

PAPÍR- ÉS SZÁMÍTÓGÉP-ALAPÚ TESZTELÉS ÖSSZEHASONLÍTÓ VIZSGÁLATA PROBLÉMAMEGOLDÓ KÖRNYEZETBEN

Molnár Gyöngyvér

SZTE Neveléstudományi Intézet, MTA-SZTE Képességkutató Csoport

Az utóbbi néhány évben számos mérés-értékeléssel foglalkozó nemzetközi szervezet és intézet (pl.: IQB, CITO, ETS, NCES, OECD) egyik fő kutatási területévé vált (l. Intel, Microsoft és Cisco Education Taskforce, 2008) a technológia alapú mérés-értékelés bevezetése, oktatási folyamatba történő integrálása. A lehetőségek minél szélesebb körben történő kihasználása, a papír-alapú tesztelésről való áttérés azonban csak fokozatosan lehetséges (Csapó, Molnár és R. Tóth, 2009). A jelen kutatás e folyamatba illeszthető, célja a mérés-értékelés médiuma befolyásoló hatásának diagnosztizálása különböző változók mentén problémamegoldó környezetben.

A technológia mérés-értékelésben betöltött szerepe

A technológia mérés-értékelésben betöltött szerepe sokrétű lehet, alkalmazásának relevanciája függ a vizsgált területtől, az adott területen a diákok technológiahasználati szokásaitól, csakúgy, mint a feladatírók, adatelemzők IKT-s kompetenciáitól. Ebből adódóan nincs egységes definíció arra vonatkozólag, hogy mit jelent a technológia-alapú mérés-értékelés. Attól függően, hogy a mérés-értékelés folyamatának melyik stádiumában, szintjén, vagy szintjein jelenik meg, továbbá a folyamatban résztvevők közül kik és mire használják az adott eszközt – aminek milyensége szintén definiáló erővel bírhat –, különböző definíciókat fogalmazhatunk meg.

A hagyományos papír-ceruza adatfelvétellel történő vizsgálatokban is már nélkülözhetetlen szerepet tölt be a számítógép. Általában azon történik a feladatok végső formába öntése, az adatok rögzítése, elemzése, a visszajelzések elkészítése stb. Ennek ellenére ezeket a méréseket a továbbiakban nem tekintjük technológia alapúnak. Technológia alapú mérés-értékelésről abban az esetben beszélünk, ha maga az adatfelvétel során is és nemcsak előtte, illetve utána kap fontos szerepet a technológia.

A papír alapú teszteléstől a technológia alapú tesztelésre történő áttérés lehetőségei

A technológia alapú mérés során leggyakrabban alkalmazott információs-kommunikációs eszköz a számítógép. A számítógépes tesztelés megvalósításának legegyszerűbb formája, amikor statikus, papír-ceruza formában is közvetíthető feladatokat oldatunk meg a diákoknak. Ennek egyik módja: a meglévő papír-ceruza tesztek eredetivel megegyező formában történő digitalizálása. Ebben az esetben a hagyományos tesztelés a feladatokat közvetítő eszközben különbözik a számítógép alapú teszteléstől, azaz a teszt feladatai papír helyett a képernyőn jelennek meg. Az adatfelvétel során a feladatokra adott válaszokat a diákok billentyű, egér, esetleg érintőképernyő segítségével adják meg (R. Tóth, Molnár és Csapó, 2008).

A tesztelés menete – a hagyományos teszteléshez hasonlóan – maradhat lineáris, azaz a feladatok azonos sorrendben jelennek meg minden egyes tesztelt személy előtt, vagy a teszt típusától függően random is közvetíthetőek ugyanazon feladatok. Utóbbi esetben az elemzések során figyelembe kell venni, hogy megváltozik minden egyes item tesztbéli pozíciója és környezete. Egy köztes megoldás, amikor az itemekből előre meghatározott klasztereket képzünk, amelyen belül az itemek helye rögzített, majd az itemek előzetes

csoportosítása után a klaszterválasztást randomizáljuk. Ebben az esetben biztosítható, hogy a tesztelés során mindenki azonos nehézségű, de különböző feladatokat kapjon.

A számítógépes tesztelés bevezetésének itemek tekintetében egy következő lépcsőfoka, amikor fokozatosan kihasználjuk a technológia adta multimédiás (hang, mozgó kép, animáció, szimuláció, interaktív szimuláció stb.) lehetőségeket, ezáltal gazdagítva az alkalmazott itemek típusát (Csapó, Molnár, Pap-szigeti és R. Tóth, 2009). A fokozatosság elvénél maradván a teljesen statikus itemeket fokozatosan felválthatják a dinamikus, interaktív, real-time és real-life itemek, amelyek papír alapú megjelenítése nem lehetséges.

