

Áttérés online tesztelésre – a mérés-értékelés új dimenziói

Az online tesztelés olyan új lehetőségeket kínál, amelyek jelentős mértékű változást eredményeznek a mérés-értékelési folyamatokban. A tanulmány célja a tesztelésben lévő lehetőségek, új mérés-értékelési módszerek bemutatása olyan képességek mérésén keresztül, amelyek technológia alapú vizsgálatának módszertana egyelőre nem kidolgozott. A kutatások közös eleme az online tesztelés és a tesztek kiközzvetítésére alkalmazott platform, az eDia, amely többek között alkalmas lineáris vagy adaptív, statikus és dinamikus, egyéni vagy kollaboratív tesztelés megvalósítására, kiszolgálva ezzel az új generációs pedagógiai mérés-értékelési igényeket.

A technológia fejlődése nemcsak mindennapi életünket, hanem az oktatási folyamatokat, ezen belül a mérés-értékelési módszereket is alapvetően megváltoztatja (Molnár, 2010). Az utóbbi években elterjedő információs és kommunikációs technológiák (IKT) számos új lehetőséget kínálnak a tanulói teljesítmények értékelésére, mérésére. A tesztelési folyamatok egyszerűsödnek, költséghatékonyabbá válnak, a tesztelési lehetőségek kiszélesednek, új típusú kérdéstechnika használatára nyílik mód, elérhetővé válik multimédiás elemek tesztbe való integrálása, új tesztelési eljárások alkalmazása (Csapó, Molnár, Pap-Szigeti és R. Tóth, 2009). A technológia ezen felül olyan képességek, tudáselemek mérés-értékelését is lehetővé teszi, amelyek eddig hagyományos eszközökkel nem voltak elérhetőek.

A technológia a mérés-értékelés folyamatában számos módon és formában jelen lehet, ezért nincs egyértelmű definíció arra vonatkozóan, hogy mit értünk technológia alapú mérésen (TBA – technology-based assessment; Molnár, 2010). A TBA elnevezés általában magában foglalja az összes mérési-értékelési rendszer alkalmazását, ahol az adatgyűjtésre valamilyen infokommunikációs eszközt használunk (Csapó, Molnár és R. Tóth, 2008). Az alkalmazott technológia függvényében a TBA számos típusát különböztetjük meg. A legelterjedtebb a számítógép alapú tesztelés (computer-based assessment), melynek legdinamikusabban fejlődő formája az internet alapú vagy online tesztelés. Az online tesztelés előnye, hogy az adatfelvétel megvalósításához csak internetkapcsolatra és internetes böngészőprogram használatára van szükség, a feladatok kiközzvetítése, a válaszok rögzítése és az adatok tárolása a központi szerveren keresztül történik. Ez nagyon egyszerűsíti a tesztelést, és megkönnyíti a tesztek módosítását, valamint az itembank fejlesztését (Csapó és mtsai, 2008).

A tanulmányban ismertetett mérésekhez az eDia online diagnosztikus mérési rendszer szolgált alapul, mely platform fejlesztése 2011-ben indult el, és használatával mostanáig több mint 30 ezer 1–6. évfolyamos tanuló tesztelése valósult meg. A rendszer változatos itemtípusok alkalmazását teszi lehetővé, a szövegek és ábrák beillesztésén túl multimédiás elemek beépítésére is lehetőséget biztosít. Alkalmazásával nemcsak a korábban már vizsgált területek, hanem új, eddig még nem vizsgált képességek mérésére is lehetőség

kínálkozik, valamint az adaptív és dinamikus itemkiosztás is megvalósulhat (Molnár és Csapó, 2013).

A tanulmány az online tesztelés lehetőségeinek öt különböző aspektusát tárja fel öt kutatás szintetizálásán keresztül. Első lépésként azt vizsgáljuk, hogy a hagyományos papír alapú tesztelésről a számítógép alapúra való áttérés milyen kérdéseket vet fel, hogyan lehet biztosítani a biztonságos áttérés folyamatát, azaz azt, hogy lehetőség szerint egyetlen tanulónak se származzon hátránya abból, hogy nem papír, hanem számítógép alapon oldja meg a tesztek. Második lépésben papír alapú feladatok digitalizált változatából összeállított feladatbank segítségével vizsgáljuk a tanulók tudásszintjéhez igazodó adaptív tesztelés hatékonyságát a tradicionális lineáris teszteléshez képest. Harmadik lépésben túlmutatunk a hagyományosnak nevezett, papír alapon is kiközvetíthető tesztek segítségével történt kutatásokon, és egy olyan, multimédiás elemeket és hiperlinkeket tartalmazó, valós keresési környezetet szimuláló számítógépes mérőeszköz kialakításának kihívásait ismertetjük, amely a tanulók IKT műveltségének közvetlen mérésére alkalmas. Negyedik lépésben ismételen egy a technológia adta lehetőségeket kihasználó mérőeszköz kidolgozásának kérdéseit és lépéseit ismertetjük, amely mérőeszköz a kreativitás diagnosztikus mérését teszi lehetővé. Végül, de nem utolsósorban, túllépve az egyéni tesztelés adta lehetőségeken, a csoportban történő kollaboratív problémamegoldó képesség technológia alapú mérésének lehetőségeit vázoljuk fel. A kísérleti elrendezés számos kérdést vet fel a technológiai kivitelezéssel, a csoportok kialakításával és azok értékelésével, a feladatszerkesztési elvekkel kapcsolatban.

A papír alapú tesztelésről a számítógép alapú tesztelésre való áttérés teljesítménybefolyásoló hatása

A TBA széles körű elterjedése lassú folyamat (Schroeders és Wilhelm, 2011), alkalmazásának gyakoriságát azonban elősegíti, ha minél több információval rendelkezünk az új típusú mérőeszközzel. A számítógép alapú mérés-értékelési rendszerekre való áttérés akkor lehet felelősségteljes, ha ismerjük a médium megváltozásával járó teljesítménybefolyásoló hatásokat. A migrációhoz megfelelő információval kell rendelkezni a médium befolyásoló erejéről a mintára, a technológiára és a teszt jellemzőire vonatkozóan (CTB/McGraw-Hill, 2003).

