért, mert néhány országban a standardizált tesztek túlzottan is használják, tévednek. A standardizált tesztek, a külső elszámoltathatóság erősítése, a pedagógus autonómiájának csökkentése vagy a piaci erőknek nagyobb teret adó oktatáspolitikák általában rosszul fordították le a PISA eredmények üzenetét. Nem véletlen, hogy éppen azok az országok csúsztak hátrébb, akik a verseny és külső kontroll erősítésében látták az eredményesség kulcsát (Egyesült Királyság, USA, Svédország, Új-Zéland, Norvégia).²³ A PISA-t versenyként felfogó és arra versennyel válaszoló reakciója éppen annyira egyszerű és rövidlátó, mint azoké, akik a referenciaként elfogadott országok egy-egy adottságát próbálják utánozni (ilyenek tekinthető a lengyel sikert a kilencévfolyamos általános iskolának betudni és azt szolgálai lemásolni).

Az eredményes országok saját erősségeikre építenek és mindig egy lépéssel előbbre járnak. A skandináv országoknak is – Pasi Sahlberg professzor szerint – nem szabad eltérniük arról az

²⁰ http://www.tani-tani.info/kisebb_es_nagyobb_tevhiteink

²¹ <https://news.tes.co.uk/b/opinion/2013/12/03/michael-barber-39-s-key-lessons-from-today-39-s-pisa-results-number-one-quot-talent-is-a-myth-quot.aspx>

²² <http://www.bbc.com/news/business-31087545>

²³

http://www.slate.com/articles/news_and_politics/the_dismal_science/2014/07/sweden_school_choice_the_county_s_disastrous_experiment_with_milton_friedman.html

útról, hogy olyan erős közoktatási rendszert tartanak fenn, amely egyenlő esélyeket nyújt minden gyermeknek.²⁴ Ugyanő a távolkeleti országok sikerét abban látja, hogy az ún. globális oktatási reform mozgalom (GERM) minden elemét, mint a verseny, standardizáció, gyakori tesztelés és privatizáció, alkalmazzák.²⁵ Ez minden bizonnyal túlzás, sőt nem is feltétlen igaz. Az ázsiai országok esetén a tanulók motiváltak és szorgalmasak, vagyis belülről irányítottak. Ugyanez mondható el a sikeres finn modell esetén, ahol a pedagógusok felkészültek és reflektálnak a munkájukra, ezáltal itt is belülről működik a kontroll. Ugyanakkor az, hogy ide eljutottak egy majdnem harminc éves folyamat eredménye, ami alatt lépcsőzetesen felépítették azt a pedagógusképzést és pályát, ami ma működik. Most is úgy tűnik, a PISA előtt járnak, amikor a kézírás helyett a gépirás tanítását helyezik előtérbe és a tantárgyakat részben komplexebb kompetenciaterületek váltánák fel. A dilemma valóban a kontroll és autonómia megfelelő arányában rejlik. Ahogyan a Fehér Könyvre reagálva az Egyesült Királyság egy oktatásért felelős korifeusa fogalmazott:

„...30 évvel ezelőtt egy olyan oktatási rendszerünk volt, amiben nagy volt az autonómia de alacsony az elszámoltathatóság mértéke. Azóta átlendültünk az ellenkező oldalra: alacsony autonómia de magas elszámoltathatóság. A jelenlegi folyamatok azt mutatják, hogy ezt ugyanúgy meg kell haladni, mint a harminc évvel ezelőtti állapotot. Egyre inkább rájövünk, hogy a kívánatos kombináció a magas szintű autonómia és elszámoltathatóság együttese. Az iskolákban és tanároknál jobban kell bízunk a saját tantervük elkészítése, tanítási stílusuk kialakítása és értékelési gyakorlatuk terén. De a szabadság ára az erős elszámoltathatóság. Csak hogy ennek jóval intelligensebb rendszernek kell lennie, mint azok a durva intézkedések, amiket eddig alkalmaztunk és amelyek eltorzították az iskolák viselkedését.” (Froese-Germain, 2010)

