

PAPÍR ÉS SZÁMÍTÓGÉP ALAPÚ TESZTELÉS NAGYMINTÁS ÖSSZEHASONLÍTÓ VIZSGÁLATA MATEMATIKA TERÜLETÉN, 1-6. ÉVFOLYAMON

Hülber László* és Molnár Gyöngyvér**

* MTA-SZTE Képességfejlesztés Kutatócsoport

** SZTE Neveléstudományi Intézet, MTA-SZTE Képességfejlesztés Kutatócsoport

Az ezredforduló óta intenzív fejlődésen ment át a pedagógiai mérés-értékelés gyakorlata, ami jelentős részben a hazai és a nemzetközi értékelési rendszerek (pl. OECD PISA, IEA PIRLS, NAEP, Országos kompetenciamérés) kiépülésének köszönhető (R. Tóth, Molnár, Latour és Csapó, 2011). A papír alapú tesztelés fejlődési lehetőségei mára kimerültek, a fejlődés elérte korlátait, a továbblépéshez, a 21. században jelentkező új mérés-értékelési igények kielégítéséhez alapvető változtatásra van szükség. Ez a felismerés jelentős kutatás-fejlesztési projektek (pl. ATCS21, Class of 2020 Action Plan; Griffin, McGaw és Care, 2012; SETDA, 2008) kezdeményezését vonta maga után, amelyek egyöntetűen a számítógép alapú tesztelésre való áttérésben jelölték meg a továbblépés irányát, feltérképezve annak lehetőségeit, elterjesztésének megvalósíthatóságát, valamint korlátait.

A hagyományos papír alapú tesztelésről a számítógép alapú tesztelésre való átállás előnye és hatékonysága a mérés-értékelés minden egyes szintjén kimutatható a teszt szerkesztés változtatosságától (Csapó, Ainley, Bennett, Latour és Law, 2012) a kiköszvetítés és adatáramlás gyorsaságán (Csapó, Lőrincz és Molnár, 2012), gazdaságosságán (Farcot és Latour, 2008) és hatékonyságán keresztül az azonnali, objektív visszacsatolás (Becker, 2004) biztosításának lehetőségéig. Számítógépen lehetővé válik azon 21. században kulcsfontosságúnak számító képességek mérése – multimédiás, dinamikus, interaktív itemek, második és harmadik generációs tesztek alkalmazásával –, amelyek korábban papír alapon nem voltak kivitelezhetőek (Molnár, 2010). Elérhetővé válik az adaptív teszt algoritmus; melynek segítségével pontosabbá válik a tudás- és képességszint-becslés (Magyar és Molnár, 2013); bővül a tesztelésbe bevonhatók köre (pl. felolvasóprogram, fejegér, szemegér használatával; Csapó, Molnár és R. Tóth, 2008) és lehetővé válik a kontextuális adatok hatékony (pl. idő, arckifejezés) rögzítése és elemzése is (Csapó, Lőrincz és Molnár, 2012). Ennek következtében a papír alapú tesztelésnél megszokott egyedüli indikátor, a teszteredmény helyett gazdag és jól strukturált, a diák tesztelés alatt mutatott viselkedésének pontosabb követését lehetővé tevő adatbázis áll rendelkezésünkre (Molnár és Lőrincz, 2012).

Azonban a nagy tétellel bíró tesztek, illetve a longitudinális kutatások esetében a felelősségteljes áttérés, a korábbi eredményekkel való összehasonlíthatóság biztosításához ismerni kell a médium megváltozásának teszteredményekre gyakorolt hatását. A konstruk-

tum-validitási kérdések az átállás első fázisában jelentkeznek, amikor a papír alapú médiummal való megfeleltetés a cél, azaz megegyező feladatokkal ugyanazt mérjük számítógépen, mint papír alapon. A további szinteken, amikor a technológia adta lehetőségeket (pl. multimédiás itemek, dinamikusan változó itemek) kihasználjuk a papír alapon nem vagy csak teljesen más módon mérhető képességterületek mérésének érdekében (dinamikus problémamegoldás, zenei képességek mérése), akkor a megfeleltetés, a mérési invariancia kérdése nem vetődik, vetődhet fel.

Jelen kutatás célja annak feltérképezése volt, hogy a matematika műveltségterületen változik-e 1–6. évfolyamos diákok teszten mutatott teljesítménye, ha a papír-alapú (PP) tesztelésről számítógép alapú (CB) tesztelésre térünk át. A munka hiánypótlónak bizonyul abból a szempontból, hogy kisiskolás diákok tesztelésére fókuszál, ahol a számítógépes gyakorlottság miatt a legnagyobbak lehetnek a különbségek. Nagy mintán vizsgáltuk a kisiskolások számítógép alapú tesztelésbe való integrálhatóságát; a kutatás során felhasznált itemek nem korlátozódtak feleletválasztós itemekre, hanem típus, tartalom és kontextus mentén a papír alapon elérhető teljes spektrumot lefedik.

A papír és számítógép alapú tesztelés összehasonlító vizsgálatai

A közvetítő eszköz teljesítménybefolyásoló hatásának kutatása az ezredforduló első évtizedének egyik legfontosabb mérés-értékelési témája volt (*Gallagher, Bridgeman és Cahalan*, 2000), ma már jól dokumentált kutatási területnek számít. Számos kis- és nagymintás (l. *Applegate*, 1993; *Ito és Sykes*, 2004), az oktatás (l. *Russel*, 1999) és a pszichológia (l. *Blazek és Forbey*, 2011) területére vonatkozó összehasonlító kutatás valósult meg, melyek közös célja a közvetítő eszköz teljesítménybefolyásoló hatásának vizsgálata volt. A kutatási eredmények sokféleségének egyik oka, hogy azok különböző feltételek mellett, eltérő tulajdonságú mintákon és különböző mérőeszközökkel valósultak meg, ami jelentős részben nehezítette és nehezíti egy közös konklúzió kialakítását. Ennek hatására minden jelentős mérés-értékeléssel foglalkozó szervezet, intézet elindította saját kutatási feltételeihez igazodó, az áttérés biztosítását megfelelő mértékben támogató és segítő kutatásait (pl. *Pearson Educational Measurement*, 2003; *Peak*, 2005). Napjainkra egyre szélesebb körben elfogadott az a nézet, miszerint a technológia terjedésével egyre kevésbé kérdéses, hogy a diákok teljesítményét befolyásolja-e az, hogy papíron vagy számítógépen oldják-e meg a feladatokat (*Mayrath, Clarke-Midura és Robinson*, 2012), azaz a médiahatás-vizsgálatok helyett a második és harmadik generációs teszteléssel kapcsolatos kérdések kerülnek a középpontba (*Way, Davis és Fitzpatrick*, 2006).

Mazzeo és Harvey (1988) úttörőnek számító metaanalízisében 30 intelligencia-, személyiség-, különböző képesség- (*aptitude*) és teljesítmény- (*achievement*) tesztekre építő összehasonlító elemzése alapján arra a következtetésre jutott, hogy a közvetítő eszköz, a tesztelés módja az időkorlátos (*speeded*) tesztek kivételével nincs hatással a diákok teljesítményére. Ezt az eredményt támasztotta alá néhány évvel később *Mead és Drasgow* (1993), valamint *Kim* (1999) metaanalízise is rögzített, fix tesztek vonatkozásában. Adap-

tív algoritmust vagy időkorlátot használó tesztek kivételével összehasonlíthatónak, felcserélhetőnek tartották a papír és számítógép alapú teszteken nyújtott teljesítményeket.

Az ezredfordulóra, a jelentősen megnövekedett technológiai alapú mérés-értékelés iránti érdeklődéssel párhuzamosan, az összehasonlító vizsgálatok száma is megnőtt. A korábbiaktól eltérően részletesebb, alaposabban dokumentált elemzéseket publikáltak, ahol a diákok átlagos teljesítményének összehasonlításán túl sor került a minta jellemzőinek (kor, nem, szocioökonómiai státusz, etnikum, számítógépes gyakorlottság; *Poggio és mtsai*, 2005); a technológiai paramétereknek (monitorméret, -felbontás; betűtípus, -méret; görgetés, visszalépés; *Waters és Pommerich*, 2007); a mért műveltségi terület (kontextus) szerepének jellemzésére is (*Kim és Huynh*, 2006). Korábban önmagában az átlagok összehasonlítása nem fejezte ki, árnyalta megfelelően az esetek legnagyobb részében eltérő mintán felvett teszteredmények közötti különbségeket, valamint a különbségekre és az azonosságokra fókuszáló elemzések feltételezték, hogy ugyanazon konstruktum mérésére került sor papír és számítógép alapon. Ezt felismerve, illetve kihasználva a legújabb elemzési technikák adta lehetőségeket, további kutatások indítására került sor (pl. *Schroeders és Wilhelm*, 2011; *Marints*, 2010), kiemelten kezelve a médiahasználat fontosságának kérdését.