A közvetítő eszköz teljesítményt befolyásoló hatását az 1990-es évektől rendszeresen kutatják (Gallagher, Bridgeman és Cahalan, 2000). Az oktatás és a pszichológia területére fókuszáló kis és nagy mintás összehasonlító kutatások közös célja a közvetítő eszköz hatásának vizsgálata volt. A kutatási eredmények azonban nagyfokú diverzitást mutatnak, amelynek hátterében az áll, hogy a minta, a mérőeszköz és az elemzési technikák tekintetében is jelentősen eltértek egymástól a vizsgálatok (Wang és Shin, 2009). Ennek okán nehéz közös, szintetizált konklúziókat megfogalmazni a számítógép teljesítményt befolyásoló hatásáról. A kutatások eredményei abból a szempontból megegyeznek, hogy minél nagyobb különbségek jelennek meg a papír alapú mérőeszközökhöz képest (pl. eltérő megjelenítés, feladatok közti lapozás hiánya), annál nagyobb az eltérés az egyes médiumokon elért pontszámok között. Az egymásnak megfelelő tesztkonstrukciók kialakításakor megbízhatóságot növelő tényező, ha követjük a nemzetközileg lefektetett online tesztelésre vonatkozó útmutatásokat, irányelveket (pl. International Test Commission, 2005; Psychological Testing Centre, 2006).

Az áttérés nehézsége számos technikai paramétertől, illetve kultúrától függő tényező is lehet (Bjerkstrand, 2009), ennek kapcsán a folyamat teljes körű biztosításához szükséges, hogy a mérési intézetek önálló összehasonlító kutatásokat folytassanak saját technológiai hátterükre alapozva releváns mintákon. Napjainkra minden egyes 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).

Az összehasonlító tanulmányok megszületése ellenére is kevés kutatás összpontosít az alkalmazott ítemtípus meghatározó szerepére (Johnson és Green, 2006), vagy az ehhez szorosan kapcsolódó teljesítménybeli különbségekre. A tanulmányok legtöbbször igaz, hogy a felhasznált ítemek nagy része feleletválasztós, így kevés információval rendelkezünk a többi altípus viselkedéséről. A vizsgált korosztályt tipikusan középiskolások vagy felsőoktatásban tanulók jelentik, olyan populáció tehát, amely feltehetőleg már gyakorlott az informatikai eszközök használatában.

A médiahatást középpontba állító kutatásunk célja 1–6. évfolyamos diákok papír, illetve számítógép alapú matematika teszten nyújtott teljesítményei alapján az ítemtípusok széles spektrumának felhasználásával olyan itemparaméterek elkülönítése, amelyekkel leírható a médiahatás természete. A paraméterek a feladathoz tartozó információk feldolgozásával, a megoldás közben működő pszichikus struktúrák jellemzőivel és a feladatmegoldó tevékenységek típusával kapcsolatosak. A különbségek megállapítása ezen paraméterek mentén lehetővé teszi, hogy ajánlásokat, keretrendszert fogalmazzunk meg a feladatok reprezentálására, a technológiai paraméterekre vonatkozóan. Célunk továbbá annak megállapítása, hogy biztosítható-e a mindenki által elfogadott papír alapú tesztlérsről a számítógép alapú tesztre történő áttérés úgy, hogy megfeleltethetők egymással a két médiumon nyújtott teljesítmények. Kutatási kérdéseink közé tartozik valamint az is, hogy hány éves kortól alkalmazhatóak megbízhatóan a számítógépes és papír alapú tesztek, illetve hogy mennyire meghatározó tényező a számítógépes gyakorlottság abban a korosztályban, amelyben a tanulók kezdik rendszeresen használni az IKT eszközöket.

A vizsgálatban egy 2010-ben országosan reprezentatív mintán ($N_p=40571$), papír alapon rögzített teszt eredményeit, valamint egy 2012-ben végzett nagymintás ($N_{SZ}=22715$) online matematika mérés adatait használtuk fel az 1–6. osztályos részmintát kiemelve. Papír alapon évfolyamonként átlagosan 30 tesztváltozat készült. Minden tesztváltozat három klaszterből (részesztből) állt, egy klaszter pedig 3-4 darab 4-5 ítemes matematika feladatot, azaz összesen körülbelül 15 ítemet tartalmazott. A papír alapú tesztek válogatott feladataiból 10 tesztváltozat készült. A teszt összeállításánál meghatározó szerepet játszott, hogy minden évfolyamon egyenlő arányban legyenek a nyílt és a zárt végű feladatok. A feladatbank lehetőségeinek és a számítógépen való azonos reprezentációs igényeknek figyelembevételével az altípusok is kiegyenlítetten szerepeltek. A nyílt és a zárt végű feladatok között mind reprezentálási módban, mind eszközhasználatban jelentős különbségek vannak, ezért információgazdag következtetések vonhatók le, ha azok reprezentatív módon hasonlíthatók össze. A különböző teszteken mutatott teljesítmények közös képességskálán történő jellemzéséhez az egyes változatokat horgony klaszterek kapcsolták össze. Az adatok skálázása kétdimenziós Rasch-modell segítségével történt. A matematika teszt mellett minden egyes diák kitöltött egy háttéradatokra (nem, szülők iskolai végzettsége, iskolai, matematika előmenetel, matematika attitűd, informatika oktatásban való részvétel) vonatkozó 7 ítemes kérdőívet.

A tesztváltozatok megbízhatósági mutatói (Cronbach- α) papír alapon 0,81, számítógép alapon 0,84 feletti értékek voltak. A számítógép alapú tesztek reliabilitás értékei átlagosan minden egyes évfolyamon magasabbnak bizonyultak, mint papír alapon. A feladatok viselkedését jellemző személyszeparációs reliabilitásmutató értékek is hasonló eredményt mutattak (papír alapon Cronbach- $\alpha=0,84$, számítógép alapon Cronbach- $\alpha=0,89$).

Az ítemek papír és számítógépen való viselkedésének összehasonlítását a két médiumon mért átlagos ítemnehézségi értékek vizsgálatán keresztül végeztük el. Egyik évfolyamon sem volt kimutatható szignifikáns különbség a feladatbank papír és számítógép alapú változatainak nehézsége között. A médiahatást pontosabban jellemzi az ítemnehéz-

ségi értékek közötti korrelációs együttthatók nagysága, amely évfolyamról évfolyamra fokozatosan erősödő kapcsolatot mutatott. Hatodik évfolyamon a papíron és számítógépen nyújtott teljesítmények közötti korreláció értéke $r=0,92$ volt.

