

# Hogyan térnek el a papír-ceruza és számítógépes teszteredmények? – Szisztematikus szakirodalom áttekintés a PISA, TIMSS és PIRLS mérésekkel kapcsolatos tapasztalatokról

*Tanulmányunk célja, hogy a nemzetközi nagymintás tanulói teljesítménymérések tapasztalatai alapján vizsgálja a médiahatás (mode effect) nagyságát, vagyis feltérképezze a papír-ceruza és az elektronikus mérési módok közötti lehetséges különbségeket. A vizsgálat alapjául a PISA, PIRLS és TIMSS méréseket választottuk, mivel a hazai nagymintás mérés, az Országos kompetenciamérés (OKM) ezek módszertanát követi. Az OKM 2022-ben első alkalommal valósult meg teljesen elektronikus formában.*

## Bevezető

Az Országos kompetenciamérés (OKM) a kétezres évek eleje óta méri a magyarországi tanulók szövegértési képességét és matematikai eszköztudását (Balázsi és mtsai., 2014). Az egyes évek eredményeinek trendjét kiegészítő mérés segítségével alakították ki (Auxné Bánfi és mtsai., 2014). A papír-ceruza alapú mérés 2010-re érte el a végső formáját, amikor a 6., 8. és 10. évfolyamos tanulók teljes populációját mérte fel és helyezte el közös képességskálán. A képességskála segítségével az egyes évfolyamok összehasonlíthatók, a 2008-ban bevezetett mérési azonosító segítségével az egyéni fejlődés is követhető.

A mérés célja olyan objektív mutatók létrehozása, melyek segítik az iskolák önértékelését és a fejlesztési irányok kijelölését, továbbá a fenntartók, partnerek (ideértve a szülőket és tanulókat is) és az oktatásirányítás informálását (Balázsi és mtsai., 2014). A mérés módszertanát az ekkor Magyarországon megjelenő nemzetközi mérések (a korábbi méréseket folytató TIMSS 1999, a PISA 2000 és a PIRLS 2001) módszertanát követve alakították ki, így nagyban hozzájárultak a nemzetközi mérési kultúra hazai megismeréséhez (Csíkos & Vidákovich, 2012).

A médiahatást értelmezhetjük a két tesztmédiumon elért teljesítmény közötti különbségként (R. Tóth & Hódi, 2011). A három nemzetközi mérés mindegyike közöl időbeli trendeket és országok közötti összehasonlításokat, ezért a papír-ceruza tesztek számítógépes tesztekre történő cseréjének következményét, az esetleges médiahatást vizsgálni és szükség szerint kontrollálni kell. A trendek folytonossága érdekében a nemzetközi mérések esetében kérdés lehet, hogy a médiahatást egységesen, vagy országonként

különböző mértékben kell-e igazítani. Felvetődhet továbbá a két módon mért konstruktumok azonosságának kérdése is.

Magyarországon elsősorban az eDia mérés-értékelési keretrendszerben megvalósuló mérésekhez kapcsolódnak médiahatás vizsgálatok. R. Tóth és Hódi (2011) hatodik évfolyamos tanulók esetében szignifikáns, átlagosan csaknem 5 százalékos eltérést talált a két módban felvett szövegértés teszteredmények összehasonlítása során, megállapítva, hogy a diákok átlagosan jobb eredményt értek el a nyomtatott médiumon, mint számítógépen. Eredményeik szerint a nem folyamatos szövegek esetén a legjelentősebb különbség a diagramleolvasási feladatoknál, a legkisebb pedig a táblázatba foglalt információk megértésénél volt tapasztalható. Herczegné Goldschmidt (2016) elemzése szerint szövegértés területen pozitív, közepes mértékű együttjárást mutattak a papír-ceruza és a számítógép-alapú teszteredmények. Vizsgálata során általánosságban arra jutott, hogy papíralapon a diákok szignifikánsan, 3–4 százalékkal eredményesebben oldották meg a tesztet, hozzáteve, hogy a tesztek itemszintű vizsgálatakor adódott olyan item, amely az online teszten bizonyult könnyebbnek. Hülber (2012) 1–6. évfolyamos tanulók körében vizsgálta a papír-ceruza és a számítógépes matematika tesztek itemeinek tulajdonságait. A nehézségi paraméterek szignifikáns, közepes együttjárást mutattak. Elemzése alapján a számítógépes teszt itemei szignifikánsan nehezebbnek bizonyultak. Egyéb jellemzők mentén az induktív gondolkodás dimenzióban és a grafikus elemet tartalmazó feladatok kapcsán volt kimutatható szignifikáns eltérés. A tanulók neme szerint nem állt fenn számottevő különbség a két médium között. Későbbi vizsgálatokban (Hülber & Molnár, 2013; Pásztor-Kovács és mtsai., 2013) egyik évfolyamon sem találtak szignifikáns teljesítménykülönbséget a két tesztmédium között.

Az OKM a korábban említett nemzetközi és hazai mérésekhez hasonlóan szintén áttért a számítógépes tesztelésre. A mérés 2022-től teljes egészében számítógépes adatfelvétellel, a természettudomány mérési területtel kiegészülve került megszervezésre (Oktatási Hivatal, 2021). A technológiai váltás lehetséges következményeinek feltárása és előkészítése érdekében készült egy nemzetközi és hazai tapasztalatokat összegző jelentés (Molnár és mtsai., 2015). A médiahatással foglalkozó fejezet rámutat a nemzetközi eredmények sokszínűségére, amit részben a vizsgálatok módszertani háttere, az eszközök, a vizsgált konstruktumok és a minták különbözőségével magyaráz. A hazai eredményekkel összevetve az OKM esetében nem számít jelentős médiahatásra, legfeljebb a jobb teljesítményű tanulók esetében kissé gyengébb eredményre. A felső tagozatos és középiskolás korcsoportban az életkor, a nem, a szocioökonómiai státusz és az IKT eszközök használata szerint nem mutatható ki jelentős médiahatás.

Mivel az OKM közöl trendeket, illetve összehasonlításokat, ezért szükséges a médiahatás nagyságának és az eredmények értelmezésére gyakorolt esetleges hatásának vizsgálata. Felvetésünk szerint egy ilyen vizsgálat előkészítésében a szükséges és lehetséges módszerek feltérképezése, a módszertanilag példaként állított nemzetközi mérések során szerzett tapasztalatok összegzése elengedhetetlen. A mérési kultúra megismertetésének szempontjából fontosnak tartjuk a nemzetközi mérések módszertani leírásainak bemutatását is, ezzel segítséget adva a számítógép-alapú mérések eredményeinek értelmezéséhez. A médiahatás témájának aktualitását adja az is, hogy 2021-ben került sor a PIRLS első, 2022-ben a PISA harmadik, és 2023-ban a TIMSS második számítógépes adatfelvételi mérésére.

## A nemzetközi mérések digitalizációja

### PISA

A Programme for International Student Assessment (PISA) a Gazdasági Együttműködési és Fejlesztési Szervezet (*Organisation for Economic Co-operation and Development*, OECD) által háromévente megszervezett tanulói teljesítménymérés, mely legutóbb 2018-ban 79 országban vizsgálta a 15 éves tanulók szövegértési és matematikai műveltségét (OECD, 2019b). A PISA mérés esetében három mérési ciklust (2009, 2012, 2015) érdemes vizsgálni a számítógépes mérés bevezetése szempontjából. 2009-ben volt először kiegészítő mérés a digitális szövegértés (*digital reading*), amit weblap-szerű nemlineáris szövegek feldolgozásával mértek, melyeknek mind a formája, mind a szituációk szerinti megoszlása különbözött a papír-ceruza szövegértés terület mérésétől (OECD, 2009, 25-26). A mérés hazai szervezője külön kötetet szentelt a digitális szövegértés területnek (Balázi & Ostorics, 2011), melyben bemutatják a két terület közti hasonlóságokat és különbségeket, valamint ismertetik a hazai és nemzetközi eredményeket. 2012-ben ismét választható mérési terület volt a digitális szövegértés, a skálát ekkor már a 2009-es digitális szövegértés skálához igazították a mérés módszertanának megfelelően.

Bár a digitális szövegértés terület skáláit az összehasonlíthatóság kedvéért a nemzetközi átlag alapján a szövegértés skálához igazították, ezt nem előzte meg semmilyen médiahatás vizsgálat. Erre utal, hogy a mérés egyik dokumentumában sem jelenik meg a médiahatás angol megfelelője, a *mode effect* kifejezés. A digitális szövegértés területen alkalmazott itemek nem papír-ceruza itemek elektronikus változatai, ezért vizsgálatunkban a két mérési területen elért eredmény összehasonlítását nem tekintettük médiahatás vizsgálatnak.

A 2012-es ciklusban a digitális olvasás mellett bevezetésre került a számítógépes matematika mérése (*computer-based assessment of mathematics*), mint választható terület (OECD, 2013). A matematika fogalmát és tartalmi keretét kibővítették, hogy illeszkedjen hozzá az új médiumon mért konstruktum is. A számítógépes matematika mérése magában foglalja a számítógépes lehetőségek használatának képességét, azonban lehetőség szerint az IKT eszközhasználati igény elmarad a matematikai műveletek mögött. Az itemek bizonyos része, de nem mindegyike, innovatív megoldásokat tartalmaz. Az itemek megoszlása a műveleti, tartalmi és kontextuális kategóriák között azonos a két felmérésben. A számítógépes matematika skálát a matematika területéhez igazították, mivel ez volt az első alkalom a terület mérésére (OECD, 2014, 253). A két matematika mérés egységére utal, hogy a külön skálák mellett kombinált skála is kialakításra került (OECD, 2014a, 40). A mérések hazai szervezője a digitális szövegértés és a számítógépes matematika területeket a fő területekkel egy kötetben mutatja be, ismertetve a matematikai keret bővítését és az itemek készítésének szempontjait (Balázi és mtsai., 2013).

A 2015. évi PISA mérés teljes egészében elektronikus formában valósult meg (OECD, 2016b), országonként meghagyva a papír-ceruza mérés lehetőségét (OECD, 2017a). Azokban az országokban, ahol ez utóbbit választották, a mérés kizárólag trend itemekkel valósult meg. Vagyis a papír-ceruza tesztek kizárólag olyan itemeket tartalmaztak, amelyek a 2012. és a 2015. évi ciklusokban is szerepeltek – mely lehetővé teszi, hogy a PISA 2015 országos eredményeit trendben és a többi országhoz képest értelmezni lehessen –, de nem tartalmazta azokat az új itemeket, amelyek a következő mérési ciklusok trend-itemei lesznek. Szintén ebben a mérési ciklusban vált először a háttérkérdőívek felvételének elsődleges módjává a számítógépes kitöltés.

## PIRLS

A Tanulói Teljesítmények Vizsgálatának Nemzetközi Társasága (*International Association for the Evaluation of Education Achievement*, IEA) által öt évente megszervezett Progress in International Reading Literacy Study (PIRLS) célja a 4. évfolyamos tanulók szövegértési képességének felmérése, melynek 2016-os mérési ciklusában 50 ország vett részt (Mullis és mtsai., 2017). 2016 óta a mérésnek része a PIRLS Literacy (Mullis & Martin, 2015), mely mérési keretében megegyezik a szövegértés méréssel, de könnyebb kérdéseket tartalmaz, így a teljesítményskála alsóbb részére irányul. Ez a mérés olyan országokban választható, ahol 4. évfolyamon még az alapvető olvasási képességek fejlesztése zajlik. Az egyező feladatok miatt a PIRLS és a PIRLS Literacy skálái összehasonlíthatók, az egyes országok eredménye összehasonlítható.

