

SZÁMÍTÓGÉPES ÉS PAPÍR-CERUZA TESZTEREDMÉNYEK ÖSSZEHASONLÍTÓ VIZSGÁLATA AZ OLVASÁS- SZÖVEGÉRTÉS TERÜLETÉN

R. Tóth Krisztina és Hódi Ágnes

Szegedi Tudományegyetem Oktatásméleti Kutatócsoport

A modern technika vívmányainak széles körű elterjedése a nagymintás, tanulói teljesítménymérést szolgáló vizsgálatok módszertanában – célcsoporttól és módszertani kerektől függetlenül – is új lehetőségeket nyitott. A feladatok számítógépes környezetbe történő átültetését célozta meg mind az OECD PISA Computer-based Assessment of Science (CBAS – természettudományi eszköztudás számítógépes mérése, OECD, 2010a), mind az Electronic Reading Assessment (ERA – digitális szövegértés mérése, I. OECD, 2011 vizsgálata. E pilotvizsgálatokban elektronikus környezetben mérték a tanulók természettudományos ismereteinek, illetve olvasási képességének alkalmazhatóságát. Szintén erre tesz kísérletet az IEA PIRLS (Progress in International Reading Literacy Study – Nemzetközi Szövegértés-vizsgálat) felmérésorozat web alapú olvasás (Web-based Reading) című pilotmérése (I. *Mullis, Martin, Kennedy, Trong és Sainsbury, 2009*) is. A nemzetközi tendenciákkal (I. *Molnár, 2010*) párhuzamosan hazánkban az SZTE Oktatásméleti Kutatócsoportja is megkezdte a számítógépes tesztelésre történő áttéréshez szükséges pilotvizsgálatokat (pl. *Csapó, Molnár és R. Tóth, 2009; Csapó, Molnár, Pap-Szigeti és R. Tóth, 2009; Hülber, 2012; Molnár és R. Tóth, 2008; R. Tóth és Hódi, 2011*)

A mérés-értékelés folyamatának technológia alapú adaptációját ösztönzi, hogy az új tesztmedium a papír-ceruza teszteléssel szemben idő- és költséghatékonyabb tesztelési és visszajelentési folyamatot biztosít, valamint olyan új itemtípusok (pl. multimédiás, interaktív itemek) és tesztalgoritmusok implementációját teszi lehetővé, amelyek nem alkalmazhatók a tradicionális tesztelés során. Továbbá az új médium és az internet tesztelési folyamatba történő integrálása lehetővé teszi, hogy új konstruktumokat mérjünk. (A számítógépes tesztelés előnyeiről részletesen I. *Csapó, Molnár és R. Tóth, 2008; van Lent, 2009; R. Tóth, 2010*)

Mindazonáltal a papír-ceruza (paper-and-pencil – PP) formátumban végzett mérés-értékelésről számítógépes (computer-based – CB) tesztelésre történő áttérés számos pszichometriai kérdést vet fel (*Johnson és Green, 2006*). Az egyik legalapvetőbb kérdés, hogy a számítógép, mint új közvetítő médium megváltoztatja-e a tesztek konstruktumvaliditását, hiszen ugyanazon teszt papíron vagy számítógépes környezetben felvéve nem minden esetben eredményez ekvivalens tanulói teljesítményeket (*Leeson, 2006*).

A közvetítő eszköz befolyásoló hatásának vizsgálata azokon a mérési területeken nem releváns, ahol az új környezetben más konstruktum vizsgálata történik (pl. *Moe*, 2009; OECD, 2010a, 2011), mint papíron. Ezzel szemben a különböző tesztkörnyezetből származó tesztpontszámok ekvivalenciájának, felcserélhetőségének biztosítása szükséges, ha a papír-ceruza adatfelvételen alapuló teszteredményeket, illetve ugyanazon mérőeszköz számítógépes adaptációjából származó teljesítményeket egyidejűleg szeretnénk elemzésekben felhasználni.

A tanulmányban az MTA-SZTE Oktatáselméleti Kutatócsoportja által végzett papír-ceruza alapú longitudinális vizsgálat (*Csapó*, 2007) első generációs (*Bennett*, 1998) technológia alapú adaptációja során kapott adatok alapján tárjuk fel a tesztet közvetítő médium tanulói teljesítményre gyakorolt hatását és annak mértékét. E vizsgálat részeként olvasás-szövegértés területén vizsgáljuk a tesztmédium hatását a mintához és a tesztek jellemzőihez kapcsolódó háttérváltozók segítségével. A bemutatott vizsgálat számos módszertani érdekességet kínál: a médiahatást feltáró vizsgálatokban újszerű, papír alapú teljesítményt alapul vevő részmintakialakítást, illetve különböző szövegtípusok mentén végzett komparatív elemzéseket. Mivel a közvetítő médium tanulói teljesítményre gyakorolt hatása keveset kutatott terület a folyamatos és nem folyamatos szövegeken, ezért e munka hiánypótlónak tekinthető.

Digitalizált és digitális szövegek értése

Kutatásunk során a diákok szövegértését *digitalizált* mérőeszközökön vizsgáltuk, mely különbözik a bevezetőben említett nemzetközi mérések módszertanától (l. PISA ERA, PIRLS web-based reading). E nagymintás tanulói teljesítménymérésre irányuló programok célja a *digitális* szövegértési képesség vizsgálata, s e konstruktum mentén hasonlítják össze különböző oktatási rendszerek működésének hatékonyságát.

A digitalizált szövegek segítségével végzett felmérés során a papír-ceruza formában is felhasználható szöveget átültetjük számítógépes környezetbe (digitalizáljuk), és ennek segítségével mérjük a diákok szövegértési képességét. Ezzel szemben a digitális szövegértés mérésekor az internet adta lehetőségek és a web 2.0 alkalmazások kerülnek előtérbe (*Balázs* és *Ostorics*, 2011). A digitális szövegértés teszttípusok hiperlinkekkel összekapcsolt weboldalakból állnak, ahol az oldalak közötti kapcsolatot menük, linkek, hivatkozások valósítják meg. Az olvasónak az oldalak között navigálva, több oldalt elolvasva kell egy kérdésre megtalálni a választ (OECD, 2009). Azt a mérési folyamatot, amikor a diákok ugyanazon papír alapú tesztet számítógépen oldják meg, a szakirodalom első generációs, míg azt a tanulói teljesítménymérést, amelyben a tesztkészítők kihasználják az infokommunikációs technológiák, valamint az internet adta lehetőségeket, második generációs tesztelésnek nevezik (*Bennett*, 1998). Az utóbbi esetben a közvetítő eszköznek nemcsak a tesztelési folyamatra – ezáltal a tanulói teljesítményekre – van hatása, hanem a mérés tartalmi keretére is. E különbségek előrevetítik, hogy a digitális szövegekkel végzett mérés-értékelés tartalmi kerete eltérést mutat az általunk referenciapontként használt olvasás-szövegértés tartalmi keretéhez képest. E változások a következő pontok

mentén összegezhetők: (1) statikusság/dinamizmus és a szövegek befolyásolhatósága, (2) szövegformátum, (3) olvasási műveletek (*Balázsai és Ostorics, 2011*).

Az OECD PISA a digitális szövegek értésének mérése során a szövegeket két nagy csoportra osztja az alapján, hogy az olvasó befolyásolhatja-e a szöveg tartalmát vagy sem. Azokat a szövegeket, amelyeket az olvasó leginkább információforrásként használ és nincs lehetősége a digitális szöveg módosítására, szerzői alapú szövegeknek nevezi, míg az olvasó aktív részvételére építő, kommunikációs eszközként működő szövegeket (pl. levelek, blogok, chatszobák, fórumok és internetes úrlapok) üzenet alapú szövegeknek. A digitális szövegértés mérésbe kerültek olyan vegyes feladatok is, amelyek egyszerre kívánták meg az olvasótól a befogadó és aktív részvételt (*Balázsai és Ostorics, 2011*). A digitalizált szövegértés mérésének tartalmi kerete statikus természeténél fogva nem terjedt ki a dinamikus, üzenet alapú szövegek megértésének és használatának mérésére.