Az adatok másodelemzése és egyéb adatokkal való összevetése a hazai sajátosságok felderítése céljából

A PISA adatok másik hasznos felhasználási terepe az adatok – sajátos, nemzeti szempontokat is előtérbe helyező - másodelemzése. Egy, az Oktatási Hivatal által kiírt pályázat keretében készült tanulmány kimutatja, hogy a magyar tanulók matematikai önhatékonysága 2003 óta szignifikánsan romlott, aminek oka, hogy a magyar diákok a zsúfolt tananyagot a környező országokhoz viszonyítva alacsonyabb matematika óraszámokban tanulják az iskolában. Ez óhatatlanul odavezet, hogy nincs idő az iskolában a gondolkodásra, így a jobb képességű tanulók sem sajátítják el a magasabb szintű tanulási stratégiákat. Ezért hiába tanulnak a magyar tanulók sokat iskolán kívül, tanulásuk nem lesz hatékony és matematikai önhatékonyságuk csökken (Csüllög-D.Molnár-Lannert, 2014).

Bár a PISA adatok oksági magyarázatokra korlátozottan alkalmasak, egyéb adatbázisokkal összevetve az összecsengő eredmények erősítik a következtetések magyarázó erejét. Magyarország ilyen szempontból kifejezetten szerencsés helyzetben van, mert létezik egy, a tanulók teljesítményét évente PISA kompatibilisan mérő teljeskörű felmérés, az Országos Kompetenciamérés. Miután ez az adatbázis egyéni szinten több mérési időpontban is összeköthető, így alkalmas oksági kapcsolatok feltárására is. Kifejezetten megtermékenyítő, ha a PISA és kompetenciamérés eredményei egymást tudják erősíteni. A kompetenciamérési eredmények és a PISA eredmények összevetésére szerencsére ma már találhatunk példát

²⁴ <http://pasisahlberg.com/is-pisa-dimming-the-northern-lights/>

²⁵ Global Educational Reform Movement, rövidítve GERM, ami angolul baktériumot jelent.

Magyarországon is. Ez egyaránt jelentheti azt, hogy a két adatbázisból származó eredmények egymást erősítik, de ennek az ellenkezőjét is. Ez utóbbira példa az, hogy a PISA eredmények romlása nem volt összhangban a kompetenciamérésben tapasztalt stagnálással.²⁶ A 2014-es Országos Neveléstudományi Konferencián a PISA adatokat magyar szempontból elemző szimpóziumon egyaránt sor került a PISA és a kompetenciamérés elemzési keretét összehasonlító, valamint a PISA egy legújabb koncepciójának, a rezilienciának iskolai szintű vizsgálatát bemutató előadásra az országos kompetenciamérés adatbázisán.²⁷

A PISA adatokban Magyarország egy dimenzióban kiugró, nálunk magyarázza a világon a legnagyobb mértékben a családi háttér a tanulói eredményeket. Ugyanezt az összefüggést a kompetenciamérés adatbázisa is visszaigazolja. Az igazi kérdés viszont az, hogy ez minek tudható be. A kompetenciamérés 2011-es és 2013-as tanulói szinten összekapcsolt adatain ennek kiderítésére is próbát tesz egy jelenleg zajló kutatás.²⁸ Az eredmények azt igazolják, hogy a pedagógusok hajlamosak a beskatulyázásra, ezáltal működtetve a szakirodalomban Pygmalion és Gólem hatásként leírt mechanizmusokat. A kompetenciamérés adatain végzett elemzés azt mutatja, hogy a pedagógusok jellemzően a hátrányos helyzetű tanulókat hajlamosak alulértékelni. Az egyik évben alulértékelt diákok között jóval magasabb a következő évben szintén alulértékelt aránya, mintha egy stigmát vinnének magukkal az általánosból a középiskolába. Úgy tűnik ráadásul, hogy a tanári értékelés hatással van a későbbi teljesítményre, hiszen a 8. évfolyamon alulértékelt tanulók teszteredménye alacsonyabb lett 10.-re, míg a felülértékelt diákoké javult. A pedagógusok elfogult félreértékelése és rossz értelemben vett differenciálása (értsd alatta, hogy nem a módszerben, hanem a tanuló iránti elvárásokban differenciálnak) a fundamentuma a magyar oktatási rendszer legnagyobb bűnének, annak hogy az nemhogy kiegyenlíti, de még fel is erősíti a családi háttér hatását.