Az összehasonlítás alapját képező változók, a vizsgálatok általános jellemzői

A közvetítő eszköz szerepét vizsgáló kutatások közös sajátossága, hogy a vizsgált korosztály jellemzően a 10 évnél idősebbek, elhanyagolható mennyiségű kutatás fókuszál a kisiskolás diákok különböző tesztkörnyezetben mutatott viselkedésének jellemzésére (l. pl. *Applegate*, 1993). A kutatások legnagyobb része középiskolás diákok papír és számítógép alapú teszteken nyújtott teljesítményét hasonlítja össze (*Bennett*, 2003). Az elemzésbe bevont háttérváltozók tekintetében leggyakoribb a nem, a szocioökonómiai háttér (SES), valamint a szülők iskolai végzettsége szerinti csoportképzés (*Wang, Jiao, Young, Books és Olson*, 2007; *Bennett, Braswell, Oranje, Sandene, Kaplan és Yan*, 2008; *Parshall és Kromrey*, 1993; *Gallagher, Bridgeman és Cahalan*, 2000).

A technológiai sokszínűség következtében megjelentek a feladatok megjelenítésének sajátosságait és a tesztelés technikai paramétereit is fókuszba állító elemzések (*Hetter, Segall és Bloxom*, 1997). Az eredmények alapján a közvetítő eszköz teljesítménybefolyásoló hatása nem szignifikáns, ha ugyanazon feladat papíron és monitoron történő megjelenítése közel azonos (pl. számítógépen nincs szükség görgetésre; *Bridgeman, Lennon és Jackenthal*, 2003), illetve a monitor felbontása, a feladat megjelenítése nem korlátozza annak olvashatóságát (*Kingery és Furuta*, 1997).

A texasi oktatási ügynökség (Texas Education Agency, 2008) főbb műveltségi területek szerinti bontásban (pl. matematika, olvasás, írás, természettudományok) szintetizálta korosztály tekintetében a közoktatásban részt vevő tanulókra vonatkozó összehasonlító elemzéseket. Az elemzésbe bevont kutatások legnagyobb része – műveltségi területtől függetlenül – összehasonlíthatónak, felcserélhetőnek tartja az eredményeket (pl. *Kim és Huynh*, 2007). A kutatások tizede a számítógép alapon megoldott tesztek (pl. *Way, Davis és Fitzpatrick*, 2006), illetve néhány elemzés a hagyományos papír alapú tesztekkel vélte nehezebbnek (pl. *Choi és Tinkler*, 2002) a számítógép alapú tesztekkel szemben.

Összehasonlító vizsgálatok a matematika területén

A matematika területén végzett összehasonlító vizsgálatokra is jellemző az eredmények sokszínűsége és az azok szintetizálására való törekvések (pl. Wang és mtsai, 2007, 2008; Kingston, 2009) megjelenése. Az 1. táblázat azon jelentős összehasonlító kutatásokat rendszerezi (1) a kutatásba bevont minta, (2) a kutatás elrendezése, (3) a vizsgált itemtípusok, (4) az elemzés során alkalmazott eljárások, (5) a kulcseredmények és (6) a további, a jelen kutatás szempontjából lényeges eredmények szerint, amelyek rögzített formátumú teszt(ek)en és legalább 100 fős általános iskolai (rész)mintán alapultak.

1. táblázat. Összehasonlító vizsgálatok a matematika műveltségterületen

Publikáció	Minta	Kutatási elrendezés	Vizsgált item-típusok	Elemzési módszer	Ekvivalencia	További, kiemelt eredmények
Choi és Tinkler, 2002	3. évf. $N_{PP,CB} \approx 800$	kevert elrendezés	FV	IRT	CB nehezebb	A számítógépes gyakorlottság befolyásolja a különbségeket.
Ito és Sykes, 2004	4-12. évf. $N_{PP}=12648$ $N_{CB}=3977$	független és összetartozó minták	FV	IRT, DIF	CB nehezebb	Preferencia tekintetében CB fölény.
Sandene és mtsai, 2005	8. évf. $N_{PP,CB}=2700$	független minták	FV, SZA	IRT	CB nehezebb	Magasabb iskolai végzettséggel rendelkező szülők gyerekeinél szignifikáns különbség a PP tesztek javára.
Zhang és Lau, 2006	8. évf. $N_{PP,CB}=801$	független minták	FV, SZA	KT, IRT	CB nehezebb	Nagyobb eltérések a szövegalkotó feladatoknál.
Wang és mtsai, 2004	2-5., 7-12. évf. $N_{PP,CB}=1744$	összetartozó minták	FV	ANOVA	ekvivalens	$Cronbach-\alpha_{CB} \geq 0,88$, $r_{CB-PP} \geq 0,8$; nincs különbség az egyes részminták viselkedésében.
Poggio és mtsai, 2005	7. évf. $N_{PP,CB}=646$	független és összetartozó minták	FV	IRT, DIF, DICE, DDF	ekvivalens	$r_{CB-PP}=0,96$
Johnson és Green, 2006	10-11 évf. $N_{PP,CB}=104$	független minták	FV	kvalitatív	ekvivalens	Számítógépen más megoldási stratégia, gyakoribb fejben számolás.
Way, Davis, és Fitzpatrick, 2006	8, 11. évf. $N_{PP,CB}=1273$	független minták	FV, SZA	IRT	ekvivalens	Nagyobb különbségek szövegalkotó feladatoknál.
Oregon Department of Education, 2007	3-10. évf. $N_{PP,CB}=1987$	összetartozó	FV	KT	ekvivalens	$r_{CB-PP}=0,75$, a legnagyobb különbségek 3. évfolyamon mutatkoztak.
Puhan, Boughton és Kim, 2007	1-12. évf. $N_{PP,CB}=1136$	független minták	FV	DIF, Cohen d	ekvivalens	Az itemek viselkedése azonos.

Megjegyzés: FV: feleletválasztós, SZA: szövegalkotó, KT: klasszikus tesztelmélet, IRT: valószínűségi tesztelmélet, DIF: különböző itemműködés (*differential item functioning*), DICE: különböző itemkategóriaműködés (*differential item category functioning*), DDF: különböző disztraktorműködés (*differential distractor functioning*).

A matematika területén végzett kutatások közel fele nehezebbnek ítélte meg a tesztet, ha azt számítógépen közvetítették ki, míg az elemzések másik fele médiafüggetlennek tartotta a tesztek viselkedését. Nem ismerünk olyan matematikára vonatkozó, a fenti feltevéleket teljesítő kutatást, amely a hagyományos papír alapú teszt megoldását vélte nehezebbnek a számítógépen megoldottal szemben.

A teljesítménybeli különbségekre rámutató elemzések közös jellemzője, hogy az eltérések mértéke minden esetben 5% alatt marad (ami a minta nagysága miatt bizonyulhatott szignifikáns különbségnek). A teszteken mutatott teljesítménybeli eltéréseket jellemzően nem befolyásolta a főképp felső tagozatos évfolyamokon tanuló diákok életkora, ugyanakkor kisiskolások körében már jelentősebb – a számítógépes gyakorlottságtól függő – viselkedési különbségekről számoltak be (Oregon Department of Education, 2007; Choi és Tinkler, 2012). Az elemzésekbe bevont itemek típusa többségében feleletválasztós (pl. Johnson és Green, 2006; Kingston, 2009) vagy olyan feleletalkotó, rövid választ igénylő itemek, amelyek javítása automatizálható (pl. Sandene, Bennett, Braswell és Oranje, 2005). Tartalmilag dominálnak az algebrai feladatok, számsorozatok, szóbeli és számokra alapozott analógiák. A tartalmi elemzések nem mutatnak ki eltérést az egyes kategóriákba sorolt itemeken mutatott viselkedésmintázatokban (pl. Ito és Stykes, 2005).

