

A JÓ NEMZETKÖZI ÖSSZEHAJONLÍTÓ VIZSGÁLATOK TÍZ ISMÉRVE

T. Neville Postlethwaite
University of Hamburg

A 21. század kezdetén számos nemzetközi felmérés áll rendelkezésre, ám a minőségbeli különbségek igen jelentősek közöttük. Némelyek láthatóan igen színvonalasak, mások minősége viszont oly mértékben megkérdőjelezhető, hogy eredményeik nem megbízhatóak. Míg a minőség garantálása elsődlegesen a kutató felelőssége, az egyes országok oktatási minisztériumainak is kötelessége biztosítani, hogy a kutatások minden szempontból színvonalasak legyenek: nem engedhetik meg, hogy az adófizetők pénzén szakszerűtlen felmérések folyjanak. 1992-ben az IEA megjelentette a *The IEA Technical Handbook* című kiadványt ([IEA Technikai Kézikönyv], *Keeves*, 1992). 1994-ben *Andreas Schleicher* meghatározta az IEA kutatástervezési és kivitelezési sztenderdeket (*Standards for the Design and Operations in IEA Studies*; *Schleicher*, 1994). Az International Academy of Education szintén megjelentetett egy kiadványt, amelyben a kutatások során követendő irányelveket foglalták össze (*Beaton és mtsai*, 1999). Ugyanekkor az IEA is megjelentette a *Technical Standards for IEA Studies* c. kiadványát ([IEA Kutatásmódszertani Szabvány], *Martin, Rust és Adams*, 1999). Ezek a publikációk részletesen tárgyalják a technikai kérdéseket. E rövid tanulmány célja tíz olyan ismerv kiemelése, amelyek alapján az olvasó minden esetben megállapíthatja, milyen színvonalú felméréssel áll szemben.

Némely pontot külön, csillaggal (*) is megjelöltünk. Ezek azok az aspektusok, amelyek kritikus fontosságúak: ha a kutatók nem kielégítően végezték a munkájukat akár csak egy ilyen szempontból is, a vizsgálatukban nem lehet megbízni. Az olvasó feladata megkövetelni a kutatótól, hogy a vizsgálatát részletesen és pontosan mutassa be.

1. A vizsgálat céljai világosan megfogalmazottak-e?

Melyek voltak az adott felmérés céljai? Egyértelműen megfogalmazták-e ezeket a jelentés írói? Világossá teszi-e a szöveg célok összefüggéseit a szakpolitikai és az elméleti problémákkal? A célokat adekvát módon képezik-e le a kutatási kérdések? Egy olyan tanulmány olvasása, amelyből nem derül ki már az elején, milyen kérdésekre próbáltak a kutatók választ keresni, mindig nehézkes. Sőt, néha az olvasónak az a kifejezett érzése támad, hogy a kutatók maguk sem tudták, mire is keresnék a választ. Megfelelően iga-

zolgák-e továbbá a dokumentumok vagy jelentések azt, hogy a kiválasztott kutatási kérdések az adott régió/ország számára valóban fontos szakpolitikai és elméleti problémákra irányulnak? (Ellenkező esetben élhetünk azzal a gyanúval, hogy a kutatók saját kedvenc témája volt a kutatás tárgya valamely, a gyakorlat számára fontos kérdés helyett.) Igazolja-e a szöveg, hogy a vizsgálatot kimondottan szakpolitikai és elméleti kérdések megválaszolására tervezték? Néhány vizsgálatban nagy gondot fordítanak arra, hogy azonosítsák a minden oktatási rendszerben egyformán fellelhető szakpolitikai kérdésköröket. A kutatási kérdéseket úgy határozzák meg, hogy azokkal megválaszolhatók legyenek a szakpolitikai kérdések, majd „üres” vagy „minta-” táblázatokat szerkesztenek, amelyekkel az eredményekről való beszámolás módját előre kijelölik. Ha a kutatási folyamatban ezt az eljárást követték, erről a jelentésben a kutatók beszámolnak. Előfordul, hogy a jelentés egy még alakulóban lévő elméleti modelltől szól. Ebben az esetben is az olvasó feladata, hogy meggyőződjék róla: az elméleti modelltől olyan specifikus kutatási kérdések következnek-e, amelyeket a kapott adatok alapján meg lehet válaszolni.

Nemzetközi vizsgálatokban néha felmerül, hogy a különböző oktatási rendszerek túlságosan eltérnek egymástól ahhoz, hogy egy közös, a teljes vizsgálatban érvényesülő kérdéscsoportot lehessen meghatározni. Saját tapasztalatom az, hogy minden ország számára fontosak a források, intézkedések és a növekedés szintjei (az inputok, a folyamatok, az eredmények tekintetében), valamint ezen szintek megoszlásának méltányossága, igazságossága az ország különböző igazgatási egységei (régiók, tartományok), illetve iskolái között.

Ezeket a kérdéseket fontos tehát feltennünk:

- Relevánsak, világosan megfogalmazottak-e a vizsgálat céljai?
- Kellően körültekintőek, adekvátak-e a kutatási kérdések?

2. Megfelelőek és összevethetőek-e a célpopulációk?

Ha például a kívánatos célpopuláció a második osztályban tanuló összes gyermek, azt kell figyelnünk, hogy az oktatási rendszerrel kapcsolatos, a vizsgálat során feltett kérdések megválaszolásához valóban ez a populáció-e a megfelelő.