A feladatok egy másik típusú dinamikusságát adhatja az automatikus itemgenerálás lehetősége, aminek következtében bizonyos típusfeladatok mindig új formában jelenhetnek meg a diákok előtt (például a szöveges feladatokban mindig más-más számértékek szerepelnek).

A számítógépes tesztelés lehetőségeinek maximális kihasználását biztosíthatja a számítógépes adaptív tesztelés. Ebben az esetben egy teljes mértékben parametrizált feladatbank áll a tesztelés háttérében, a feladatok kiválasztása egyedi, a vizsgázó korábbi válaszainak függvényében történik. Ennek következtében az adatfelvétel során minden tanuló többségében a saját képességszintjének megfelelő feladatokat old meg. A számítógépes adaptív tesztelés (Computerized Adaptive Testing – CAT; Wainer, 2000; Becker, 2004; Eggen és Straetmans, 2009) a teljesítmények pontosabb meghatározását teszi lehetővé, miután megvalósítja a személyre szabott mérés-értékelést. Minden egyes diáknál a tesztelési folyamat során megoldandó itemek a számára optimális nehézséggel bíró itemek (a flow-élmény, l. Csíkszentmihályi, 1997) sávjában marad. Mindez előnyösen hat a diákok motivációjára, aminek a tesztek gyakori alkalmazásánál meghatározó jelentősége van (Csapó, Molnár és R. Tóth, 2008).

A technológia adatfelvétel során történő alkalmazása lehetőséget biztosít a diákok konkrét válaszáin kívül további adatok gyűjtésére is. A logfájlok segítségével információt kaphatunk többek között arról, hogy az egyes itemek megoldásához a diákoknak mennyi időre volt szüksége, rögzíthetjük az egér mozgását, a billentyűk lenyomása között eltelt időt, a kattintások számát, amelyek további adatokat szolgáltatnak a tesztelés menetéről és a diákok képességszintjéről (Csapó, Molnár és R. Tóth, 2008).

A közvetítő eszköz befolyásoló hatása

A 21. században már nem kérdés többé, hogy a technológia által nyújtott lehetőségek rövid időn belül megváltoztatják az oktatás és értékelés folyamatát. A változások egyik, eddig leghangszúlyosabb iránya (1) a hagyományos mérés-értékelés hatékonyabbá tétele (Bridgeman, 2009 idézi Csapó, Latour, Bennett, Ainley és Law, 2009), egy (2) másik megközelítés a mérőeszközök autentikusabbá tétele (Pellegrino, Chudowosky és Glaser, 2004), (3) harmadrészt a technológia lehetővé teszi olyan képességek mérését, amelyek papír-alapú mérése nem, vagy nagyon nehezen megvalósítható (pl.: technológiailag gazdag környezetben a problémamegoldó gondolkodás vizsgálata, l. Bennett, Persky, Weiss és Jenkins, 2007; OECD, 2008); (4) lehetővé teszi a diák és a teszt közötti dinamikus interakció megvalósulását (Csapó, Latour, Bennett, Ainley és Law, 2009); továbbá (5) a logfájlok elemzése lehetőséget biztosít a mérés dinamikájának azonosítására (pl.: hányszor módosította a diák a választát, mennyi idő kellett az egyes feladatok megoldásához, milyen úton jutott el a válaszig; l. Ainley, 2006; Hadwin, Wynne és Nesbitt, 2005; R. Tóth, 2009).

A technológia alapú mérés-értékeléssel foglalkozó publikációk döntő többsége a fent nevezett lehetőségek ellenére még a technológia adatfelvétel során történő alkalmazásának lehetőségeit, a papír-alapú méréshez viszonyított relevanciáját, összehasonlíthatóságát vizsgálják. Számos, a közvetítő eszközre vonatkozó kérdést vetnek fel (pl.: Ugyanazt méri-e a

papír-ceruza, mint a számítógépes tesztelés? Összehasonlíthatóak-e egymással ugyanazon teszt papír és számítógép alapú változatának eredményei? Milyen média típusával összefüggő tényezők befolyásolják az eredményeket, ha azonosíthatók ezek?), amelyek mind a közvetítő eszköz befolyásoló szerepét vizsgálták különböző kontextusban, különböző itemformátumok, különböző kutatási elrendezés, mintavétel alkalmazásával. A kutatások céljából adódóan a kutatások tervezői törekedtek arra, hogy az összehasonlítandó tesztek közel azonosak legyenek mind papír, mind számítógépes formátumban. Ebből adódóan az alkalmazott itemformátumok döntő többsége feleletválasztós (multiple choice) item volt.