Az összehasonlító vizsgálatok során alkalmazott kutatási elrendezés két alapvető módja a független (a minta egyik része papíron, másik része számítógépen oldja meg a tesztet), illetve az azonos (a minta minden tagja papíron és számítógépen is megoldja a tesztet) minta alkalmazása. Független minta alapú kutatási elrendezés esetén kulcsfontosságú a részminták – az eredményeket befolyásolható háttérváltozók szerint (pl. életkor, nem, szocioökonomiai státusz, tanulmányi előmenetel; Poggio és mtsai, 2005) – történő összehasonlíthatósága. Azonos mintára építő kutatási elrendezésben az összehasonlítás alapját képező tesztek, az első tesztelési periódus során bekövetkező tanulási hatás kiküszöbölése, esetleges ekvivalens tesztváltozatok kidolgozása jelenthet nehézséget. Mindez rotált (kevert) adatfelvételi módot alkalmazva (a minta egyik fele először a papír, majd a számítógép alapú tesztet oldja meg, a másik fele fordítva) elkerülhető probléma (pl. Poggio és mtsai, 2005).

Az alkalmazott elrendezési mód meghatározza és esetlegesen behatárolja az elemzésekhez használható eljárások körét, melyek között tipikusan a nyerspontszámok (Russel, 1999) és/vagy a skálázott képességszintek (Way, Davis és Fitzpatrick, 2006; Choi és Tinkler, 2002) összehasonlítása szerepel. Ugyanakkor az összehasonlíthatóság kérdésében nem elég önmagában csak az átlagos teljesítmények összehasonlítását elvégezni. Az itemek szintjén jelentkezhet olyan médiahatás, amit elfed az összpontszámok különbsége (Pommerich, 2004).

Az összehasonlító vizsgálatok értékeléséhez hozzátartozik, hogy a legtöbbjük Amerikában zajlott, az amerikai mérési hagyományokat követve (feleletválasztós itemek, sztenderdizált tesztek), amelyek nem teszik lehetővé egyrészt a teszteredmények tartalom szerinti, másrészt a különböző feladattípusokon mutatott viselkedésmintázatok és feladat-megoldási stratégiák (Johnson és Green, 2006) elemzését. A közös következtetés levonását tovább nehezíti, hogy az egyes kutatások mintája, felépítése, a vizsgált konstruktum, az alkalmazott elemzési eljárások, illetve azok dokumentációja eltérő

(Wang és Shin, 2009). Mindezt nemzetközi szinten minden jelentősebb mérés-értékeléssel foglalkozó intézet – melynek vannak papír alapú mérési hagyományai – elindította a saját mérési koncepcióján, tesztjein, mintáján nyugvó médiahatás-vizsgálatokat, amelyek célja, hogy biztosítsák a mérési eredmények összehasonlíthatóságát. Annak ellenére, hogy a technológia széles körű terjedésének és használatának következtében körvonalazódik a tendencia, miszerint háttérbe szorulnak a gyakorlottságból eredő problémák, és a papír alapú tesztek elhagyva előtérbe kerül a diákok által határozottan jobban preferált (Way, Davis és Fitzpatrick, 2006), technológia adta plusz lehetőségeket kihasználó számítógépes tesztek alkalmazása.

A fenti áramlatba illeszkedik a tanulmányban bemutatott kutatás. A 20 éves mérési hagyományokkal rendelkező szegedi műhely kutatásainál a papír alapú adatfelvételt fokozatosan felváltják a számítógép alapú adatfelvételi módok. A korábbi eredményekkel való összevethetőség biztosítása érdekében ezért kiemelt fontosságúvá vált az összehasonlító elemzések végzése. A tanulmány célja a matematika műveltségterületen, tág életkori intervallumban, változatos itemtípusok alkalmazásával azon itemjellemzők azonosítása, amelyek felelősek a megegyező, illetve eltérő, különböző tesztkörnyezetben érvényesülő médiahatásért.

A kutatás célja

A kutatás alapvető célja 1–6. évfolyamos diákok papír, illetve számítógép alapú matematikateszten mutatott viselkedése alapján azon itemjellemzők azonosítása, amelyekkel leírhatóak a médiahatástól független és médiahatástól függő feladatok. Célunk annak feltérképezése, hogy biztosítható-e a hagyományos, mindenki által elfogadott papír alapú tesztelésről a számítógép alapú tesztelésre történő áttérés úgy, hogy összehasonlíthatóak maradjanak a két tesztkörnyezetben nyújtott teljesítmények.

Az áttérés egyik lényeges kérdése annak meghatározása, hogy hány éves kortól alkalmazhatók ekvivalens módon a számítógépes és papír alapú tesztek, alkalmazhatóak-e már az iskolába lépés kezdetén vagy esetleg csak idősebb korban. Specifikálhatóak-e olyan itemtípusok, amelyek alkalmazása ajánlott, mert nem módosítja a teszt viselkedését, vagy meghatározhatóak-e olyan itemtípusok és itemjellemzők, amelyek tipikusan eltérő viselkedéshez, más-más feladat-megoldási stratégia alkalmazásához vezetnek papír és számítógép alapú környezetben. A kutatás célja (1) 1–6. évfolyamos diákok számítógép alapú teszteken nyújtott teljesítményének megbízhatósága, (2) ugyanazon tulajdonságokkal jellemezhető itemek papír és számítógép alapú környezetben való viselkedésének összehasonlítása, (3) azon itemjellemzők meghatározása, amelyekkel leírhatóak a médiahatás-független és -függő feladatok, valamint (4) a papír és számítógép alapon eltérő itemviselkedést indukáló tartalmi, formai és technikai jellemzők azonosítása.

Módszerek

Minta és adatfelvétel

A papír alapú adatfelvétel mintáját egy 2010-ben 1–6. évfolyam tekintetében országosan reprezentatív mintán lezajlott kutatás képezte ($n=40\,571$). A számítógép alapú adatfelvétel két tanévre rá 22 715 diák részvételével történt, évfolyamonként legalább 3000 diákkal (1. 2. táblázat). A két adatfelvétel mintájának azonosságát diákszintű minta-illesztéssel biztosítottuk. Ennek következtében a két független, de már illesztett mintára tekinthetünk úgy, mint azonos mintára.

Az iskolák önkéntes alapon vettek részt a kutatásban. A számítógép alapú adatfelvétel során az iskolák saját infrastruktúrájukat, számítógépeiket és internethálózatukat használták. A tesztelés az eDia (elektronikus Diagnosztikus mérési rendszer) platform segítségével valósult meg (Molnár és Csapó, 2013), lebonyolításához internetes böngészőn és internetkapcsolaton kívül más, előzetes program telepítésére nem volt szükség. A mérés előtt felhívtuk a felügyelő tanárok figyelmét arra, hogy igény esetén biztosítsanak minden tanuló részére jegyzetpapírokat.

2. táblázat. A kutatás eredeti és illesztett mintája

<i>Évfolyam</i>	<i>Számítógép alapú adatfelvétel</i>	<i>Papír-ceruza alapú adatfelvétel (eredeti)</i>	<i>Papír-ceruza alapú adatfelvétel (illesztett)</i>
1.	3048	7924	3048
2.	3459	7017	3459
3.	3702	6366	3702
4.	4179	6749	4179
5.	4502	6809	4502
6.	3825	5706	3825
<i>Összesen</i>	<i>22 715</i>	<i>40 571</i>	<i>22 715</i>

Mérőeszközök

Papír alapon évfolyamonként átlagosan 30 tesztváltozatot dolgoztunk ki. Minden tesztváltozat három klaszterből (résztesztből) épült fel. Egy klaszter 3-4 darab 4-5 ítemes matematikafeladatot, azaz összesen körülbelül 15 ítemet tartalmazott. A papír alapú tesztek válogatott, számítógépen is a papír alapú megjelenítéshez hasonló módon alkalmazható feladataiból – biztosítva a feladattípus és itemparaméterek szerinti sokféleséget és kiegyenlített elemszámot – 10 tesztváltozat készült. Korábbi kutatások eredményei alapján (Hülber, 2012), a részletesebb elemzések elvégezhetősége érdekében, a feladatok szelekciója során külön hangsúlyt fektettünk a grafikus itemek megfelelő számban történő megjelenésére is. A feladatok digitalizálása során törekedtünk azok tulajdonságainak,

kinézetének megőrzésére (l. 1. ábra). Összességében mindkét médiumon 184 feladat (879 item) kiközvetítésére került sor (3. táblázat).

Karikázd be annak a műveletsornak a betűjelét, amelynek a lépéseit mutatja az ábra! Pl.: **(a)**
Húzd át annak a betűjelét, amelyik nem az ábrázolt lépésekhez kapcsolódik! Pl.: **-a)**
 A besatírozott részek a részeredményeket és a végeredményt jelölik.

a) $30 + 20 \cdot 3 - 10 : 2$
 b) $(30 + 20) \cdot 3 - 10 : 2$
 c) $(30 + 20 \cdot 3 - 10) : 2$
 d) $30 + (20 \cdot 3 - 10) : 2$
 e) $[(30 + 20) \cdot 3 - 10] : 2$

1. lépés
 2. lépés
 3. lépés
 4. lépés

Jelöld Igaznak annak a műveletsornak a betűjelét, amelynek a lépéseit mutatja az ábra!
Jelöld Hamisnak annak a betűjelét, amelyik nem az ábrázolt lépésekhez kapcsolódik!
 A besatírozott részek a részeredményeket és a végeredményt jelölik.