Amikor országok közötti összehasonlításokról van szó, egymásnak megfeleltethető populációkat vetettek-e össze? Például, ha egy adott évfolyam teljesítményét hasonlították össze, belekerült-e a célpopulációba az összes, ezen az évfolyamon tanuló diák? Kisebbségi tanulói csoportok kihagyása ezekben az esetekben igen elterjedt, vagy azért, mert kevesen vannak (és aránytalanul nagy költségeket emésztene fel tőlük is adatokat gyűjteni – pl. távoli, elszigetelt területeken élnek), vagy mert fogyatékkal élők iskolájába járnak (pl. látás- vagy halláskárosult diákok). E tanulókra tipikusan „kizárt” populációként szokás utalni és a csoport nagysága elfogadhatónak tekinthető, amíg a „kívánt” célpopuláció 5%-ánál kevesebb tanulót érint. A „kizárt” populáció nélkül tekintett „kívánt” populáció adja a „meghatározott” populációt. Ami *nem* elfogadható, az az, ha az egyik országban 2%-ot hagynak ki, egy másikban pedig 14-et. Tárgyalják-e a jelentések egyrészt

a diák- és iskola-szintű kizárások eltérő mértékét, másrészt e kizárások várható hatásait az átlagok összehasonlítására és az országokénti eloszlások alakulására?

Gyanakodnunk kell az olyan vizsgálatra, amely nem említ kizárt diákokat. Az a kutató, aki tudja a dolgát, mindig jelzi a kizárások mértékét és okát. Ha ilyen információ a tanulmányban nem található, akkor a legvalószínűbb az, hogy erre egyáltalán nem is figyeltek, tehát az olvasó nem tudhatja, pontosan mit hasonlítottak össze mivel. Ez a rossz vizsgálat egyértelmű ismertetőjegye.

Ugyanez érvényes az életkori csoportok összehasonlítására is. Az iskolai osztályok helyett életkori csoportok választásának egyik oka, hogy egy adott időszakon belül (pl. az egy naptári évben) születettek teljesítményét kívánjuk mérni. Ez a megközelítés arra keresi a választ, hogy az oktatási rendszer hogyan birkózott meg egy adott kohorsz oktatásával. Az olyan rendszerekben, ahol az évismétlők aránya magas, elképzelhető például, hogy a 13–14 éves tanulók több évfolyamon szóródnak. Felmerülhet ezzel kapcsolatban az az érv egyes oktatási rendszerekben, hogy a három osztállyal elmaradók számára az adott tesztek túl nehezek lesznek, ezért ezeket a tanulókat „ki kellene zárni”. Ilyenkor vagy nincs a teszteknek „alsó lefedése” (amely esetben a teszt nem alkalmas a diákok összességének a vizsgálatára), vagy pedig e tanulók teljesítménye vagy 0, vagy véletlenszerű pontszámot ér. E probléma kezelésének egyik lehetséges módja az „5% alatt” elv követése.

A felteendő kérdések közül a következőket emeljük ki:

- Tárgyalja-e a jelentés a „kizárt” populációt, valamint az enélkül tekintett „meghatározott” populációt?
- A kívánt népesség 5% alatt maradt-e a kizárt populáció?
- Valóban összevethetőek voltak-e a vizsgált csoportok?*

3. Megfelelő volt-e a minta kiválasztása?

A mintavétel fő célkitűzése, hogy a célpopuláció minden tagjának meghatározott, nullától különböző esélye legyen a mintába kerülésre. Ügyeltek-e a kutatást végzők erre? Mivel – különböző okok miatt – általában eltérés mutatkozik a tervezett és a valós minta között, szokásos eljárás súlyokat számítani és alkalmazni az egyes mintavételi rétegek között esetlegesen felmerülő aránytalanságok korrigálására. Minden tanulmányt kétkedéssel kell kezelni, ahol nem közlik, ez hogyan történt. A magyarázatnak mindig szerepelnie kell, akár lábjegyzetben, akár egy technikai fejezet formájában. Minél nagyobb különbségek mutatkoznak a vizsgált iskolák között, annál nagyobb számú iskolának kell szerepelnie a mintában. A p érték az a statisztikai adat, amellyel megmutatjuk az iskolák közötti különbséget. Megadja-e a jelentés a p értékét?

Ha azt tervezzük, hogy a rendszer valamely szegmensét vagy a tanulók egy adott csoportját mélyebb elemzésnek vetjük alá, ehhez e csoportból több tanulót kell bevonni, ennek pedig kihatása lesz a minta teljes méretére. Lennie kell továbbá egy olyan táblázatnak, amely ismerteti a tervezett és végleges minták nagyságát (megadva mindkét esetben az iskolák és a tanulók számát). A részvételi aránynak (a résztvevő iskolák ará-

nya megszorozva a résztvevő tanulók arányával) 85% felettinek kell lennie (lásd még a 7. pontot).

A minták alapján a populációra adott becsléseknek mindig van mintavételi hibája, aminek azonban elfogadható mértékűnek kell lennie az eredményekre alapozott szakpolitikai döntések tekintetében. Az 1960-as évek közepe óta számos nagy nemzetközi vizsgálatban sztenderd eljárás olyan minta kiválasztása, amelynek a mintavételi pontossága legalább eléri, vagy meghaladja egy pedagógiai eredményességvizsgálat 400 fős random mintájáét. A mintavételi pontosság e szintje lehetővé teszi, hogy a teszt itemeinek eredményeiben (a helyes válaszok százaléka) a mintavételi hiba ne legyen magasabb egy sztenderd hiba 2,5%-ánál, illetve két sztenderd hiba 5%-nál. Ez azt jelenti például, hogy a populációra adott 50%-os becslés esetén 20 esetből 19-ben biztosak lehetünk abban, hogy e becslést 50% valódi értéke 45 és 55% között van. Mivel szinte minden országban kétlépcsős mintavételt alkalmaznak (először az iskolák mintáját választják ki, majd a tanulókat az egyes iskolákon belül) nagyon fontos ennek figyelembe vétele a sztenderd hiba számításakor. Sokan elkövetik azt a hibát, hogy az SPSS-szel olyan sztenderd hibaszámítást futtatnak le, amely egylépcsős, egyszerű random mintát feltételez. Ennek eredménye pontatlan sztenderd hiba lesz, mert nem veszi figyelembe a minta kétlépcsős jellegét és kisebb értéket mutat, mint amekkora az valójában. Így ha közlik a kutatók az átlagok különbségét (például nemek közötti, vagy városi–vidéki összehasonlításokban), olyan különbségeket mutatnak ki, amelyek valójában nem szignifikánsak. A jó kutatás ismérve, hogy a sztenderd hibát megfelelően számítják ki és ennek módját is közlik.