A feladatokhoz tartozó karakterek számának teljesítménybefolyásoló hatása évfolyamról évfolyamra csökkent, és nem különbözött szignifikánsan az egyes médiumokon. A grafikus elemek használatának nehézségi indexre gyakorolt hatása pozitívnak bizonyult mindkét tesztkörnyezetben, minden évfolyamon. A táblázatokat tartalmazó feladatok viselkedése mind átlagosan, mind évfolyamonkénti bontásban médiafüggetlen volt.

A feladatok kontextusa és az azok megoldásához szükséges pszichikus struktúra mindkét médiumon azonos mértékben befolyásolta a feladatok nehézségét ($F_{\text{Kon-textus}}=1,07$, $p=0,38$; $F_{\text{Pszí.Str.}}=1,73$, $p=0,19$). A feladatokat matematikai tartalom szerinti csoportosításban az első öt évfolyamon azonos eredménnyel oldották meg papír és számítógép alapon ($F_{\text{Tartalom}}=2,44$, $p=0,23$), a hatodik évfolyamon megjelenő statisztikai tartalmú feladatok nehezebbnek bizonyultak számítógépes környezetben ($|t|=2,06$, $p<0,05$).

A feladatok típusa az alternatív választás feladatok esetében volt meghatározó, az első három évfolyamon átlagosan könnyebben oldották meg azokat a diákok számítógépen, felsőbb évfolyamokon nem jelentkezett szignifikáns különbség. Az előzetes feltételezésekkel ellentétben a szövegalkotó feladatokban egyik évfolyamon sem kerültek hátrányba a diákok számítógép alapon a gépelési gyakorlottságból eredő különbségek miatt.

A feladatra adandó válaszok karakter száma nem bizonyult médiahatást generáló tényezőnek. Alsóbb évfolyamokon a hosszabb választ igénylő feladatok tipikusan nehezebbnek bizonyultak, de ez a befolyásoló erő azonos volt papír és számítógép alapon.

Az eszközhasználat tekintetében első évfolyamon átlagosan könnyebbek voltak azok a feladatok, ahol a diákoknak csak egeret kellett használniuk a feladat megoldásához ($|t|=3,17$, $p<0,01$), azonban a parciális korrelációs vizsgálatok arra mutattak rá, hogy a különbséget a feladattípus befolyásoló hatása okozza (az alternatív típusú feladatok megoldásához elegendő az egér használata).

A technológia korábban nem tapasztalt, innovatív lehetőségeket biztosít a mérés-értékelés folyamatában. A számítógépes tesztelés pedagógiai alkalmazásához azonban meg kell vizsgálnunk, hogy lehetnek-e nem kívánatos mellékhatások (Csapó, Molnár és R. Tóth, 2008). Jelen kutatás a kisiskolás diákokat is involválva nagymintás adatfelvétel eredményeire alapozva vetette össze a papíron és számítógépes formában is megjeleníthető feladatokon nyújtott teljesítményeket. Az eredmények arra engednek következtetni,

A feladatok típusa az alternatív választás feladatok esetében volt meghatározó, az első három évfolyamon átlagosan könnyebben oldották meg azokat a diákok számítógépen, felsőbb évfolyamokon nem jelentkezett szignifikáns különbség. Az előzetes feltételezésekkel ellentétben a szövegalkotó feladatokban egyik évfolyamon sem kerültek hátrányba a diákok számítógép alapon a gépelési gyakorlottságból eredő különbségek miatt. A feladatra adandó válaszok karakterszáma nem bizonyult médiahatást generáló tényezőnek. Alsóbb évfolyamokon a hosszabb választ igénylő feladatok tipikusan nehezebbnek bizonyultak, de ez a befolyásoló erő azonos volt papír és számítógép alapon.

hogy a számítógép összességében legalább annyira megbízható közvetítő eszköz, mint a papír alap. Különböző feladattípusok alkalmazásával elérhető a teszteredmények szám-szerű felcserélhetősége, miután a feladatokon belül megjelenő médiahatás már a teszt szintjén kiegyenlítődik. Az azonos típusú feladatok (pl.: alternatív választás) alkalmazása nagyobb odafigyelést és az eredmények transzformációját, újraskálázását igényli, de egyéb tesztyszerkesztési elvekkel is egybecseng a változatos feladattípusok alkalmazása. A kutatás nem talált olyan feladatjellemzőt, amely egyértelműen jelentős mértékű médiahatást eredményezett volna, ezért használata korlátozásra szorulna. Eredményeink szerint a kisiskolás diákok (1–2. osztály) elsősorban a technológiai jártasságuk hiányának köszönhetően igényelhetnek külön figyelmet az átállás jelenlegi stádiumában.

Lineáris tesztelésről adaptív tesztelésre való áttérés hatékonyságvizsgálata

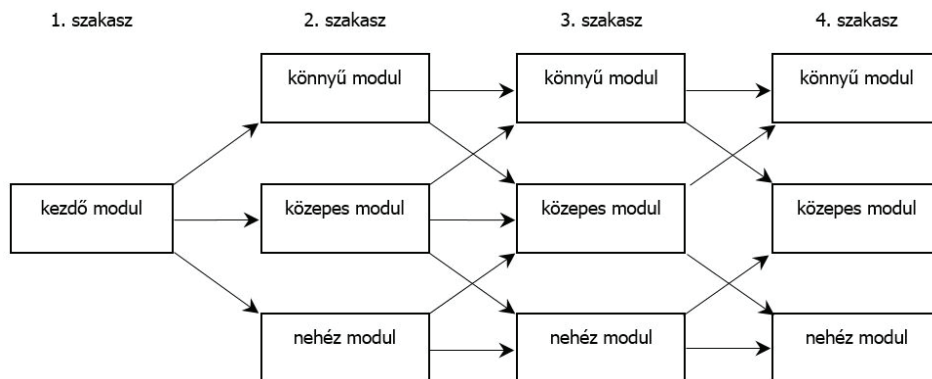
A számítógépes adaptív tesztelés (computerized adaptive testing) alkalmazása jelentős mértékben megváltoztatja a hagyományos lineáris tesztelés során alkalmazott mérés-értékelés folyamatát (Csapó, Molnár és R. Tóth, 2008; Kárpáti, Molnár, Tóth és Főző, 2008; Molnár, 2011). Adaptív tesztelés során a hagyományos, lineáris tesztelésnél megszokottól eltérően a tanulók saját képességszintjükhez illeszkedő feladatokat kapnak (Magyar, 2012). Ezáltal a tesztelés a szóbeli vizsgáztatáshoz válik hasonlóvá (Csapó és mtsai, 2008; Reckase, 2009), ahol a tanár a tanuló előző kérdésekre adott válaszait figyelembe véve teszi fel a következő kérdést (Zenisky, Hambleton és Luecht, 2010). Amennyiben a tanuló jól válaszol, nehezebbet kap, rossz válasz esetén pedig könnyebb kérdés következik. Ezáltal a tesztelés személyre szabottá, a tanulók képességszintjéhez alkalmazkodóvá válik (Keng, 2008). Az adaptív tesztelés továbbá a teljesítmény sokkal pontosabb mérését teszi lehetővé, mivel a feladatok nehézségi szintjét a tanuló képességszintjéhez igazítja, ezáltal növeli a kinyerhető információ nagyságát mind a tanulókra, mind az itemekre vonatkozóan (Molnár, 2013; Tian, 2007).