A digitalizáció szempontjából a 2016-os és 2021-es méréseket vettük figyelembe, mivel 2016-ban került bevezetésre az ePIRLS kiegészítő mérés (Mullis & Martin, 2015), 2021-ben pedig először volt elsődleges az elektronikus adatfelvétel (Mullis & Martin, 2019). A mérés dokumentumaiból kiderül, hogy a választható ePIRLS mérés célja kissé eltér a papíralapú mérésétől, mivel az iskolai környezetben folytatott interneten történő olvasás szövegértését vizsgálja (Mullis & Martin, 2015, 22). Ehhez igazodva mind a teszt felépítése, mind az összetétele különbözik a főmérésétől. Egyrészt, a lineáris, nyomtatott szövegek helyett internet-szerű, weboldalakat szimuláló felületen „böngésznek” a tanulók. Ennek következtében a mérés tartalmaz – a szövegértés mellett, vagy annak részeként – bizonyos stratégiai elemeket, például egy még meg nem tekintett oldal információjának megbecsülését, vagy a korábban megtekintett oldalakon talált információkra való emlékezést. Másrészt, a mérés céljához igazodva, a szövegek teljes egészében információkeresési célúak, eltérően a főmérésétől, ahol a szövegek egyik felének célja az élményszerzés, a másik felének pedig az információszerzés (Mullis & Martin, 2015, o. 14, Exhibit 2.). A terület eredményeit az itemek szintjének meghatározásával a szokásos PIRLS skálához igazítják (Martin és mtsai., 2017, 13. fejezet), ezáltal a két terület eredményei összehasonlíthatók. A szervezők nem végeztek médiahatás-vizsgálatot a két mérés összehasonlításával. A két terület eltérő tartalma és mérése alapján a PIRLS és ePIRLS méréseket vizsgálatunkban nem tekintettük ugyanazon mérés papír-ceruza és számítógépes változatának, az eredmények összevetését nem fogadtuk el médiahatás-vizsgálatnak. A 2016-os ePIRLS mérésben Magyarország nem vett részt, így a méréshez kapcsolódó hazai jelentésben az ePIRLS-t meg sem említik (Balázi és mtsai., 2017).

A PIRLS 2021 mérési ciklusban az adatfelvétel a résztvevő országok felében digitális (digitalPIRLS), másik felében papír-ceruza (paperPIRLS) formában valósult meg (Mullis & Martin, 2019, 57). Az összehasonlíthatóság érdekében a két mérés tartalmában megegyezik. A digitalPIRLS és ePIRLS területek különbségét erősíti meg, hogy az ePIRLS továbbra is külön mérési területként szerepel, azonban a szervezők kifejezik a terület integrálása iránti szándékukat (Mullis & Martin, 2019, 67), például PIRLS és ePIRLS hibrid tesztfüzetekkel.

A PIRLS 2021 mérés összefoglaló eredményei várhatóan 2023 májusában jelennek meg, ezért vizsgálatunkban ennél a mérésnél előkészítő dokumentumokra, esetleg a médiahatást előzetesen vizsgáló elemzésekre számítottunk.

## TIMSS

Szintén az IEA szervezésében négy évente kerül sor a Trends in International Mathematics and Science Study (TIMSS) vizsgálatokra, mely 4. és 8. évfolyamon méri a tanulók matematikai és természettudományi tudását (Mullis & Martin, 2017). A legutóbbi,

2019-es mérésben 64 ország vett részt (Mullis és mtsai., 2020). A TIMSS mérésnek 2015 óta része a könnyebb feladatokat tartalmazó TIMSS Numeracy, mely 4. évfolyamon méri a matematika területet olyan országokban, ahol ezen az évfolyamon még az alapvető számolási készségek fejlesztése zajlik. A két mérés eredményeit, hasonlóan a PIRLS és PIRLS Literacy méréshez, azonos teljesítmény-skálán lehet megjeleníteni.

A digitalizáció szempontjából a 2019. évi mérési ciklus jelentős, mivel e mérés során vezették be a számítógépes mérési rendszert. Ugyanebben a mérési rendszerben kerül megszervezésre 2021-től a PIRLS mérés is. A TIMSS mérés 2019-ben magában foglalta az elsődleges számítógépes mérést (eTIMSS) – amiben az országok több, mint fele vett részt –, a papír ceruza mérési módot (paperTIMSS) és a TIMSS Numeracy mérést. Mindhárom nemzetközi tanulói teljesítménymérés hazai szervezője az Oktatási Hivatal.

### *Összegző tapasztalatok*

Az említett három nemzetközi nagymintás teljesítménymérés az utóbbi mérési ciklusok során a papír-ceruza tesztfüzeteket számítógépes tesztekre cserélte. A különböző mérések esetében eltérő indoklással találkozhatunk. A munka világában történő helytálláshoz szükséges kompetenciák mérését célul tűző PISA esetében bizonyos 21. századi képességek mérésére (mint az internetes szövegek értése vagy az interaktív matematikai eszközök használata) alkalmasabbnak találták a számítógépes megvalósítást (OECD, 2017c). A PIRLS esetében a különböző nehézségű mérések (PIRLS és PIRLS Literacy) és a 2016-ban bevezetett ePIRLS egységes szervezési és megvalósítási lehetősége indokolta a váltást (Mullis & Martin, 2019). A TIMSS esetében a természettudomány terület mérését kiegészítő innovatív, kísérleti situációkat szimuláló problémamegoldás terület bevezetése, valamint a különböző nehézségű mérések egységes keretben való szervezése volt a fő indok (Mullis & Martin, 2017).

A számítógépes mérések a TIMSS és PIRLS esetében a csoport adaptív (*group adaptive*) (Mullis és mtsai., 2021; Mullis & Martin, 2019), a PISA esetében a többszakaszos adaptív mérések (*multistage adaptive testing*) irányába mozdultak el (OECD, 2019a). A csoport adaptív mérés a különböző szintű vagy célú mérési anyagok egységes elektronikus rendszerben történő fejlesztését, szerkesztését teszi lehetővé. A mérés is ugyanebben az egységes rendszerben valósul meg, azaz a tanuló országának (vagy valamely egyéb jellemzőjének) ismeretében a rendszer a megfelelő, például a könnyebb feladatokat tartalmazó TIMSS Numeracy tesztfüzet kitöltését teszi lehetővé. A többszakaszos adaptív tesztelés ezzel szemben minden résztvevőhöz azonos nehézségű kezdő teszt-szakaszt rendel, és a kitöltés sikeressége alapján a folytatáshoz könnyebb vagy nehezebb teszt-szakaszt sorsol. Ebben az értelemben a csoport adaptív mérés inkább logisztikai vagy mérésszervezési eszköznek tekinthető, mivel maguk a kitöltött „füzetek” hagyományos lineáris tesztek.

### **Kutatási kérdések, módszerek**

Vizsgálatunkban az alábbi két kutatási kérdésre keressük a választ:

1. Milyen tudás halmozódott fel a PISA, PIRLS és TIMSS mérések digitalizációjával, azon belül a médiahatással kapcsolatban a hazai és nemzetközi tudományos publikációkban és mérési dokumentumokban?
2. Mekkora és milyen irányú a médiahatás a PISA, PIRLS és TIMSS mérések esetében?

A vizsgálat módszerül a szisztematikus szakirodalmi áttekintést (Rother, 2007) választottuk. A kutatás során a PRISMA elveit követtük (Page és mtsai., 2021). Az adatbázisokban folytatott keresés célja a lehető legtöbb és legrelevánsabb forrás felfedése és szintetizálása előre jól meghatározott keresési és kizárási kritériumok alapján. A keresést 2021. december 2-án hajtottuk végre.

### *Adatbázisok*

A hazai szerzők által publikált eredmények keresésére három adatbázist használtunk. Az egyik a magyar folyóiratok tartalomjegyzékeinek kereshető adatbázisa (MATARKA), a másik a Magyar Tudományos Művek Tára (MTMT), harmadik pedig az Arcanum Digitális Tudománytára (arcanum.com/hu) volt. A nemzetközi adatbázisok esetében az oktatási, neveléstudományi témájú forrásokat is tartalmazókat vizsgáltuk. A választást az intézményi hozzáférés befolyásolta. A keresésbe bevont adatbázisok az EBSCO (kivéve a Green és Legal adatbázist), az ERIC, a JSTOR, a ProQuest, a Science Direct és a Web of Science voltak. Ezek mindegyike lehetővé teszi a részletes keresést, összetett logikai kifejezések szerinti keresést, a megjelenés évére és a lektorált (*peer review*) forrásokra történő szűkítést.

A nemzetközi mérések dokumentumait a szervezők oldaláról gyűjtöttük össze. A TIMSS és PIRLS mérések szervezője megegyezik, a dokumentumok egyetlen felületről érhetőek el (<https://timssandpirls.bc.edu/isc/publications.html>). A PISA méréshez kapcsolódó publikációk két helyen, a mérés saját oldalán (<https://www.oecd.org/pisa/publications/>) és a szervező publikációit tartalmazó felületen ([https://www.oecd-ilibrary.org/education/pisa\\_19963777](https://www.oecd-ilibrary.org/education/pisa_19963777)) is elérhetőek. Mi az előbbit használtuk, mivel itt mérésenként összegyűjtve szerepelnek a kapcsolódó információk. A magyar nyelvű mérési dokumentumokat az Oktatási Hivatal honlapjáról (oktatas.hu) a mérések saját aloldaláról gyűjtöttük le, de nem vontuk be azokat az egyéb elemzéseket, amelyek más aloldalon nem lektorált tudományos közlemény formájában jelentek meg. Elfogadtuk más országok mérési dokumentumait is, ha azok angol vagy magyar nyelven jelentek meg, és szerepelnek valamely nemzetközi adatbázisban.

A dokumentumok jellemzően a mérések keretrendszerét, a mérés eredményeit és a technikai részleteket közlik. A keretrendszerek azonosítják az egyes méréseket, azok céljait, a méréshez kialakított elméleti háttérrel és előre vetítik a mérés bizonyos módszertani és technikai jellemzőit. Az eredmények jellemzően a mérés utáni évben jelennek meg, és az egyes országok eredményei, valamint a háttérkérdőívek alapján számított statisztikák mellett tartalmaznak bizonyos mérésmódszertani leírásokat is. A technikai leírások kifejezetten a mérés és elemzés megismerésének és megismételhetőségének érdekében készülnek, és tartalmaznak minden olyan információt, ami az itemek fejlesztésétől, az adatok rögzítésén és tisztításán át az elemzés elméleti háttérig releváns lehet.

### *Beválogatási és kizárási kritériumok*

A keresési időszak kezdetét a digitális adatfelvételek időpontjához igazítottuk. Mivel a mérés médiumának cseréje már mind a három nemzetközi mérés esetében megtörtént – a PISA 2015-ben, az eTIMSS 2019-ben, a digitalPIRLS 2021-ben alkalmazta teljeskörűen ezt a megvalósítást –, ezért feltételeztük, hogy mind leírások, mind kapcsolódó független vizsgálatok is születtek már a témában. Azt is feltételeztük, hogy a 2015-ös PISA mérés előkészítése jóval a mérés előtt megtörtént, ezért az adatbázisokban folytatott keresés kezdeti időpontját 2010. január 1-ben határoztuk meg. A keresési időszak záró időpontja 2021. november 30. volt. A mérések dokumentumai esetében a PISA 2009, TIMSS 2019

és PIRLS 2016 mérési ciklusokkal kezdtük a vizsgálatot, tekintet nélkül a dokumentum megjelenési évére.