A digitalizált olvasás-szövegértés tesztek esetében a PISA tartalmi keretében háromféle szövegformátum különül el: folyamatos (élményszerző), nem folyamatos (információszerző) és kevert szöveg (OECD, 2010b). Utóbbiban a folyamatos elemek segítik az olvasót a nem folyamatos ábrák, címkék, képek által közvetített információ megértésében és felhasználásában. Ezek a formátumok megtalálhatók mind a hagyományos, mind a digitális tesztkörnyezetben. A hipertextus olvasásértésének vizsgálatára a szövegformátumok köre még egy kategóriával bővült. Ezek a többszörös formátumú szövegek, amelyek alatt a „több, egymáshoz lazán kapcsolódó, asszociációs gondolkodási útvonalon, linkek segítségével elérhető szövegeket értjük. A többszörös szövegek a folyamatos, a nem folyamatos és a kevert formátumú szövegek értelmezésének integrálását igénylik” (*Gonda, 2011, online*). A dinamikus tesztkörnyezetben multimédiás anyagokkal is találkozhatunk. A többszörös szövegekben az itemek helyes megválaszolásához szükséges információ diagramokkal, legördülő menüpontokkal és egyéb grafikai elemekkel gazdagított környezetből nyerhető ki akár több különböző weboldaltól vagy egy adott weboldalhoz tartozó lapokról (OECD, 2009).

A papír alapú és a digitalizált olvasás során alkalmazott gondolkodási műveletek (információ visszakeresése, integrálás és értelmezés, reflexió és értékelés) működésére, valamint tartalmára is hatással volt a második generációs mérési platform megjelenése. A multimediális elemekkel, hipertextussal gazdagított dinamikus szövegek tartalmának értelmezése, értékelése előtt az olvasónak hozzá kell férnie a kívánt információhoz, így a digitális szövegértés egy új gondolkodási művelettel, a hozzáféréssel egészült ki. Azonban papír-alapú tesztek megoldása esetén ebben a hozzáférést az olvasó csak kismértékben befolyásolhatja, hiszen a rendelkezésre álló eszközök (tartalomjegyzék, oldalszámozás, névmutató, navigációs gombok) egyszálú lineáris olvasást tesznek lehetővé, míg a többszálú lineáris olvasás során a dinamikus környezetben az olvasó szabadon válogathatja ki és rendezheti az információt (*Balázsai és Ostorics, 2011*).

Az olvasási műveleteket tekintve a második generációs olvasástesztelésnél megfigyelhető a reflexió és értékelés műveleti szint tartalmának gazdagodása is. A nyitott webes környezetben olvasás/böngészés során az olvasó számos szüretlen információforrással találkozhat, melyek szelektálása, illetve hitelességük ellenőrzése kiemelkedő fontos-

sággal bír. A digitalizált olvasás platformjának zárt természete miatt ez a gondolkodási művelet ilyen jelleggel nem rendelkezik (*Balácsi és Ostorics, 2011*).

Jelen kutatásunk középpontjában elsősorban a médiahatás vizsgálata állt digitalizált szövegek segítségével. Mivel a digitalizált olvasás-szövegértésnek nem célja innovatív (hiperlinkek, menük stb.) tesztelési elemek tesztelési folyamatba történő beiktatása, ezért mérésünkben a PISA papír alapú olvasás-szövegértés tanulói teljesítményméréshez alkalmazott tartalmi keretét használtuk fel (OECD, 2010b).

Nemzetközi összehasonlító tanulmányok a médiahatás tanulói teljesítmény alapján történő vizsgálatáról

A tesztfelvétel különböző módjainak a tanulók válaszaikra gyakorolt lehetséges hatását az 1970-es évektől tanulmányozzák (*Gaskill és Marshall, 2007*). Az elmúlt 25 évben több mint 300 nemzetközi tanulmány készült a tesztkörnyezet megváltozásából adódó teljesítménykülönbségekről, az eltérések okainak feltárásáról, illetve a számítógépes és a papír-ceruza tesztelésből származó adatok összehasonlításáról (*Wang, Jiao, Young, Brooks és Olson, 2008*). A kutatások egy része nem igazol szignifikáns különbséget a papír-ceruza és a számítógépes teszteredmények között (*Bodmann és Robinson, 2004; Horkay, Bennett, Allen, Kaplan és Yan, 2006; Johnson és Green, 2006; Poggio, Glasnapp, Yang és Poggio, 2005; Puhan, Boughton és Kim, 2007; Wang, Jiao, Young, Brooks és Olson, 2007, 2008*). Ugyanakkor számos vizsgálat szignifikáns különbséget tárt fel a két teszt-médiumon elért teljesítmények között (*Bennett, Braswell, Oranje, Sandene, Kaplan és Yan, 2008; Csapó és mtsai, 2009; Higgins, Russell és Hoffmann, 2005; Pomplun, Frey és Becker, 2002*), amit *Clariana és Wallace (2002)* alapján médiahatásnak (test mode effect) nevezünk.

A tesztkörnyezet megváltozásából adódó teljesítménykülönbségek vizsgálata során az egyik legintenzívebben kutatott terület a minta, illetve a részminták tulajdonságainak teljesítményre gyakorolt hatása. A minta tulajdonságai dimenzióban leggyakrabban vizsgált változók a számítógéppel kapcsolatos tényezők – például mennyire jártas a diák a számítógép használatában –, a demográfiai adatok közül a kor, a nem, az etnikai hovatartozás, továbbá a szocioökonómiai státusz. Az ezen faktorok mentén végzett vizsgálatok eredményei sem azonosak. A tesztelésben részt vevő diákok demográfiai adatai alapján képzett részcsoportok eredményeinek vizsgálata során *Gallagher, Bridgeman és Cahalan (2002)* kismértékű eltérést igazolt számos amerikai tudásszintmérő program (pl. Graduate Record Examination – GRE, the Graduate Management Admissions Test – GMAT, a SAT I: Reasoning test, the Praxis Professional Assessment for Beginning Teachers, valamint a Test of English as a Foreign Language – TOEFL) eredményeit felhasználva. Mindazonáltal a demográfiai adatok egyes komponenseit vizsgáló tanulmányok közül *MacCann (2006), Clariana és Wallace (2002, 2005)* munkája alapján a tesztalanyok neme mint változó nincs kapcsolatban a papír-ceruza és a számítógépes környezetben elért teljesítménybeli különbséggel. *Kingston (2009)* az általános és a középiskolások teljesítménykülönbségére fókuszáló metaanalízise szerint az általános és a

középkölszintű korosztályú tanulók populációján a kornak mint változónak szintén nincs prediktív ereje a nyomtatott és a számítógépen elért tesztpontszámok különbsége tekintetében.

A szocioökonómiai státusz (SES) vizsgálata során az alacsonyabb SES-indexszel rendelkező diákok papír-ceruza és számítógépes környezetben elért teljesítménye között szignifikáns (egy százalékos) a különbség, ugyanakkor a magasabb státuszú diákok esetében nem volt kimutatható médiahatás (MacCann, 2006). Ezzel szemben egyéb, szintén SES alapján képzett részmintákon végzett vizsgálatok (Pomplun, Ritchie és Custer, 2006; Sandene, Horkay, Bennett, Allen, Braswell, Kaplan, és Oranje, 2005) eredményei azt tükrözik, hogy a SES nem függ össze a papír-ceruza és a számítógépes teljesítménykülönbségekkel. E kutatásokban a SES-indexet az ingyenebéd kritériumában határozták meg.