Az adaptív teszteknek sokféle típusa létezik attól függően, hogy az adaptivitás az itemek, feladatok vagy részesztek szintjén valósul meg (Magyar, 2012; Molnár, 2013). Könnyű alkalmazhatósága és jó mérésmetodikai tulajdonságai miatt gyakran használt típus a többszakaszos adaptív teszt, mely félúton helyezkedik el a hagyományos és az item alapú tesztek között. A tesztelés során több szakaszban itemek helyett modulok kerülnek kiosztásra, melyek tulajdonképpen különböző nehézségi szintű rövid fix tesztek. Egy szakaszon belül két vagy több modul lehet, melyek nehézségi szintjükben különböznek. Miután a diák végez egy-egy modullal, képességszintje becslésre kerül, és ez alapján kap a következő szakaszban újabb modult (Zenisky, Hambleton és Luecht, 2010; Magyar, 2013a, b). Ez a technika egyrészt a kérdéseket a tanuló képességszintjéhez igazítja, másrészt lehetőséget ad az itemek sorrendjének előzetes meghatározására (Luecht, 2011).

Kutatásunkban sor került adaptív és lineáris tesztváltozatok mérési pontosságának összehasonlítására a kinyert információk és a standard hibák tekintetében, és annak a kutatási kérdésnek a megválaszolására, hogy többszakaszos adaptív tesztek esetében a különböző képességszinteken több információ nyerhető-e ki a tesztből, mint hagyományos, lineáris tesztek esetén. Miután a tesztben alkalmazott modulok nehézségi szintje illeszkedett a tanulók képességszintjéhez, hipotézisünk értelmében a többszakaszos teszt mindegyik képességszinten nagyobb információt közvetít a tanulók teljesítményéről, mint a lineáris tesztváltozat. A mérési hibák nagyságát összehasonlítva feltételeztük, hogy az adaptív teszt minden képességszinten kisebb standard hibát mutat.

A vizsgálat 10–14 éves tanulók részvételével zajlott 2012 őszén (N=158). A mérőeszköz induktív gondolkodást mérő, korábbi kutatásokban már papír alapon alkalmazott és kalibrált itembankból (*Csapó és Molnár, 2012*) került összeállításra. Az adatfelvétel két fázisban történt. Az első fázisban a tanulók számára kétféle tesztváltozat (adaptív és lineáris) véletlenszerűen került kiközvetítésre. A mérés második fázisa két hét elteltével zajlott, az először lineáris tesztet megoldók adaptív változatot kaptak, az adaptívát megoldók pedig lineárisat. A feladatok megoldására mindkét esetben 45 perc állt rendelkezésre. A tesztet a tanulók a saját iskolájukban, az iskola saját internethálózatán keresztül oldották meg.

A tesztelés során alkalmazott lineáris teszt 28 itemből állt, és széles itemnehézségi skálán mozgó feladatokat tartalmazott. Az adaptív változat 1-3-3-3 szerkezetű négy-szakaszos adaptív tesztet jelentett, és szintén 28 itemből állt. A teszt 10 itemes kezdő modullal indult, mely vegyesen tartalmazott könnyebb és nehezebb itemeket, majd 6-6 itemből álló modulok követték azt három különböző nehézségi szinten (könnyű, közepes és nehéz). Ezzel összesen 17 különböző tesztváltozat jött létre (1. ábra).

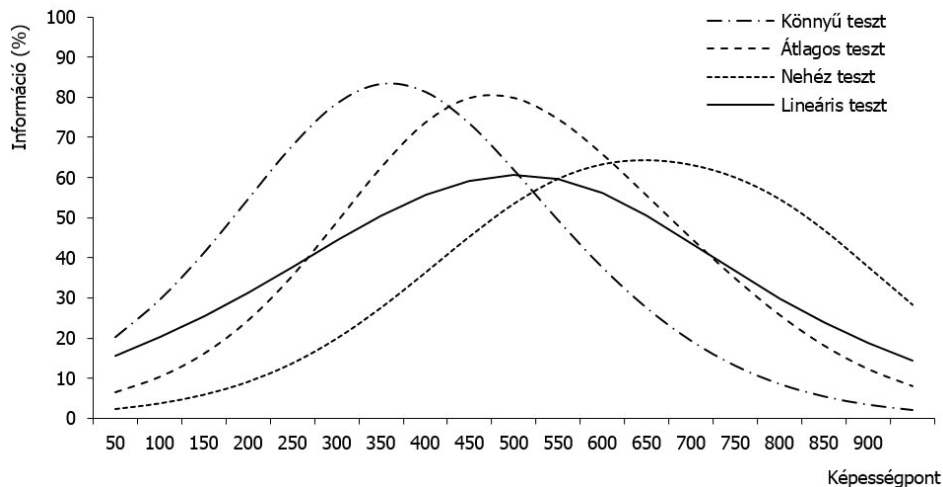


1. ábra. Az 1-3-3-3 szerkezetű adaptív teszt

A tanulók tesztben elért pontszámainak meghatározása és az elemzések a valószínűségi tesztelmélet felhasználásával a ConQuest programmal történtek. A képességszinteket 500-as átlagú, 100-as szórású skálára transzformáltuk. A tesztekben kinyert információk nagyságának jellemzésére a teszt információs görbét alkalmaztuk, mely a tesztet megoldó tanulók átlagos képességszintje és az itemek nehézségi szintje közötti különbséget fejezi ki. Minél közelebb van egymáshoz ez a két érték, annál nagyobb a tesztelés során kinyert információ mennyisége.