Az adatbázisokban végzett keresés esetén további feltételként szabtuk meg, hogy a találatok lektorált (*peer review*) publikációk legyenek. A mérések saját dokumentumainál ezt a feltételt nem kötöttük ki, mivel ezek nem *peer review* publikációk, de a mérések integritását biztosító fontos dokumentumok, melyeket tudományos munkákban is hivatkoznak, illetve saját kutatás tervezésekor is figyelembe vesznek.

A publikációk nyelvét magyar és angol nyelvben határoztuk meg. A megszorítás miatt előfordulhat, hogy bizonyos régiókra vonatkozó eredmények kiszorulnak a vizsgálatból, azonban álláspontunk szerint a mérések nemzetközi jellege miatt a mérések szervezői által készített dokumentumok és a releváns tudományos eredmények mindenképpen bekerülnek.

A közlemények formai szempontból lehetnek tudományos cikkek, tanulmánykötetben megjelent tanulmányok és könyvfejezetek. Kizárásra kerültek a recenziók, a könyvismertető, az interjú, a konferencia-előadások és az absztraktok. További feltétel, hogy a tételek lektoráltak legyenek. Ezt a külföldi adatbázisok esetében a keresési beállításokban szabályoztuk, a magyar keresés esetében a megjelenés helyének ellenőrzésével. Erre folyóiratok esetében az MTMT folyóiratkereső felületét alkalmaztuk, mely feltünteti a folyóirat tudományos és lektorált jellegét.

A publikációk bevonására tartalmi és szervezési kritériumokat is meghatároztunk. Azok a publikációk kerültek bevonásra, melyek empirikus kutatás eredményét vagy ilyen kutatások összegzését közlik. A kutatás minimális mintanagyságát kvantitatív vizsgálat esetében 100 főben, kvalitatív vizsgálat esetében 15 főben határoztuk meg. További kritériumként szabtuk meg, hogy a kutatás célja a papír-ceruza és számítógépes adatfelvétel közötti, a mérés módjából következő eltérések vizsgálata vagy a két mérési mód összehasonlítása legyen. További megkötés, hogy a kutatás szorosan kapcsolódjon a három vizsgált nemzetközi méréshez, azaz eredeti adatbázisokat vagy eredeti feladatokat használjon fel, illetve a lebonyolítás szorosan kapcsolódjon a felmérésekhez. A két mérési mód vizsgálata kizárja azokat a választható területeket és méréseket, melyek kizárólag számítógépes adatfelvétellel valósultak meg.

### *Kulcsszavak*

A kereséshez a kulcsszavakat két szempont alapján választottuk ki. Egyrészt fókuszáltunk a kiválasztott három nemzetközi tanulói teljesítménymérésre, melyek mozaikszavai (PISA, TIMSS, PIRLS) megfelelő indikátorok a mérésekhez kapcsolódó összes forrás megtalálásához. A felmérések teljes nevét nem tartottuk fontosnak bevásztani, mivel a szövegekben ugyan szükségszerűen megjelenik, de a címekben, az absztraktokban és a kulcsszavak között a rövid változat a jellemző, a teljes szövegben pedig biztosan felbukkan. A hazai adatbázisokban történt keresés esetében nem számítottunk feldolgozhatatlan számú találatra, ezért a kulcsszavakat nem bővítettük tovább.

A mérésekhez kapcsolódó teljes nemzetközi szakirodalom várható gazdagsága miatt a nemzetközi adatbázisokban történt keresést további kulcsszavakkal egészítettük ki. Egyrészt a médiahatás angol megfelelőit vontuk be, melyek azonban különbözőek az egyes mérések esetében. A magyar médiahatás szó angol eredetije a „*mode effect*”, ami a papír-ceruza és a számítógépes adatfelvétel eredménye közötti szisztematikus eltérésre utal. A kifejezés általánosan elterjedt, a magyar forrásokban is ez szerepel angol eredetiként (lásd például R. Tóth & Hódi, 2011). Ezt a kifejezést használja a PISA (OECD, 2017b, 152-162), az erre vonatkozó vizsgálatot „*mode study*” kifejezéssel jelöli. Az összefoglalóban (OECD, 2017c) kötőjeles formában jelenik meg („*mode-effect*”), azonban összesen három alkalommal szerepel. A TIMSS mérés az „*item equivalence*”

kifejezést használja (von Davier és mtsai., 2020), ami arra utal, hogy a mérési módok közötti megfeleltetést olyan itemekkel lehet biztosítani, melyek egyformán viselkednek a két adatfelvételi felületen. Az erre vonatkozó vizsgálatot az „*item equivalence study*” kifejezéssel jelöli.

A mérések és a médiahatás kulcsszavaival („*mode effect*” és/vagy „*item equivalence*”) a nemzetközi adatbázisokban igen kevés, jellemzően 10 alatti találatot kaptunk. Ez alapján úgy döntöttünk, hogy a témára vonatkozó kulcsszavakat a számítógépes adatfelvételre vonatkozó „*computer-based*” kulcsszóval bővítjük, mivel médiahatás-vizsgálat esetében ennek a szónak – a mérések eredeti papír-ceruza (*paper-based*) adatfelvételi módjához képest – biztosan meg kell jelennie. Ez a találatok számának jelentős növekedését hozta. Az egyes adatbázisok, keresőszavak és alkalmazott szűkítő feltételek kombinációját az 1. táblázat tartalmazza.

1. táblázat Az egyes adatbázisokban futtatott keresések kulcsszavai és beállításai

Adatbázis	Kereső kifejezés	Egyéb kritérium
Arcanum	SZO=(pisa timss pirls) AND DATE=(2010--)	
MATARKA	PISA OR TIMSS OR PIRLS	2010-től
MTMT	PIRLS	Jelleg: Tudományos Év: >=2010
MTMT	PISA	Jelleg: Tudományos Év: >=2010
MTMT	TIMSS	Jelleg: Tudományos Év: >=2010
EBSCO ERIC JSTOR ProQuest Science Direct Web of Science	(PISA OR TIMSS OR PIRLS) AND („mode effect” OR „item equivalence” OR „computer-based”)	2010-től lektorált

### A szakirodalomkeresés folyamata és eredménye

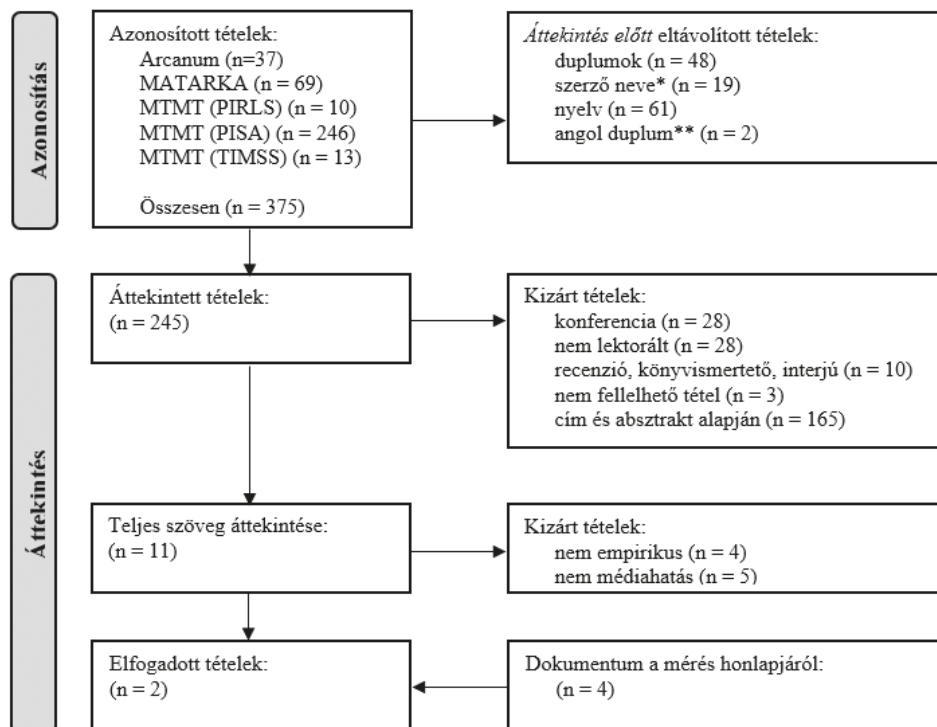
A keresések eredményét a Zotero (zotero.org) hivatkozáskezelő szoftverbe gyűjtöttük, és Microsoft Excel szoftver felhasználásával a metaadatok alapján szelektáltuk.

A hazai áttekintés kizárási folyamatát az 1. ábra mutatja. A találatok egy részében a „PISA” keresőszó a szerző nevében szerepelt. Más része nem magyar vagy angol nyelvű tétel volt. Egy tétel latin nyelvűként szerepelt, a szöveg ellenőrzése után magyar nyelvűnek bizonyult, ezért ezen kritérium alapján nem került kizárássra. Két tétel esetében a magyar nyelvű tétellel tartalmilag egyező, azonos szerzőtől származó angol nyelvű publikációt találtunk, az angol nyelvű tételeket kizártuk. A duplumok és ezen tételek eltávolítása után a tételeket a publikáció címe alapján is ellenőriztük. Az áttekintés során a folyóiratok címe alapján kizártuk a nem lektorált tételeket, továbbá formai szempontok alapján a konferenciaanyagokat, recenziókat, könyvismertetőket és interjúkat is. Három tételt online formában nem sikerült megtalálni. Cím és absztrakt alapján 165 tételt zártunk ki, ebből 44 tétel a kiválasztott mérésekkel foglalkozik, azonban nem a médiahatás témáját vizsgálja.

A teljes szöveg áttekintésére 11 publikáció esetében került sor, ezek esetében a téma érinthette a médiahatás kérdését. A tételek harmada nem empirikus megközelítésű volt,



a tételek fele pedig nem vizsgálja a médiahatás kérdését. Velkey (2018) munkája foglalkozik az adatfelvételi módok közötti különbséggel és összefoglal néhány médiahatással foglalkozó eredményt, de nem végez empirikus kutatást a témában, ezért ezt a tételt kizártuk az elemzésünkől. A hivatkozott munkákat áttekintettük további forrásokért, de új forrást nem találtunk.



1. ábra A magyar katalógusokban fellelt tételek PRISMA folyamatábrája. Saját ábra (Page és mtsai., 2021) alapján.

Jelmagyarázat \*: kulcsszó a szerző nevében; \*\*: magyar nyelvű szöveggel tartalmilag egyező angol nyelvű szöveg

A befoglalási kritériumoknak végül két tétel felelt meg, mindkettő az Oktatási Hivatal mérésekhez kapcsolódó dokumentuma. Az Oktatási Hivatal oldalán további négy, hasonló dokumentumot találtunk, melyek a megfelelő mérésekhez kapcsolódnak, és információval szolgálhatnak a kiválasztott három mérés médiahatásával kapcsolatban. Az így kapott hat tétel adatait a 2. táblázat tartalmazza. Mivel ezek mindegyike a mérésekhez kapcsolódó jelentés, ezért a nemzetközi mérési dokumentumokkal együtt dolgoztuk fel őket.