☐ Igaz ☐ Hamis a) $30 + 20 \cdot 3 - 10 : 2$
☐ Igaz ☐ Hamis b) $(30 + 20) \cdot 3 - 10 : 2$
☐ Igaz ☐ Hamis c) $(30 + 20 \cdot 3 - 10) : 2$
☐ Igaz ☐ Hamis d) $30 + (20 \cdot 3 - 10) : 2$
☐ Igaz ☐ Hamis e) $[(30 + 20) \cdot 3 - 10] : 2$

1. lépés
 2. lépés
 3. lépés
 4. lépés

1. ábra

Ugyanazon feladat papír és számítógép alapú megjelenítése

A matematikateszt mellett minden diák kitöltött egy háttéradatokra (nem, szülők iskolai végzettsége, iskolai, matematikai előmenetel; matematikaattitűd, informatikaoktatásban való részvétel) vonatkozó 7 ítemes kérdőívet.

3. táblázat. A tesztváltozatok feladattípusonkénti elemszáma

Zárt végű itemek		Nyílt végű itemek	
Altípusok	N	Altípusok	N
Választás	314	Kiegészítés	167
Hozzárendelés	62		
Rendezés	48	Teljes válasz	288
Összesen	424	Összesen	455

Eljárások

A kutatás során függetlenminta-elrendezést alkalmaztunk. Miután a két adatfelvétel mintájának tulajdonságai jelentős mértékben eltértek egymástól, az eredmények megbízhatósága érdekében indokolt volt mintaillesztési eljárás alkalmazása. A számítógép alapú adatfelvételben részt vevő minden diákhoz illesztettünk egy évfolyam, nem, régió, anya, apa iskolai végzettsége, valamint kerekített tanulmányi átlag vonatkozásában azonos jellemzőkkel bíró tanulót a papír alapú adatfelvétel mintájából. A diákszintű illesztés eredményeként a tanulók 70%-ához legalább öt szempont szerint illeszkedő párt rendeltünk.

A különböző teszteken mutatott teljesítmények összehasonlítását az azokat összekötő horgonyitemek, illetve a valószínűségi tesztelméleti modellek alkalmazása tette lehetővé. A tesztek felépítő klaszterek, résztesztek speciális elrendezése biztosította a horgonyzás megfelelő stabilitását. A számítógépen felhasznált mind a 10 klaszter három-három különböző tesztben szerepelt, egyszer a teszt elején, majd a közepén és a végén. Ezáltal összességében mind a 10 összeállított teszt, ha áttételesen is, de minden másik tesztrel horgonyozhatóvá vált, valamint kiküszöböltük a feladat tesztben elfoglalt helyének esetleges torzító hatását. A papír alapú tesztek horgonyzása hasonló technikával történt (erről részletesebben l. Molnár, 2013).

A különböző, de horgonyitemekkel összekötött teszteken mutatott teljesítmények összehasonlítását, közös képességskálára konvertálását, az adatok skálázását a kétdimenziós Rasch-moddal végeztük, feltételezve, hogy a feladatok azonossága ellenére nem feltétlen teljesül az a kitétel, miszerint papír és számítógép alapon teljesen azonos konstruktum mérésére kerül sor. A tesztek megbízhatóságának leírására a klasszikus tesztelméletben használt Cronbach- α mellett az azzal analóg, de horgonyitemekkel összekötött tesztek közös jellemzésére is alkalmas személyszeparációs reliabilitásmutatót (l. Molnár, 2013b) alkalmaztuk.

A közvetítő eszköz teljesítménybefolyásoló és ezzel párhuzamosan itemviselkedést meghatározó hatását a következő szempontok szerint elemeztük: (1) a feladatokhoz tartozó információk feldolgozása (információ típusa, mennyisége, elrendezése), (2) a feladatmegoldás során szerepet játszó (pszichikus, tartalmi, műveleti, kontextuális jellemzők), illetve (3) a feladatmegoldó tevékenységhez kötődő itemparaméterek (a rögzíteni kívánt információ) mennyisége, minősége és a rögzítés módja (l. 4. táblázat).

4. táblázat. Az összehasonlítást biztosító paraméterek rendszere (Csíkos és Csapó, 2011; Vidákovich, 2012 alapján)

Információ feldolgozása	információ mennyisége	1-2 karakter 1-2 szó egy vagy több mondat	
	információ típusa	szimbólumok számok betűk vegyes grafikus elemek táblázatok	
	információ elrendezési módja	lineáris szöveg szövegdobozok szöveg + grafikus elemek	
Feladatmegoldás	tartalom	számok, műveletek, algebra relációk, függvények geometria kombinatorika, valószínűség-számítás, statisztika	
	kontextus	rutinfeladat (szöveges vagy nem) realisztikus feladat (autentikus vagy nem)	
	pszichikus struktúrák	szaktudományi, alkalmazási dimenzió	gondolkodási dimenzió
		ismeret megértés alkalmazás magasabb szintű műveletek	rendszerező képesség kombinatív képesség deduktív gondolkodás induktív gondolkodás
Feladatmegoldó tevékenység	feladattípus	zárt végű feladatok nyílt végű feladatok	
	válasz jellemzői (szövegalkotó feladatoknál)	válasz hossza válasz típusa	
	eszközhasználat	csak egérhasználat egér- és billentyűzethasználat	

Eredmények

A tesztek megbízhatósága

A tesztek belső konzisztenciája megfelelő volt mind a hat évfolyamon papír (Cronbach- $\alpha \geq 0,86$) és számítógép alapon is (Cronbach- $\alpha \geq 0,91$; l. 5. táblázat). A számítógép alapú tesztek megbízhatósági mutatói átlagosan minden évfolyamon magasabbnak bizonyultak, mint papír alapon. A feladatok viselkedését egyben jellemző személyszeparációs reliabilitásmutató értéke is ezt támasztja alá (papír alapon=0,84, számítógép alapon=0,89).

5. táblázat. A tesztváltozatok megbízhatósági mutatói

Évf.	Számítógép alapú adatfelvétel			Papír alapú adatfelvétel		
	Cronbach- α értékek átlaga	Legkisebb Cronbach- α érték	Személy- szeparációs reliabilitás	Cronbach- α értékek átlaga	Legkisebb Cronbach- α érték	Személy- szeparációs reliabilitás
1	0,94	0,93	0,91	0,88	0,87	0,88
2	0,93	0,91	0,89	0,87	0,85	0,85
3	0,89	0,80	0,91	0,85	0,80	0,85
4	0,91	0,84	0,88	0,86	0,82	0,84
5	0,87	0,81	0,85	0,83	0,81	0,82
6	0,91	0,88	0,89	0,86	0,84	0,82

A papír és a számítógép alapú tesztek viselkedése

A tesztek papíron és számítógépen való viselkedésének összevetését, a két médiumon mért átlagos itemnehézségi értékek kapcsolatának vizsgálatával végeztük el. Átlagosan egyik évfolyamon sem volt kimutatható szignifikáns különbség a feladatbank papír és számítógép alapú változata, nehézsége között (5. táblázat). Az azonos tulajdonsággal jellemezhető itemek évfolyamonkénti átlagos médiumfüggését pontosabban jellemzi az itemnehézségi értékek közötti korrelációs együtthatók nagysága. Az iskoláztatás elején, az első három évfolyamon $r=0,70$ ($p<0,01$) körüliek, majd negyedik évfolyamtól kezdődően fokozatosan erősödnek a korrelációs értékek. Hatodik évfolyamra a papír és a számítógépen nyújtott teljesítmények közötti összefüggés értéke $r=0,92$ ($p<0,01$; 6. táblázat).