Első lépésben a tesztek mérési pontosságának meghatározására a reliabilitásmutatókat használtuk. A lineáris teszt esetében a megbízhatósági mutató (Cronbach- α = 0,83) valamivel alacsonyabb értékű volt, mint az adaptív teszt esetében számolt WLE személyszeparációs mutató (0,85) (bővebben *Molnár, 2013*).

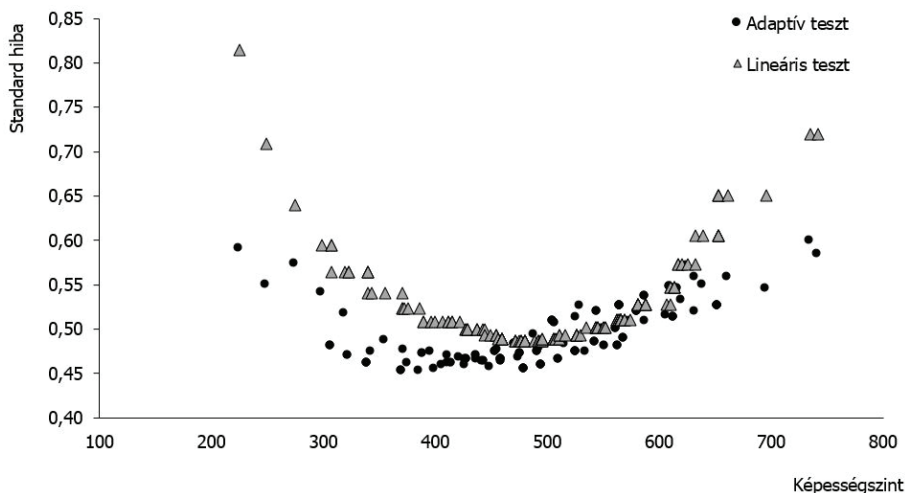
Mivel minden tanuló megoldotta mindkét tesztet, lehetőség volt a becsült képességszintek pontosságának személyek szintjén való meghatározására és összehasonlítására. A tesztből kinyert információk mennyisége az adaptív tesztben mindegyik modul esetén (könnyű, közepes, nehéz részteszt) nagyobb volt, mint a lineáris tesztben. Az információs görbéket összehasonlítva a különböző nehézségű modulok görbéi minden képességszinten a lineáris teszt görbéje felett helyezkednek el (2. ábra).



2. ábra. A lineáris és az adaptív teszt különböző nehézségű moduljainak információs görbéi

Az adaptív tesztben a könnyű modulból 87%, a közepesből 79%, a nehéz modulból pedig 68% volt a kinyert információk átlagos mennyisége, míg a lineáris tesztben 62%. A közepes szintű modulnál az információ szakasról szakaszra nőtt (66%, 81% és 89% a 2., 3. és 4. szakaszban), ami arra mutat, hogy a nem ebbe a képességszintbe tartozó tanulók fokozatosan átkerültek a könnyű, illetve a nehéz modulokba. A teszt végére a diákok részvételi aránya közel azonos lett a különböző szintű modulokban (Könnyű: 37%, Közepes: 34%, Nehéz: 29%).

A tesztek mérési pontosságának másik mutatója a mérési hiba nagysága. A teszt egészére vonatkozóan a lineáris teszt esetén szignifikánsan magasabb volt a mérési hiba (SE átlag: 0,53), mint adaptív teszt esetén (SE átlag: 0,49) ($t=-7,54$, $p<0,01$). A standard hiba változását a különböző képességszintek függvényében vizsgálva annak mértéke az alacsony és a magas képességtartományokban tér el jelentősen, azaz elsősorban ezekben a képességtartományokban mér jelentősen kisebb hibával az adaptív teszt (3. ábra).



3. ábra. Az adaptív és a lineáris teszt standard hibái a képességszintek függvényében

A kutatás során adaptív és lineáris tesztkörnyezetben vizsgáltuk online tesztek mérési pontosságát. A mérési pontosságot a tesztekből kinyert információk mennyiségének, valamint a tesztek standard hibáinak összehasonlításával jellemeztük. Az eredmények alapján megállapítható, hogy az adaptív teszt a képességskála teljes szintjén több információt és alacsonyabb mérési hibát szolgáltat, mint a lineáris tesztváltozat. A két teszt esetén elsősorban az alacsony és a magas képességtartományokon jelentős az eltérés, itt helyenként több mint 30% a különbség a tesztekben kinyert információk között, és a standard hiba is jelentősen megnő a lineáris tesztben. Eredményeink hozzájárulhatnak az adaptív tesztek minél szélesebb területen való alkalmazásának elterjedéséhez.

Valós keresési környezetet szimuláló, az információs és kommunikációs technológiai műveltség mérésére alkalmas mérőeszköz kidolgozásának kihívásai

Az értékesnek számító tudás ezredforduló utáni erőteljes megváltozása (Anderson, 2010) magával hozta az IKT műveltség fejlesztésének (Molnár, 2011) és technológia alapú, szimulációs mérésének igényét (Ainley, Fraillon, Gebhardt és Schulz, 2012; Mayrath, Clarke-Midura és Robinson, 2012). A technológia rapid változásának következtében az IKT műveltség fogalmi meghatározása az elmúlt három évtized során jelentős mértékű változáson ment át (Binkley, Erstad, Herman, Raizen, Ripley és Rumble, 2012; NAGB, é. n.; Partnership for 21st Century Skills, 2008; Tongori, 2012). Számos nemzetközi kutatás is felhívta a figyelmet az IKT műveltség komponenseinek (meghatározás, hozzáférés, rendszerezés, integrálás, értékelés, kommunikáció, létrehozás) mérésére alkalmas, felső- (Katz és Macklin, 2007) és középfokú oktatásban (Ainley és mtsai, 2012; Law, Lee és Yuen, 2010; Zelman, Shmis, Avdeeva, Vasilev és Froumin, 2011) alkalmazható számítógép alapú mérőeszköz létrehozásának szükségességére.