2. táblázat A PIRLS, PISA és TIMSS nemzetközi mérések hazai szervezőjénél (Oktatási Hivatal) fellelt technikai és összegző jelentések listája a megjelenés évének sorrendjében.

Forrás	Cím	Adatbázis
Balázi & Ostorics, 2011	PISA2009 Digitális szövegértés. Olvasás a világhálón	MTMT
Balázi és mtsai., 2013	PISA 2012 Összefoglaló jelentés	OH
Ostorics és mtsai., 2016	PISA 2015 Összefoglaló jelentés	OH
Balázi és mtsai., 2017	PIRLS 2016 Összefoglaló jelentés a 4. évfolyamos tanulók eredményeiről	MTMT
Oktatási Hivatal, 2019	PISA 2018 Összefoglaló jelentés	OH
Palincsár és mtsai., 2020	TIMSS 2019 Összefoglaló jelentés	OH

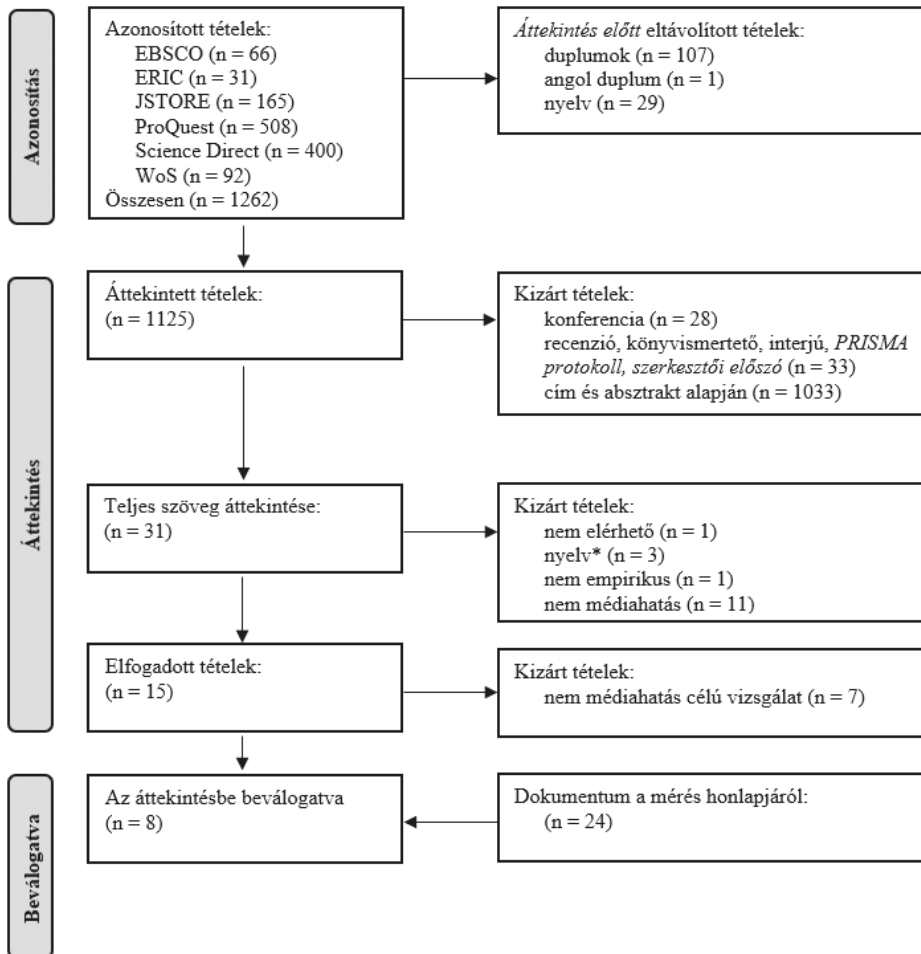
A nemzetközi adatbázisokban talált tételek szelekciós folyamatáról a 2. ábra tájékoztat. Áttekintés előtt a duplumokat és a nem angol nyelvű találatokat zártuk ki. A hivatkozáskezelőből exportált nyelvet – ahol eltért az előre meghatározottaktól vagy hiányzott – és a publikáció címét együttesen használtuk fel a nyelv szerinti kizárás eldöntésére.

Az áttekintés során 61 tételt zártunk ki a publikáció típusa alapján. Itt új típusként megjelentek a PRISMA protokollok és a szerkesztői előszavak, melyek az adatbázisokban cikként kerültek kategorizálásra. A cím és absztrakt alapján kizárt 1033 publikáció között 107 foglalkozott valamelyik kiválasztott méréssel, de témája nem a médiahatás vizsgálata volt. Ide tartoznak a trendvizsgálatok, a PISA csak számítógépes adatfelvétellel mért választható területeivel, például a komplex problémamegoldás (*Complex Problem Solving*) vagy a kollaboratív problémamegoldás (*Collaborative Problem Solving*) méréseivel kapcsolatos vizsgálatok, a számítógépes adatfelvétel naplófájljainak vizsgálatai, a tanulói kérdőív IKT eszközök használatával kapcsolatos válaszainak kutatása. Ezen a ponton még nem zártuk ki azokat a publikációkat, melyek a PISA 2009 és 2012 választható digitális szövegértését vizsgálták, feltéve, hogy a főmérés szövegértés területével vetették össze. 27 tétel a médiahatás vizsgálatára irányult, de nem valamelyik kiválasztott mérés adatait vagy keretét használta fel. A kódolás minőségének ellenőrzésére a tételek közel 20 százalékát (221 tétel) másodkódolásnak vetettük alá. Mindösszesen 5 tétel esetében különbözött a kódolók véleménye, a Cohen-kappa (Cohen  $\kappa = 0,69$ ) alapján ez jelentős egyezést jelent (Landis & Koch, 1977). Az eltérően kódolt tételeket a kódolók a teljes szöveg áttekintése során kizárták.

A tételek teljes szövegének vizsgálatára 31 tételt fogadtunk el. A teljes szöveg egy esetben nem volt elérhető. A tételek közül három angol címmel és absztrakttal szerepelt, de a teljes szöveg idegen nyelvűnek bizonyult (bolgár, koreai és német), így a nyelv alapján összesen 32 (áttekintés előtt 29, teljes szöveg áttekintésekor további 3) tétel került kizárásra. A feldolgozott 27 publikációból egy nem empirikus munka, további 11 kutatás célja pedig nem a médiahatás vizsgálatára irányult.

A fennmaradt 15 publikáció két markáns csoportba volt osztható. Az egyikbe olyan publikációk tartoznak, melyek kifejezetten a kutatási kérdésben meghatározott médiahatást vizsgálják, azaz itemek papír-ceruza és számítógépes adatfelvétel közötti jellemzőinek összehasonlítására irányulnak. A másik csoport olyan tételekből áll, melyek elsősorban a lineáris és nemlineáris (jellemzően online vagy digitális) szövegek olvasása közötti különbségeket vizsgálják különböző háttértényezők (például IKT használat) szerint, és a kétféle szövegértési folyamatot a kétféle adatfelvétellel mérik. Jellemzően a PISA 2009 és 2012 digitális szövegértés részterületen és a szövegértés főterületen elért eredményt, illetve a 2016. évi PIRLS és ePIRLS mérések eredményét vetik össze. A szövegek kódolását két kódoló végezte, két tétel esetében volt eltérő vélemény (kizárando vagy nem

médiahatás célú kapcsolatvizsgálat). A Cohen-kappa (Cohen  $\kappa = 0,89$ ) alapján ez majdnem tökéletes egyezés (Landis & Koch, 1977). Mivel a két kategória egyike sem került bele a végső feldolgozásba, ezért a vélelmes tételek esetében nem volt szükséges egyezésre jutni. Az elemzésbe beválogatott nyolc publikáció adatait a 3. táblázat tartalmazza.



2. ábra A nemzetközi adatbázisokban és a mérések dokumentumai között fellelt tételek PRISMA folyamatábrája. Saját ábra (Page és mtsai., 2021) alapján.

Jelmagyarázat \*: angol cím és absztrakt, teljes szöveg nyelve alapján kizárt

3. táblázat A nemzetközi adatbázisokban folytatott keresés eredménye.

Forrás	Cím	Kiadvány	Adatbázis(ok)
Fishbein és mtsai., 2018	The TIMSS 2019 Item Equivalence Study: Examining Mode Effects for Computer-Based Assessment and Implications for Measuring Trends	<i>Large-Scale Assessments in Education</i> , 6	ERIC, ProQuest, WoS
Hamhuis és mtsai., 2020	Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students?	<i>British Journal of Educational Technology</i> , 51(6), 2340–2358	EBSCO, ERIC, WoS
Jerrim, 2016	PISA 2012: how do results for the paper and computer tests compare?	<i>Assessment in Education: Principles, Policy &amp; Practice</i> , 23(4), 495–518	EBSCO, ERIC, WoS
Jerrim és mtsai., 2018	PISA 2015: how big is the 'mode effect' and what has been done about it?	<i>Oxford Review of Education</i> , 44(4), 476–493	EBSCO, ERIC, WoS
Kroehne és mtsai., 2019	Construct Equivalence of PISA Reading Comprehension Measured With Paper-Based and Computer-Based Assessments	<i>Educational Measurement: Issues &amp; Practice</i> , 38(3), 97–111	EBSCO, WoS
Robitzsch és mtsai., 2020	Reanalysis of the German PISA Data: A Comparison of Different Approaches for Trend Estimation With a Particular Emphasis on Mode Effects	<i>Frontiers in Psychology</i> (Vol. 11)	WoS
Zehner és mtsai., 2019	Unattended consequences: how text responses alter alongside PISA's mode change from 2012 to 2015	<i>Education Inquiry</i> , 10(1), 34–55	EBSCO, ERIC, ProQuest, WoS
Zehner és mtsai., 2020	PISA reading: Mode effects unveiled in short text responses	<i>Psychological Test and Assessment Modeling</i> , 62(1), 85–105	ProQuest

A nemzetközi mérések szervezőinek honlapjáról összesen 24 dokumentumot gyűjtöttünk össze a mérés éve és a kötet címe alapján. A dokumentumok összegzett jellemzőit a 4. táblázat tartalmazza. Ehhez a 24 forráshoz vettük hozzá a hazai szervező honlapján talált hat dokumentumot. A mérési dokumentumok jellemzően a mérések keretrendszerét, a mérés eredményeit és a technikai részleteket közlik. A három különböző célú kiadvány kiegyensúlyozottan szerepel a gyűjtésben, azonban a mérések közül a PISA nagyobb arányt képvisel. Ennek oka, hogy míg a TIMSS és PIRLS mérések lényegében egy-egy elektronikus ciklussal szerepeltek a keresésben meghatározott időintervallumban, addig a PISA esetében két digitális kiegészítő és két teljesen digitális mérés szerepelt.

4. táblázat A nemzetközi mérések saját dokumentumainak összegzett jellemzői. Az egyes cellákban a mérések adott célú dokumentumainak száma található.