A számítógéphasználattal kapcsolatos változók közül a számítógépes jártasság (computer familiarity) szerinti mintaképzés eredményei sem egybehangzóak. Míg Clariana és Wallace (2002), R. Tóth, Molnár és Csapó (2009) vizsgálataikban nem igazolták a számítógépes jártasság és a papír-ceruza és számítógépes teszteredmények különbségének kapcsolatát, addig a Bennett és munkatársai (2008), valamint a Horkay és munkatársai (2006) által végzett kutatások eredményei alapján a számítógépes jártasság hatással van az online környezetben elért teljesítményekre. A tradicionális és az elektronikus adatfelvételtől származó eredmények különböznek a tanulók számítógépes jártasságának függvényében. Az eredmények inkonzisztenciájának hátterében az áll, hogy a számítógépes jártasságnak többféle értelmezése létezik a szakirodalomban (Sandene és mtsai, 2005), ebből adódóan több módon mérhető.

A médiahatás-vizsgálatok középpontjában a mintatulajdonosságokon túl a tesztmédiium technológiai sajátosságaiból fakadó jellemzők állnak. Leeson (2006) a felhasználói felület olvashatóságát és interaktivitását vizsgálta a médiahatás tekintetében. A felület olvashatósága szempontjából potenciális különbséget okozó tényezőként említi a monitorméretet, a képernyőfelbontást, a betűtípust és a betűméretet, a sorok hosszát, azok számát, a szóközök és a sortávolság mértékét. Az interaktivitás oldaláról például a görgetés (scroll) és a visszalépés (item review) változók lehetséges hatását mutatja be.

A felhasználói felület olvashatóságát nagymértékben befolyásolja a monitorméret és a képernyőfelbontás (Bridgeman, Lennon és Jackenthal, 2003). A képernyőfelbontás a megjelenített szöveg betűméretére, valamint a megjelenített szöveg hosszúságára is hatást gyakorol, hiszen nagyobb képernyőfelbontás esetén egyszerre több szót és több sort lehet megjeleníteni a monitoron. Amennyiben a felbontás állandó, a képernyő nagysága nincs hatással a megjelenített tartalom (szöveg) hosszúságára, csak a méretére (Bridgeman és mtsai, 2003).

A felhasználói felületet az interaktivitás oldaláról tekintve a szövegek görgetését többnyire negatív hatású faktorként jellemzik (pl. Mead és Drasgow, 1993; Leeson, 2006). Ennek oka az, hogy papíron vagy számítógépen a lapozás funkciót választva az adott bekezdés fix helyen marad, így a diákok meg tudják jegyezni az információ helyét, míg a görgetés közben elveszítik azt. Hosszabb, egy monitoron nem megjeleníthető szövegek digitalizálásánál két lehetőség adódik: egy az egyben digitalizáljuk a szöveget, így a tanulóknak görgetniük kell az olvasás során, vagy a szöveget felosztjuk több részre,

ezáltal az elektronikus teszteléskor a tesztmegoldóknak navigációs gomb segítségével kell lapozniuk a szövegrészek között. Függetlenül attól, hogy a lapozás vagy a görgetés funkciót választjuk, figyelembe kell venni, hogy a képernyőn egyszerre megjelenő információ hatást gyakorol a teszteredményre. Számos vizsgálat (pl. *Bridgeman* és mtsai, 2003) igazolta, hogy a médiahatás mértéke nem jelentős, amennyiben az egy itemre vonatkozó összes információ megjelenik a képernyőn; ezzel összhangban szignifikáns a médiahatás, ha a tanulóknak navigálniuk (tipikusan görgetniük) kell a képernyőn az itemek megválaszolásához (*Pommerich*, 2004). Az adatfelvétel során *Leeson* (2006) további potenciális változóként említi az előző itemre történő visszalépést, mely tényezőnek többnyire nincs hatása az eredménykülönbségekre, ugyanakkor a diákok preferálják a visszalépés lehetőségét.

A médiahatás-vizsgálatok harmadik kérdésköre a tesztjellemzők hatásvizsgálata. E dimenzióban a feladatok megoldásához szükséges tevékenységek, itemtípusok (l. *Kingston*, 2009), illetve válaszadó stratégiák (question answering strategies) is felelősek lehetnek a két környezetben elért teljesítmények különbségéért. Például olvasás-szövegértés tesztek esetén a teljesítménykülönbségeket magyarázhatják, hogy a tanulók online környezetben nem alkalmazhatják a támogató vagy funkcionális olvasási stratégiákat (*Mokhtari* és *Reichard*, 2002). A matematika tudásszintmérő teszteknel, ha a feladat megoldásához bonyolultabb számításra van szükség, negatívan befolyásolja az online környezetben elért teljesítményeket, hiszen a diákok hiába használhatnak papírt a feladatmegoldáshoz, nem élnek a lehetőséggel (*R. Tóth*, 2008). Továbbá a kérdések típusa, a kérdésben szereplő számok is kapcsolatban állhatnak a tanulók feladatmegoldási hajlandóságával, így e változók is felelősek lehetnek a médiahatásért (*Johnson* és *Green*, 2006).

A bemutatott eredmények következtében a szakirodalom alapján nem fogalmazható meg egy egységes, minden területre, kontextusra, itemformátumra és mintára kiterjedő konklúzió. Az eredmények inkonzisztenciája háttérben az áll, hogy a különböző vizsgálatok más-más konstruktumot, különböző mintán, különféle mintakialakítással és itemformátum segítségével mérnek (*Wang* és *Shin*, 2009). E nagyszámú változórendszer hatása és eltérő eredménye miatt az elektronikus tesztelésre történő átállás csak fokozatosan történhet, mert az új médium újabb változókat von be a tesztelés folyamatába, melyek hatását kontrollálni kell (*Csapó* és mtsai, 2009).

Az empirikus vizsgálat céljai, kérdései

Az empirikus vizsgálat célja, hogy összehasonlítsa a papír-ceruza és a számítógépes tesztkörnyezetben felvett olvasás-szövegértés teszteken elért tanulói teljesítményeket, a tesztet közvetítő eszköz tanulói teljesítményekre gyakorolt hatását, illetve jellemezze a tesztmédium megváltoztatásával leginkább érintett tanulói csoportokat. Arra kerestük a választ, hogy:

- 1) statikus (interaktív elemeket nem tartalmazó) teszteknel, a hagyományos és számítógépes formátumban elért teszteredmények jelentősen eltérnek-e egymástól,

- illetve amennyiben az eltérés szignifikáns, papíron vagy számítógépen sikeresebbek-e diákok?
- 2) különböző szövegtípusok esetén van-e számottevő különbség a papír-ceruza és a számítógépes környezetben elért teszteredmények között? Amennyiben igen, mely szövegtípusnál jelentkezik a legjelentősebb eltérés?
 - 3) különböző változók (nem, településtípus, IKT-háttérváltozók és tanulói? teljesítmény) mentén képzett részminták között azonosítható-e szignifikáns különbség? Amennyiben igen, a minta mely jellemzői magyarázzák a teljesítménybeli eltéréseket?

Módszerek

Minta és adatfelvétel

A mintát hatodik évfolyamos diákok (N=449) alkották. Minden diák egy papír és egy számítógépes formátumú olvasás-szövegértés tesztet oldott meg. A vizsgálatban 7 régió, 14 iskola 25 osztálya vett részt. Az iskolák önként jelentkeztek a pilotmérésre, mert az elektronikus tesztelés lebonyolításához megfelelő infrastruktúrával (számítógépparkkal és internetkapcsolattal) kellett rendelkezniük az intézményeknek. A tanulók közel fele (49,6%) fiú volt. A nemek aránya megegyezik az országos reprezentatív longitudinális minta jellemzőivel.