Az IKT műveltség mérésével kapcsolatban a Molnár és Kárpáti (2012) által publikált összegző elemzés az IKT műveltség számos aspektusát feltáró, többségében papír alapú, kérdőíves méréseket tekintti át jellemzően a különféle korosztályok használati szokásait, hardvereszközökhöz való hozzáférést, illetve használati gyakoriságát vizsgáló IKT műveltségi méréseken keresztül. Hazai viszonylatban úttörő munkának számít az IKT műveltség számítógépes szimuláción alapuló online teljesítménymérése.

A fejlesztések célja egy 5–10. évfolyamos diákok mérésére alkalmas, multimédiás elemeket és hiperlinkeket tartalmazó, valós keresési környezetet szimuláló számítógépes mérőeszköz kidolgozása. A nemzetközi meghatározások és modellek értelmében az IKT műveltség hét komponensének működtetésében meghatározó szerepet játszik a problémamegoldó gondolkodás (Tongori, 2013). A mérőeszköz életszerű szituációkba ágyazott, fokozatosan nehezedő problémákból áll.

Az IKT műveltség mérésével kapcsolatban a Molnár és Kárpáti (2012) által publikált összegző elemzés az IKT műveltség számos aspektusát feltáró, többségében papír alapú, kérdőíves méréseket tekintti át jellemzően a különféle korosztályok használati szokásait, hardvereszközökhöz való hozzáférést, illetve használati gyakoriságát vizsgáló IKT műveltségi méréseken keresztül. Hazai viszonylatban úttörő munkának számít az IKT műveltség számítógépes szimuláción alapuló online teljesítménymérése.

A diákok egy szimulált, azaz a valós weboldalakhoz megjelenésében és működésében rendkívül hasonló, de a követhetőség érdekében kontrollált környezetben dolgoznak a teszt problémáinak megoldásakor. A kutatás tervezése során külön kihívást jelentett az adatfelvételkor rendelkezésre álló, a hazai iskolarendszer felépítéséhez illeszkedő, egy tanórát átfogó 45 perces időkorlát, miután a nemzetközi, azonos konstruktum mérésére alkalmas adatfelvételek általában 120–150 percet vesznek igénybe. A problémák felépítése, a keresési környezet komplexitása jelentős mértékben csökkenti a tesztben felhasználható problémák, és ezzel párhuzamosan a független ítemek számát, így külön kihívást jelent a megfelelő mennyiségű információ kinyerését lehetővé tevő teszt kidolgozása.

Az IKT műveltség mérésére szolgáló mérőeszköz fejlesztése jelentős előrelépést jelent az autentikus, valós infokommunikációs közeget szimuláló, interaktív, technológia alapú – és azon belül is online – mérések terén.

A kreativitás innovatív, diagnosztikus mérését lehetővé tevő online mérőeszköz kidolgozásának kérdései

A TBA a kreativitás új szemléletű vizsgálatának megvalósítását, azaz egy innovatív, a 21. század követelményeinek megfelelő mérőeszköz kidolgozását is lehetővé teszi. A kreativitás jelentősége rendkívül felértékelődött napjainkban. A felgyorsult gazdasági, társadalmi és technikai fejlődés megköveteli az újszerű, eredeti gondolatokat és megoldásokat, a modern munkaerőpiac számos szférájában elengedhetetlen a kreativitás az érvényesüléshez. Változó világunkban nem könnyű feladat annak bejósolása, hogy a jövő generációinak milyen problémákkal kell majd szembenéznük. A kiszámíthatatlanságra, az új kihívásokra való felkészítésnek egyik alkalmas eszköze a kreativitás fejlesztése, melyben központi szerepet tölthet be az iskola (Piirto, 2011). A hatékony fejlesztéshez elengedhetetlen a pontos diagnózis felállítása, az iskolai kontextusban is használható mérőeszközök kidolgozása.

A területen végzett eddigi kutatások azt igazolják, hogy a kreativitás rendkívül összetett jelenség, tanulmányozása többféle nézőpontból is lehetséges (lásd például Plucker és Renzulli, 1999). Jelen tanulmányban főként a divergens gondolkodással foglalkozunk, azonban fontos megjegyeznünk, hogy a divergens gondolkodás (Guilford, 1967) csak egy – bár az egyik legtöbbször kutatott és jelentős – aspektusa a kreativitás vizsgálatának. Ennek ellenére ez az a terület, ahol markánsan érzékeltetni tudjuk a kreativitás papír alapú tesztelésének nehézségeit és az online tesztelésben rejlő lehetőségeket.

Az intelligenciatesztekben mért konvergens gondolkodással ellentétben, ahol jellemzően egy vagy kevés a helyes megoldások száma, a divergens gondolkodást mérő feladatoknak számos jó megoldása lehet, a cél az, hogy minél nagyobb számú és eredetibb választ adjunk egy adott problémára. A kitöltőnek például az lehet a feladata, hogy sorolja fel minél több konvencionálistól eltérő használati lehetőségét a gyufának vagy a könyvnek, előre megadott rajzokat kell minél kreatívabb formában kiegészítenie, vagy különböző absztrakt ábrákról és képekről kell eredeti ötleteket generálnia arra vonatkozóan, hogy vajon mit ábrázolhatnak (lásd például Torrance, 1966; Wallach és Kogan, 1965). A válaszokat leggyakrabban három skálán értékeljük, ezek: *fluencia*, ami a gondolatáramlás könnyedségére utal, és egyenlő a megadott válaszok számával; *flexibilitás*, amely a szempontváltás képességét jelenti, azon kategóriák számával egyenlő, amelyekbe a válaszok besorolhatóak; valamint az *originalitás*, ami egy ötlet eredetiségére, nem szokványos jellegére vonatkozik, és egy adott válasz gyakoriságát jelenti. Ezek a nyílt végű feladatok igen nagyszámú választ eredményeznek, amelyek kezelése nehézkes a papír alapú tesztelési módszerekkel, különösen, ha nagyobb mintán kívánunk méréseket megvalósítani. A teszt kiértékelőjének minden választ manuálisan kell kódolnia

és értékelnie egy tesztkiértékelő kézikönyv segítségével (magyar nyelven lásd például *Barkóczy és Zétényi*, 1981; *Zétényi*, 1989). Továbbá ahhoz, hogy kényelmesen végezhessünk statisztikai elemzéseket, az adatokat digitalizálni is szükséges. A tesztfelvétel és a kiértékelés a hagyományos papír alapú tesztelést igénybe véve rendkívül időigényes, a visszacsatolás csak jelentős késéssel érkezik meg, így ez a módszer diagnosztikus célokra nehezen alkalmazható.