A közvetítő eszköz teljesítményre gyakorolt hatásával kapcsolatos kutatások többsége területtől, kontextustól függetlenül nem mutatott ki jelentős teljesítménybeli eltérést a két médián mutatott teljesítmény között. A vizsgálatok között számos kis és nagymintás mintavétel is található (l. Wang, Jiao, Young, Brooks és Olson, 2007, 2008; Singleton, 2001; Csapó, Molnár és R. Tóth, 2009; Walt, Atwood és Mann, 2008; Kim és Huynh, 2008; Puhan, Boughton és Kim, 2008; Horkay, Bennett, Allen, Kaplan és Yan, 2006; Johnson és Green, 2006; Poggio, Glasnapp, Yang és Poggio, 2005). Ezen egybehangzó eredmények ellenére a kutatások kisebb része utal például az alkalmazott itemtípus meghatározó szerepére (Johnson és Green, 2006), vagy szintén az ezzel szorosan összefüggő teljesítménybeli különbségekre [általában a számítógép alapú tesztverzió bizonyult nehezebbnek, főképp azokon a területeken, ahol nagyobb hatást tudott gyakorolni a két médián történő olvasási képesség fejlettségi szintje közötti különbség, l. Bennett és mtsai (2008); vagy a szövegalkotásbeli eltérés Horkay, Bennett, Allen, Kaplan és Yan (2006)]. A leggyakrabban vizsgált háttérváltozók közül a számítógéphasználat gyakorisága, fejlettségi szintje általában nem bizonyult erős teljesítménymódosító tényezőnek (l. pl.: O'Dwyer, Russell, Bebell és Tucker-Seeley, 2008).

A kutatások egy másik része nemcsak a közvetítő médiát, hanem annak jellemző tulajdonságait, azok tesztelést és a tesztelés eredményének befolyásoló szerepét is összehasonlította [pl.: a monitor mérete, felbontása mennyire meghatározó, l. Bridgeman, Lennon és Jackenthal (2003) illetve McDonald (2002)].

A továbbiakban kitérünk egy olyan terület számítógép-alapú vizsgálatának lehetőségeinek ismertetésére, ahol a terület adottságai és az új mérési lehetőségek gyökeresen megváltoztatják a korábbi, hagyományos, területre vonatkozó mérés-értékelés modelljeit.

A problémamegoldó gondolkodás számítógép alapú nemzetközi vizsgálatai

A problémamegoldó gondolkodás vizsgálati lehetőségeit jelentős mértékben megváltoztatta és kitágította a technológia mérés-értékelésbe történő integrálása. A terület adottságai lehetővé teszik a technológia adta lehetőségek teljes spektrumának kihasználását. Az információs és kommunikációs technológiák adatfelvétel során történő alkalmazása megteremti a lehetőséget a problémamegoldás dinamikus, komplex és életszerű mérésére. Ennek következtében a nemzetközi szervezetek és nagy intézetek empirikus vizsgálataiban is megjelenik a problémamegoldó gondolkodás különböző dimenzióinak számítógép alapú mérése, fokozatosan eltávolodva a papír-alapú vizsgálatoktól.

A teljesség igénye nélkül például az OECD méréseiben, ahol a nemzetközi vizsgálatok 2006-ig kizárólag papír alapúak voltak, a tervek szerint 2012-ben az „Általános problémamegoldás” (General Problem Solving) számítógépes felmérésére kerül sor (Csapó, 2009). A megvalósítási lehetőségek között felmerült a dinamikus problémamegoldás (l. Blech és Funke, 2005; Wirth és Funke, 2005; Greiff és Funke, 2008), melynek során a diákoknak a számítógéppel való interakció révén azonosítani kell egy véges állapotú automata viselkedési szabályait, majd az így megszerzett tudást különböző feladatok megoldásában kell hasznosítani (Csapó, 2009).

Szintén az OECD szervezésében jelenleg előkészületi fázisban van a felnőttkori készségek vizsgálatára irányuló, 2011-re tervezett felmérés, a PIAAC (Programme for the International Assessment for Adult Competencies). A mérésben 27 ország 16 és 65 év közötti polgárai vesznek részt, minimum 5000 ember országonként (OECD, 2008). Három fő műveltségi területen vizsgálják a felnőttek képességeinek fejlettségi szintjét: szövegértés és használat, számolás és technológiailag gazdag környezetben végzett problémamegoldás. A szóbeli interjúkat kérdezőbiztosok végzik, a tesztek a TAO (Testing Assisté par Ordinateur – számítógépes tesztelés) platform (Plichart, Jadoul, Vandenabeele és Latour, 2004; Latour és Farcot, 2008) közvetíti.

Németországban a DIPF TBA (Deutsches Institut für Internationale Pädagogische Forschung Technology Based Assessment) projekt, illetve a dinamikus változó környezetben alapuló MicroDYN megközelítés említendő (Greiff és Funke, 2008), utóbbi várhatóan a PISA 2012-es PS modul alapját is képezi. Ennek előzménye volt a 2003-ban német opcióként választható analitikus és dinamikus problémamegoldás modul (Wirth és Klieme, 2003).