6. táblázat. A papíron és a számítógépen kiköszvetített itemek átlagos nehézségi értékei közötti kapcsolatok

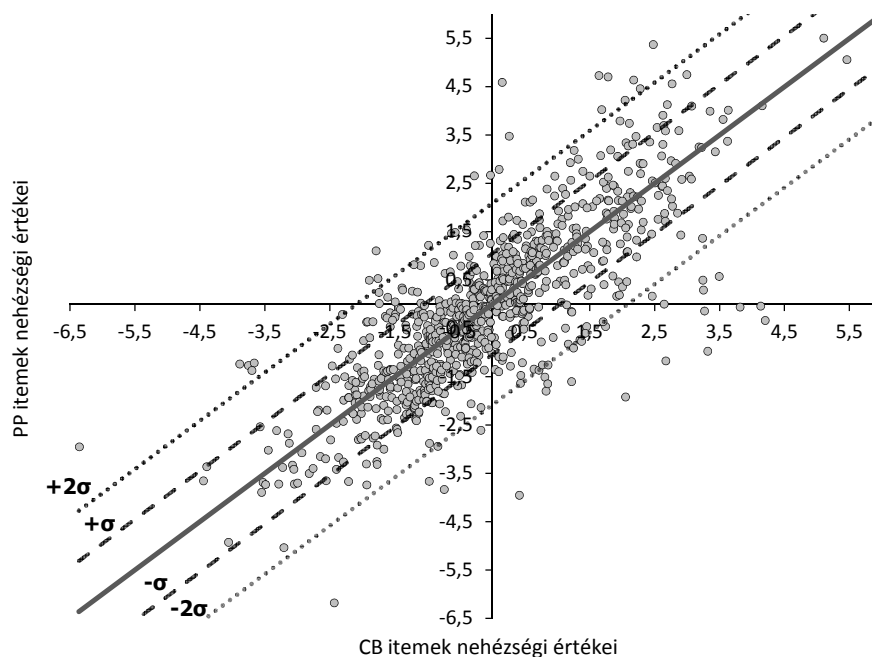
Évfolyam	PP-CB itemnehézségi értékek	
	r	t (p)
1	0,70	1,14 (0,26)
2	0,72	0,90 (0,37)
3	0,69	1,05 (0,30)
4	0,77	0,18 (0,86)
5	0,85	0,03 (0,97)
6	0,92	0,15 (0,90)

Megjegyzés: Minden r érték $p<0,01$ szinten szignifikáns.

Az itemek tipikus médiahatás-független és médiahatás-függő tulajdonságai

A 2. ábra az itemek viselkedését mutatja nehézségi indexük mentén a két médian. Minden alakzat egy itemet reprezentál. Egy item annál könnyebb, minél magasabb a ne-

hézségi indexe és annál nehezebb, minél alacsonyabb. A médiafüggetlen itemek, melyek nehézségi indexét nem befolyásolta jelentős módon a közvetítő eszköz, a szaggatott vonalak között (egy szóráson belül) helyezkednek el. Az itemek 77%-a ebbe a kategóriába sorolható. A szaggatott vonal felett, illetve alatt elhelyezkedő itemek lehetnek felelősek az eltérő viselkedésmintázatokért. A szaggatott vonal felett lévő itemek papír alapon bizonyultak könnyebbnek, míg a szaggatott vonal alattiak számítógép alapon. Ezen itemeket nehézségi indexük közötti különbség szerint tovább csoportosítottuk: ahol a különbség mértéke a különbségek egy és két szórása közötti tartományában van és ahol az eltérés mértéke már meghaladja a két szórás nagyságát.



2. ábra

Az itemnehézségi mutatók változása a közvetítő eszköz függvényében

Az itemek tipikus médiahatás-független és médiahatás-függő tulajdonságainak azonosításához összehasonlítottuk a papír és a számítógép alapon szignifikánsan különböző itemnehézségi indexszel rendelkező ($N=195$) itemek tulajdonságait a médiahatást nem mutató itemek ($N=685$) jellemzőivel. Az egyes itemjellemzők előfordulásának százalékos gyakoriságának különbségeiből alkottuk meg a tipikusan azonosan, illetve eltérően viselkedő itemek profilját. Első és második évfolyamon azon feladatok, amelyekre e tulajdonságok közül legalább egy illett – kombinatorikai tartalmú (26%), magasabb szintű műveleteket tartalmazó (33%), permutálást (27%) igénylő, teljes válasz (34%) –, papír alapon szignifikánsan könnyebbnek bizonyultak. A zárójelben feltüntetett százaléérté-

kek azt mutatják, hogy az adott paraméterrel bíró feladattípus mekkora százalékos aránnyal szerepelt gyakrabban szóráson túl, mint a médiahatást nem eredményező egy szórás értéken belül. Alátámasztva korábbi eredményeinket (Hülber, 2012), a grafikus elemeket tartalmazó (43%) és/vagy alternatív típusú (38%) feladatok számítógépes környezetben nem jelentettek akkora nehézséget a diákok számára, mint papír alapon. Harmadiktól hatodik osztályig a jelentősebb médiahatást mutató feladatok között algebra tartalmú (28%), ismeretet számon kérő (29%), szimbólumokkal dolgozó (47%) és/vagy rutinfeladatok (42%) szerepeltek, melyek megoldása papír alapon könnyebb volt a diákoknak, mint számítógépen. Leginkább médiumfüggetlennek a kiegészítés típusú (1%) feladatok mutatkoznak.

A különböző itemviselkedést befolyásoló formai, tartalmi és technikai jellemzők

A feladatokhoz tartozó karakterek számának médiahatást okozó hatása évfolyamról évfolyamra csökkent. Első évfolyamon még közepes erősségű összefüggés figyelhető meg mind papír, mind számítógép alapon a feladatban szereplő karakterszám mennyisége és a feladat nehézsége között, majd ez fokozatosan csökken, és negyedik évfolyamtól már egyik közvetítő eszköz esetén sem mutatható ki összefüggés a feladat nehézségi szintje és a feladatban található szöveg hossza között. Hasonló jelenség figyelhető meg a feladatokra adandó válaszok hossza és az itemnehézségi paraméterek összefüggése között. Az évfolyamon belüli összefüggések erőssége minden évfolyamon azonos volt, azaz a hipotézisünkkel ellentétben számítógép alapon nem számítanak nehezebbnek a több szöveget és/vagy hosszabb választ igénylő feladatok (l. 7. táblázat).

7. táblázat. A feladatnehézségi értékek és a feladatmegoldáshoz szükséges szövegmenyiség, valamint a feladatra adott válasz terjedelme közötti összefüggés erőssége számítógép és papír alapon

Évfolyam	r _{–karakterszám_nehézségi index}		z (p)	r _{–válaszkarakterszám_nehézségi index}		z (p)
	CB	PP		CB	PP	
1.	0,48**	0,56**	0,94 (0,35)	0,42*	0,34*	0,76 (0,45)
2.	0,26**	0,24**	0,18 (0,43)	0,30*	0,39*	0,86 (0,39)
3.	0,28**	0,20*	0,69 (0,49)	0,23*	0,12	0,33 (0,33)
4.	n.s.	n.s.	n.s.	0,23*	0,17	0,54 (0,59)
5.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
6.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Megjegyzés: * p<0,05 szinten szignifikáns, ** p<0,01 szinten szignifikáns, n.s. nem szignifikáns.

A grafikus elemek teljesítménybefolyásoló hatása kizárólagosan első évfolyamon volt kimutatható, ott könnyebbnek bizonyultak ($|t|=2,01$, $p<0,05$) a grafikus elemeket tartalmazó számítógép alapú feladatok, mint a hasonló tulajdonságokkal bíró papír alapúak. Magasabb évfolyamon a grafikus elem használatának nehézségi indexre gyakorolt hatása

azonosnak bizonyult a két tesztkörnyezetben. A táblázatokat tartalmazó feladatok viselkedése mind átlagosan, mind évfolyamonkénti bontásban médiafüggetlen volt.

A feladatok kontextusa és a feladat megoldásához szükséges pszichikus struktúra mindkét közvetítő eszköz esetében azonos mértékben befolyásolta a feladatok nehézségét. A többszempontos varianciaanalízis csoportosító változói az altípusok és az évfolyam voltak ($F_{\text{kontextus}}=1,07$, $p=0,38$; $F_{\text{pszi.str.}}=1,73$, $p=0,19$). Az első öt évfolyamon a feladatok matematikai tartalom szerinti csoportosításban is azonosan viselkedtek papír és számítógép alapon ($F_{\text{tartalom}}=2,44$, $p=0,23$), azonban a hatodik évfolyamon belépő statisztikai tartalmú feladatok megoldása már nehezebbnek bizonyult számítógépes környezetben, mint papíron ($|t|=2,06$, $p<0,05$).