A problémák kezelése érdekében több kutatás is irányult a számítógép alapú mérés-értékelésben rejlő lehetőségek kiaknázására. Az eredmények arra mutatnak, hogy az online tesztelési módszerek jelentős előrelépést jelentenek a verbális válaszokat elváró divergens gondolkodás tesztek felvételében és kiértékelésében (*Cheung és Lau*, 2010; *Kwon, Goetz és Zellner*, 1998; *Lau és Cheung*, 2010; *Palaniappan*, 2012; *Pretz*, 2008). Ezek a rendszerek úgy működnek, hogy az online felületen megadott válaszokat a szoftver összehasonlítja egy olyan adatbázissal, ahol a válaszok már paraméterezettek, és hozzárendeli a válaszokhoz szükséges értékeket, majd tanuló szinten kiszámolja a teljesítményt. A biztató eredmények ellenére azonban további kutatások szükségesek az online tesztelésre történő áttálláshoz, például keveset tudunk a számítógépes környezet teljesítménybefolyásoló hatásairól (*Lau és Cheung*, 2010), valamint olyan manipulatív feladatok működéséről, amelyek nem (csak) verbális választ várnak el a teszt kitöltőjétől (*Kwon és mtsai*, 1998).

A kutatásban egy online felületen működő divergens gondolkodást mérő teszt kifejlesztésére vállalkoztunk, amely későbbi nagymintás vizsgálatok kivitelezésére is szolgálhat. A tesztfeladatok kidolgozásához a már említett nemzetközi, valamint hazai mérőeszközöket vettük alapul (*Barkóczy és Zétényi*, 1981; *Mednick*, 1965; *Torrance*, 1966; *Wallach és Kogan*, 1965).

A kidolgozott itemekhez kapcsolódó előzetes adatfelvétel alapján elvégeztük a válaszok skálázását, amely segítségével létrehozható a válaszkategóriákat és azok originalitását jellemző értékeket tartalmazó adatbázis. Az adatbázis segítségével lehetőség nyílik a későbbi mérések válaszáinak automatikus kiértékelésére, jelentősen csökkentve ezáltal a tesztelés idő- és költségigényét. Miután a teszt olyan feladatot is tartalmaz, amelyben meghatározott idő alatt kell a tanulóknak szavakat begépelniük, lehetőségünk adódik a gépelési jártasság hatásainak elemzésére is.

Méréseink jelentősen hozzájárulnak a kreativitás online diagnosztikus értékelésének hazai megvalósításához, továbbá a fejlesztésére irányuló hatékony módszerek és eszközök létrehozásához. A teszt továbbfejlesztése során külön hangsúlyt fektetünk az interaktív, manipulációt igénylő feladatok előtérbe helyezésére, valamint a kontextuális információk, például a szemmozgás vagy a kurzor mozgásának követésére, vagy az arc kifejezések monitorozására (*Csapó, Lőrincz és Molnár*, 2012), elősegítve a kreativitás mögött húzódó kognitív és emocionális folyamatok feltérképezését.

A kollaboratív problémamegoldó képesség online mérésének megvalósíthatósága

Az egyéni tesztelés adta lehetőségeken túllépve végül, de nem utolsó sorban a csoportban történő kollaboratív problémamegoldó képesség technológia alapú mérésének opcióit vizsgáljuk fel, még jobban kitágítva ezzel a TBA adta lehetőségek körét. A mérőeszköz kidolgozásának szükségességét az a tény indokolja, hogy a 21. század potenciális munkavállalói az elmúlt évtizedekben tapasztalható rohamos léptékű gazdasági és társadalmi változások következtében teljesen más elvárásokkal szembesülnek, mint elődeik ötven évvel ezelőtt (*Kozma*, 2009). Ahhoz, hogy minden szempontból felkészült, produktív munkaerő lépjen ki az iskolarendszerekből a munkaerőpiacra, elengedhetetlen például,

hogy a diákok fejlett problémamegoldó képességre, kiváló együttműködő képességre, illetve csoportban zajló, ún. kollaboratív problémamegoldó képességre tegyenek szert. A kollaboratív problémamegoldó képességet fejlesztő programok hatékonyságvizsgálatahoz nélkülözhetetlen a megfelelő mérőeszközök kidolgozása. A több komplex mérési terület egyesítését igénylő képesség vizsgálatát a legnagyobb mérés-értékeléssel foglalkozó nemzetközi szervezetek, projektek is célul tűzték ki (*Binkley és mtsai, 2012; OECD, 2013*), jelenleg azonban hazai szinten nem, illetve nemzetközi szinten is elenyésző számú empirikus mérési adattal rendelkezünk csupán a képességre nézve (*O'Neil, Chuang és Chung, 2003; Greiff, 2012*).

A kor követelményeinek megfelelően olyan eljárások kialakítása érdemes, amelyek kihasználják a modern IKT-eszközök nyújtotta lehetőségeket az egyre inkább korszerűtlennek minősített papír-ceruza teszteléssel szemben (*Csapó, Ainley, Bennett, Latour és Law, 2012*). A nagymintás mérések kivitelezésére tervezett tesztek előállításához – amelyekre az oktatásban égető igény jelentkezik standardizált tesztek hiányában – szintén olyan mérőeljárásokban célszerű gondolkodnunk, amelyek minimális humán erőforrást igényelnek a kiértékeléshez, ennek megfelelően kellően költséghatékony megoldást jelentenek. A technológia alapú mérések azáltal, hogy megteremtik az automatikus kódolás lehetőségét, a legkézenfekvőbb kondíciót képezik.