5. Az elemzés korlátai és a PISA tudás társadalmasítása

A nagymintás, tesztkérdéseket is alkalmazó vizsgálatok elemzésére ma még világszerte nincs elegendő kapacitás és kompetencia. Ennek a tudásnak a bővülését és a módszertanban való elmélyülést szolgálják a technikákkal foglalkozó kézikönyvek, manuálok, képzések, amelyek egyelőre nagyrészt az USA-ban elérhetők. Bizonyos szoftverekbe már beépültek ezek az új technikák, mint a SAS vagy HLM6. Magyarország még gyerekcipőben jár, pedig nálunk nem pusztán a nemzetközi adatbázisok, de az országos kompetenciamérés adatai is nagyon jó inputot szolgáltatnának. Ezen adatbázisok elemzése terén mostanában kezdtünk nagyobb sebességre kapcsolni, de a megfelelő módszertani tudás és háttértámogatás még hiányzik ahhoz, hogy ez valóban megbízható módon és eredményesen működjön.

A nagymintás teljesítménymérések terepén Magyarország Európában jóval nagyobb szerepet is játszhatna, ha sikerülne kihasználnunk előnyös adottságainkat. Az Oktatási Hivatal Köznevelési Mérési Értékelési Osztálya – megfelelő költségvetési támogatással – olyan központtá válhatna, amely segíthetné az ilyen jellegű adatbázisok elemzéséhez szükséges

²⁶ <http://nahalkaistvan.blogspot.hu/2014/04/lehetett-e-latni-elore-2012-evi-pisa.html>

²⁷ <http://onk2014.unideb.hu/program/>, Németh Szilvia és Papp Z. Attila előadása

²⁸ „Oktatási törzsindikátorok és mérési eredmények alapján készülő indikátorok rendszerének fejlesztése a TÁMOP-3.1.8-09/1-2010-0004 azonosító számú projekt keretében” A középiskolák összehasonlító elemzése a KIR bázisán című kötetben Lannert Judit – Csüllög Krisztina: A diákok ambícióinak és teljesítményének alakulása és változása az általános iskolától a középiskoláig. A továbbtanulás mint a középiskolai munka egy lehetséges indikátora tanulmánya, kézirat

kompetenciák kifejlődését. Már eddig is voltak ilyen kezdeményezések képzések és célzott pályázati kiírások formájában, de sokkal többre lenne szükség, amire jelenleg kapacitásszűke miatt nincs lehetőség. Amennyiben az oktatáspolitikai felismerné az ebben rejlő lehetőséget és megfelelő (folyamatos és kiszámítható) költségvetést rendelne az oktatási adatbázisokat létrehozó intézményekhez, akkor ez lehetőséget adna arra is, hogy

- részletesebb magyar PISA jelentés születhessen
- az országos kompetenciamérés részletes technikai leírása is megszülessen,
- hogy létrejöhessenek az elektronikus diagnosztikus (edia) rendszerben olyan adatbázisok, amelyeket másodelemezni lehet,
- elkészüljön az érettségi adatbázis kutatásra alkalmas változata,
- képzéseket és tréningeket tudjanak tartani ebben a témában az adott intézmények,
- pályázati kiírások szülessenek, amelyek a több diszciplínából összeálló kutatói csapatok munkáját támogatják (erre jó példa az MTA szakmódszertani pályázata),
- szakmai háttértámogatást nyújtsanak a PISA, OKM és érettségi adatbázisok elemzéséhez.