Angliában a NAEP (National Assessment of Educational Progress) kutatások keretein belül került sor a problémamegoldás technológiailag gazdag környezetben történő mérésére (PS in TRE – Problem Solving in Technology-Rich Environments; Bennett, Persky, Weiss és Jenkins, 2007).

A problémamegoldó gondolkodás hagyományos és számítógép alapú mérésének összehasonlító vizsgálata

A kutatás célja

A jelen kutatás célja más, mint a fent említett nemzetközi projekteké. A projekt első fázisában még nem törekedtünk dinamikus elemek megjelenítésére, a hangsúly a közvetítő eszköz befolyásoló hatásának meghatározásán, illetve a különböző (klasszikus és valószínűségi) tesztelméleti elemzések, lehetőségek áttekintésén, továbbá az item és személyszintű problémák azonosításán volt.

Módszerek

Az adatfelvételt 2009 tavaszán 6. évfolyamos diákok részvételével (n=598) végeztük. A minta egyik része papír (PP), másik része számítógép (CB) alapon oldotta meg a 28 ítemes problémamegoldó feladatlapot.

Az online adatfelvétel a TAO platform segítségével történt. A rendszer alapját a Luxemburgi Egyetem munkatársai által kidolgozott, több európai ország kutatói által átvett, adaptált, folyamatosan fejlesztett (továbbá a PISA felmérésekben is alkalmazott) nyílt forráskódú, moduláris rendszerű, online tesztelésre alkalmas platform adta.

A kontextus és az alkalmazott itemtípusok következtében a papír- és a számítógép-alapú feladatlap nem volt teljesen azonos, egyes problémák megoldása során más-más típusú itemet alkalmaztunk a két médium esetében. A papír-alapon feleletalkotó típusú kérdéseket feleletválasztó típusú kérdésekkel helyettesítettük online környezetben.

Az adatfelvétel során megtartottuk a teszt linearitását, azaz minden egyes diák ugyanabban a sorrendben kapta ugyanazokat a feladatokat. Az alkalmazott online platform azonban nem tette lehetővé, hogy a papír-alapú tesztfüzet egy oldalán található feladatok az online adatfelvétel során egyszerre jelenjenek meg a képernyőn. A tesztfeladatok közötti navigálást egy előre és visszafelé lépést lehetővé tevő navigációs gombbal valósítottuk meg. A feleletválasztós kérdésekre adott választ papíron a diákok karikázással adták meg, számítógépen rádiógomb használatával, kattintással kellett kiválasztatni a helyesnek ítélt

megoldást. A feladatlap megoldására mindkét esetben egy tanítási óra állt a diákok rendelkezésére.

Az elemzés során mind klasszikus, mind valószínűségi tesztelméleti eszközöket, azon belül a Rasch modell függvényeit használtuk. Összehasonlítottuk a teszt reliabilitásmutatóját mind a papír, mind a számítógép alapú adatfelvétel tekintetében. Az itemszintű elemzésekhez itemnehézségi mutatót és különböző itemparamétereket számoltunk. A személyszintű különbségek azonosítását a tesztösszpontszám, illetve a személyparaméterek összevetésével végeztük (1. táblázat).

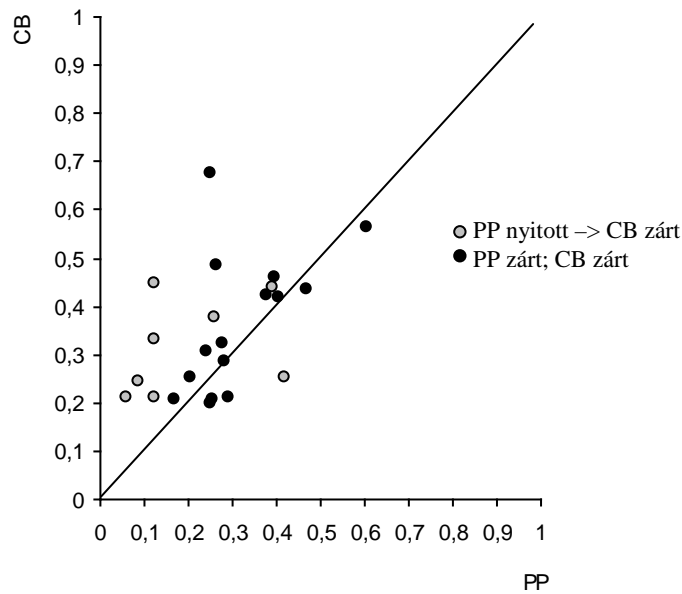
1. táblázat. Az elemzések mátrixa

	Klasszikus	IRT
Reliabilitás	Cronbach- α	Cronbach- α
Itemek	Item nehézség	Item paraméter
Személyek	Test pontszám	Személy paraméter