Az olvasónak tehát fel kell tennie a kérdést: „Olyan módon történt-e a mintavétel, hogy a keletkező sztenderd mintavételi hibák elfogadhatóak legyenek a vizsgálat céljait tekintve?” Általában a mintavételi eljárásokban jártas kutatók tanulmányaikban rendre közlik a mintaválasztás lépéseit, valamint a pontos mintavételi hiba-értékeket. Amennyiben ez az információ hiányzik, egyértelműen gyanússá válhat a vizsgált minta. Ugyancsak általános gyakorlat az eredmények ismertetésekor a mintavételi hiba-értékek szerepeltetése a táblázatokban. Amennyiben ezek hiányoznak, tanácsos gyanakvással tekinteni az adott tanulmányra.

A gyanakvást indokolja, ha a valójában mért minta túl kicsi (túl nagy az eltérés a tervezett és a valós minta között), vagy ha a kizárt populáció meghaladja az 5%-ot, vagy ha a megfelelő p nem ismert stb. – ezekben az esetekben a bármely változóra számított átlagok és szórások hibásak lehetnek.

Előfordul néha, hogy a mintavétel helyes kivitelezése mellett szignifikáns különbségeket kapunk. Nagy mintáknál azonban rendszerint szignifikáns különbségek mutatkoznak. A valódi kérdés ekkor az, hogy pedagógiai szempontból van-e jelentősége ezeknek az eltéréseknek. Például, ha egy teszten csupán egyetlen itemben találunk szignifikáns különbséget, az nem említésre méltó, releváns eredmény. Tehát a szignifikáns különbségeket pontosan kell kiszámítani és kellő körültekintéssel kell értelmezni.

A mintavétellel kapcsolatos kérdések:

- Feltüntették-e a szerzők a mintavétel konfidencia-határát?*
- A tárgyalt mintavételi eljárásokhoz megadják-e a p értékét?*
- Nagyobb volt-e a részvételi arány (iskolák x tanulók), mint 85%?*
- Számítottak és alkalmaztak-e mintasúlyokat?*

- Megadták-e a mintavételi hiba-értékeket minden becsült értékre?*
- Különbséget tesz-e a jelentés a statisztikailag szignifikáns és a pedagógiai relevanciával bíró eltérések között?*

4. Jól kidolgozott és kipróbált-e a teszt?

A teszteknek nyilvánvalóan alkalmasnak kell lenniük a kijelölt konstruktum mérésére. Amennyiben a reliabilitás és validitás igazolása hiányzik, az olvasó joggal fog gyanút, akár országos, akár nemzetközi felmérésről van szó.

Általában a tesztek annak mérésére szolgálnak, amit az iskolarendszer egy adott pontjáig eljutva a tanulóknak már el kellett sajátítaniuk. Alkalmanként előfordul, hogy azt kell mérniük, mire lesz szüksége a diákoknak a társadalomba való kilépésükkor. Bármelyikről is legyen szó, elengedhetetlen annak bizonyítása, hogy a teszt betölti azt a szerepet, amire szánták.

Először is természetes, hogy pontos leírást adjon a szöveg arról, hogy mit ért például olvasáson vagy matematikán (vagy bármin, amit a vizsgálat mért) az iskolarendszer vizsgált pontján. Ha ez hiányzik a jelentésből (legalább függelék formájában ott kell lennie), okunk van kétségbe vonni a teljes vállalkozást. Másodszor: természetes egy teszt-váz vagy egy értékelési rendszer kidolgozása. Ez számos formát ölthet, de rendszerint egy táblázat, amely függőlegesen a tartalmakat, vízszintesen a kognitív viselkedési mintázatokat rendezi el, s a táblázat minden egyes cellája így egy-egy pedagógiai célt jelenít meg. Szokásos gyakorlat ennek a közlése a jelentésben.

Amennyiben a vizsgálat célja annak mérése, hogy az adott pillanatig a tanulók mit sajátítottak el, akkor a mérőeszközöknek le kell fedniük a résztvevő ország(ok) megcélzott tantervi tartalmait. Ez általában egy kétlépcsős folyamat, amely az érintett országok curriculumainak tartalomelemzésével kezdődik (tantervi utasítások, tankönyvek, vizsgák, illetve az alapján, hogy a tanárok milyen tartalmak tanításáról számolnak be). A második lépés az elsőre épül: a tartalomelemzés alapján teszt-váz(ak) kidolgozása. Míg számos tantervi célkitűzést nemzetközi szinten is mindenütt megtalálhatunk, mindig lesznek olyanok is, amelyek az országoknak csak egy részében érvényesül. Végül a vizsgált tananyagot gyakran területekre bontják. A szövegértés esetében például gyakran alkalmazzák a narratív próza–kifejtés–dokumentum felosztást. Ezeket a vizsgált részterületeket pontosan meg kell határozni.

Némely esetben a tanulmány az eredmények olyan aspektusaira összpontosít, mint például elég jól olvasnak-e a tanulók „a társadalomban való boldoguláshoz”, vagy „ahhoz, hogy felsőbb osztályba léphessenek”. Ezekben az esetekben először minden országban szakértői munkacsoportnak kell definiálnia, hogy valójában mi szükséges az ilyen típusú eredményekhez. Ez jelentős munkát igényel, de alapos feltárásnak kell lennie.

Más esetekben az a szokás, hogy a vizsgált évfolyamra vagy korcsoportra jellemző készségek és kompetenciák hierarchikusan rendezett halmazát adják meg. Minden szinten meghatározzák, mit tudnak a tanulók. Erre mutat példát az 1. táblázat, amely a SACMEQ vizsgálat hatodik évfolyamra vonatkozó anyagához tartozik. Itt nyolc szintet

különítettek el, de vizsgálták a hatodik évfolyam diákjait és a tanáraikat. Fontos, hogy minden szinthez tartozzon megfelelő item.