Mindenekelőtt pedig a felsőoktatásban kellene sokkal nagyobb mértékben megjelennie az ilyen adatbázisok elemzéséhez szükséges kompetenciáknak. A pedagógusképzésben is meg kellene ennek jelennie a módszertani képzés során legalább illusztratív jelleggel, hogy később valóban olyan multidiszciplináris kutatói csapatok tudjanak összeállni, akik tudnak és hajlandók egymással kommunikálni és nem saját diszciplínájuk elefántcsonttornyából tekintenek le másokra.

A módszertani tudás bővítése azért is fontos, mert a survey technikákhoz szokott kutatók is – világszerte - számtalan hibát elkövetnek a PISA adatok elemzése során. A módszertan nagyon új, így az ehhez szükséges kompetenciák még viszonylag szűkösek. Ezt mutatja az a tény is, hogy az OH által kiírt pályázat, amely a kompetenciamérés eredményeinek elemzését szolgálta, a meghirdetett hét csomagból csak három kelt el, megfelelően felkészült kutatói csapatok hiányában.²⁹

Használati útmutató kutatóknak

A nagymintás teljesítményt mérő vizsgálatok eltérnek a survey típusú felvételektől. A minta nem egyszerű véletlen mintavétellel készül és a tanulóktól nem kérdezik meg az összes kognitív item kérdéseit. Éppen ezért ilyen adatfelvételi design esetén különös jelentősége van a mintavételi súly használatának, a teljesítmény adatok elemzésének speciális technikáinak és a hibabecslések módjának. A súlyok az ilyen jellegű adatbázisoknál úgy vannak megállapítva, hogy a megfigyelési egységek (iskola, pedagógus vagy tanuló) különböző valószínűséggel kerülnek kiválasztásra. A nyers súlyok úgy kerülnek meghatározásra, hogy a súlyok összege kiadja a megfigyelési egységek összegét. A súly inverzét rendelik oda ezután a megfigyelési egységhez, ami a kiválasztás valószínűségét reprezentálja. Amennyiben például egy tanuló súlya 10, akkor a kiválasztásának a valószínűsége 1/10. Amennyiben az elemző többszintű elemzést végez, akkor a különböző szintekhez a súlyokat manuálisan rendeli hozzá. Mind az IEA, mind a PISA felvételek esetén megpróbálják minimalizálni a tanulók által kitöltendő tesztkérdések mennyiségét, ezért a tesztfüzeteket komplex módon rotálják. Ezzel csökkentik a tanuló tesztelési idejét, de ez a módszer nem teszi lehetővé az egyéni képességszintek hagyományos becslését. Ehelyett használják a plauzibilis értéket, ahol a kevés számú item

²⁹ Iletve részben a kiírás hibájából, amely mindegyik csapat esetén impact faktoros matematikust várt el, viszont a módszerekhez szintén értő közgazdászok bevonását nem szorgalmazta.

alapján végzett becsléssel hozzák létre a populációsintű jártasságot. Ez az eljárás a tanulói teljesítményt missing-ként kezeli, a háttérkérdőívek és a limitált mennyiségű itemek alapján egy ún. többszörös imputálással közelítik meg az adott populáció vagy alcsoportok tanulói képesség eloszlását (Rubin, 1987). Random módon öt fix értéket rendelnek így hozzá egy-egy tanulóhoz. Ezek az értékek nem valódi teljesítmény pontok, éppen ezért elemzésük sem magától értetődő. Amennyiben egy kutatót a tanulói teljesítmények és egy bizonyos tényező (pl. tanulásra szánt idő) kapcsolata izgatja, akkor jár el korrekten, ha mind az öt értékre lefuttatja a regressziós modelljét és utána egy bizonyos eljárással kombinálja az eredményeket (Rubin, 1987). Két hibát követnek el gyakran, egyik, amikor csak az egyik értékre futtatják le a modellt és az alapján elemeznek, vagy ha az öt érték átlagát veszik. Mind a két eljárás alulbecsli a standard hibát és így olyan összefüggés is szignifikánsnak tűnhet, ami valójában nem az (Rutkowski et al, 2010). A szórásbecslés hagyományos technikái sem működnek ilyen esetekben. Az IEA és az OECD erre a replikáció (replication) módszereket alkalmazza. További fontos szempont, hogy mindegyik adatfelvétel keresztmetszeti jellegű, tehát oksági magyarázatokra nem alkalmas. Az ilyen jellegű adatbázisok elemzéséhez a következőket ajánlják (Rutkowski et al, 2010):