Eredmények és az eredmények értelmezése

A feladatlap reliabilitásmutatója különbözött egymástól a két környezetben (Cronbach- α _PP=0,78; Cronbach- α _CB=0,58), ami ellentmond a korábbi eredményeknek (Csapó, Molnár és R. Tóth, 2009; Csapó, Molnár, Pap-szigeti és R. Tóth, 2009). Ennek oka lehet egyrészt az alkalmazott itemtípusok, illetve a kontextus különbözősége, másrészt a problémák megoldásához szükséges információk megjelenítése. Utóbbi több olvasást igényelt a diákoktól, mint a korábbi, a közvetítő eszköz befolyásoló hatását vizsgáló kutatások során alkalmazott tesztek feladatainak megoldásához szükséges volt. Ez alapján megállapítható, hogy az ismeretek alkalmazhatóságának vizsgálatára kidolgozott problémamegoldó gondolkodást vizsgáló teszt pontosabb képet ad a diákok e képességbeli fejlettségéről, ha azt papír alapú formában közvetítjük, azaz számítógép alapon más típusú itemek kidolgozása szükséges.

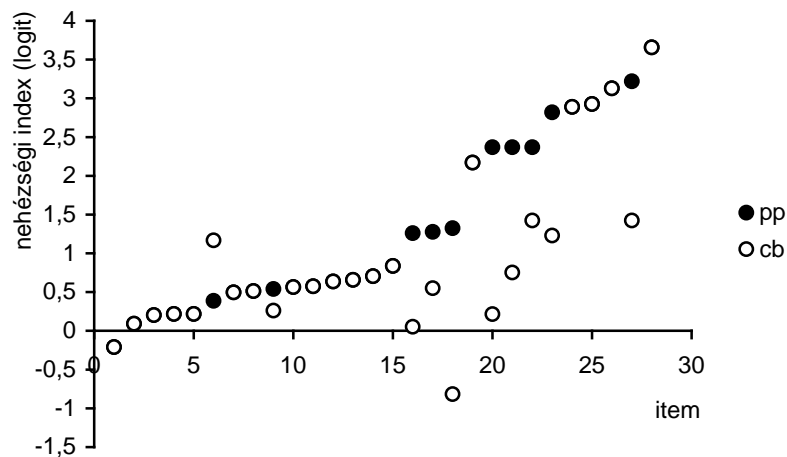
Az 1. ábra az itemek viselkedését mutatja nehézségi indexük fényében a két médián. A szürke alakzatok azokat az itemeket reprezentálják, ahol itemtípusbeli váltás történt a papír és a számítógép-alapú megjelenítés között, míg a fekete alakzatok esetében az item típusa nem változott. Egy szürke alakzattal reprezentált feladat kivételével, ahol számítógép alapon egy lépéssel többet kellett a diákoknak tenni a megoldás eljutásáig, mint papír alapon, minden egyes szürkével jelzett feladat könnyebbnek bizonyult feleletalkotó és digitális formában. Ennek oka feltehetőleg az item típusának megváltoztatásában és nem a közvetítő eszköz változásában keresendő. Egy item, amelynek, bár itemtípusát nem változtattuk meg (fekete alakzattal jelölt), feltűnően máshogy viselkedett számítógépes formában. Ennek oka a képernyőn és papíron egyszerre megjelenítendő információmennyiségben lévő különbségben lehet. Papír alapon több zavaró információ közül kellett a diákoknak kiválasztani a megoldáshoz szükséges adatokat, mind számítógép alapon. Korábbi papír-alapon végzett kutatási eredmények arra utalnak, hogy minél több információt kell összegyűjteniük a diákoknak és azok minél több megjelenési formában és helyen találhatóak, annál kisebb a helyes megoldás valószínűsége, annál nehezebb az adott probléma (Molnár, 2006).



1. ábra

Az itemnehézségi mutatók változása a közvetítő eszköz fényében

A X^2 -próba alapján az itemparaméterek pontosabb meghatározását érjük el, ha a papír- és számítógép-alapú itemek skálázását egy kétdimenziós modell segítségével végezzük el. A felállított modellben az egyik dimenzióban a papír-alapon, míg a másikban a számítógép-alapon felvett itemek vannak. A skálázás után az itemparaméterek közötti korreláció szignifikáns ($r=0,76$; $p<0,01$). Az itemnehézségi indexeket egymásra vetítve mutatja a 2. ábra. Azon itemeket, ahol nem változott az itemtípus, illetve amelyek nehézségi indexei a korábbi elemzések alapján szignifikánsan nem különbözött egymástól azonos itemekként, anchor itemekként kezeltük. Ezekhez történt a többi item kétdimenziós skálázása. Az eredmények egyrészt alátámasztják a korábban tapasztaltakat, másrészt finomabb felbontásban mutatják a nehézségi paramétereket.