Általában az országok közötti eltérések kisebbek az olyan tárgyak tekintetében, mint a szövegértés vagy az idegen nyelvek, szemben például a matematikával, a történelemmel, vagy a társadalomismerettel. Ha azonban a vizsgálat a nemzeti tantervek közös tartalmaira irányul, mégis egyetértésnek kell lennie a nemzetközi teszt szerkezetében, és ennek minden egyes ország esetében le kell fednie a tantervi tartalmak nagy részét.

1. táblázat. A SACMEQ-felmérés a hatodik évfolyamon

Szint	Szövegértés	Matematika
1.	<i>Olvasás előtti szint:</i> Mindennapi tárgyakat és konkrét jelenségeket ábrázoló képek párosítása szavakkal, rövid írott instrukciók követése.	<i>Elemi számolás:</i> Egyműveletes összeadás és kivonás. Egyszerű alakzatok felismerése. Számképfelismerés. Számlálás egész számokkal.
2.	<i>Kezdő olvasás:</i> Prepozíciókat és absztrakt fogalmakat is érintő képek és szavak összekapcsolása; segítő technikák alkalmazása (hangos kiolvasás, egyszerű mondatstruktúrára, ismerős szavakra támaszkodás) a kifejezések értelmezésében olvasás során.	<i>Kezdő számolás:</i> Számosságokkal végzett kétműveletes összeadás és kivonás, ellenőrzéssel és becsléssel. Ismert tárgyak hosszának becslése. Gyakori kétdimenziós alakzatok felismerése.
3.	<i>Alapvető olvasás szintje:</i> Jelentés megértése (szavak és kifejezések párosítása, mondatkiegészítés, egymást követő szavak összeillesztése) rövid, egyszerű szövegekben (tovább- és újraolvasással).	<i>Alapszintű számolás:</i> Verbális információ (mondat, egyszerű ábra, táblázat) átalakítása egy aritmetikai művelet ismétlődő lépéseinek végzésével. Grafikus információ törtekké alakítása. Egész számok helyiértékének megértése a tízes számkörben. Egyszerű, mindennapos mértékegységek ismerete.
4.	<i>Tartalom megértése:</i> A szöveg különböző részeiben szereplő információ összekötése és értelmezése tovább- és újraolvasással.	<i>Alapműveleti számolás:</i> Verbális vagy grafikus információ átalakítása egyszerű aritmetikai feladattá. Különböző aritmetikai műveletek végzése (megfelelő sorrendben) egész számok, törtek és/vagy tizedestörtek körében.
5.	<i>Értelmező olvasás:</i> A szöveg különböző részeiben szereplő információ összekapcsolása és értelmezése tovább- és újraolvasással, valamint társítása külső információval (felidézett lexikai ismeretekkel), amely kiegészíti és kontextualizálja a jelentést.	<i>Gyakorlott számolás:</i> Verbális, grafikus vagy táblázatos formában megadott információ aritmetikai formába alakítása a feladat megoldásához. Többműveletes feladatok megoldása (a megfelelő műveleti sorrend követésével), amelyekben gyakori mértékegységek és/vagy egész és vegyes számok szerepelnek. Alapvető mértékegységek átváltása egymásba (pl. méter centiméterre).
6.	<i>Következtető olvasás:</i> Hosszabb, bonyolultabb (narratív, dokumentum vagy kifejtő) szövegek tovább- és újraolvasása, hogy a szöveg különböző részeitől származó információk összekapcsolásával azonosítsa az író célját.	<i>Matematikai készségek:</i> Többműveletes feladatok megoldása (a megfelelő műveleti sorrend követésével) törtszámok, arányok, tizedestörtek körében. Verbális és grafikus információ szimbolikus, algebrai, egyenlettel felírt formába alakítása adott matematikai feladat megoldásához. Az eredmény ellenőrzése és becslése külső (a feladatban nem megadott) tudás felhasználásával.

1. táblázat folytatása

Szint	Szövegértés	Matematika
7.	<u>Analitikus olvasás:</u> Hosszabb (narratív, dokumentum vagy kifejtő) szöveg különböző részeiben szereplő információ összekapcsolása és értelmezése tovább- és újraolvasással az író személyes meggyőződéseinek (értékek, előítéletek és/vagy részrehajlás) feltárására.	<u>Problémamegoldás:</u> Információ kikeresése és átalakítása (pl. mértékváltás) táblázatból, grafikomból, vizuális és szimbolikus reprezentációból többlépcsős feladat azonosításához, majd megoldásához.
8.	<u>Kritikus olvasás:</u> Hosszabb (narratív, dokumentum vagy kifejtő) szövegben szereplő információ azonosítása azért, hogy a szöveg különböző részeiben szereplő információkat összekapcsoljon és így következtetéseket vonjon le, illetve értékelje, az író mit feltételezett a témáról és az olvasó jellemzőiről (pl. kor, tudás, személyes meggyőződések: értékek, előítéletek és/vagy részrehajlás).	<u>Absztrakt problémamegoldás:</u> Expliciten meg nem fogalmazott, verbális vagy grafikus információba ágyazott matematikai probléma jellegének azonosítása, majd ennek szimbolikus, algebrai vagy egyenlettel felírt formába alakítása a probléma megoldásához.