Mind a hagyományos, mind a számítógépes tesztfelvétel 2009 tavaszán zajlott. Az online tesztelés a TAO-platformokon (Testing Assisté par Ordinateur – számítógépes tesztelés, *Farcot és Latour, 2008*) keresztül valósult meg. A diákok mindkét tesztet tanórán oldották meg; az adatfelvételt pedagógusok végezték. A tanulók az iskolában elérhető számítógépeken oldották meg az elektronikus tesztet. A tanulási hatás ellensúlyozására a mintát véletlenszerűen felosztottuk: egyes diákok előbb a számítógépes tesztváltozatot, majd a nyomtatott tesztet, míg a diákok egy másik csoportja először a papír-alapú változatot, majd az online tesztet oldotta meg.

Részminták képzése

Részmintákat képeztünk a nemek, IKT-háttérváltozók (mióta használ számítógépet, milyen gyakran internetezik/számítógépezik, mennyire szeret számítógépezni, illetve mióta tanul számítástechnikát az iskolában) és a papír-ceruza formátumban elért átlagos teljesítmény alapján. A papír-ceruza formátumban elért átlagos teljesítmény alapján öt részmintát alakítottunk ki a kvantilisok mentén (1. táblázat). Amennyiben a mintabontás határán több tanuló teljesítménye megegyezett, minden azonos teljesítményű tanulót ugyanabba a részmintába soroltunk. Ebből adódik a részminták eltérő elemszáma. Az így képzett tanulói csoportok átlagteljesítménye a Dunnett T3 eljárás alapján szignifikánsan különbözik egymástól.

1. táblázat. Részminták képzése a papír-ceruza átlagteljesítmény alapján

Részminták	Teljesítmény-intervallumok (%)	Gyakoriság (fő)
1. részminta	0 – 63,88	78
2. részminta	63,89 – 74,99	106
3. részminta	75,00 – 80,55	77
4. részminta	80,56 – 88,88	109
5. részminta	88,89 – 100,0	79
Összesen	100	449

Mérőeszközök

A vizsgálatban két különböző tesztváltozatot alkalmaztunk. Az egyiket papír-ceruza alapon vettük fel az MTA-SZTE Oktatásméleti Kutatócsoport vizsgálatának keretében, a másikat online környezetben oldották meg a diákok.

A teszteredmények összehasonlíthatósága miatt a tradicionális és a számítógépes formátumú olvasási képességet mérő tesztek párhuzamos tesztváltozatok, azaz (1) ugyanazokat az instrukciókat tartalmazzák, (2) a feladattípusok és a megoldásukhoz szükséges műveletek is azonosak, (3) mindkettő 36 itemből áll, valamint (4) minden diák megoldotta a papír-ceruza és a számítógépes tesztváltozatot is (Alderson, Clapham és Wall, 2000).

A tesztek készítésekor a nemzetközi olvasásvizsgálatok módszertanára támaszkodtunk. Mind a papír-ceruza, mind a számítógépes teszt két résztesztből áll, melyek szövegtípológiai paraméterei mentén a két mérőeszköz szövegtípusok szintjén ekvivalens (2. táblázat).

2. táblázat. A PP és CB olvasás-szövegértés teszt szövegtípológiai paraméterei (Vári, 2003 nyomán)

Szövegparaméterek	1. részteszt	2. részteszt
Forma	Nem folyamatos	Folyamatos
Szövegfajta	Dokumentum	Leíró
Közléshelyzet/olvasási szituáció	Tanulási célú	Személyes célú

Az első részteszt nem folyamatos szövegeket tartalmazott, mellyel három különböző dokumentumtípus (táblázat, térkép, diagram) formájában megjelenített információk de-kódolását és megértését térképeztük fel. E részteszt első feladata táblázatba szervezett információk, a további feladatok (két térkép adatainak összevetése, diagram adatainak le-olvasása) képi tartalmak megértésére irányultak. A második résztesztben egy folyamatos szöveg segítségével mértük a diákok digitalizált olvasási képességét. Szövegnyelvészeti változókat tekintve mind a nyomtatott, mind a monitoron közvetített folyamatos szöveg két részre tagolódt, hét bekezdésből és 16 mondatból állt, a papíron található szöveg

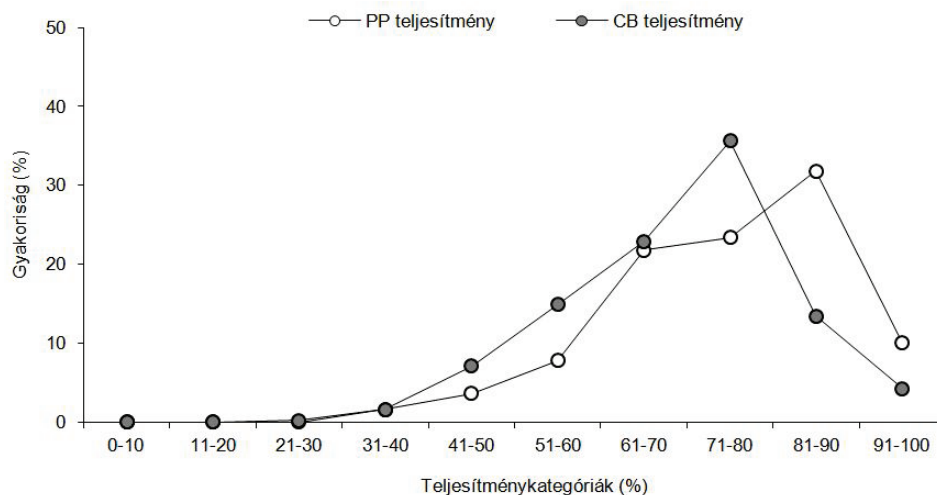
282, a monitoron megjelenített 281 szót tartalmazott. Az online szöveg mondatai átlagosan 6,4, míg a papír alapú szöveg mondatai 5,9 szó hosszúak voltak.

Mindkét teszt lineáris, azaz a feladatok ugyanabban a sorrendben szerepeltek, azonban a PP- és a CB-szöveg megjelenítésüket tekintve eltértek egymástól. Ennek oka, hogy az alkalmazott mérőeszköz változtatás nélkül történő digitalizálása során a diáknak használniuk kellett volna mind a horizontális, mind a vertikális görgetősávot. Így míg PP alapon a diákok egyidejűleg láthatták a teljes tesztlapot, addig online környezetben a szövegek felosztása és egyszerre egy megjelenő item prezentációja mellett döntöttünk. A szakirodalmi áttekintésben bemutatott technológiai paraméterek hatásának csökkentésére a teszt megjelenítését standardizáltuk, az egy monitoron megjelenő szöveg sorainak és azon belüli szavak számát előre rögzítettük, a betűtípus, a szóközök száma, a sortávolság, továbbá a betűszín (l. *Pommerich*, 2004) mindkét tesztváltozaton megegyezett, így e változók hatását nem vizsgáljuk a szakirodalmi áttekintésben. Ezáltal igyekeztünk biztosítani, hogy az iskolában használt legkisebb méretű monitorokon is ugyanannyi információ és ugyanolyan elrendezésben jelenjen meg a feldolgozandó szöveg, mint a nagyobb képernyőjű monitorokon.

A tesztváltozatok kizárólag zárt feladatokat tartalmaztak, mert a szöveges válaszok újabb változókat (pl. gépelési sebesség, számítógépes jártasság a billentyűzet használatában) vonnak be a számítógépes mérésbe. A diákok válaszaikat rádiógombok és jelölőnégyzetek segítségével adták meg, valamint a feladatok közötti váltást navigációs gombokkal végezték.

Eredmények

A nyomtatott formátumban felvett olvasási képesség fejlettségét vizsgáló teszt reliabilitásmutatója (Cronbach- α) 0,76, míg az online tesztváltozaté 0,70. A diákok papír-alapon ($x_{PP}=74,91\%$, $sd=12,97$) átlagosan jobban teljesítettek, mint online formátumban ($x_{CB}=69,22\%$, $sd=12,72$; $t=8,89$, $p<0,01$). A papír-ceruza és a számítógépes teljesítmények eloszlását tekintve (1. ábra) a papír-alapú teszten elért eredmények eloszlásfüggvénye jobbra tolódott a számítógépes környezetben felvett adatok alapján képzett görbéhez képest. A 80% fölötti eredményt elérő diákok száma tradicionális tesztelés során kétszer annyi volt, mint a számítógépes teljesítmény alapján. Ugyanakkor az alacsonyabb képességszintű tanulók kategóriába (41–60%) papír-alapon felvett teszt eredményei alapján feleannyian tartoznak, mint a számítógépes adatfelvételtől származó adatok szerint.