2. ábra

Az itemparaméterek változása kétdimenziós skálázás esetén a közvetítő eszköz függvényében

A személyszintű klasszikus tesztelméleti módszerekkel történő elemzések során minden egyes diákhöz hozzárendeltük százalékos teljesítményét. A közvetítő eszköz függvényében különbözött a diákok átlagos százalékos formában jellemzett teljesítménye ($x_{PP}=24\%$, $sd_{PP}=15\%$; $x_{CB}=34\%$, $sd_{CB}=14\%$, $p<0,001$), ami szintén ellentmond a korábbi tapasztalatainknak, viszont egybeesik a projekt keretein belül végzett másik, olvasási képesség fejlettségi szintjét mérő (Hódi és R. Tóth, 2009) vizsgálattal. A diákok azonos szórás mellett átlagosan magasabb teljesítményt mutattak számítógépes környezetben, mint papíron. A fiúk papír alapon szignifikánsan ($p<0,05$) magasabb teljesítményt értek el, míg ez az előny számítógépes formában nem volt kimutatható.

A teljesítménybeli különbségek egyik oka az alkalmazott itemtípusok és a kontextus, a problémák megoldását segítő információk mennyiségének eltérésében keresendő. Ez annak következménye, hogy bár az egyes itemek esetén ugyanannyi információt kaptak a diákok mindkét környezetben, de a PP teszten az egy oldalra kerülő összes megoldandó problémához tartozó információ közül kellett válogatni a diákoknak, míg számítógépes környezetben egyszerre kevesebb zavaró információt lehetett megjeleníteni a képernyő mérete következtében. A számítógépes környezetben másképpen viselkedő itemek a PP feladatlapon általában nyílt kérdésként szerepeltek, amelyek számítógépes analóg párja feleletválasztós item volt. Minél közelebb volt az adott probléma számítógépes megjelenítési formája a PP változathoz, annál kevésbé volt befolyásoló a közvetítő eszköz hatása.

A klasszikus tesztelméleti módszerek azonban nem adnak pontos képet arról, hogy különböző diákok, különböző teszten mutatott teljesítménye hogy viszonyul egymáshoz. A teljesítmények árnyaltabb képét kapjuk, ha kihasználjuk a valószínűségi tesztelméleti eszközöket és azon itemeket, amelyek paraméterei nem különböztek szignifikánsan egymástól a két médián, horgony itemekként kezelve, közös képességskálán jellemezzük a diákok teljesítményét. A képességszintek közös képességskálán történő jellemzése már nem mutatott ki szignifikáns különbséget a két részminta teljesítmény között ($x_{PP}=-1,27$, $sd_{PP}=0,96$; $x_{CB}=-1,22$, $sd_{CB}=0,77$, n. s.).

Az elemzések eredménye közötti különbségek felhívják a figyelmet arra, hogy az esetleges kisebb különbségek pontosabb leírásához lehetőség szerint a mérés minél több változóját figyelembe kell venni és az elemzések alapját szolgáltató modellbe integrálni. Az eredmények rávilágítottak arra, hogy minél összetettebb egy item, minél több olvasnivaló tartozik egy feladathoz, minél inkább kihasználjuk a technológia adta lehetőségeket, ugyanakkor a terjedelmi korlátokba ütközünk, annál nagyobb eltérés lesz hasonló itemek itemparamétereiben. Éppen ezért, ha számítógép-alapú eredményeinket össze kívánjuk vetni korábbi papír-alapú képességvizsgálatok eredményeivel, számolnunk kell az itemek esetleges különböző viselkedésével. A korábbi eredmények és jelen kutatás elemzései is arra engednek következtetni, hogy az egyszerűbb, papír alapon is megjeleníthető, ugyanakkor kevés olvasást igénylő feleletválasztós kérdések itemparaméterei a leginkább közvetítő eszköz független itemek.

A tanulmányban ismertetett és hasonló nemzetközi kutatások eredményei is valószínűsítik, hogy a papír-alapú tesztelésről a számítógép-alapú tesztelésre történő áttérés folyamán még jó ideig párhuzamosan fog zajlani a PP és CB-alapon történő adatfelvétel. Míg szükséges a korábbi eredményekkel történő összehasonlítás, addig a mérés hasonló típusú tervezése, a papír alapú teszteléstől való fokozatos eltávolodás fényében nélkülözhetetlen.