A teszt vázában rögzített minden egyes célkitűzéshez ki kell dolgozni itemeket. Az itemek formátumának egységesnek és jól alátámasztottnak kell lennie. Továbbá az itemeket ki kell próbálni és ki kell elemezni. Feleletválasztó itemeknél meg kell győződni arról, hogy a disztraktorok tartalmilag, de diagnosztikus erejüket és elterelő hatásukat tekintve is megfelelőek. A nyílt kérdéseket, amelyekre a választ a tanulóknak maguknak kell megalkotniuk, próbamérésen be kell mérni, hogy biztosíthatjuk, megbízhatóan értékelhető válaszok halmazát hívják elő. Amennyiben skálaértékeket rendelünk a válaszokhoz, elengedhetetlen, hogy egyetértés legyen a skála értelmezésében, abban, hogy a skála egyes pontjai milyen tanulói teljesítményt jelentenek az adott feladatban. Az itemek megfelelő voltát a résztvevőknek el kell fogadniuk, a teszt megbízhatóságát pedig igazolni kell. Ha a vizsgálat célja az idő előrehaladtával bekövetkező – például a legutóbbi és a jelenlegi felmérés közötti – változás mérése, akkor a két időpontban megfelelő mennyiségű közös itemet kell felvenni az eltelt idő alatt történő változás megbízható mérése. Végül minden egyes itemet meg kell vizsgálni a lehetséges értelmezési különbségek azonosítása céljából minden egyes résztvevő országban. Az itemek pszichometriai mutatóinak hasonlóknak kell lennie a résztvevő országok nagy részében. Amennyiben egymást átfedő tesztekkel álló battériát alkalmaznak, a kipróbálás fázisában igazolni kell, hogy a közös itemek valóban lehetővé teszik az ugyanarra a skálára történő kalibrálást.

Esetenként életszerű feladatokon nyújtott teljesítmény mérése is szükséges lehet a vizsgált terület célkitűzései teljes körének lefedésére. Ilyen feladatok kidolgozásakor tekintettel kell lenni a méréshez rendelkezésre álló (rendszerint) korlátozott időkeretre, az egyszerű, nagy számban és a résztvevő országok lehetőségein belül hozzáférhető eszközök használatának igényére, valamint arra a kíváncságra, hogy a különböző országokban egyaránt megbízhatóan értékelhető válaszokat hívjanak elő e feladatok. Ha rotálták a részteszteket, igazolni kell, hogy ez megfelelően történt. Például a részteszteken lenniük kell közös itemeknek, hogy azok transzformálhatóak legyenek egy közös skálára.

Végül a tesztek validitását alá kell támasztani. Ha végeztek érvényességvizsgálatot a kutatók, akkor beszámolnak róla, az érvényesség milyen aspektusait vizsgálták. Annak eldöntése már az olvasó feladata, hogy az adott eljárások meggyőzőek-e vagy sem. Ha ilyen vizsgálatra nincs utalás, azt el sem végezték. Ebben az esetben nem lehet képünk a tesztek validitásáról, ezért az eredményeket gyanakvással kell szemlélnünk. Az érvényesség elemzésében nemzetközi szinten az egyik fontos lépés az, hogy a résztvevő országok megvizsgálják, a teszt mely itemei vonatkoznak az adott ország tantervének elemeire. A kutatók ezután külön kiszámítják az egyes országokhoz tartozó nemzeti tantervi pontszámot és a teszt összpontszámát (amelyben minden item szerepel, függetlenül attól, hogy szerepel-e a tantervben vagy sem). Ezt követően minden résztvevő országot pontszámok sorával jellemeznek: az összpontszámmal, A ország tantervi pontszámával, B ország tantervi pontszámával és így tovább. Különböző nemzetközi vizsgálatok kimutatták, hogy az országok sorrendje nem változik jelentős mértékben, bármelyik pontszámot is használjuk a rendezéshez. Ez azt jelzi, hogy az alkalmazott teszt jó, mivel a kimeneti változót minden résztvevő ország számára egyformán kielégítő módon méri.

A tesztfejlesztéssel kapcsolatos kérdések:

- Jól és meggyőzően tárgyalja-e a szöveg a teszttel vizsgálni kívánt tartalmat?*
- Jól meghatározza-e minden egyes vizsgált tartalom részterületeit?*
- Meggyőző-e a folyamat, amelynek során megtörtént a létező tantervek elemzése, illetve a társadalom által megkívánt készségek azonosítása?
- Meggyőző-e az itemek kidolgozásának folyamata?
- Kipróbálták és elemezték-e az itemeket?*
- Hogyan történt a skálák kidolgozása?*
- Meggyőző-e a validitás vizsgálata?*
- Elég magasak-e a tesztek reliabilitásmutatói?*

5. Megbízhatóak-e a kérdőívek és az attitűdskálák?

Általánosan elterjedt nézet, hogy kérdőívet és attitűdskálát könnyebb kidolgozni, mint teszteseteket. Ez a nézet azért hibás, mert a tesztfejlesztéshez teljes, kidolgozott technológia áll rendelkezésre, ezzel szemben a kérdőívek fejlesztését sokkal szerényebb háttér segíti. A kérdőívek és attitűdmérő eszközök kidolgozásának titka: próbamérés, próbamérés, próbamérés. Ha nem történt próbamérés, akkor a mérőeszközök vélhetően nem jók.

A *kérdőívnek* tartalmaznia kell minden, a vizsgálat kezdetén felvetett kutatási kérdések megválaszolásához szükséges indikátort. Az indikátorok egy része a változók azon csoportjába tartozik, amely „származtatott változó” néven ismert, vagyis amelyet egy vagy több kérdéssel megszerzett információ alapján konstruálunk meg. Ezek között lehetnek egyszerű, arányokat jellemző változók, ám számos változóból álló faktorok is. Majdnem minden esetben tartozik skála az egyes kérdésekhez, illetve a származtatott változókhoz. A kérdéseket olyan egyszerűen kell megfogalmazni, hogy a válaszadásra kért tanulók mindegyike (a gyengébbek és a kiválóak egyaránt) könnyen megértse azokat. Minden kérdést be kell mérni és elemezni, hogy biztosítható legyen: pontos és meg-

bízható információt szolgáltatnak az indikátorokhoz és a származtatott változókhoz. A származtatott változók és konstrukciójuk módja normális esetben a függelékben szerepel a reliabilitást jellemző adatokkal együtt.