1. ábra
A PP-teljesítmények és a CB-teljesítmények eloszlása

Szövegtípusok analízise

A szövegtípusok szerinti papír-ceruza és számítógépes adatfelvételtől származó eredmények összehasonlítását klasszikus és valószínűségi tesztelméleti eszközökkel is elvégeztük. A tanulók teljesítményét szövegtípusonkénti és tesztkörnyezet szerinti bontásban egy négydimenziós személyitem-térkép (2. ábra) segítségével ábráztuk. A személyitem-térkép lehetőséget nyújt arra, hogy egyidejűleg ábrázzuk a diákok egy-egy teszten (részteszten) elért teljesítményét és a tesztitemek nehézségét. A diákok képességszint szerinti eloszlását a 2. ábrán az 1–4. dimenzió oszlopokban található x-ek tükrözik, ahol minden x négy diákot jelöl. Az ábra jobb oldali oszlopában található számok a tesztitemek sorszámaikat jelölik.

Az első dimenzió a nyomtatott formában nem folyamatos szövegen, a második dimenzió a papír-ceruza formában folyamatos szövegen, a harmadik dimenzió az online adatfelvétel nem folyamatos szövegeken, a negyedik a szintén online környezetben, de folyamatos szövegen elért eredmények alapján kirajzolódó képességszint eloszlást mutatja. Az adatok alapján a nem folyamatos szövegeket tartalmazó részeszt szövegeit papír-ceruza környezetben könnyebben dolgozzák fel a diákok, mint online (átlagos képességparaméterek: $x_{PP}=1,72$, $sd_{PP}=0,04$; $x_{CB}=1,54$, $sd_{CB}=0,04$, $p<0,05$). Ugyanakkor a folyamatos szövegeken elért teljesítmények között még jelentősebb a különbség a két adatfelvétel során (átlagos képességparaméterek: $x_{PP}=1,04$, $sd_{PP}=0,05$; $x_{CB}=0,51$, $sd_{CB}=0,03$, $p<0,05$).

Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasás-szövegértés területén

	PP nem foly.	PP folyamatos	CB nem foly.	CB folyamatos	+ item
4	X				
	X				
	X				
	X				
	XX	X	X		
	XX	X	XX		
3	XXX	X	X		
	XXXX	XX	XX		
	XXXX	XX	XXX		19
	XXXXX	XXX	XXXXX		
	XXXX	XXX	XXXXX	X	43
	XXXXXX	XXX	XXXXX		
2	XXXXX	XXXX	XXXXXX	X	48
	XXXXXXXX	XXX	XXXXXXXX	X	12, 32, 50
	XXXXXXXX	XXXXX	XXXXXXXX	XX	52, 54
	XXXXXXXX	XXXXX	XXXXXXXX	XXX	7
	XXXXXXXX	XXXXXX	XXXXXXXX	XXXX	
	XXXXXX	XXXXX	XXXXXX	XXXX	
	XXXXXX	XXXXXX	XXXXXX	XXXXX	18
	XXXXX	XXXXX	XXXXX	XXXXX	14, 33, 56, 60
1	XXX	XXXXXX	XXXXX	XXXXXXXX	15, 25, 53, 58
	XXX	XXXXXX	XXX	XXXXXXXX	41, 64
	XXXX	XXXXX	XXXX	XXXXXXXX	8, 17, 35, 47
	XXX	XXXX	XX	XXXXXXXX	66
	XX	XXXX	XXX	XXXXXXXXXX	2, 29, 34, 59, 67
	X	XXXX	XX	XXXXXXXX	6, 63, 69
	X	XXX	XX	XXXXXXXX	
0	X	XX	X	XXXXXXXX	21, 24, 57
		XXX		XXXXX	28, 36, 61, 65, 70
	X	XX	X	XXXX	16, 26, 49, 51
		XXX	X	XXX	13, 20
		X		XXX	40
		XX		X	22, 68
		X		X	9, 45
-1				X	1, 5, 46
				X	31, 55
		X			23, 27, 37, 62
					2, 10, 11, 30, 38
		X			
-2					
					4
					44
					39, 42

2. ábra

Olvasás-szövegértés fejlettsége papír-ceruza és számítógépes környezetben folyamatos és nem folyamatos szövegeken (minden 'x' 4 diákat reprezentál)

A nem folyamatos szövegek esetén papír-ceruza formátumban a diagramleolvasási feladat bizonyult a legkönnyebbnek, míg számítógépes környezetben a táblázatos formában olvasható szöveg megértésében voltak a diákok a legsikeresebbek. A legnagyobb

problémát mindkét formátumban a térképek összevetéséből származó információk megértése jelentette.

A két médiumon elért tanulói szintű teljesítményeket összevetve a legjelentősebb különbség ($t=10,78$) a diagramleolvasási feladatoknál (3. táblázat) azonosítható. Szintén szignifikáns ($t=4,74$, $p<0,01$) a médiahatás mértéke a két térképen található információk összevetését célzó feladaton. Ugyanakkor mindössze 2,2%-os az eltérés ($t=2,80$) a táblázatos formában prezentált szöveg megértésében a papír-ceruza és az online formátumban elért teljesítmények alapján. Az eredményeket magyarázhatja, hogy mind a diagramleolvasási, mind a térképösszevetési feladat közös jellemzője, hogy képi formában közöl információt, ugyanakkor a táblázatok szöveges információt jelenítettek meg.

1. táblázat. A nem folyamatos részteszt három dokumentumtípusán elért papír-alapú (PP) és számítógépes (CB) átlagteljesítményeinek összehasonlítása t-próbával

<i>Dokumentumtípus</i>	<i>Átlagok különbsége (%) (PP-CB)</i>	<i>t</i>	<i>p</i>
Táblázat	2,20	2,80	0,01
Térkép	7,28	4,74	0,00
Diagram	13,05	10,78	0,00
<i>Teljes részteszt</i>	<i>4,91</i>	<i>7,28</i>	<i>0,00</i>

A folyamatos, leíró szövegek megértése szintén jelentősen könnyebbnek bizonyult papír-ceruza formában, mint online (átlagos képességparaméter: $x_{PP}=1,04$, $sd_{PP}=0,05$; $x_{CB}=0,51$, $sd_{CB}=0,03$, $p<0,01$). A teljesítménykülönbségeket (mindkét szövegtípus esetén) magyarázhatja továbbá, hogy a tanulók online környezetben nem alkalmazhatták az úgynevezett támogató vagy funkcionális olvasási stratégiákat. Míg papíron a tanulók kiemeléssel, aláhúzással, jegyzeteléssel, esetleg kezükkel végigvezetve segíthetik a szöveg megértését, a szövegrészletek összevetését (pl. diagram oszlopainak összehasonlításakor), addig a jelen vizsgálatban számítógépen prezentált mérőeszközön ez nem volt kivitelezhető.

Részmintákra irányuló elemzések

A papír-ceruza és a számítógépes környezetben elért átlagos teljesítmények kapcsolata szignifikáns ($p<0,01$) mind a teljes teszt ($r=0,44$), mind a részteszt szintjén. Ugyanakkor a szövegtípusonkénti elemzések megerősítették, hogy a két tesztmédiumon elért eredmények erősebb kapcsolatot mutatnak a nem folyamatos ($r=0,45$) szövegek esetén, mint a folyamatos szövegeket ($r=0,20$) tartalmazó részteszteken. Azonban ezek a korrelációs értékek a tanulók átrendeződését vetítik elő a két tesztmédium esetén, amit különböző részminták segítségével vizsgáltunk.