A feladatokra adandó válasz típusa médiahatás szempontjából meghatározó tényezőnek bizonyult. Az alternatív választást kívánó feladatokat az első három évfolyamon átlagosan könnyebben oldották meg a diákok számítógépen ($|t_{\text{évf1}}|=4,16$, $p<0,001$; $|t_{\text{évf3}}|=4,16$, $p<0,001$; második évfolyamon kevesebb item állt rendelkezésre $N_{\text{alt.vál}}=9$), felsőbb évfolyamokon nem volt kimutatható médiahatás ($t_{\text{évf4}}=-0,64$, $p=0,52$; $t_{\text{évf5}}=0,19$, $p=0,85$; $t_{\text{évf6}}=1,15$, $p=0,26$). Hipotézisünkkel ellentétben a feleletalkotó feladatokban egyik évfolyamon sem kerültek hátrányba a diákok a számítógépes gyakorlottságból eredő különbségek miatt.

A nyílt végű, szövegalkotó választ kívánó feladatokat tovább csoportosítottuk az adandó válasz típusa (szöveg, szám, szimbólum) szerint. A szöveg és a szám válaszként történő bevitele egyik évfolyamon sem jelentett gondot a diákok számára, ám a harmadik évfolyamon belépő szimbólumok használata már igen. Szignifikánsan magasabban teljesítettek a harmadik évfolyamos diákok papír alapon az e típusú feladatokon, mint számítógépen ($|t|=5,15$, $p<0,01$). A különbség negyedik évfolyamon már nem volt detektálható.

A feladatra adandó válaszok hossza (karakterszám) szerinti kategorizáció nem bizonyult médiahatást generáló faktornak. Évfolyamonkénti bontásban különböző mértékben befolyásolta a feladat nehézségét (alsóbb évfolyamokon a hosszabb választ igénylő feladatok tipikusan nehezebbnek bizonyultak, mint felsőbb évfolyamokon, majd ötödik évfolyamtól kezdve semmilyen mértékben nem befolyásolta a feladat nehézségét), de ez a befolyásoló erő azonos volt papír és számítógép alapon.

Az eszközhasználat tekintetében (csak egerhasználatot igényel a válaszolás vagy billentyűzet dominanciájú) az első évfolyamon átlagosan könnyebbek voltak azok a feladatok, ahol a diákoknak elegendő volt egeret használni a feladat megoldáshoz ($|t|=3,17$, $p<0,01$). Magasabb évfolyamokon már nem jelentkezett az eszközhasználat e típusú befolyásoló hatása. Azonban az eredmények alaposabb elemzése – a korrelációs ($r=0,88$) és a parciális korrelációs együtthatók ($r=0,65$, $p<0,01$) összevetése – rávilágított arra, hogy ez a hatás a feladattípus befolyásoló hatása, miután tipikusan az alternatív válaszlehetőségeket felkínáló feladatban elegendő a csak eger használata, azaz a válaszadáshoz szükséges eszközhasználat tekintetében sem mutatható ki különbség a papír és a számítógép alapú eredmények között.

Az eredmények értelmezése

A kutatásban alkalmazott tesztek és feladatbank szintű jószágmutatók értékei alapján a számítógép alapú tesztelés kisiskolás korban is megbízhatóan alkalmazható, illetve átlagosan a számítógép alapú tesztek jószágmutatói magasabbnak bizonyultak minden évfolyamon. Nemcsak a papír, hanem a számítógép alapú eredmények és az azokból levonható következtetések is általánosíthatóak. Ugyanakkor a papír és a számítógépes környezetben nyújtott teljesítmények felcserélhetősége csak abban az esetben elfogadott, ha (1) a különböző médiumokon elért pontok sorrendje közelít egymáshoz vagy (2) az átlagok, a szórások és a pontszámok eloszlása közel azonos vagy (3) mindez újraskálázással elérhető (AERA, APA és NCME, 1999). A bemutatott kutatásban egyik évfolyamon sem volt kimutatható átlagos szignifikáns teljesítménybeli különbség a papír, valamint a számítógépen megoldott tesztek vonatkozásában. Azonban a feladatsztű, már részletesebb elemzés rávilágított az évfolyamok között tapasztalható viselkedésbeli különbségekre.

Már első évfolyamon is magas volt a korrelációs ($r=0,70$) együtttható értéke a papír és a számítógép alapú teszteredmények között. A diákok teljesítményét alapvetően nem határozta meg a közvetítő eszköz típusa, pedig a tanulók ebben az évben kezdték meg tanulmányaikat, jellemzően nem gépesített tanulási-tanítási környezetük. Legtöbbjüknek ez volt az első számítógép alapú tesztelése, sőt sokak jelen kutatásban találkoztak először számítógéppel. A papíron és számítógépen elért eredmények közötti korreláció fokozatosan erősödött, majd ötödik évfolyamon már $r=0,85$ -ös, hatodik évfolyamon $r=0,92$ -es értéket vett fel. Ez alapján megállapítható, hogy a felső tagozatosok esetében – feladattípustól, feladatjellemzőtől függetlenül – már egyértelműen kijelenthető az eredmények felcserélhetősége. Ez a megállapítás egyrészt egybecseng Poggio és munkatársai (2005) kutatási eredményeivel, akik 7. évfolyamos diákok eredményei alapján $r=0,96$ kapcsolatot állapítottak meg a diákok papír és számítógépen elért teszteredményei között. Az évfolyamonkénti bontásban kimutatott tendencia illeszkedik Choi és Tinkler (2002), illetve Oregon Department of Education (2007) elemzéseire, akik amellet, hogy évfolyamonként növekvő korrelációs együttthatókat azonosítottak, felhívták a figyelmet a kisiskolás korosztályban esetlegesen előforduló eltérő viselkedésmintázatokra.

Várhatóan a kor előrehaladtával, a technológiai eszközök oktatási-tanulási folyamatba való integrációjával és elterjedtségével, valamint hazánkban az informatika tantárgy kötelező megjelenésével párhuzamosan már kisgyermek korban sem jelent plusz kihívást a számítógép alapú feladatmegoldás (Oregon Department of Education, 2007). A fent említett különbségek végleg eltűnnek, és a tradicionális papír alapú tesztelés egyértelműen kiváltható a számos előnnyel bíró számítógép alapú teszteléssel.

Az itemtípusok teljesítménybefolyásoló szerepére fókuszáló eredmények hiánypótlónak számítnak, ugyanis mind feleletválasztó (zárt), mind feleletalkotó (nyitott) típusú feladatokon mutatott viselkedésmintázatokat is összehasonlítottunk az elemzések során. Az amerikai standardizált tesztelési hagyományok miatt ugyanis nemzetközi szinten is kevés információ áll rendelkezésre a nyílt végű, szövegalkotó feladattípusok számítógép alapú kiközvetítésével kapcsolatban (Bennett, 2003). Ennek oka, hogy a feleletválasztó feladatok előnyét kihasználva automatizálható javításuk, az eredmények azonnal rendelkez-

kezésre állnak, nem sérül az értékelés objektivitása, illetve a kutatások fókuszja közel kizárólagosan a feleletválasztó feladatokra tevődött.

A feleletválasztós feladatokon számítógépes környezetben magasabb teljesítményt értek el az alacsonyabb évfolyamos diákok, mint papír alapon. Ennek oka lehet, hogy ebben az életkorban még eltérő megoldási stratégiákat alkalmaznak a diákok a két környezetben. A feladatok üresen, válaszolatlanul hagyása papír alapon gyakoribb, mint számítógépen, ahol inkább tippelnek, mint válasz nélkül mennek tovább (*Johnson és Green, 2006*). A nagyobb válaszadási hajlandóság más-más stratégiák alkalmazására utal. A papír alapú környezetet erősebben kötik a tétellel bíró iskolai tesztekhez, ezért a kevésbé tudatos kognitív folyamatok a korábbi iskolai tapasztalatok miatt papír alapú környezetben (*William, 1999*) más stratégiák követését idézik elő.

A nyílt végű, feleletalkotó feladatok számítógépes megjelenítése, a válaszok billentyűzet segítségével történő bevitele sem okozott szignifikáns teljesítménycsökkenést előidéző különbséget a már megszokott papír alapon ceruzával, tollal történő válaszadáshoz képest. A jelentkező médiahatást nem a beviteli eszköz kezelésének problémája, hanem a feladat megoldásához szükséges műveletek komplexitása okozta. A bonyolultabb, több jegyzetelést, köztes lépések felvázolását kívánó kombinatorikai tartalmú, magasabb szintű műveleteket és/vagy permutálást igénylő feleletalkotó feladatok ezért papír alapon könnyebbnek, számítógépen nehezebbnek bizonyultak. Ennek oka, hogy számítógépen a diákok több műveletet próbálnak fejben elvégezni, hiába áll esetleg rendelkezésükre jegyzeteléshez papír (*Johnson és Green, 2006*). Az ilyen viselkedésmintázat okozhatta a fent jellemzett típusú feladatoknál a kutatásban tapasztalt teljesítménybeli különbségeket.