Az *attitűdmérő eszközök* (amelyek esetenként a kérdőívek részei) kiemelt attitűd-dimenziókat vizsgálnak. Ezeket a dimenziókat pontosan ismertetni kell. Az egyes attitűd-iteket általában a célpopuláció körében végezett speciális kismintás vizsgálatok során gyűjtik össze. Ezeket is bemérik és elemzik. Gyakori, hogy körülbelül háromszor annyi item szükséges a próbaméréshez, mint amennyit végül az attitűdmérő eszközben felhasználnak. A végleges skála reliabilitását és validitását is igazolni kell az adott vizsgálat szempontjából. Az attitűdskála kidolgozásának ismertetésekor fontos annak tárgyalása, a kutatók hogyan jutottak el a megadott számú válaszlehetőséghez és mely itemeket invertáltak azért, hogy a tanulók ne válaszolhassanak mindegyikre ugyanúgy.

A kérdések, amelyeket fel kell tennünk:

- Tárgyalja-e a szöveg a vizsgálat valamennyi kutatási kérdését lefedő itemkészlet kidolgozását biztosító folyamatot?*
- Az attitűd-itekekben alkalmazott állítások valóban abból a populációból származnak-e, amely a mérés célpopulációja?
- Megtörtént-e a kérdőívek és attitűdmérő eszközök bemérése többszöri próbamérésekkel?*
- Bemutatja-e a szöveg a származtatott változókat?*
- Ahol ez szükséges, bemutatja-e a szöveg az eszközök skáláit?*

6. A központi munkanyelvet használó, több nemzetet átfogó vizsgálatokban hitelesítették-e a fordításokat e nyelvről a többire?

Nyilvánvaló, hogy minden item lefordításának és ellenőrzésének alapos hitelesítési folyamatban kell történnie azt elérendő, hogy azok mindegyik nyelven hasonló nyelvi nehézségi szinten legyenek. Ehhez részletesen kidolgozott eljárások állnak rendelkezésre, és ha a kutatók elvégezték valamelyiket, akkor ezt bizonyosan dokumentálják is. Igaz, ez a verifikációs eljárás meglehetősen költséges. Ha azonban nem került rá sor, az olvasó nem tudhatja, mennyire összevethetőek a tesztek és kérdőívek itemei. Nemzetközi vizsgálatokban a nemzetközi kérdőívek itemeit gyakran adaptálni kell a nemzeti változatok készítésekor, ez utóbbiakat így a nemzetközi központnak alaposan ellenőriznie kell. Ennek elmulasztása esetén találkozhatunk olyan problémákkal, mint amikor a nemzetközi kérdőív osztálynagyságra vonatkozó kérdéseit a spanyol változatban négyzetméterekben, nem pedig a tanulók számával kellett megválaszolni.

Itt a következőt fő kérdés merül fel:

- A fordítás átesett-e alapos hitelesítési folyamaton?*

7. Megfelelően végezték-e el az adatgyűjtést?

Az adatgyűjtés szakasza bármely vizsgálatban kulcsfontosságú. Az adatgyűjtés célja, hogy a mintába választott összes válaszadót megvizsgáljuk, és hogy mindannyian válaszoljanak minden kérdésre a kérdőíven és minden itemre a tesztben, amire tudnak. Általában útmutatót szokás összeállítani azok számára, akik az egyes országokban irányítják az adatgyűjtést. Ezek az útmutatók azért szükségesek, hogy biztosítható legyen: az adatgyűjtés olyan módon történik, amely minden egyes adatgyűjtési helyen egységes feltételek megeremtésével érvényes adatokat szolgáltat.

A National Centre Manual (alkalmanként National Research Co-ordinator/NRC Manual vagy National Project Manager/NPM Manual [Országos Központi; Országos Kutatási Koordinátori; Országos Projektmenedzseri Útmutató] le kell fedjen minden lehetséges, az adatgyűjtés során figyelembe veendő részletet. Ebbe beletartoznak az iskola- és tanuló specifikációk, amelyek biztosítják, hogy a megfelelő iskolát válasszák ki, illetve a megfelelő tanulókat mérjék (nem pedig másokat), és a megfelelő tanárokat válasszák ki (amennyiben használnak tanári kérdőíveket vagy tesztek). Rendszerint egy második útmutató készül az adatgyűjtők számára, amely részletesen leírja, hogy mit kell tenni minden egyes kiválasztott iskolában. Egy harmadik, adatfelvételi útmutató pontosan meghatározza, (a) mit kell tennie és mondania a tesztet felvevő biztosnak az adatfelvétel folyamata alatt; (b) az eljárásokat és időkereteket a mérőeszközök felvételéhez; (c) hogyan kell összegyűjteni és a központba visszajuttatni a mérőeszközöket. Cél, hogy nagyon kevés iskola maradjon ki a méréstől (lehetőség szerint egy sem), és minél kevesebb tanuló hiányozzon. A vizsgálat beszámolójában meg kell adni a hiányzó iskolák és diákok arányát. Azt szokás mondani, hogy iskolai szinten 10%-nál, tanuló szinten 20%-nál nem lehet nagyobb a távolmaradási arány. Mivel azonban nincsenek teljességgel érvényes eljárások a hiányzó adatok kezelésére, ezeket az értékeket a legnagyobb megengedhető aránynak kell tekinteni.

Egyes felmérésekben nem fordítanak elég gondot arra, hogy minél kisebb legyen a meg nem válaszolt kérdések aránya. Fontos, hogy a kutatóközpont biztosítsa, olyasvalaki gyűjtse össze a tesztek/kérdőíveket, aki ellenőrzi az íveket ennek kiszűrésére, még mielőtt azok elhagynák az iskolát. Így lehetővé válik a meg nem válaszolt kérdések kitöltetése a mérőeszközök visszaküldése előtt.

Nagy léptékű felméréseknél gyakori az adatfelvétel minőségbiztosítása. Ekkor speciálisan kiképzett mérőbiztosokat küldenek ki véletlenszerűen kiválasztott iskolákba az adatfelvétel megfigyelésére, s így biztosítják, hogy az adatfelvétel megfelelően folyjon. A biztosok ellenőrzik, hogy a megfelelő tanulóktól folyik az adatgyűjtés, ellenőrzik az adatfelvételi helyiségben a diákok elhelyezkedését, hogy ne csalhassanak és így tovább.