A nemek közötti teljesítménykülönbségek vizsgálata a két tesztkörnyezetben

A fiúk és a lányok teljesítménye a teljes teszt és a részteszt szintű (4. táblázat) elemzések alapján nem különbözött egymástól sem papír alapú, sem az online formátum esetében. A nem folyamatos szövegek három dokumentumtípusa alapján végzett elemzések szerint a táblázatba foglalt információk megértésében a lányok mindkét környezetben jobban teljesítenek ($t_{PP}=2,70$, $t_{CB}=2,74$, $p<0,05$). Ugyanakkor a fiúk sikeresebbek online környezetben a térképek összevetését célzó feladatban. A diagramleolvasásban nincs jelentős különbség a fiúk és lányok teljesítménye között.

4. táblázat. Nemek közötti különbségek a szövegtípusok szerinti bontásban

Szövegtípus	PP		CB	
	<i>t</i> (fiúk-lányok)	<i>p</i>	<i>t</i> (fiúk-lányok)	<i>p</i>
<i>Nem folyamatos szövegek</i>	-1,20	<i>n. s.</i>	-1,45	<i>n. s.</i>
Táblázat	-2,70	0,01	-2,74	0,01
Térkép	0,85	<i>n. s.</i>	2,03	0,04
Diagram	0,46	<i>n. s.</i>	-1,69	<i>n. s.</i>
<i>Folyamatos szöveg</i>	-0,09	<i>n. s.</i>	0,08	<i>n. s.</i>

Megjegyzés: *n. s.* nem szignifikáns

A médiahatás vizsgálata nemek alapján

A nemek alapján képzett részmintákra irányuló elemzések mind a fiúk, mind a lányok médiafüggő viselkedését igazolják a teljes teszt ($t_{fiúk}=5,11$, $t_{lányok}=7,04$, $p<0,05$), a résztesztek és a különböző dokumentumtípusok (térkép, diagram, táblázat) szintjén (5. táblázat). Mindkét nem papír alapon jelentősen jobb teljesítményt nyújtott, mint online formátumban. A fiúknál és a lányoknál is a diagramleolvasási feladatok megválaszolásánál a legjelentősebb a különbség. Továbbá mindkét nemnél a táblázatban található információk megértésében igazolódott a legkisebb médiahatás, aminek mértéke már a fiúk részmintáján elhanyagolható.

5. táblázat. Papír-ceruza és számítógépes formátumból származó teljesítménykülönbségek nemek szerinti bontásban a szövegtípusok alapján

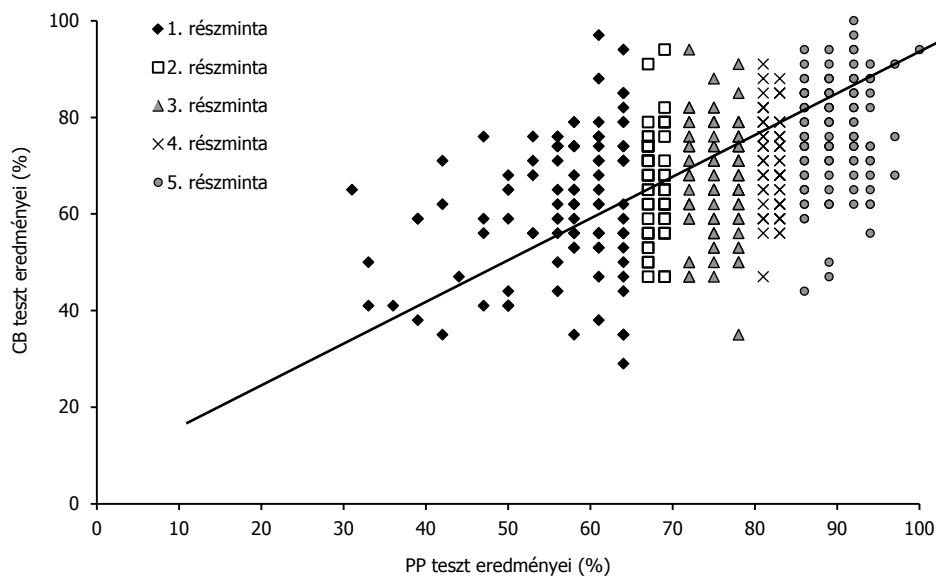
Szövegtípus	Fiúk		Lányok	
	<i>t</i> (PP-CB)	<i>p</i>	<i>t</i> (PP-CB)	<i>p</i>
<i>Nem folyamatos szövegek</i>	4,84	0,00	5,17	0,00
Táblázat	1,77	<i>n. s.</i>	2,68	0,01
Térkép	2,66	0,01	3,64	0,00
Diagram	7,86	0,00	6,08	0,00
<i>Folyamatos szöveg</i>	3,25	0,00	5,43	0,00

Megjegyzés: *n. s.* nem szignifikáns

A médiahatás vizsgálata teljesítmény-alapú mintafelosztás segítségével

A papír-ceruza és a számítógépes teljesítményeket egy közös térben ábrázolva pontfelhőként jelenítettük meg (3. ábra). A pontfelhő egy pontja egy tanulót reprezentál. A vízszintes tengelyen a papír alapon, a függőleges tengelyen az elektronikus környezetben elért teljesítményt ábrázoltuk százalékban kifejezve. Optimális helyzetben a két formátumon elért teszteredmények megegyeznek. Ebben az esetben a ponthalmaz a vastagított vonalon vagy annak közvetlen közelében helyezkedik el. Azonban a tanulók egy rész-halmaza távolabb helyezkedik el az elvárt egyenestől, azaz teljesítményükre jelentős hatást gyakorol a tesztet közvetítő médium.

A különbözőképpen viselkedő tanulói csoportok részleteiből vizsgálatát a papír-ceruza teljesítmény alapú mintafelosztás segítségével végeztük. A kialakított öt rész-minta diákjait eltérő alakzattal jelöltük a 3. ábrán. A papír alapon leggyengébb teljesítményt nyújtók (rombusszal jelölt diákok) többsége a vastagított vonal felett helyezkedik el, e rész-halmaz tanulói jelentősen jobb teljesítményt értek el online, mint a nyomtatott teszt-lapon (6. táblázat).



3. ábra

A papír alapú és a számítógépes teszteredmények összevetése teljesítmény alapú mintafelosztás segítségével

Ugyanakkor a 3–5. rész-minta (3.: háromszög, 4.: kereszt, 5.: kör) diákjainak többsége papír alapon szignifikánsan jobb teljesítményt nyújtott, mint elektronikus környezetben. A második rész-minta diákjainak papír-ceruza és számítógépes teljesítménye nem tér el jelentős mértékben, teljesítményük médiafüggetlennek bizonyult.

A résztesztek szintjén végzett elemzések alapján a folyamatos és a nem folyamatos szövegtípuson is az előzőekben bemutatott tendencia rajzolódik ki: az alacsonyabb teljesítményűek számára előnyös a számítógépes tesztelés bevezetése, míg a jobb képességűeknek a hagyományos tesztelés járul hozzá a jobb olvasásteljesítményhez.

6. táblázat. Papíron és számítógépen elért tesztpontszámok eltérése teljesítmény szerinti részminták között; Tukey-b próba segítségével

Rész- minták	n	Eltérően viselkedő tanulói csoportok PP és CB teljesítménykülönbségei (%)			
		1	2	3	4
1.	78	-7,55	n. s.	n. s.	n. s.
2.	106	n. s.	1,87	n. s.	n. s.
3.	77	n. s.	n. s.	8,67	n. s.
4.	109	n. s.	n. s.	11,17	11,17
5.	79	n. s.	n. s.	n. s.	13,43

Megjegyzés: n. s. nem szignifikáns

A számítógépes háttérváltozók hatása a PP és a CB eredménykülönbségekre

A számítógéphez fűződő háttérváltozók és a papír-ceruza és számítógépes teszteredmények különbsége közötti kapcsolat meglétét regresszióelemzéssel teszteltük. Azonban az analízis nem igazolt összefüggést a PP és CB eredménykülönbség és az internet, illetve számítógép-használat gyakorisága között, valamint aközött, hogy a tanuló mióta használ számítógépet, mennyire szeret számítógépezni, illetve mióta tanul számítástechnikát az iskolában.