Mérés	Mérési keret	Mérés eredményei	Technikai jellemzők	Egyéb	Összesen	Hazai dokumentum
PIRLS	2	2	1	-	5	1
TIMSS	2	1	1	-	4	1
PISA	4	5	4	2	15	4
Összesen	8	8	6	2	24	6

### Eredmények

A szövegértés és a digitális szövegértés eredményeket háttértényezők mentén összehasonlító publikációkat a teljes szöveg áttekintése során kizártuk jelen elemzésből. A matematikai műveltség papír-ceruza és számítógépes eredményeinek vizsgálatával foglalkozó forrásokról – a mérési keret azonossága miatt – a kutatás célja szerint döntöttünk. Amennyiben a digitális/számítógépes készségek és egyéb háttértényezők kapcsolatvizsgálatára irányult, akkor kizártuk, amennyiben a tartalmi keret egyezésén alapult, és a médiahatás vizsgálatára irányult, akkor belefoglaltuk jelen elemzésbe.

Nem találtunk olyan publikációt, mely ugyanabban a vizsgálatban több nemzetközi mérés adatait is felhasználja, ezért az eredményeket a mérések szerinti tagolásban mutatjuk be, összevonva a hazai és nemzetközi mérési dokumentumok és a nemzetközi adatbázisok találatainak eredményeit. A hazai mérési dokumentumok egy kivétellel a méréshez kapcsolódó összefoglaló jelentések, a hatodik kifejezetten az egyik részterület, a PISA 2009 digitális olvasás eredményeivel foglalkozik. Az egyes jelentések jellemzően szűkszavúan, egy-egy bekezdés erejéig foglalkoznak a médiahatás kérdésével. A nemzetközi adatbázisokból leválogatott 8 tétel esetében csak azokat a kutatási kérdéseket és eredményeket tárgyaljuk, amelyek a médiahatás vizsgálatával foglalkoznak.

### PISA

A 2015. évi PISA mérés az ezt vállaló országokban teljes egészében elektronikus formában valósult meg (OECD, 2016b). A hazai szervező az új mérési móddal kapcsolatban arról tájékoztat, hogy a médiahatást a próbamérés során vizsgálták, és nemzetközileg egységesen a trendek kiszámításakor figyelembe vették (Ostorics és mtsai., 2016). A 2018-as mérés összefoglalója (Oktatási Hivatal, 2019) a 2015 előtti és utáni eredmények összehasonlításánál óvatosságra int, valamint hivatkozik egy hazai elemzésre (Lak, 2020), amelynek célja a médiahatás magyarországi vizsgálata, azonban ez a publikáció nem lektorált tudományos folyóiratban került megjelentetésre, ezért elemzésünkbe nem vontuk be.

A mérési keret (OECD, 2017d) a médiahatással kapcsolatban említ korábbi vizsgálatokat és azok ellentmondásos eredményeit, valamint megemlíti az OECD PIAAC médiahatással kapcsolatos eredményét is. A dokumentum szerint a médiahatás vizsgálatát a 2015-ös mérés próbamérése során végezték, a leírás külön forrásként található meg (OECD, 2016a), és ebben található a médiahatás első PISA definíciója. E szerint „a médiahatás (*mode effect*) arra a megfigyelésre utal, ahol az egyik módban (például papír alapon) bemutatott feladatok másképp működhetnek, mintha egy másik módban (számítógép alapú) mutatják be őket” (OECD, 2016a, 4). A trend itemek vizsgálata klasszikus és modern tesztelméleti (*Item response theory*, IRT) modellek segítségével

történt. Az itemparaméterek korrelációja mind a nehézség, mind a meredekség paraméterek esetében 0,9 feletti. A médiahatást becsülő IRT modellek összehasonlítása alapján nem szükséges országonként vagy személyenként, elegendő itemenként alkalmazni a paraméterek eltolását. Ez egyrészt jelenti a mért konstruktumok azonosságát, másrészt az itemek paramétereinek egyedi vizsgálatát. A trend itemek nagy része skála invariánsnak bizonyult, azaz mindkét paramétere megegyezik a két mód között, ami lehetővé teszi a két mód összekötését. A metrikusan invariáns itemek, azaz amelyek esetében a nehézség különbözött, mindkét irányú médiahatást mutattak. Az ország-mód és a nem-mód kereszthatást regressziós modellekkel tesztelték, szignifikáns ( $p < 0,05$ ) kereszthatást nem találtak, azaz az itemenként végzett korrekció továbbra is biztosítja az egyes nemek és országok közti összehasonlíthatóságot. Felhívják ugyanakkor a figyelmet a kis mintanagyságra (országonként kb. 400 papír-ceruza és 600 számítógépes tesztkitöltés), mint az eredmények érvényességének korlátjára.

A főmérés technikai leírása (OECD, 2017a) szerint az itemeket a próbamérés eredményei alapján elemezték. Szintén bemutatják a próbamérés során végzett vizsgálatot. A főmérés trend itemeinek nagy része invariánsnak bizonyult, az összes item 90 százaléka legalább metrikusan invariáns volt (ld. 5. táblázat). Az itemek paramétereinek eltolása és az invariáns itemek segítségével a 2015. évi mérés eredményeit a korábbi trendhez kötötték. Itt jegyezzük meg, hogy a 2015. évi mérés során a skálák számításának módját két további ponton, az IRT modell megválasztásában és a korábbi ciklusok skáláinak felhasználásában is megváltoztatták. A PISA 2018 dokumentumai nem tartalmaznak médiahatásra vonatkozó új információt, azokban jellemzően a 2015-ös mérés dokumentumaiban közölt információkat közlik.

5. táblázat A közös és egyedi paraméterezésű itemek százalékos aránya a PISA 2015 egyes mérési területein (Forrás: OECD, 2017b. 225.) Rövidítések: PBA: papír-ceruza teszt (paper based assessment), CBA: számítógépes teszt (computer based test).

	Matematika	Szövegértés	Természettudomány trend
% egyedi paraméter (csoportra jellemző)	2,16%	3,01%	2,62%
% egyedi paraméter (néhány csoportra jellemző)	3,36%	7,98%	7,68%
% metrikusan invariáns közös/nemzetközi paraméterek	33,22%	30,33%	20,96%
% skála invariáns közös/nemzetközi paraméterek	61,25%	58,68%	68,74%
Mód és az itemek száma a PISA 2015 főmérésben	PBA item: 83 CBA item: 81	PBA item: 103 CBA item: 103	PBA item: 85 CBA item: 85

A nemzetközi adatbázisban történt keresés eredményeként kapott 8 találat közül 6 (Jerrim, 2016; Jerrim és mtsai., 2018; Kroehne és mtsai., 2019; Robitzsch és mtsai., 2020; Zehner és mtsai., 2019, 2020) foglalkozik a PISA méréshez kapcsolódó médiahatás vizsgálatokkal. Ezek körében is két jól elkülöníthető csoportot alakíthatunk ki. A Jerrim (2016; 2018), Kroehne és munkatársai (2019), valamint Robitzsch és munkatársai (2020) nevével jelzett források a próbaméréshez hasonló vizsgálatokat végeznek, a médiahatást az itemek jellemzőinek és az egyes területeken mért teljesítmények összehasonlításával elemzik. Zehner és munkatársai (2019, 2020) egymásra épülő kutatásokat mutatnak, melyek a nyílt válaszok bizonyos jellemzőinek a két tesztmédiium közötti

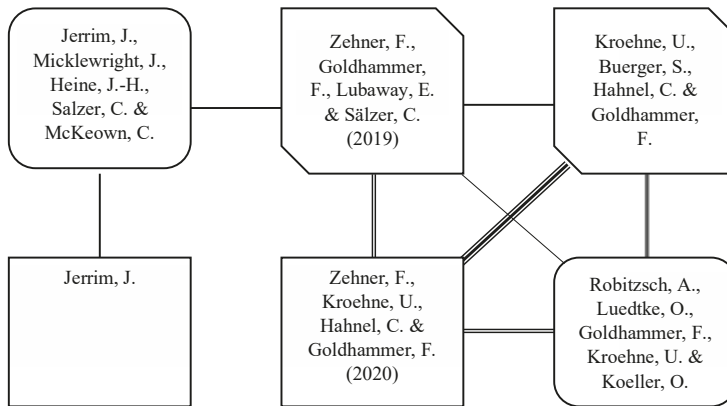
különbözőségét vizsgálják. Az egyes források módszertani jellemzőit és fő eredményét a 6. táblázat tartalmazza. A mérési adatok forrása négy esetben kizárólag Németország. Az érintett mérésekben megjelennek 2012. és 2015. évi főmérések, a 2015-ös mérés próbamérése, valamint egy, a 2012. évi főméréshez kapcsolódó német kiegészítő vizsgálat. Ennek során a 2009-es ciklus 30 papír-ceruza alapú szövegértés itemét számítógépre is adaptálták, és egy almintán felvették a mérés másnapján. A vizsgálatokban használt elemzési módszerek magas minőségűek, a mintaelemszámok szintén megfelelőek.

6. táblázat. A nemzetközi adatbázisban történt keresés eredményeként kapott kutatások módszertani jellemzői és fő eredménye.

A jelzések és rövidítések: \*: A PISA 2012 méréshez kapcsolódó német médiahatás-vizsgálat, PISA 2009-es itemek papír-ceruza és számítógépes változatával; LPM: lineáris valószínűségi (linear probability) modell regresszió; OLS: lineáris regresszió legkisebb négyzetek módszerével (ordinary least squares); SEM: strukturális egyenletek módszere (structural equation modelling); IRT: modern tesztelméleti modell (item response theory); GLMM: általánosított lineáris vegyes modell (general linear mixed-model) regresszió

Forrás	Mérés	Terület	Mintanagyság	Módszer	Eredmény
Jerrim, 2016	2012 matematika és szg. mat.	32 gazdasági egység	~200000 fő	Ország szintű korreláció, LPM	Eltérés az országok harmadánál (10-20 pont)
Jerrim és mtsai., 2018	2015 próbamérés	Németország, Svédország, Írország	3438 fő	OLS regresszió	Kis mértékű, területenként különböző eltérés (10-20 pont)
Kroehne és mtsai., 2019	2012* szövegértés	Németország	856 fő	SEM	Egyező konstruktum, Kis, egyetlen médiahatás
Robitzsch és mtsai., 2020	2015 próbamérés	Németország	1023 fő	IRT, Jackknife szimuláció	Kis mértékű, területenként eltérő médiahatás (10-20 pont)
Zehner és mtsai., 2019	2012, 2015 szövegértés	Németország	43396 válasz	GLMM	Hosszabb és több információt tartalmazó válaszok
Zehner és mtsai., 2020	2012* szövegértés	Németország	7495 válasz	GLMM	Hasonló eredmény, kisebb különbség

A források közötti kapcsolatot a közös szerzőkkel jelezve (3. ábra) kirajzolódik a PISA méréshez kapcsolódó források mögötti szakmai együttműködés hálózata, melynek magját Goldhammer és Kroehne adják 4, illetve 3 szerzőséggel.