Az adatgyűjtés eredményeképpen legalább 85%-os válaszadási arányt kell elérni (ahogyan ezt a 3. pontban tárgyaltuk).

A felteendő kérdések:

- Ismerteti-e a tanulmány az útmutatókat?
- Ismerteti-e az alanyok kiválasztásának dokumentumait (iskola- és tanuló specifikációt)?

- Megfelelően magas volt-e a részvételi arány (a mintába utólag, pótlásként bevásztott iskolák nélkül)?*
- Elenyésző-e a hiányzó adatok mennyisége?*
- Alkalmaztak-e minőségbiztosítási eljárást az adatgyűjtési folyamatban?

8. Megfelelően folyt-e az adatrögzítés, az adatok megtisztítása, az értékelés és a minta súlyozása?

Az adatokat többnyire az országos központ számítógépein rögzítik. Az adatfelvitelhez szükséges szoftvert szinte mindig a kutatók biztosítják. Gyakori, hogy a mérőeszközök eredményeinek 10%-át a kutatók újra rögzítik hitelesítés céljából. A jó adatrögzítő szoftver már az adatok bevitelkor elvégez bizonyos ellenőrzéseket, amelyek révén a javítás az adatrögzítés folyamán azonnal megtörténhet. Ezt azután további ellenőrzések sokasága követi mind az országos, mind a nemzetközi vizsgálatok során. Amikor számos ország vesz részt egy felmérésben, az adatok megtisztítására ugyanazt a protokollt kell követniük. Ha mindegyikük saját, külön adattisztítási szabályokat követ, igen nehéz az összehasonlíthatóság biztosítása. Minden rögzített adatállományban lesznek olyan „extra” hibák, amelyeket a rögzítő szoftver nem vesz észre, legyen az bármennyire jó. Konzisztencia-ellenőrzéssel kiszűrhetők a kérdőívek azon kérdései, amelyeknél a válaszadók véstettek valamilyen hibát. Ezeket a „problémákat” jelzik az országos központoknak, akik a kérdések tisztázása érdekében felveszik a kapcsolatot a válaszadó iskolával, majd a „helyes” válaszokat visszajuttatják a nemzetközi adatfeldolgozó központba, ahol elvégzik a szükséges változtatásokat. Az adatok tisztításának e folyamata hosszú időt vehet igénybe, különösen, ha magas a résztvevő országok száma. Meg kell említenünk azt is, hogy mértéktelenül hosszú időt emészt fel az adatok megtisztítása abban az esetben, ha egy ország adatállományában nyilvánvaló a gondatlanság az adatgyűjtés és/vagy -rögzítés folyamatában.

Elvárható, hogy az olvasó figyelmét felhívja a jelentés azokra a változókra, ahol olyan sok a hiányzó adat, hogy azokat nem lehetett bevonni az elemzésekbe. Ha sok olyan változó van, ahol 20% feletti a hiányzó adat, az olvasónak minden oka megvan a gyanakvásra. Az is figyelmet érdemel, hogyan kezelték az hiányzó adatok problémáját. Ennek számos módja van (például becsült értékek hozzárendelése); bármelyiket is használják a kutatók ezek közül, azt jelezniük kell a tanulmányban. Ha nem esik arról szó, hogyan kezelték a hiányzó adatokat, az olvasó gyanakvása jogos.

Ha a mérőeszközben szerepeltek nyitott itemek, ezeket értékelni kell, majd skálákat kell szerkeszteni. Ismét fontos, hogy az értékelési eljárásokat ismertessék a kutatók, rendszerint egy függelékben vagy egy külön jelentésben.

Végül kiválasztás különböző valószínűségű tényezői miatt (az adatgyűjtés hiányosságai, a mintavételi rétegek közötti aránytalanság, pontatlan mintaválasztási terv, hiányzó adatok stb.) mintasúlyokat kell számítani. Mivel a felmérésekre épülő vizsgálatokban szinte mindig előfordul valamilyen hiba, elengedhetetlen a mintasúlyok használata. Elvárható, hogy ezek kiszámításának módját megadja jelentés, vagy – kivételes esetben –

megindokolja, ezekre miért nem volt szükség. Ha hiányzik a mintasúlyok számításának bemutatása, nagyon valószínű, hogy ez nem történt meg, ezért a változók becsült átlagai és szórásai nem jók.

Az ide vonatkozó kérdések:

- A konzisztenciát is ellenőrző adatrögzítő szoftvert használtak-e?*
- Végeztek-e további ellenőrzéseket?*
- Volt-e nagyszámú olyan változó, amelynek esetében a hiányzó adatok aránya meghaladta a 20%-ot?
- Számítottak és használtak-e mintasúlyokat?*

9. Megfelelő volt-e az adatok elemzése?

Jellemzően minden kutatási beszámolóban találkozunk egyváltozós és többváltozós elemzésekkel is. Természetesen az elemzéseknek igazodniuk kell a kutatási kérdésekhez, de ezek általában annyira összetettek, hogy egyváltozós elemzéseknél többet kívánnak.

Némely elemzés egyszerű, mások komplexek. Szokásos eljárás üres táblázatok kidolgozása a vizsgálat munkálatainak kezdetén, amelyek lefedik a kutatási kérdéseket és irányítják az elemzések elvégzését, amelyek eredményei majd kitöltik a celláikat. Ha az olvasó maga nem elég járatos az adatok elemzésében, tanácsos szakemberek segítségét kérnie annak eldöntéséhez, hogy a feltett kérdések megválaszolására az elvégzett elemzések alkalmasak voltak-e.