Összegzés, további feladatok

A különböző tesztmédiákon történő olvasás-szövegértés eltérő tanulói teljesítményeket eredményezhet, amit médiahatásnak nevez a szakirodalom. E tanulmányban először bemutattuk a papír-ceruza és a számítógépes teszteredményeket összehasonlító nemzetközi vizsgálatok fontosabb eredményeit, a vizsgálatok dimenzióit. Ezt követően ismertettük a választott teljes olvasás-szövegértés teszt, valamint a különböző szövegtípusokat tartalmazó résztesztek szintjén végzett komparatív vizsgálatunk eredményeit.

A PP és CB teszteredmények összehasonlítása alapján a 6. évfolyamos diákok jobb eredményt értek el a nyomtatott médiumon, mint számítógépen. A szövegtípusok mentén végzett elemzések is ezt az eredményt támasztották alá. A nem folyamatos szövegek esetén a diagramleolvasási feladatoknál a legjelentősebb a különbség, míg a legkisebb a táblázatba foglalt információk megértésénél.

A minta jellemzői alapján képzett részminták segítségével feltártuk a tesztkörnyezet megváltoztatásával leginkább érintett tanulócsoportokat. Megállapítottuk, hogy a fiúk és a lányok teljesítménye sem papír alapon, sem számítógépes formátumban nem különbözött jelentős mértékben. A számítógépes háttérváltozók bevonásával végzett regresszióelemzés sem igazolt kapcsolatot az IKT-változók és a médiahatás mértéke között. Ugyanakkor a teljesítmény alapján történt mintafelosztás segítségével végzett összehasonító vizsgálattal három eltérően viselkedő részmintát azonosítottunk: (1) a papír alapon alacsony teljesítményt nyújtó tanulók jelentősen jobb eredményt értek el a papírceruza teszt párhuzamos, számítógépes változatán; (2) a PP alapon átlagos, illetve jó teljesítményű tanulók rosszabb eredményt értek el számítógépes környezetben, végül (3) a diákok harmadik csoportjának médiafüggetlen viselkedését támasztotta alá a vizsgálat.

A kutatás keretein belül számos további kérdés fogalmazódott meg. Mivel a számítógépes jártassággal kapcsolatos változók egyike sem magyarázza a médiahatást, illetve a részminták tanulóinak teljesítménykülönbségét, érdemes megvizsgálni, hogy a számítógép, mint innovatív tesztelési eszköz milyen motivációs hatással bír az egyes tanulói csoportokra. Jövőbeni vizsgálataink alkalmával a számítógépes teszteléssel kapcsolatos tanulói preferenciákat egy kvalitatív kérdőív segítségével mérjük fel.

Köszönetnyilvánítás

A tanulmány a „Diagnosztikus mérések fejlesztése” című TÁMOP 3.1.9-08/1-2009-0001 kutatási program keretében, az SZTE Oktatásméleti Kutatócsoport és az MTA-SZTE Képességfejlesztés Kutatócsoport támogatásával készült. Köszönjük Molnár Évának, hogy rendelkezésünkre bocsátotta az olvasás-szövegértés teszteket.

Irodalom

- Alderson, J. C., Clapham, C. és Wall, D. (2000): *Language test construction and evaluation*. Cambridge University Press, Cambridge.
- Balázi Ildikó és Ostorics László (2011): *PISA 2009 Digitális szövegértés - olvasás a világhálón*. Oktatási Hivatal, Budapest.
- Bennett, R. E. (1998): *Reinventing assessment: Speculations on the future of large-scale educational testing*. Policy Information Center, Educational Testing Service, Princeton, NJ.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. és Yan, F. (2008): Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6. 9. sz. Letöltési ideje: 2011. május 15, <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/178>.
- Bodmann, S. M. és Robinson, D. H. (2004): Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research*, 31. 1. sz. 51–60.
- Bridgeman, B., Lennon, M. L. és Jackenthal, A. (2003): Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, 16. 3. sz. 191–205.
- Clariana, R. és Wallace, P. (2005): Test mode familiarity and performance-gender and race comparisons of test scores among computer-literate students in Advanced Information Systems Courses. *Journal of Information Systems Education*, 16. 2. sz. 177–182.

Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasás-szövegértés területén

- Clariana, R. és Wallace, P. (2002): Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, **33**. 5. sz. 593–602.
- Csapó Benő (2007): Hosszmetszeti felmérések iskolai kontextusban – az első átfogó magyar iskolai longitudinális kutatási program elméleti és módszertani keretei. *Magyar Pedagógia*, **107**. 4. sz. 321–355.
- Csapó, B., Molnár, Gy. és R. Tóth, K. (2009): Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In: Scheuermann, F. és Björnsson, J. (szerk.): *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*. Office for Official Publications of the European Communities, Luxemburg. 113–118.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papíralapú teszteléstől a számítógépes adaptív tesztelésig. A pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 3–4. sz. 3–16.
- Csapó Benő, Molnár Gyöngyvér, Pap-Szigeti Róbert és R. Tóth Krisztina (2009): A mérés-értékelés új tendenciái: a papír- és számítógép-alapú tesztelés összehasonlító vizsgálatai általános iskolás, illetve főiskolás diákok körében. In: Kozma Tamás és Perjés István (szerk.): *Új kutatások a neveléstudományokban 2008*. MTA Pedagógiai Bizottsága, Budapest. 99–108.
- Farco, M. és Latour, T. (2008): An open source and large-scale computer based assessment platform: A real winner. In: Scheuermann, F. és Pereira, A. G. (szerk.): *Towards a research agenda on computer-based assessment: Challenges and needs for European Educational Measurement*. European Commission Joint Research Centre, Ispra. 64–67.
- Gallagher, A., Bridgeman, B. és Cahalan, C. (2002): The effect of computer-based tests on racial-ethnic and gender groups. *Journal of Educational Measurement*, **39**. 2. sz. 133–147.
- Gaskill, J. és Marshall, M. (2007): *Comparisons between paper and computer-based tests: Foundation Skills Assessment 2001–2006 data*. Kelowna, BC: SAEE. <http://www.sae.ca/pdfs/038.pdf>.
- Gonda Zsuzsa (2011): A nyomtatott és a digitális szövegek megjelenítése az interaktív táblán. *Anyanyelv-pedagógia*, **13**. 1. sz. Letöltés ideje: 2011. szeptember 6. <http://www.anyanyelv-pedagogia.hu/index.php?issue=13>.
- Higgins, J., Russell, M. és Hoffmann, T. (2005): Examining the effect of computer-based passage presentation on reading test performance. *Journal of Technology, Learning, and Assessment*, **3**. 4. sz. Letöltés ideje: 2011. május 15. <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/196>.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B. és Yan, F. (2006): Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, **5**. 2. sz. Letöltés ideje: 2011. május 15. <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/180>.
- Hülber László (2012): Matematika műveltségterületen különböző típusú itemek viselkedésének összehasonlító elemzése számítógép és papír alapú környezetben. *X. Pedagógiai Értékelési Konferencia*. Szeged, 2012. április 26–28. 122. o.
- Johnson, M. és Green, S. (2006): On-line mathematics assessment: The impact of mode on performance and question answering strategies. *Journal of Technology, Learning, and Assessment*, **4**. 5. sz. Letöltés ideje: 2011. május 15. <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/191>.
- Kingston, Neal M. (2009): Comparability of computer- and paper-administered multiple-choice tests for K–12 populations: A Synthesis. *Applied Measurement in Education*, **22**. sz. 22–37.
- Leeson, H. V. (2006): The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, **6**. 1. sz. 1–24.
- MacCann, R. (2006): The equivalence of online and traditional testing for different subpopulations and item types. *British Journal of Educational Technology*, **37**. 1. sz., 79–91.
- Mead, A. D. és Drasgow, F. (1993): Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, **114**. sz. 449–458.