6. Összegzés

A PISA hatalmas vállalkozás, amelynek üzemeltetése igényli a média és a politikusok folyamatos érdeklődésének fenntartását. A folyamatos érdeklődés fenntartását szolgálják a rangsorok, amelyek hatásos és egyszerű üzenetet hordoznak. A PISA jelentősége ugyanakkor messze nem a rangsorokban áll, hanem abban, hogy az adatok mélyebb elemzésével többet tudhatunk meg arról, hogyan működnek az oktatási rendszerek. A PISA adatok elemzéséhez szükséges módszertani tudás ugyanakkor ma még szűkös. Érthető az oktatáskutatók és pedagógusok szorongása, ha érzékelik, hogy valami befolyásolja az életüket, de nem igazán értik ennek működését, módszertanát. Éppen ezért a PISA-tudás társadalmasítására lenne szükség, mint ahogy ez a PISA-t elsősorban módszertani szempontból bírálók is javasolják.

A PISA-t a globális hatása miatt bírálók nagy része elsősorban nem a PISA-val, hanem a saját országának a PISA adatokra adott válaszával elégedetlen. Két rossz válasz képzelhető el a PISA rangsor kapcsán, az egyik, amikor nem vesznek róla tudomást és elhallgatják (ahogy a franciák tették a legelső időkben), vagy amikor azt versenyként felfogva direkt módon a versenyt erősítik saját rendszereikben (pl. USA, Svédország). A PISA rangsorban előljárók általában a PISA előtt járnak (Lengyelország oktatási reformja már 1997-ben elindult, vagy a finnek pedagóguspolitikája több évtizedes fejlődés eredménye), és nem cibálják az oktatási rendszerüket az aktuális PISA adatok tükvében, vagy másolják más országok példáját. Használják és értelmezik a PISA adatokat, de csak egy indikátoraként az oktatásnak és folyamatosan kutatják-elemzik saját maguk is az oktatásukat. A sikeres országokra jellemző még, hogy belülről irányítottak, legyen ez a belső motor akár a tanuló motiváltsága és szorgalma (távol-keleti országok), vagy akár a pedagógusok felkészültsége, akik folyamatosan elemzik saját munkájuk eredményét (finnek). A középmezőnyben találhatjuk a kívülről irányított országokat, amelyek elsősorban a külső kontroll vagy a piaci mechanizmusok felerősítésével válaszoltak a romló adatok láttán. Az utolsó harmadban pedig az irányítatlan országokat találjuk. Erre legjobb példa India, ahol egyáltalán nem ritka, hogy a pedagógus meg sem tartja az órát.

Nem kérdéses, hogy habár jelenleg a PISA-t ért kritikákra ingerülten válaszol az OECD, hosszabb távon ezek minden bizonnyal megtermékenyítőleg hatnak a PISA további

fejlődésére, hiszen az OECD érdekelt abban, hogy az adatait egyre értőbben elemezzék. Ebben még van tere a fejlődésnek, például a technikai leírásban részletesebben meg kellene mutatni, hogy hogyan alakították ki a teljesítményszinteket, illetve, hogy a tartalmi kerethez hogyan rendelik hozzá az itemeket.

Magyarország a nagymintás teljesítménymérések terepén akár európai nagyhatalommá is válhatna. Az Oktatási Hivatal által gondozott országos kompetenciamérés, vagy a szegedi egyetemen az elektronikus diagnosztizáláson munkálkodó műhely egyaránt jó alapot adna ehhez. Természetesen ahhoz, hogy ebbe az irányba elmozduljunk az oktatáspolitikának fel kellene ismernie ennek jelentőségét és nagyobb költségvetéssel támogatni ezen műhelyek és általában az oktatási adatbázisokat létrehozó intézmények munkáját.