Összességében a feladattípusokra vonatkozó elemzések is alátámasztják korábbi megállapításunkat, miszerint a papír alapú tesztelés biztonsággal kiváltható számítógép alapú teszteléssel. Ha a tesztelés, hasonlóan a korábbi papír alapú teszteléshez, változatos feladatformákat tartalmaz, akkor az egyes médiumokon elért teszteredményekben nem tapasztalható eltérés, viszont ha csak zárt, vagy csak nyitott, feleletalkotó feladatokat, vagy kizárólag egyszerű vagy csak bonyolult, magasabb rendű műveleteket kívánó feladatokat alkalmazunk számítógépen, akkor érdemes külön figyelmet fordítani a teljesítmények médiafüggő változására.

A feladatok médiafüggő viselkedését befolyásoló további technikai, formai tényezők közé soroltuk például a feladatok hosszát, a grafikai elemek megjelenését vagy a rögzítendő válasz típusát. A feladatok hossza, a feladatokhoz tartozó karakterek számának médiahatást generáló ereje évfolyamról évfolyamra csökkent. Ennek egyik oka lehet, hogy alsóbb évfolyamokon még problémát jelenthet, ha a feladat teljes egészében nem fér rá egy képernyőképre, hanem görgetést igényel. Ennek jelentősége egyrészt a technológiai jártasság miatt évfolyamról évfolyamra csökken, másrészt feltételezhetjük, hogy az alacsonyabb évfolyamokon a rövid távú memória jelentősebb (OECD, 2012) teljesítménybefolyásoló erővel bír, mint magasabb évfolyamokon, mert azt még a kevésbé automatizált olvasás nyelvi elemei foglalhatják le. Ezért érdemes alsóbb évfolyamokon a görgetést igénylő feladatokat mellőzni.

A grafikus elemeket, színes képeket tartalmazó feladatok megoldása számítógépen könnyebb volt, mint papír alapon. Ez, hipotézisünk szerint, a diákok feladatmegoldó motivációjának növekedésével magyarázható. A színesebb, kontrasztosabb, ezért életsze-

rúbb feladatokat szívesebben oldják meg, mint a hasonló, grafikus elemeket tartalmazó, de fekete-fehér, szürkébb feladatokat. Ez a jelenség abban az esetben is megfigyelhető volt, amikor a képnek a megoldás szempontjából nem volt szerepe, csak színesítette a feladatot.

Ha a feladatra adandó válasz különböző szimbólumokat (például relációs jelek, zárójelek, százalékjel) is tartalmazott, azok billentyűzettel, illetve egér segítségével történő rögzítése problémát jelentett a diákoknak. Ennek oka, hogy ezek bevitele különböző billentyűkombinációkkal lehetséges, amelyek ismerete és használata nem egyértelmű a gépelésben tapasztalatlan diákok számára. Ebben az esetben, meglátásunk szerint, megoldást jelenthet a szimbólumok bevitelét biztosító dedikált ikonok használata.

Összefoglalás

A technológia korábban nem tapasztalt lehetőségeket biztosít a mérés-értékelés folyamatában. Ahogy fokozatosan mindennapi életünk részévé vált és válik, a tanulási és tanítási tevékenységekben is egyre nagyobb szerepet tölt be, használata már kisiskolás korban is természetessé válik. Ezzel párhuzamosan fokozatosan háttérbe szorulnak a tradicionálisnak számító tanulási-tanítási módszerek, beleértve a hagyományosnak nevezhető papír alapú tesztelést is. Megjelent az igény (Lent, 2009) az innovatívabb, több lehetőséget biztosító, a diákok számára motiválóbb környezetet adó számítógép alapú mérési-értékelési rendszerek kidolgozására. Azonban a továbblépéshez, a számítógépes tesztelés pedagógiai alkalmazásához ki kell szűrni a nemkívánatos mellékhatásokat (Csapó, Molnár és R. Tóth, 2008).

Jelen kutatás tág életkori intervallumban, 1–6. évfolyamos diákokra fókuszálva, nagymintás adatfelvétel eredményeire alapozva elemezte és hasonlította össze a mind papíron, mind számítógépes formában is megjeleníthető feladatokon nyújtott teljesítményeket. Az eredmények értelmében a számítógép alapú tesztek összességében legalább annyira megbízhatóak, mint a papír alapú tesztek. Változatos feladatformák alkalmazásával biztosítható a papír és számítógép alapú teszteredmények számszerű felcserélhetősége, miután a feladatokon belül megjelenő médiahatás már teszt szintjén kiegyenlítődik. Azonos típusú feladatok alkalmazása nagyobb odafigyelést és az eredmények transzformációját, újráskálázását igényli, miután abban az esetben halmozottan jelentkezik a számítógép alapú tesztelés során tapasztalt plusz tesztmegoldási motiváció, ami magasabb teljesítménnyel is párosul, vagy esetleg egy konkrét matematikai alapú problémakört kiemelve, a papír alapú számoláshoz képest a fejben számolás stratégiájának alkalmazása, ami alacsonyabb teljesítménnyel jár együtt. Nem definiálható olyan feladatparaméter, ami egyöntetűen jelentős mértékű médiahatást eredményezett volna, ezért használata korlátozásra szorulna. A kutatás eredményei szerint egyedül a kisiskolás diákok körében alkalmazott tesztek igényelhetnek külön figyelmet a technológiai jártasság és az átállás jelen fázisában.

A kutatást a TÁMOP 3.1.9/11 kutatási program és az Oktatásméleti Kutatócsoport támogatta.

Irodalom

- AERA, APA és NCME (1999). *Standards for educational and psychological testing*. American Educational Research Association, Washington, D.C.
- Applegate, B. (1993): Construction of geometric analogy problems by young children in a computer-based test. *Journal of Educational Computing Research*, **9**. 1. sz. 61–77.
- Becker, J. (2004): *Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT)*. Digitális disszertáció. Freie Universität, Berlin
- Bennett, R. E. (2003): *Online assessment and the comparability of score meaning*. 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. 4–38.
- Bjerkestrand, O. (2009): The European coherent framework of indicators and benchmarks and implications for computer-based assessment. 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, Luxembourg. 24–29.
- Blazek, N. L. és Forbey, J. D. (2011): A comparison of validity rates between paper and pencil and computerized testing with the MMPI-2. *Assessment*, **18**. 63–66.
- 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.
- Choi, S. W. és Tinkler, T. (2002): Evaluating comparability of paper and computer based assessment in a K-12 setting. Előadás. Annual Meeting of the National Council on Measurement in Education, 2002. április 15., New Orleans, Amerikai Egyesült Államok.
- Csapó Benő, Lőrincz András és Molnár Gyöngyvér (2012): Innovative assessment technologies in educational games designed for young students. In: Ifenthaler, D., Eseryel, D. és Ge, X. (szerk.): *Assessment in game-based learning: foundations, innovations, and perspectives*. Springer, New York. 235–254.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papír alapú teszteltől a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 3–4. sz. 3–16.
- Csapó, B., Ainley, J., Bennett, R., Latour, T. és Law, N. (2012): Technological issues of computer-based assessment of 21st century skills. In: McGaw, B. és Griffin, P. (szerk.): *Assessment and teaching of 21st century skills*. Springer, New York. 143–230.
- Csíkos Csaba és Csapó Benő (2011): A diagnosztikus matematika felmérések részletes tartalmi kereteinek kidolgozása: elméleti alapok és gyakorlati kérdések. In: Csapó Benő és Szendrei Mária (szerk.): *Tartalmi keretek a matematika diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 141–168.
- Farcot, 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. (2000): *The effect of computer-based tests on racial/ethnic, gender and language groups* (GRE Board Professional Report No. 96–21P). Educational Testing Service, Princeton, NJ.
- Hetter, R. D., Segall, D. O. és Bloxom, B. M. (1994). A comparison of item calibration media in computerized adaptive testing. *Applied Psychological Measurement*, **18**. 3. sz. 197–204.
- Hülber László (2012): A papír és a számítógép alapú tesztelés összehasonlító vizsgálata különböző item paraméterek mentén. *Iskolakultúra*, **12**. 12. sz. 13–26.