- Moe, E. (2009): Introducing Large-scale Computerized Assessment – Lessons Learned and Future Challenges. In: Scheuermann, F. és Björnsson, J. (szerk.): *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*. Office for Official Publications of the European Communities, Luxemburg. 51–56.
- Mokhtari, K. és Reichard, C. A. (2002): Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, **94**. 2. sz. 249–259.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, 7–8. sz. 22–34. o.
- Molnár Gyöngyvér és R. Tóth Krisztina (2008): A számítógép- és papír-alapú tesztelés eredményeinek összehasonlító vizsgálata 5. évfolyamon. VIII. Országos Neveléstudományi Konferencia, Budapest. 2008. november 13–15. 131.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L. és Sainsbury, M. (2009): *PIRLS 2011 Assessment Framework*. TIMSS & PIRLS International Study Center Lynch School of Education, Boston College, Amsterdam. Letöltés ideje: 2010. március 15.
http://timss.bc.edu/pirls2011/downloads/PIRLS2011_Framework.pdf
- OECD (2009): *PISA CBAS analysis and results – Science performance on paper and pencil and electronic tests*. OECD, Paris.
- OECD (2010a): *PISA Computer-Based Assessment of Student Skills in Science*. OECD, Paris.
- OECD (2010b): *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science*. Letöltés ideje: 2011. december 1.
<http://www.oecd.org/dataoecd/10/61/48852548.pdf>
- OECD (2011): *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI)*. Letöltés ideje: 2012. január 5. <http://dx.doi.org/10.1787/9789264112995-en>
- Poggio, J., Glasnapp, D. R., Yang, X. és Poggio, A. J. (2005): A comparative evaluation of score results from computerized and paper and pencil mathematics testing in a large scale state assessment program. *Journal of Technology, Learning, and Assessment*, **3**. 6. sz. Letöltés ideje: 2007. november 12.
<http://escholarship.bc.edu/jtla/vol3/6/>
- Pommerich, M. (2004): Developing Computerized Versions of Paper-and-Pencil Tests: Mode Effects for Passage-Based Tests. *Journal of Technology, Learning, and Assessment*, **2**. 6. sz. 1–44.
- Pomplun, M., Frey, S., és Becker, D. F. (2002): The Score Equivalence of Paper-and-Pencil and Computerized Versions of a Speeded Test of Reading Comprehension. *Educational and Psychological Measurement*, **62**. 2. sz. 337–354.
- Pomplun, M., Ritchie, T. és Custer, M. (2006): Factors in paper-and-pencil and computer reading score differences at the primary grades. *Educational Assessment*, **11**. 2. sz. 127–143.
- Puhan, P., Boughton, K. és Kim, S. (2007): Examining differences in examinee performance in paper and pencil and computerized testing. *Journal of Technology, Learning, and Assessment*, **6**. 3. sz. Letöltés ideje: 2011. március 15. <http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/172>
- R. Tóth Krisztina (2010): Papír-ceruza és számítógépes tesztek eredményeinek összehasonlító vizsgálata. In: Vajda Zoltán (szerk.): *Bölcsész-műhely 2009*. JATEPress, Szeged. 125–136.
- R. Tóth Krisztina (2008): Az online teszteléssel kapcsolatos attitűdök és eredmények a háttérváltozók tükrében. VIII. Országos Neveléstudományi Konferencia, Budapest, 2008. november 13–15. 56. o.
- R. Tóth, K. és Hódi, Á. (2011): Comparing Students Reading Comprehension Achievement along Different Text-types in Paper-based and Computerized Environment. Előadás: American Educational Research Association - Annual Meeting, New Orleans, Louisiana, USA, 2011. április 8–12.
- R. Tóth, K., Molnár, Gy. és Csapó, B. (2009): Online assessment of reasoning skills. 13th European Conference for the Research on Learning and Instruction. Amsterdam, 2009. augusztus 25–29. 238. o.

Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasás-szövegértés területén

- Sandene, B., Horkay, N., Bennett, R., Allen, N., Braswell, J., Kaplan, B. és Oranje, A. (2005): *Online Assessment in Mathematics and Writing: Reports From the NAEP Technology-Based Assessment Project, Research and Development Series (NCES 2005-457)*. U.S. Government Printing Office, U.S. Department of Education, National Center for Education Statistics. Washington, DC.
- Vári Péter (2003, szerk.): *PISA-vizsgálat 2000*. Műszaki könyvkiadó, Budapest.
- Van Lent, G. (2009): Risks and Benefits of CBT versus PBT in High-Stakes Testing. In: Scheuermann, F. és Björnsson, J. (szerk.): *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing*. Office for Official Publications of the European Communities. Luxemburg. 83–91.
- Wang, S., Jiao, H., Young, M., Brooks, T. és Olson, J. (2008): Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, **68**. 1. sz. 5–24.
- Wang, S., Jiao, H., Young, M., Brooks, T. és Olson, J. (2007): A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, **67**. 2. sz. 219–238.
- Wang, S. és Shin, C. D. (2009): Comparability of Computerized Adaptive and Paper-Pencil Tests. *Test, Measurement & Research Service. Bulletin*, 13. sz. 1–7.

R. Tóth Krisztina és Hódi Ágnes

ABSTRACT

KRISZTINA R. TÓTH AND ÁGNES HÓDI: A COMPARATIVE STUDY OF LEARNERS' READING COMPREHENSION ACHIEVEMENT ON COMPUTER-BASED AND PAPER-AND-PENCIL TESTS

The integration of computers into learner assessment has opened up new opportunities in measuring learners' reading skills. The OECD-PISA Electronic Reading Assessment and the PIRLS Web-based Reading Initiative show that large-scale learner assessment programs have also recognised the importance of the digital era and the role computers play in reading literacy. In line with international trends, the Hungarian Educational Longitudinal Programme also intends to shift from paper-and-pencil tests to a computerized environment because of the benefits of modern technology. However, the introduction of computerized tests requires a sophisticated comparative analysis of learner performance in both environments and must be carried out regarding the effect which testing format might have on learner performance. The main objectives of the present study are: (1) to provide an overview of international comparative studies and their results with regard to the test mode effect; (2) to investigate achievement differences on reading tests containing different text types administered on screen and on paper; and (3) to identify learners most affected by the testing medium. This mode effect study is based on a paper-and-pencil test and its computerized parallel test version, which is designed to measure the reading comprehension skills of Hungarian schoolchildren in Year 6. 449 learners participated in the study carried out in 2009. Our results provide empirical evidence for the test mode effect. Participants achieved better performance when reading texts on paper than on screen. Analysis that aimed at examining mode effect on different text types indicated significant differences in learners' performance on continuous and non-continuous test types as well. Learners performed better on both texts presented in the printed medium than online. With regard to gender analysis, the delivery media had a significant impact on both boys' and girls' achievement as they performed significantly better on the paper test than in an electronic environment. However, learners' computer experience did not have a notable effect on the difference in their reading performance achieved on paper or screen. Furthermore, a subgroup-level analysis, carried out by dividing the sample into subgroups on the basis of achievement, identified groups of learners that exhibited varying behaviours.

Magyar Pedagógia, 111. Number 4. 313–332. (2011)

Levelezési cím / Address for correspondence: R. Tóth Krisztina és Hódi Ágnes, Szegedi Tudományegyetem Neveléstudományi Intézet, H–6722 Szeged, Petőfi S. sgt. 30–34.