Köszönetnyilvánítás

A tanulmány a K75274 OTKA kutatási program, az Oktatásméleti Kutatócsoport és az SZTE MTA Képességkutató Csoport keretében készült. A tanulmány írása idején a szerző Bolyai János Kutatási ösztöndíjban részesült.

Irodalom

- Ainley, M. (2006): Connecting with learning: motivation, affect and cognition in interest processes. *Educational Psychology Review*, 18 (4), 391-405
- 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., 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. <http://escholarship.bc.edu/jtla/vol6/9/>
- Bennett, R. E., Persky, H., Weiss, A. R. és Jenkins, F. (2007): Problem solving in technology-rich environments: A report from the NAEP Technology-Based Assessment Project (NCES 2007-466). Washington, DC: National Center for Education Statistics, US Department of Education. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007466>
- Blech, C. és Funke, J. (2005): Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology. Bonn: Deutsches Institut für Erwachsenenbildung. http://www.die-bonn.de/esprid/dokumente/doc-2005/blech05_01.pdf
- Bridgeman, B. (2009): Experiences from Large-Scale Computer-Based Testing in the USA. 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. 39-44.
- 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. 191-205.
- Csapó Benő (2009): Nemzetközi kutatási-fejlesztési programok a technológiai alapú értékelés elősegítésére. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 94. o.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papír alapú tesztek a számítógépes adaptív tesztlésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 3-4. sz. 3-16.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2009): Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In: Friedrich Scheuermann, Julius Björnsson (szerk.): The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing. Luxemburg: Office for Official Publications of the European Communities. pp. 113-118.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina: A papír alapú tesztek a számítógépes adaptív tesztlésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 2008. 3-4. sz. 3-16.
- Csapó Benő, Molnár Gyöngyvér, Pap-szigeti Róbert és R. Tóth Krisztina (2009): A mérés-értékelés új tendenciái, a papír alapú tesztléstől az online tesztlésig. In: Perjés István és Kozma Tamás (szerk.): Új kutatások a neveléstudományokban. Hatékony tudomány, pedagógiai kultúra, sikeres iskola. Magyar Tudományos Akadémia, Budapest. 99-108.

- Csapó, B., Latour, T., Bennett, R., Ainley, J. és Law, N. (2009): Technological Issues of Computer-Based Assessment of 21st Century Skills. Kézirat.
- Csapó, B., Molnár, G. és R. Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills. A pilot study for introducing electronic testing 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.
- Csíkszentmihályi Mihály (1997): Flow. Az áramlat: a tökéletes élmény pszichológiája. Akadémiai Kiadó, Budapest.
- Eggen, T. és Straetmans, G. (2009): Computerised adaptive testing at the entrance of primary school teacher training college. In: Sheuermann, 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. 134-144.
- Greiff, S. és Funke, J. (2008): Measuring complex problem solving: The MicroDYN approach. Heidelberg: unpublished manuscript. Available: http://www.psychologie.uni-heidelberg.de/ae/allg/forschun/dfg_komp/Greiff&Funke_2008_MicroDYN.pdf
- Hadwin, A., Winne, P. és Nesbit, J. (2005): Roles for software technologies in advancing research and theory in educational psychology. *British Journal of Educational Psychology*, 75. 1-24.
- Hódi Ágnes és R. Tóth Krisztina (2009): Olvasási képesség mérése számítógépes környezetben. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 96-97. o.
- 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. <http://escholarship.bc.edu/jtla/vol5/2/>
- Intel, Microsoft and Cisco Education Taskforce (2008): Transforming Education: Assessing and Teaching the Skills Needed in the 21st Century: A Call to Action. http://www.google.hu/search?q=Intel%2C+Microsoft+and+Cisco+Education+Taskforce.+%282008%2C+September%29.+Transforming+education%3A+Assessing+the+skills+needed+in+the+21st+century%3A+A+call+to+action&ie=utf-8&oe=utf-8&aq=t&client=firefox-a&rlz=1R1GGGL_en__HU346. Letöltés ideje: 2009. 01. 05.
- Johnson, M. és Green, S. (2006): On-line Mathematics Assessment: The Impact of Mode on Performance and Question Answering Strategies. *The Journal of Technology, Learning, and Assessment*, 4. 5. sz. <http://escholarship.bc.edu/jtla/vol4/5/>
- Kim, D. és Huynh, H. (2008): Comparability of Computer and Paper-and-Pencil Versions of Algebra and Biology Assessments. *The Journal of Technology, Learning, and Assessment*, 6. 4. sz. <http://escholarship.bc.edu/jtla/vol6/4/>
- Latour, T. és Farcot, M. (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. Office for Official Publications of the European Communities, Luxemburg. 64-67.
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers and Education*, 39. 3. sz. 299-312.