7. Irodalom:

Adams, Ray (2011): Comments on Kreiner 2011: Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment
Baird, Jo-Anne; Isaacs, Talia; Johnson, Sandra; Stobart, Gordon; Yu, Guoxing; Sprague, Terra & Daugherty, Richard (2011): Policy effects of PISA. Oxford University Centre for Educational Assessment.

Balázi Ildikó, Ostorics László, Szalai Balázs, Szepesi Ildikó, Vadász Csaba (2013): PISA 2012 : Összefoglaló jelentés. Oktatási Hivatal, Budapest

Blum, A.;Guérin-Pace, F. (2000): De Lettres et des Chi_res. Des tests d'intelligence à l'évaluation du _savoir lire_, un siècle de polémiques. Paris:Fayard.

Brown, Giorgina; Micklewright, John; Schnepf, Sykel Viola; Waldmann, Robert (2005): Cross-national surveys of learning achievement: how robust are the findings?, IZA Discussion Papers, No. 1652

Carvalho, Luis Miguel; Costa, Estela (2014): Seeing education with one's own eyes and through PISA lenses: considerations on the reception of PISA in European countries. Institute of Education, University of Lisbon. Portugal.

Csüllög Krisztina–Molnár D. Éva–Lannert Judit (2014): A tanulók matematikai teljesítményét befolyásoló motívumok és stratégiák vizsgálata a 2003-as és 2012-es PISA mérésekben. In: Hatások és különbségek, Oktatási Hivatal, Budapest.

Education at a Glance (2010). OECD Indicators. Paris: OECD.

Froese-Germain, Bernie (2010): The OECD, PISA and the Impact on Educational Policy. Virtual Research Centre.

Goldstein, Harvey (2004): International comparisons of student attainment: some issues arising from the PISA study. Assessment in Education.

Hopmann, T. Stefan; Brinek, Gertrude (2007): Introduction: PISA According to PISA – Does PISA Keep What Promises?, LIT Verlag Vienna, Berlin.

Kreiner, Svend (2010): Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment. Dept. of Biostatistics, University of Copenhagen.

Molnár Gyöngyvér (2005): Az objektív mérés lehetősége: a Rasch-modell. Iskolakultúra 2005/3

Mons, Nathalie, Pons, Xavier; van zanten, Agnés; Pouille, Jeremy (2009): The reception of PISA in France. Knowledge and Regulation of the Educational System. KNOWandPOL

Mortimore, Peter (2009). Alternative Models for Analysing and Representing Countries' Performance in PISA. Brussels. [Expert paper commissioned by the Education International Research Institute]

www.ei-ie.org/research/en/documentation.php

PISA 2009 Results (2011): Students on Line. Digital Technologies and Performance– Volume VI, OECD

Rubin, D. (1987). Multiple imputation for nonresponse in sample surveys. New York: John Wiley.

Rutkowski, Leslie; Gonzales, Eugenio; Joncas, Marc; von Davier, Matthias (2010): International large-Scale Assessment Data: issues in Secondary Analysis and Reporting. Educational Researcher, Vol. 39, No. 2, pp. 142–151

Sjoberg, Svein (2012): PISA: Politics, fundamental problems and intriguing results. English manuscript. La Revue, Recherches en Education Numéro 14.

Sahlberg, Pasi (2008). Real Learning First: Accountability in a Knowledge Society. Paper prepared for the Alberta Teachers' Association conference on Leadership in Educational Accountability: Sustaining Professional Learning and Innovation in Alberta Schools, Edmonton, Alberta, Canada, April 18-19, 2008.

www.pasisahlberg.com/downloads/Real%20Learning%20First%20FINAL.pdf

Wuttke, Joachim (2007): Uncertainties and Bias in PISA. In: PISA zufolge PISA – PISA According to PISA. Halt PISA, was es verspricht? Does PISA Keep What it Promises? Ed.: S.T.Hopmann, G.Brinek and M. Retzl. Reihe Schulpädagogik und Pädagogische Psychologie, Bd.6. Wien: Lit-Verlag

Yamamoto, Kentaro (2002): Estimating PISA students on the IALS prose literacy scale. Educational Testing Service