- Ito, K. és Sykes, R. C. (2004): Comparability of scores from norm-reference paper-and-pencil and web-based linear tests for grades 4–12. Előadás. Annual meeting of the American Educational Research Association, 2004. április 12–16., San Diego, Amerikai Egyesült Államok.
- 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. 4–33.
- Kim, D. és Huynh, H. (2006): Comparison of student performance between paper-and-pencil and computer-based testing in four content areas. Előadás. Annual meeting of National Council on Measurement in Education, 2006. április, San Francisco, Amerikai Egyesült Államok.
- Kim, D. és Huynh, H. (2007): Comparability of computer and paper-and-pencil versions of Algebra and Biology assessments. *Journal of Technology, Learning and Assessment*, **6**. 4. sz. 4–29.
- Kim, J. P. (1999). Meta-analysis of equivalence of computerized and P&P tests on ability measures. Előadás. Annual Meeting of the Mid-Western Educational Research Association, 1999. október 13–16., Chicago, Amerikai Egyesült Államok.
- Kingery, D. és Furuta, R. (1997): Skimming electronic newspaper headlines: A study of typeface, point size, screen resolution, and monitor size. *Information Processing and Management*, **33**. 685–696.
- Kingston, N. M. (2009): Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. *Applied Measurement in Education*, **22**. 1. sz. 22–37.
- Lent, v. 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, Luxembourg. 83–91.
- Lottridge, S. M., Nicewander, W. A., Schulz, E. M. és Mitzel, H. C. (2010): Comparability of paper-based and computer-based tests: A review of the methodology. In: Winter, P. C. (szerk.): *Evaluating the comparability of scores from achievement test variations*. Council of Chief State School Officers, Washington. 119–152.
- Magyar Andrea és Molnár Gyöngyvér (2013): Adaptív és rögzített formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*, **113**. 3. sz. 181–193.
- Martins, N. (2010): Measurement model equivalence in web- and paper-based surveys. *Southern African Business Review*, **14**. 3. sz. 77–107.
- Mayrath, M. C., Clarke-Midura, J. és Robinson, D. (2012): *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Information Age, Charlotte.
- Mazzeo, J. és Harvey, A. L. (1988): *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature*. (College Board Report 88-8). College Entrance Examination Board, New York.
- Mead, A. D. és Drasgow, F. (1993): Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, **3**. 114. sz. 449–458.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, 7–8. sz. 22–34.
- Molnár Gyöngyvér (2013a): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér (2013b): Területspecifikus komplex problémamegoldó gondolkodás fejlődése. In: Molnár Gyöngyvér és Korom Erzsébet (szerk.): *Az iskolai sikerességet befolyásoló kognitív és affektív tényezők értékelése*. Nemzedékek Tudása Tankönyvkiadó, Budapest. 161–180.
- Molnár Gyöngyvér és Csapó Benő (2013): Az eDia online diagnosztikus mérési rendszer. Előadás. XI. Pedagógiai Értékelési Konferencia, Szeged, 2012. április 11–13.
- Molnár, G. és Lőrincz, A. (2012): Innovative assessment technologies: Comparing ‘face-to-face’ and game-based development of thinking skills in classroom settings In: Chen, D. (szerk.): *International*

- Proceedings of Economics Development and Research. Management and Education Innovation*. Vol. 37. IACSIT Press, Singapore. 150–154.
- OECD (2012): PISA 2015 Proposed screen design for computer-based cognitive items. OECD, Kézirat.
- Parshall, C. és Kromrey, J. D. (1993): Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect. Előadás, Annual Meeting of the American Educational Research Association, 1993. április, Atlanta, Amerikai Egyesült Államok.
- Peak, P. (2005): Recent trends in comparability studies. Pearson Educational Measurement. http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE-8484-07AACAE2DA75/0/TrendsCompStudies_r0505.pdf. Utolsó letöltés: 2013. szeptember 16.
- Pearson Educational Measurement (2003): *Virginia standards of learning web-based assessments comparability study report – Spring 2002 administration: Online & paper tests*. Pearson, Austin.
- 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 stat assessment program. *Journal of Technology, Learning and Assessment*, **3**. 6. sz. 4–29.
- 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.
- 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. 4–19.
- R. Tóth Krisztina és Hódi Ágnes (2011): Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasásszövegértés területén. *Magyar Pedagógia*, **111**. 4. sz. 313–332.
- R. Tóth Krisztina, Molnár Gyöngyvér, Thibaud Latour és Csapó Benő (2011): Az online tesztelés lehetőségei és a TAO platform alkalmazása. *Új Pedagógiai Szemle*, **61**. 1–2–3–4–5. sz. 8–22.
- Russell, M. (1999): Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives*, **7**. 20. sz. 1–47.
- Sandene, B., Bennett, R. E., Braswell, J. és Oranje, A. (2005): Online assessment in mathematics. In: Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan, B. és Oranje, A. (szerk.): *Online assessment in mathematics and writing: Reports from the NAEP technology-based assessment project* (NCES 2005-457). National Center for Education Statistics, US Department of Education, Washington, DC.
- Schroeders, U. és Wilhelm, O. (2011): Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, **71**. 5. sz. 849–869.
- Texas Education Agency. (2008): *A review of literature on the comparability of scores obtained from examinees on computer-based and paper-based tests*. Texas Education Agency (TEA) Technical Report Series, Texas.
- Vidákovich Tibor (2012): A feladatok paraméterezése. Kézirat.
- Wang, S. és Shin, C. D. (2009): Comparability of computerized adaptive and paper-pencil tests. *Test, Measurement & Research Service. Bulletin*, **13**. sz. 1–7.
- 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., 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.
- Waters, S. D. és Pommerich, M. (2007): Context effects in internet testing: A literature review. Előadás. 22nd Annual Conference of the Society for Industrial and Organizational Psychology, 2007. április 7., New York City, Amerikai Egyesült Államok.
- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): Score comparability of online and paper administrations of Texas assessment of knowledge and skills. Előadás. Annual meeting of National Council on Measurement in Education, 2006. április, San Francisco, Amerikai Egyesült Államok.

Wiliam, D. (1999): The half-second delay: what follows? Előadás. European Conference on Education Research, 1999. szeptember, Lahti, Finnország.

Zhang, L. és Lau, C. A. (2006): A comparison study of testing mode using multiple-choice and constructed-response items – Lessons learned from a pilot study. Előadás. AERA, 2006. április 7–11., San Francisco, Amerikai Egyesült Államok.

ABSTRACT

LÁSZLÓ HÜLBER AND GYÖNGYVÉR MOLNÁR: COMPARING PAPER-AND-PENCIL AND ONLINE TEST PERFORMANCE IN MATHEMATICS: FINDINGS FROM A LARGE-SCALE ASSESSMENT IN YEARS 1 TO 6

There is little doubt today that a sizeable percentage of educational assessment is computer-based (CB). However, when computer-based assessment replaces paper-and-pencil (PP) testing, a number of questions arise regarding issues of equivalence. This paper compares results from PP and CB testing to identify domains and item formats where the two media may influence achievement. Mathematics tests comprising various item types and connected by anchor items were administered in PP and CB modes to six age groups from Years 1 to 6 in Hungarian schools (N=40 571 and 21 895, respectively). Online data collection was carried out on the eDia platform. The internal consistencies of the tests were good: Cronbach α was over .86 and .91 in PP and CB modes, respectively. Strong correlations were found between the total scores on the two versions of the test, and they showed an increasing trend over time, indicating that paper- and computer-based test performances become more similar with age ($r_{\text{grade } 1}=.70$; $r_{\text{grade } 6}=.92$). This paper argues that the media effect is related to the item format, type, complexity, length and content used on the tests; however, no single parameter can be identified generally, which could have resulted in a steadily significant media effect so that its use should be restricted. The children in the lower years performed higher on multiple-choice items in the CB environment because of the higher motivation for testing that resulted in less missing data. The average scores on open-ended CB tasks proved to be lower for items requiring calculation and complex operations and/or higher-level thinking skills. The length of the tasks, specifically scrolling through texts to find an answer, had an effect on the learners in the lower age groups, but if the tasks contained colourful pictures test-takers achieved higher scores on CB than on PP. Results indicate that if the test contains various item types, formats, contexts and levels of complexity, no derivation can be experienced in the test results on the different mediums. If the test contains similar items, e.g. only simple closed or open-ended tasks without any illustration, we should pay special attention to changes in performance depending on the media.

Magyar Pedagógia, **113**. Number 4. 243–263. (2013)

Levelezési cím / Address for correspondence:

Hülber László, MTA-SZTE Képességfejlesztés Kutatócsoport, H-6722 Szeged, Petőfi S. sgt. 30–34.

Molnár Gyöngyvér, SZTE Neveléstudományi Intézet, MTA-SZTE Képességfejlesztés Kutatócsoport, H-6722 Szeged, Petőfi S. sgt. 30–34.