- Molnár Gyöngyvér (2006): Az ismeretek alkalmazhatóságának korlátai: komplex problémamegoldó gondolkodás fejlettsége 7. és 11. évfolyamon. Magyar Pedagógia. 4. sz. 329-344.
- O'Dwyer, L., Russell, M., Bebell, D. és Tucker-Seeley, K. R. (2008): Examining the Relationship between Students' Mathematics Test Scores and Computer Use at Home and at School. The Journal of Technology, Learning, and Assessment, 6. 5. sz. <http://escholarship.bc.edu/jtla/vol6/5/>
- OECD (2008): Issues arising from the PISA 2009 field trial of the assessment of reading of electronic texts. Document of the 26th meeting of the PISA Governing Board. OECD Directorate for Education, Paris.
- Pellegrino, J., Chudowosky, N. és Glaser, R. (2004): Knowing What Students Know: the Science and Design of Educational Assessment. National Academy Press, Washington, DC.
- Plichart, P., Jadoul, R., Vandenabeele, L. és Latour, T. (2004): TAO, a Collective distributed computer-based assessment framework built on semantic web standards. In Proceedings of the International Conference on Advances in Intelligent Systems – Theory and Application AISTA2004, In cooperation with IEEE Computer Society, November 15-18, 2004. Luxembourg, Luxembourg.
- Poggio, J., Glasnapp, D. R., Yang, X. és Poggio, A. J. (2005): A Comparative Evaluation of Score Results from Computerized and Paper & Pencil Mathematics Testing in a Large Scale State Assessment Program. The Journal of Technology, Learning, and Assessment, 3. 6. sz. <http://escholarship.bc.edu/jtla/vol3/6/>
- Puhan, G., Boughton, K. A. és Kim, S. (2008): Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing. The Journal of Technology, Learning, and Assessment, 6. 3. sz. <http://escholarship.bc.edu/jtla/vol6/3/>
- R. Tóth Krisztina (2009): A számítógépes és papír-ceruza tesztek eredményeinek összehasonlító vizsgálata háttérváltozók alapján. IX. Országos Neveléstudományi Konferencia, Veszprém, 2009. november 19-21. 97. o.
- R. Tóth Krisztina, Molnár Gyöngyvér és Csapó Benő: A számítógépes tesztelés lehetőségei. VI. Pedagógiai Értékelési Konferencia. Szeged, 2008. április 11-12. 84. o.
- Singleton, C. (2001). Computer-based assessment in education. Educational and Child Psychology, 18. 3. sz. 58-74.
- Wainer, H. (2000, szerk.): Computerised Adaptive Testing: A Primer. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Walt, N., Atwood, K. és Mann, A. (2008): Does Survey Medium Affect Responses? An Exploration of Electronic and Paper Surveying in British Columbia Schools. The Journal of Technology, Learning, and Assessment, 6. 7. sz. <http://escholarship.bc.edu/jtla/vol6/7/>
- 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.
- Wirth, J. és Funke, J. (2005): Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme. In: Klieme, E., Leutner, D. és Wirth, J. (szerk.): Problemlösekompetenz von Schülerinnen und Schülern. VS Verlag für Sozialwissenschaften, Wiesbaden. 55-72.

Wirth, J. és Klieme, E. (2003): Computer-based assessment of problem solving competence.
Assessment in Education: Principles, Policy & Practice, 10. 3. sz. 329-345.

ABSZTRAKT

A technológia alapú mérés-értékelés oktatásba történő bevezetése csak fokozatosan lehetséges. A jelen kutatás célja a mérés-értékelés médiuma befolyásoló hatásának diagnosztizálása problémamegoldó környezetben. Az adatfelvételt 2009 tavaszán 6. évfolyamos diákok részvételével (n=598) végeztük, utóbbi a TAO platform segítségével történt. A diákok átlagosan magasabb teljesítményt mutattak számítógépes környezetben. A fiúk papír alapon magasabb teljesítményt értek el, míg ez az előny számítógépes formában nem volt kimutatható. A teljesítménybeli különbségek egyik oka az alkalmazott itemtípusok, illetve a problémák megoldását segítő információk mennyiségének eltérésében keresendő.

The introduction of technology-based assessment into education is only gradually possible. The purpose of this paper is to study the mode-effect in problem-solving environment. The PP and CB data collection took part in spring 2009. The sample for the study were drawn from 6th grade students (N=596). The CB data collection was accomplished with the TAO platform. Results showed that students achieved higher in CB problem solving environment than in PP format. Regarding gender analyses, the delivery media had significant impact for boys' achievement at test level. Boys' achievement was higher in PP test, while no achievement differences were found by gender in CB mode. The highest medium effect was noticeable in the case of items having different item types in PP and CB mode, most probably the difference is in these cases the role of item type and not the impact of medium.