3. ábra A PISA méréshez kapcsolódó médiahatás-vizsgálatok kapcsolatai a közös szerzők alapján. A kapcsolati vonalak száma a közös szerzők számára, az alakzatok formája a közös adatforrásra utal.  
(Forrás: saját ábra)

A PISA méréshez kapcsolódó források mindegyike hasonló megállapításokra jut. A papír-ceruza és számítógép alapon mért konstruktumok nem különböznek egymástól, legalábbis nem olyan mértékben, ami veszélyeztetné a mérés folytonosságát. Ezt egyrészt az itemparaméterek, másrészt a kétféle adatfelvételben elért teljesítménypontok magas korrelációja alapján állítják, amit Kroehne és munkatársai (2019) strukturális egyenletek módszerével is ellenőriz. Az itemeknek kis része érintett a médiahatás miatti eltérő paraméterekkel. A teljesítmény szerinti eltérés szignifikáns, a két tesztfelvételi mód között jellemzően 10 és 20 pont között mozog, és a próbamérésen végzett elemzés alapján nem egyformán érinti az egyes mérési területeket. Ez az eltérés az évek közötti összekötés hibája 4–6-szorosának (OECD, 2014a, 281.), az iskolai évfolyamok közötti különbség (41 pont) 25–50 százalékának, illetve a Magyarország 2012-es matematika eredménye (477 pont) és az OECD átlag (494 pont) közötti szignifikáns különbségnek feleltethető meg (OECD, 2014a, 46-47.). Ezek az eredmények megfelelnek a mérés dokumentumaiban szereplő információknak, ugyanakkor alátámasztják, hogy az egyes országokban különböző mértékű médiahatással lehet számítani, amit a trendek nem vesznek figyelembe.

A nyílt válaszok elemzése (Zehner és mtsai., 2019, 2020) nem egyformán kapcsolódhatnak a médiahatáshoz. A 2019-es publikáció a PISA 2012 és 2015 közös szövegértés itemeire adott szöveges válaszokat hasonlította össze, így az eredmény a populációkból származó különbséget is hordozhatott. A cikk ennek megfelelően médiahatás helyett ciklushatásnak nevezi a jelenséget. A 2020-as publikáció ugyanezzel a kutatási kérdéssel foglalkozik, kifejezetten a médiahatásra fókuszálva a 2012-es kiegészítő mérés nyílt válaszain, és a korábbihoz hasonló, bár kevésbé markáns eredményt kaptak. A nyílt válaszok értékelésének szempontjai a megjelenő elemek száma és az információtartalom voltak. A vizsgálatok a számítógépes adatfelvétel válaszaiban több válasz-elemet és valamivel több információt találtak.



### PIRLS

A nemzetközi adatbázisokban és a mérési dokumentumok között nem találtunk a méréshez szorosan kapcsolódó, médiahatás vizsgálatára irányuló kutatást vagy leírást. A lehetséges publikációk jellemzően a lineáris és nemlineáris (internetes) szövegértés tanulói eredményeit vizsgálják különböző háttértényezők (például nem vagy IKT használat) mentén. Ezeket a kutatásokat a mérések keretbeli különbségei alapján nem tekintettük médiahatás vizsgálatnak és kizártuk a további elemzésből.

### TIMSS

A TIMSS 2019 mérés hazai jelentése a bevezető fejezetben tájékoztat az elektronikus adatfelvételtől (Palincsar és mtsai., 2020). Ez alapján a papír-ceruza és elektronikus tesztek tartalmában és felépítésben egyeztek (nagyra részben a papír-ceruza TIMSS 2015 itemeit vették alapul), valamint a nemzetközi eredmények és a trend összehasonlíthatósága érdekében az elektronikus mellett hozzávetőlegesen feleakkora mintán papír-ceruza adatfelvétel is történt. A trendeket két lépésben kapcsolták össze: először a hagyományos adatfelvételt a 2015. évi méréshez, majd az elektronikus adatfelvételt a 2019. évi papír-ceruza mérés eredményeihez. Ez az információ minden tekintetben egyezik a mérés eredményeinek nemzetközi dokumentumával (Mullis és mtsai., 2020), amely szintén nem ad további technikai részleteket. Ugyanitt az eredményeket közlő táblázatokban külön-külön szerepelnek az eTIMSS és a paperTIMSS országok eredményei.

A mérési keret dokumentuma (Mullis & Martin, 2017) tájékoztat az egyes mérések tesztfüzeteinek összeállításáról. Ez alapján a méréshez 14 (az innovatív, például húz-és-vidd – *drag-n-drop* – itemektől eltekintve) teljesen egyező tesztfüzetet alakítottak ki, azaz a TIMSS 2019 mérést úgy tervezték meg, hogy 1) a mérőeszközök csak a közvetítő médiumban különböznek, ezáltal alkalmasak a médiahatás mérésére, és 2) a trendek és a két mérési mód megbízható összehasonlíthatósága érdekében ezt a médiahatást figyelembe is vették az eredmények közlésekor.

A médiahatás számítását és a skálák összekötésének módját a technikai leírás (Martin és mtsai., 2020) ismerteti. A dokumentum szerint a tartalmi validitás és a médiahatás vizsgálatára 2017-ben, a próbamérést megelőző évben került sor az eTIMSS Item Equivalence Study keretében. A vizsgálat további célja volt az itemek elektronikus változatának operatív tesztelése és ezzel a próbamérés előkészítése. Az eredményeket tudományos cikk formájában jelentették meg (Fishbein és mtsai., 2018), ami éppen a nemzetközi adatbázisban történt keresés egyik találat. A technikai leírás ezért nem is közöl részletes eredményeket, hanem a cikkekre hivatkozik.

Fishbein és munkatársai (2018) vizsgálata 25 országban, két mérési területen, két évfolyamon zajlott 26 000 tanuló bevonásával. Magyarország nem vett részt a vizsgálatban. A mérés konstrukcióját (matematika, illetve természettudomány) egyezőnek találták a két mérési módban. A híd-itemek 80 százalékát invariánsnak, azaz a két mérési módon megegyező mérési tulajdonságúnak találták. Az eltérő itemek jellemzően valamilyen technikai probléma vagy innovatív eszköz miatt lehettek különbözőek. Az invariáns itemek százalékos megoldottságai összességében kismértékű médiahatást mutattak, a papír-ceruza mérést könnyebbnek jelezve. Meredekségben és a hiányzó/el nem ért válaszokban nem volt különbség. A két mérési mód alapján számított teljesítményben matematikából 14, természettudományból 7 pontnyi átlagos különbség rajzolódott ki, ami szignifikáns és összemérhető az elfogadható mértékű mérési hibával (annak 1–2-szerese), vagy az évfolyamok közötti különbséggel (ami az általános iskolai évfolyamok közötti különbség negyede), vagy a Magyarország 2019-es matematika eredménye (523 pont) és a TIMSS

középérték (500 pont) különbségének felével. Ez indokolja a főmérés során az elektronikus adatfelvételt kiegészítő papír-ceruza adatfelvételt és a médiahatás figyelembevételét a skálák igazításakor. Háttérjellemzők szerinti különbséget nem találtak.

A TIMSS méréshez kapcsolódó másik találat Hamhuis és munkatársainak (2020) az Item Equivalence Study-ra irányuló kritikai vizsgálata. Az eredeti adatfelvétel holland adatainak másodelemzését az motiválta, hogy a holland általános iskolákban elterjedt a tablethasználat, ami a számítógép mellett az eTIMSS másik elfogadott elektronikus adatfelvételi médiuma. A mintát 25 iskolából 532 tanuló alkotta, mindegyikük 4. évfolyamos. A két módban elért eredmények között nem találtak szignifikáns különbséget ( $p > 0,05$ ), ami felveti a holland TIMSS trend felülbecslését a 2019-es felmérésben. Ugyanakkor az eredeti tanulmánytól eltérően kis különbséget találtak a nemek között.

A 2019-es főmérés adatain végezett, médiahatás-vizsgálattal kapcsolatos elemzéseket a technikai leírás (Martin és mtsai., 2020) tartalmaz. A 10–12. fejezetek részletesen bemutatják az itemek ellenőrzésének módját és a skálázási eljárást, amit a kiegészítő papír-ceruza adatfelvétel tesz lehetővé. Ennek során az eTIMSS méréshez képest feleakkora mintán a TIMSS 2015 trend-itemeinek papír-ceruza alapú kitöltése is megtörtént. A trend-itemek több, mint 80 százaléka bizonyult invariánsnak (Martin és mtsai., 2020, 12.54), ezek adták a skálázás alapját. A 13. fejezet az egyes országokra számított médiahatást és annak nagyságát vizsgálja. Az invariáns itemeken számított százalékos eredmények alapján kis számú országban található szignifikáns ( $p < 0,05$ ) eltérés, ami nem jelenik meg mindkét mérési területen. Magyarország esetében 3 és 7 pont közötti az eltérés a papír-ceruza és a számítógépes adatfelvétel eredménye között, ami egyik területen vagy évfolyamon sem jelent szignifikáns eltérést.

## Összegzés

Magyarország jelenleg három nemzetközi tanulói teljesítménymérésen vesz részt. A PISA, TIMSS és PIRLS mérések különböző korosztályokat mérnek szövegértés, matematika és természettudomány területeken. Mindhárom mérés az utóbbi mérési ciklus(ok) során papír-ceruza mérési módról számítógépes adatfelvételre váltott, amit több tényező indokolt. Az érvek között szerepel a 21. századi képességek bevonása a mérésbe (OECD, 2017d), az innovatív, a terület tartalmi elemeit jobban mérő itemek lehetősége (Mullis & Martin, 2017) vagy a különböző szintű mérések kombinálásának lehetősége (Mullis & Martin, 2019). A felmérés módjának váltása miatt, a trendek további számíthatóságának és összehasonlíthatóságának érdekében, szükséges a különböző mérési ciklusokban alkalmazott mérőeszközök azonossága, vagy legalábbis a különbségekből származó lehetséges eltérések pontos kontrollálása.

A fentiekhez módszertanában hasonló, három évfolyamon teljes körű hazai mérés, az OKM a 2022. évtől szintén elektronikus formában valósult meg (Oktatási Hivatal, 2021). Az egymást követő évek eredményének összehasonlítása csak akkor lehetséges, ha a mérési mód cseréje esetén is azonos skálára kerülnek az egyes eredmények. A mérés médiumának cseréje kapcsán kutatásunk céljaként tűztük ki a nemzetközi tanulói teljesítménymérések hazai és nemzetközi dokumentumainak áttekintését, valamint a papír-ceruza tesztelésről a számítógép-alapú tesztelésre való áttérés következményeivel foglalkozó elemzések, azaz a médiahatás-vizsgálatok összegyűjtését és elemzését.

A mérési dokumentumok alapján a digitális olvasás mérési területeket (PISA 2009 és 2012, ePIRLS 2016 és 2021) nem tekinthetjük a papír-ceruza szövegértés tesztek számítógépes változatának, mivel ezekben eltérő típusú vagy célú szövegek feldolgozását követelik a tanulóktól (Mullis & Martin, 2015; OECD, 2009). A számítógépes matematikai műveltség mérése (PISA 2012) a papír-ceruza mérés tartalmi keretében került

értelmezésre, ezért amennyiben a kutatás célja és módszertana megfelelő, megszorításokkal alkalmas lehet a médiahatás vizsgálatára.