Tanulságos lehet az olvasók számára néhány példa nem megfelelő elemzésekre, amelyek tipikusan megtalálhatók gyenge minőségű vizsgálatokban. Előfordul, hogy egy táblázat egy-egy cellájában (vagyis a teljesítmény valamely aspektusára) annak ellenére adnak meg átlagot, hogy a vonatkozó itemek száma nem engedné különálló skála felvételét. Előfordul az is, hogy értékelik a kapott lineáris korrelációs együtthatót anélkül, hogy megfontolnák és ellenőriznék, megtartja-e az a szignifikanciáját akkor is, ha más változók, például a tanulók szocioökonómiai státusának vagy az iskola földrajzi helyének (városi–vidéki) hatását kizárják. Harmadik példánkban iskolák elemzésére végeztek többváltozós elemzéseket 100 változóval, miközben összesen 150 iskola szerepelt a mintában. (Ebben az esetben legalább hatszor annyi iskolára lenne szükség, mint változóra.)

Ha az adatok elemzése során új konstruktumok (vagy faktorok) jelennek meg, akkor azokat ismertetni kell.

A gyenge minőségű vizsgálatokban *értelmezési* hibák is találhatóak. Előfordul, hogy a jelentések szerzői nyilvánvalóvá teszik gondatlanságukat, amikor megfelelnek arról, hogy a korreláció nem jelent feltétlenül oksági összefüggést. Máskor azt nem veszik figyelembe, hogy az észlelésre vonatkozó kérdésekre kapott válaszok nem feltétlenül rajzolják ki a valós helyzetet (például a tanári kérdőívvel gyűjtött adatok arról, hogy a tanárok szerint mik az iskola céljai).

Már ejtettünk szót annak fontosságáról, hogy minden becsült értékhez megadjuk a sztenderd mintavételi hiba mértékét. Ennek kiszámítására ma már igen jó programok állnak rendelkezésre, így elvárható, hogy valóban minden becsült érték mellett ott szerepel-

jen a sztenderd mintavételi hiba a táblázatokban és ábrákon. Ha ez elmarad, akkor nem lehetünk egészen biztosak benne, hogy a kutatók tudták, mit csinálnak.

Néhány kérdés, amit fel kell tennünk:

- A tesztpontszámok (mind a teszt-összpontszámok, mind résztesztek összpontszámú) kiszámításához elegendő számú item állt-e rendelkezésre? Ha nem, kétség merülhet fel azzal kapcsolatban, a kutatók értették-e a dolgukat.*
- Figyelembe vették-e az összes fontos változót a változók kapcsolatainak elemzésekor?*
- Szerepel-e a sztenderd mintavételi hiba minden egyes becsült adat esetében?*

10. Jól megírt-e a jelentés?

A jelentésnek világosan érthetőnek kell lennie, és minden egyes felvetett szakpolitikai kérdéssel külön-külön kell foglalkoznia. A tárgyalt adatok forrásainak mindig egyértelműnek kell lenniük, akárcsak az elemzések értelmezése során alkalmazott érveknek. Nyilvánvalónak kell lennie, ha a vizsgálat során (mint az esetenként előfordul) először a legfőbb egyváltozós elemzések eredményeit adják meg, majd a kutatási kérdéseket csoportokba rendezve külön jelentésekben tárgyalják.

Fontos, hogy még a jelentés végső formába öntése előtt a kutatók visszajelzést kapjanak a munkájukról. Ez részint más kutatók véleménye, részint a jelentésben tárgyalt eredmények majdani felhasználóié, részint pedig a vizsgálatban érintetteké, például iskolai igazgatóké, tanároké. Abban az esetben, ha a felhasználó az oktatási minisztérium, tanácsos, hogy a kutatók a publikálás előtt a minisztérium felelős munkatársaival megbeszéljék a javasolataikat. (Ha így történt, ezt jelzik a jelentésben.) Ugyancsak hasznos a minisztérium számára, ha a kutatók az eredményeket nem csak téma szerint csoportosítják, hanem költség (alacsony, közepes, magas költségek), illetve a megvalósításhoz szükséges idő (rövid, közepes, hosszú távú javaslatok) szerint is.

Végül szokás az adatállomány archiválása és hozzáférhetővé tétele, hogy mások elemezhesék azt maguk, s ily módon meggyőződhesenek a kutatók állításának igazságtartalmáról és felderíthessék, milyen további kérdések megválaszolása lehetséges a gyűjtött adatokból. Fontos, hogy az archívumot a jelentés publikálását követően gyorsan (vagy épp azzal egy időben) és felhasználóbarát módon tegyék hozzáférhetővé.

Irodalom

Beaton, A. E., Postlethwaite, T. N., Ross, K. N., Spearritt, D. és Wolf, R. M. (1999): *The Benefits and Limitations of International Educational Achievement Studies*. International Academy of Education and International Institute for Educational Planning, Paris.

Keeves, J. P. (1988, szerk.): *Educational Research, Methodology, and Measurement: An International Handbook*. 1. kiadás. Pergamon Press, Oxford.

Keeves, J. P. (1992, szerk.): *The IEA Technical Handbook*. IEA, The Hague.

T. Neville Postlethwaite

Keeves, J. P. (1997, szerk.): *Educational Research, Methodology, and Measurement: An International Handbook*. 2. kiadás. Pergamon Press, Oxford.

Martin, M., Rust, K. és Adams, R. (1999): *Technical Standards for IEA Studies*. IEA, Amsterdam.

Schleicher, A. (1994): *Standards for the Design and Operations in IEA Studies*. IEA, The Hague.

Fordította: Braunitzer Gábor

ABSTRACT

T. NEVILLE POSTLETHWAITE: TEN POINTS TO ENSURE THAT CROSS-NATIONAL STUDIES ARE OF GOOD QUALITY

There are several cross-national studies in existence at the beginning of the 21st century. Some have very high standards of quality and others are of very questionable quality such that the results cannot be trusted. The aim of this short article is to highlight ten points that readers should look for in any study in order to be able to judge the technical soundness of the research. These include the aims, the target population, the sampling, the development, construction and pre-test (as well as the verification of the translations) of the instruments, data collection and handling, analyses, reports and archival access.

Magyar Pedagógia, **104**. Number 1. 5–18. (2004)

Levelezési cím / Address for correspondence: T. Neville Postlethwaite, University of Hamburg