A médiahatás-vizsgálat a PISA mérés esetében a 2015. évi mérés próbamérésének keretében (OECD, 2016a, 2017a), a TIMSS 2019 esetében pedig önálló vizsgálatról történt (Fishbein és mtsai., 2018). Utóbbi bemutatása lektorált tudományos folyóiratban cikként jelent meg. A két vizsgálat közös eleme, hogy a teljes mérésre koncentrálnak, fókuszában a 1) két mérés konstruktum-azonosságának vizsgálata, 2) az itemek szintjén történő médiahatás-vizsgálat, és 3) a teljesítmények szintjén történő médiahatás-vizsgálat, egyszersmind a trendek folytonosságának biztosítása áll. Mindkét vizsgálat megerősíti, hogy a két tesztelési módban mért konstruktum azonosnak tekinthető, csak az itemek kis részében adódik szignifikáns különbség az itemparaméterekben. Míg a PISA mérés a két módban nem invariáns itemek esetében mindkét irányú médiahatásról beszámol, addig a TIMSS vizsgálat jellemzően a számítógépes adatfelvételt találta nehezebbnek. Az itemek meredekségében nem mutattak ki médiahatást, azaz az itemek viselkedése azonos a két módban. Mindkét mérés esetében szükségesnek találták valamilyen korrekció alkalmazását, azonban az itemek kis részénél határoztak meg országspecifikus item paramétereket. A médiahatás összességében 10–20 pontnyi eltérést mutat az egyes területeken (és évfolyamokon), ami nyomán a trendek számításában érvényesítik a médiahatás korrekcióját. Ekkora különbség körülbelül 6–8 helyezésnek felel meg a mérések rangsoraiban (a középértékek közelében). A PISA esetében a próbamérés eredménye alapján alkalmazott nehézség paraméterrel, a TIMSS esetében a méréssel együtt felvett papír-ceruza kiegészítő méréssel számították ki a szükséges korrekciókat.

A mérések hazai szervezője dokumentumaiban tájékoztat a módszertani változásokról, és felhívja a figyelmet az ebből származó bizonytalanságra, önálló vizsgálatot lektorált folyóiratban nem publikált. A hazai tudományos irodalom áttekintése során számos publikációt találtunk mind a mérésekkel (például az *Educatio* 2015/2 tematikus száma), mind a médiahatás vizsgálatával kapcsolatban (például az *eDia*-hoz kapcsolódóan Herczegné Goldschmidt, 2016; R. Tóth & Hódi, 2011), vagy általában a médiahatás részeként a konstruktum-validitásról (például Hülber, 2012). Olyan forrást, mely a médiahatást a három nemzetközi mérés kontextusában vizsgálja – a fent említett jelentéseken kívül – nem találtunk.

A nemzetközi adatbázisokban végzett szisztematikus keresés során nyolc publikációt találtunk, melyek a médiahatást valamely mérés adatbázisai vagy feladatai segítségével vizsgálja. A PIRLS méréshez nem kapcsolódik elemzés, aminek elsődleges oka lehet, hogy az első elektronikus megvalósítás 2021-ben zajlott. A TIMSS esetében a korábban említett vizsgálaton kívül egy, a PISA-hoz kapcsolódóan hat kutatást találtunk. Ezek jellemzően abból a szempontból vizsgálják a kérdést, hogy a mérésekben alkalmazott egy-egy skála-korrekció megfelel-e az egyes országok esetében. Az eredmények, akárcsak a médiahatás vizsgálatának általános eredményei, eltérő képet mutatnak. Úgy tűnik, hogy a médiahatás bizonyos országoknál negatív, másoknál pozitív irányú lehet a papír-ceruza mérés eredményeihez képest (Jerrim, 2016), de az is előfordul, hogy nincs kimutatható médiahatás (Hamhuis és mtsai., 2020). Ennek következménye lehet, hogy az egyes országok trendje alá- vagy felülbecsli a valóságos eredményt, bár ez a torzítás jellemzően nem jelentős. A médiahatás Zehner és munkatársai (2019, 2020) kutatása alapján kimutatható a számítógépes adatfelvétel szöveges válaszainak nagyobb elem-gazdagságában és információtartalmában.

A feldolgozott mérési dokumentumok és cikkek alapján úgy véljük, hogy az OKM esetében a mérés médiumának cseréjét előkészítő vizsgálatok (Molnár és mtsai., 2015) után is szükség lehet a médiahatás empirikus feltérképezésére, alkalmasint annak figyelembevételére a 2022-es eredmények kiszámításában. Ehhez leginkább az elektronikus adatfelvétellelhez illesztett papír-ceruza kiegészítő teszt a legalkalmasabb. A nemzetközi

mérések eredményei alapján megfontolandó mérési területenként és évfolyamonként külön-külön végezni a vizsgálatot és a korrekciót. Várhatóan az itemek viselkedése nem, csak nehézsége különbözik a két mérési módban, és a papír-ceruza teszttel megegyező megjelenésű számítógépes itemek jelentős része invariáns lesz. Az OKM esetében is javasoljuk a trendek folytatólagos eredményeinek vagy a tanulók különböző mérési módban felvett adatainak óvatos értelmezését, ahogy – elsősorban a PISA esetében – a nemzetközi mérések hazai szervezője és a módszertani változásokat tárgyaló publikációk jelzik.

### Korlátok és kitekintés

A kutatás korlátai többfélék, melyek egyrészt a szisztematikus áttekintéssel kapcsolatosak, másrészt a fellelt szakirodalmak limitációiból származnak. A hazai áttekintés korlátja, hogy nincs kifejezetten erre a célra kialakított adatbázis, ezért lehetséges, hogy a keresés során kimaradtak olyan publikációk, melyek címében nem szerepel egyik mérés mozaikszava sem, vagy nem szerepel a MATARKA adatbázisban. Ez utóbbi hiányosságot az Arcanum adatbázisában folytatott szövegszerű kereséssel küszöböltük ki.

A nemzetközi adatbázisokban folytatott keresés során olyan adatbázisokban folytattunk keresést, melyekhez intézményi hozzáférésünk volt. Ezek mindegyike elfogadott, megbízható forrás a neveléstudomány területén, de nem teljes körű. További korlát, hogy kizárólag angol vagy magyar nyelvű forrásokat fogadtunk el, ami bizonyos régiókban folytatott vagy publikált kutatások kizárását jelenthette, ugyanakkor a teljes szövegek áttekintése során talákoztunk távol-keleti, kelet-európai, dél-afrikai és dél-amerikai adatokon végzett vagy ottani szerzőségű publikációkkal. A befoglalási és kizárási kritériumok alapján az úgynevezett szürke irodalomba (*grey literature*) tartozó források (Dobó, 2000) nem kerültek bele a válogatásba. Ennek részben kutatási kapacitással kapcsolatos okai vannak, másrészt igyekeztünk magas minőségű, lektorált, tudományos folyóiratban megjelent forrásokat felkutatni.

Az elemzésbe bevont források a PISA esetében jellemzően németországi adatok feldolgozását jelentette, a mérési dokumentumokon felül vizsgált nyolc tanulmány majd mindegyike európai kontextusban vizsgálódik. A távol-keleti és dél-amerikai szempontok egy publikációban jelentek meg (Jerrim, 2016). A különböző mérések adatain folytatott összehasonlítások esetében nem zárható ki a minták populációból adódó különbsége, ami további bizonytalanságot eredményez a médiahatás meghatározásában. További probléma a PISA mérés esetében, hogy a számítógépes mérési mód bevezetésével egy időben más módszertani változások is történtek. Egyrészt az itemparaméterek paraméterezése (egyparaméteres helyett kétparaméteres modell illesztése), másrészt a skálák kialakítása (az itemparaméterek kialakításakor minden korábbi mérési ciklust figyelembe vettek), harmadrészt az el nem ért itemek kezelése (helytelen válasz helyett az el nem ért itemek nem számítottak bele a pontszámításba) is megváltozott. Ezek egy része szintén vizsgálat tárgyát képezte (Robitzsch és mtsai., 2020), azonban hatással lehetnek a médiahatás meghatározására is.

**T. Kárász Judit**

*ELTE Eötvös Loránd Tudományegyetem Neveléstudományi Doktori Iskola  
ELTE Eötvös Loránd Tudományegyetem Neveléstudományi Intézet  
Károli Gáspár Református Egyetem, Bölcsészettudományi Kar, Pszichológiai Intézet*

**Szék Krisztián**

*ELTE Eötvös Loránd Tudományegyetem Neveléstudományi Intézet*

## Köszönetnyilvánítás, támogatás

Köszönöm Takács Szabolcsnak a másodkódolásban nyújtott segítséget.

„Az Innovációs és Technológiai Minisztérium ÚNKP-21-3 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.”



## Irodalom

- Auxné Bánfi I., Balázi I., Balkányi P., Balogh V. K., Gyapay J., Lak Á. R., Ostorics L. I., Palincsar I., Rábainé Szabó A., Rózsa C., Szabó L. D., Szepesi I., Szipőcsné Krolopp J., & Vadász C. (2014). *Országos kompetenciamérés—Technikai leírás*. Jelentés. Budapest: Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktat/meresek/orszmer2012/OKM\\_Technikaileiras.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktat/meresek/orszmer2012/OKM_Technikaileiras.pdf)
- Balázi I., Balkányi P., Ostorics L., Palincsar I., Rábainé Szabó A., Szepesi I., Szipőcsné Krolopp J., & Vadász C. (2014). *Az Országos kompetenciamérés tartalmi keretei—Szövegértés, matematika, háttérkérdőívek*. Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktat/meresek/orszmer2014/AzOKMtartalmikeretei.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktat/meresek/orszmer2014/AzOKMtartalmikeretei.pdf)
- Balázi I., Balkányi P., & Vadász, C. (2017). *PIRLS 2016 Összefoglaló jelentés a 4. Évfolyamos tanulók eredményeiről* (0 kiad.). Oktatási Hivatal; MTMT. <https://m2.mtmt.hu/api/publication/3340776>
- Balázi I., & Ostorics, L. (2011). *PISA2009 Digitális szövegértés. Olvasás a világhálón* (0 kiad.). Oktatási Hivatal; MTMT. <https://m2.mtmt.hu/api/publication/2411815>
- Balázi I., Ostorics L., Szalay B., Szepesi I., & Vadász C. (2013). *PISA 2012 Összefoglaló jelentés* (o. 80). Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktat/nemzetkozi\\_meresek/pisa/pisa2012\\_osszefoglalo\\_jelentes.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktat/nemzetkozi_meresek/pisa/pisa2012_osszefoglalo_jelentes.pdf)
- Csíkos C., & Vidákovich T. (2012). A matematika-tudás alakulása az empirikus vizsgálatok tükrében. In Csapó B. (Szerk.), *Mérlegen a magyar iskola* (o. 83–130). Nemzeti Tankönyvkiadó.
- Dobó, K. (2000). Az elektronikus szürke irodalom új formái, témái és felhasználása—2000. 46. Évf. 4. Szám—Elektronikus Periodika Archívum. *Könyvtári Figyelő*, 46(4), 581–585.
- Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 Item Equivalence Study: Examining Mode Effects for Computer-Based Assessment and Implications for Measuring Trends. *Large-Scale Assessments in Education*, 6. DOI: [10.1186/s40536-018-0064-z](https://doi.org/10.1186/s40536-018-0064-z)
- Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS’ paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology*, 51(6), 2340–2358. DOI: [10.1111/bjet.12914](https://doi.org/10.1111/bjet.12914)
- Herczegné Goldschmidt Z. (2016). Papíralapú és számítógép-alapú tesztelés összehasonlító vizsgálata olvasás-szövegértés területén 4. Évfolyamos diákok körében. *Iskolakultúra*, 26(6), Art. 6. DOI: [10.17543/iskkult.2016.6.30](https://doi.org/10.17543/iskkult.2016.6.30)
- Hülber L. (2012). A papír- és a számítógép alapú tesztelés összehasonlító vizsgálata különböző item paraméterek mentén. *Iskolakultúra*, 22(12), Art. 12.
- Hülber L., & Molnár G. (2013). Papír és számítógép alapú tesztelés nagymintás összehasonlító vizsgálata matematika területén, 1-6. Évfolyamon. *Magyar Pedagógia*, 113(4), 243–263.
- Jerrim, J. (2016). PISA 2012: How do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy & Practice*, 23(4), 495–518. DOI: [10.1080/0969594x.2016.1147420](https://doi.org/10.1080/0969594x.2016.1147420)
- Jerrim, J., Micklewright, J., Heine, J.-H., Salzer, C., & McKeown, C. (2018). PISA 2015: How big is the ‘mode effect’ and what has been done about it? *Oxford Review of Education*, 44(4), 476–493. DOI: [10.1080/03054985.2018.1430025](https://doi.org/10.1080/03054985.2018.1430025)
- Kroehne, U., Buerger, S., Hahnel, C., & Goldhammer, F. (2019). Construct Equivalence of PISA Reading Comprehension Measured With Paper-Based and Computer-Based Assessments. *Educational Measurement: Issues & Practice*, 38(3), 97–111. DOI: [10.1111/emip.12280](https://doi.org/10.1111/emip.12280)
- Lak Á. R. (2020). A 2012. És 2015. Évi magyar PISA-populációk összehasonlítása az Országos kompetenciamérés segítségével. *Köznevelési Elemzési Jelentések*, 3(1), 1–5.

- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. DOI: [10.2307/2529310](https://doi.org/10.2307/2529310)
- Martin, M. O., Mullis, I. V. S., & von Davier, M. (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/methods/pdf/TIMSS-2019-MP-Technical-Report.pdf>
- Molnár, G., Magyar, A., Pásztor-Kovács, A., & Hülber, L. (2015). *A mérési-értékelési rendszer elektronikus alapokra helyezésével kapcsolatos helyzetfelmérés* (o. 97). Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/unios\\_projektek/tamop318/OKM\\_kutatasi\\_eredmenyek2015/meresi\\_ertekelesi\\_rendszer.pdf](https://www.oktatas.hu/pub_bin/dload/unios_projektek/tamop318/OKM_kutatasi_eredmenyek2015/meresi_ertekelesi_rendszer.pdf)
- Mullis, I. V. S., & Martin, M. O. (Szerk.). (2015). *PIRLS 2016 Assessment Framework* (Second Edition). TIMSS & PIRLS, International Study Center, Lynch School of Education, Boston College. [https://timssandpirls.bc.edu/pirls2016/downloads/P16\\_Framework\\_2ndEd.pdf](https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf)
- Mullis, I. V. S., & Martin, M. O. (Szerk.). (2017). *TIMSS 2019 Assessment Frameworks*. TIMSS & PIRLS. <https://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks.pdf>
- Mullis, I. V. S., & Martin, M. O. (Szerk.). (2019). *PIRLS 2021 Assessment Frameworks*. TIMSS & PIRLS; Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/pirls2021/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (Szerk.). (2017). *PIRLS 2016 International Results in Reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. <http://timssandpirls.bc.edu/pirls2016/international-results/wp-content/uploads/structure/CompletePDF/P16-PIRLS-International-Results-in-Reading.pdf>
- Mullis, I. V. S., Martin, M. O., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. <https://timss2019.org/reports/wp-content/themes/timssandpirls/download-center/TIMSS-2019-International-Results-in-Mathematics-and-Science.pdf>
- Mullis, I. V. S., Martin, M. O., & von Davier, M. (Szerk.). (2021). *TIMSS 2023 Assessment Frameworks*. TIMSS & PIRLS International Study Center. [https://timssandpirls.bc.edu/timss2023/frameworks/pdf/T23\\_Frameworks.pdf](https://timssandpirls.bc.edu/timss2023/frameworks/pdf/T23_Frameworks.pdf)
- OECD. (2009). *PISA 2009 Assessment Framework—Key Competencies in Reading, Mathematics and Science*. <https://www.oecd.org/pisa/pisaproducts/44455820.pdf> DOI: [10.1787/9789264062658-en](https://doi.org/10.1787/9789264062658-en)
- OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD. DOI: [10.1787/9789264190511-en](https://doi.org/10.1787/9789264190511-en)
- OECD. (2014a). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)* (Köt. 1). OECD Publishing. DOI: [10.1787/9789264208780-en](https://doi.org/10.1787/9789264208780-en)
- OECD. (2014b). *PISA 2012 Technical Report* (<https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>). OECD Publishing. <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- OECD. (2016a). Annex A6 The PISA 2015 field trial mode-effect study. In *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD.
- OECD. (2016b). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD. DOI: [10.1787/9789264266490-en](https://doi.org/10.1787/9789264266490-en)
- OECD. (2017a). *PISA 2015 Technical Report* ([https://www.oecd.org/pisa/data/2015-technical-report/PISA2015\\_TechRep\\_Final.pdf](https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf)). OECD Publishing. [https://www.oecd.org/pisa/data/2015-technical-report/PISA2015\\_TechRep\\_Final.pdf](https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf)
- OECD. (2017b). *PISA 2015 Technical report*. OECD. <https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>
- OECD. (2017c). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. OECD.
- OECD. (2017d). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving, revised Edition*. OECD Publishing. DOI: [10.1787/9789264281820-en](https://doi.org/10.1787/9789264281820-en)
- OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. DOI: [10.1787/b25efab8-en](https://doi.org/10.1787/b25efab8-en)
- OECD. (2019b). *PISA 2018 Results (Volume I): What Students Know and Can Do*. OECD Publishing. DOI: [10.1787/5f07c754-en](https://doi.org/10.1787/5f07c754-en)
- Oktatási Hivatal. (2019). *PISA 2018 Összefoglaló jelentés* (o. 99). Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktatas/nemzetkozi\\_meresek/pisa/PISA2018\\_v6.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/pisa/PISA2018_v6.pdf)
- Oktatási Hivatal. (2021, 0 15). *A digitális országos mérések általános leírása*. [https://www.oktatas.hu/kozneveles/meresek/digitalis\\_orzsagos\\_meresek/altalanos\\_leiras](https://www.oktatas.hu/kozneveles/meresek/digitalis_orzsagos_meresek/altalanos_leiras)
- Ostörics L., Szalay B., Szepesi I., & Vadász C. (2016). *PISA 2015 Összefoglaló jelentés* (o. 90). Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktatas/nemzetkozi\\_meresek/pisa/PISA2015\\_osszefoglo\\_jelentes.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/pisa/PISA2015_osszefoglo_jelentes.pdf)
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical*

*Epidemiology*, 134, 103–112. DOI: [10.1016/j.jclinepi.2021.02.003](https://doi.org/10.1016/j.jclinepi.2021.02.003)

Palincsár I., Szalay B., Szepesi I., Ostorics L., & Vadász C. (2020). *TIMSS 2019 Összefoglaló jelentés* (o. 184). Oktatási Hivatal. [https://www.oktatas.hu/pub\\_bin/dload/kozoktat/nemzetkozi\\_meresek/timss/TIMSS2019.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktat/nemzetkozi_meresek/timss/TIMSS2019.pdf)

Pásztor-Kovács A., Magyar A., Hülber L., Pásztor A., & Tongori Á. (2013). Áttérés online tesztlésre – a mérés-értékelés új dimenziói. *Iskolakultúra*, 23(11), 86–100.

R. Tóth, K., & Hódi, Á. (2011). Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasás-szövegértés területén. *Magyar Pedagógia*, 111(4), 313–332.

Robitzsch, A., Luedtke, O., Goldhammer, F., Kroehne, U., & Koeller, O. (2020). Reanalysis of the German PISA Data: A Comparison of Different Approaches for Trend Estimation With a Particular

Emphasis on Mode Effects. In *Frontiers in Psychology* (Köt. 11). FRONTIERS MEDIA SA. DOI: [10.3389/fpsyg.2020.00884](https://doi.org/10.3389/fpsyg.2020.00884)

Rother, E. T. (2007). Systematic literature review X narrative review. *Acta Paulista de Enfermagem*, 20, v–vi. DOI: [10.1590/s0103-21002007000200001](https://doi.org/10.1590/s0103-21002007000200001)

Velkey, K. (2018). A 2015-ös lengyel PISA-eredmények és ami mögöttük van. *Educatio*, 27(2), 332–340. DOI: [10.1556/2063.27.2018.2.14](https://doi.org/10.1556/2063.27.2018.2.14)

Zehner, F., Goldhammer, F., Lubaway, E., & Sälzer, C. (2019). Unattended consequences: How text responses alter alongside PISA's mode change from 2012 to 2015. *Education Inquiry*, 10(1), 34–55. DOI: [10.1080/20004508.2018.1518080](https://doi.org/10.1080/20004508.2018.1518080)

Zehner, F., Kroehne, U., Hahnel, C., & Goldhammer, F. (2020). PISA reading: Mode effects unveiled in short text responses. *Psychological Test and Assessment Modeling*, 62(1), 85–105. Publicly Available Content Database.

### Absztrakt

Az Országos kompetenciamérés (OKM) a Magyarországon működő legnagyobb volumenű tanulói teljesítménymérés. A mérés húsz éve méri a közoktatásban tanulókat matematikai eszköztudás és szövegértési képesség területen korábban két (6. és 8.), később három (6., 8. és 10.) évfolyamon. Az OKM első alkalommal 2022-ben – követve a nemzetközi tanulói teljesítménymérések trendjét – valósult meg elektronikus adatfelvételi móddal. A PISA, TIMSS és PIRLS mérések, melyekben Magyarország jellemzően indulás óta részt vesz, a legutóbbi mérési ciklusok során a papír-ceruza tesztek elektronikusra váltották. A fenti négy mérés mindegyikének célja országok vagy évfolyamok összehasonlítása, trendek bemutatása, melyek folytonosságának biztosítása szükségessé teszi a papír-ceruza és a számítógép-alapú mérési módok lehetséges különbségének, azaz a médiahatásnak (*mode effect*) a vizsgálatát. Mivel az OKM jellemzően a nemzetközi tanulói teljesítménymérések módszertanát követi, feltételeztük, hogy a nemzetközi mérésekhez kapcsolódó médiahatás-vizsgálatok – akár a mérések saját dokumentumai, akár független elemzések – jó információforrások lehetnek a hazai mérés esetében. Kutatásunkban arra kerestük a választ, hogy a PISA, TIMSS és PIRLS mérésekhez kapcsolódóan a hazai térben, a nemzetközi adatbázisokban és a mérések saját dokumentumaiban milyen tudományosan megalapozott eredmények érhetőek el a papír-ceruza és a számítógép-alapú tesztelés közötti különbségekről. Vizsgálatunkat szisztematikus szakirodomelemzés segítségével végeztük el. A mérések dokumentációja jellemzően 10–20 pontos médiahatást mutatott ki, amit a trendek folytonossága érdekében – néhány ország kivételével – egységes módon vettek figyelembe a trendek kialakításakor. A kritikai vizsgálatok ugyanakkor rámutatnak, hogy a médiahatás eltérő lehet az egyes országokban, ami a trendből levont hamis következtetéshez vezethet.

**Kulcsszavak:** tanulói teljesítménymérés, médiahatás, szisztematikus szakirodalmi áttekintés