Egy valid, megbízható technológia alapú mérőeszköz megszerkesztése elsősorban a képesség komplexitásából adódóan rendkívül problematikusnak bizonyul, a képességet leíró két fő komponens, azaz a kognitív és szociális részképességek halmaza ugyanis eltérő jellege miatt teljesen más mérési hagyományokkal rendelkezik. A problémamegoldó komponens vizsgálatának technológia alapú mérési gyakorlata egyre kiterjedtebb szakirodalmi leírással rendelkezik (pl.: *Greiff, Wüstenberg, Holt, Goldhammer és Funke, 2013; Molnár, Greiff és Csapó, 2013; Molnár, Greiff, Wüstenberg és Fischer, 2014*), ezek a vizsgálatok megfelelő módszertani háttérrel biztosítanak a számítógépen exponált problémák megoldásának elemzéséhez. Nagyobb kihívást jelent a kollaboratív komponens számítógépes analízise. A csoporttagok interakcióinak, kommunikációs mintázatainak, konfliktuskezelési kultúrájuknak vizsgálata ugyanis a számítógépes adatrögzítést követően jellemzően humán elemzés segítségével történik, elsősorban tartalomelemzés jelenti a szokásos eljárást. A számítógéppel támogatott kollaboratív tanulás kutatásait tanulmányozva jelentős számú tartalomelemzésre kialakított kódrendszerrel találkozhatunk (pl.: *Henri, 1992; Gunawardena, Lowe és Anderson, 1997; Chiu, 2000*).

Az online tesztelés az egyéb, internetkapcsolatot nem igénylő technológia alapú mérésekkel szemben számos további haszna mellett az azonnali visszacsatolás jelentős pedagógiai szempontú előnyével is rendelkezik, ahhoz azonban, hogy ez megvalósulhasson, szükséges, hogy a mérés jól megalapozott automatikus kódoló rendszerre épüljön. A szakirodalom egyetlen online mérési kísérletről számol be a kollaboratív problémamegoldó képességet illetően (*Griffin és Care, 2012*), ez a mérőeszköz azonban egyelőre nem képes automatikus kódolásra, ennek megfelelően azonnali visszajelzést sem tud adni a kitöltő teljesítményéről.

A szegedi műhely keretein belül egy olyan kollaboratív problémamegoldó képességet feltérképező online mérőeljárás kidolgozását tűztük ki célul, amelyben úgy nevezett előre definiált, kötött üzenetek alkalmazásának segítségével automatikus kódoló rendszert alakítunk ki, így úttörő módon az azonnali visszacsatolás lehetőségét is megteremtjük. Eltérést mutat mérőeszközünk a többi létező, illetve születőben lévő mérőeszközzel szemben (*Hsieh és O'Neil, 2002; Rosen és Tager, 2013; OECD, 2013*) abban a tekintetben is, hogy az előre definiált üzenetek kicserélése mellett a szabadon megfogalmazott üzenetek váltását is megengedi. A tervezett kutatás alapját egy általunk fejlesztett teszt biztosítja, amelynek problémáit a 12 és 14 éves tanulónak az eDia platformon keresztül csoportokban kell megoldaniuk, chaten kommunikálva egymással. A társalgó szoftver mint kolla-

boráló partner alkalmazása helyett (OECD, 2013; Rosen és Tager, 2013) a diákok tehát egymással működnek együtt négy fős csoportokban, utóbbi feltétel szintén különbözik az eddigi elrendezésekben alkalmazott páros, illetve három fős csoportmunkáktól (Hsieh és O'Neil, 2002; Griffin és Care, 2013; Rosen és Tager, 2013). Végezetül, generalizációs szempontokat figyelembe véve, újfent egyedülálló módon a diákoknak a tesztelés alatt több csoportban is lehetőségük nyílik a kollaborációra, minden egyes problémát egy újabb csoportban oldhatnak meg. A csoporttagok kiválasztása és rotálása random módon történik szerverünk segítségével. Kutatásunkkal megteremtjük az infokommunikációs társadalmunkban egyre inkább releváns kollaboratív problémamegoldó képesség újabb, innovatív megoldásokban gazdag online diagnosztikus mérési lehetőségét.

Összegzés

A tanulmányban ismertetett kutatások azt bizonyítják, hogy az online tesztelés egyrészt olyan tudás- és képességterületek mérését teszi lehetővé, amelyek korábban papír alapon nem voltak kivitelezhetőek, másrészt korábban is kutatott területek innovatív mérési lehetőségét teremti meg. A bemutatott kutatások eredményei szerint a papír alapú mérésről az online mérésre való átállás nem okoz jelentős mértékű változást a diákok teljesítményében, ugyanakkor előnyös, ha a tanulók minél fiatalabb korban ismerkednek meg a technológia alapú mérésekkel, és gyakorolják be használatukat.

Az új technológia alkalmazása olyan új, eddig papír alapon nem, vagy csak nehezen megvalósítható tesztelési eljárások megvalósítását teszi lehetővé, mely a lineáris teszteléshez képest jóval pontosabb, precízebb. Ez az eljárás az adaptivitás, mely alkalmazásával részben vagy egészében személyre szabható a feladatsor, és a tanulók számára nagyobb információértékű visszajelzés adható. Eredményeink szerint elsősorban az alacsony és a magas képességtartományok esetén jelentősen nagyobb a közvetített információ mennyisége, vagyis a mérés precizitása.

A megváltozott technológia a 21. században kulcsfontosságúnak számító képességek autentikus vizsgálatát is lehetővé teszi, így például a korábbi önbevalláson alapuló kérdőívek használata helyett/mellett az IKT műveltség valódi fejlettségi szintjének feltérképezése is megvalósulhat. A folyamatban lévő, valós környezetet szimuláló mérőeszközön alapuló kutatás jelentős eredményekkel járulhat hozzá a nemzetközi vonatkozó kutatásokhoz.

A TBA lehetővé teszi egyrészt már korábban is kutatott területek újszerű mérését-értékelését is, így a kreativitás innovatív, a 21. század elvárásainak megfelelő, azonnali

A tanulmányban ismertetett kutatások azt bizonyítják, hogy az online tesztelés egyrészt olyan tudás- és képességterületek mérését teszi lehetővé, amelyek korábban papír alapon nem voltak kivitelezhetőek, másrészt korábban is kutatott területek innovatív mérési lehetőségét teremti meg. A bemutatott kutatások eredményei szerint a papír alapú mérésről az online mérésre való átállás nem okoz jelentős mértékű változást a diákok teljesítményében, ugyanakkor előnyös, ha a tanulók minél fiatalabb korban ismerkednek meg a technológia alapú mérésekkel, és gyakorolják be használatukat.

viSSzacsatolást biztosító vizsgálatát, másrészt korábban hagyományos technikákkal nem vizsgálható képességek fejlődésének feltérképezését. A kollaboratív problémamegoldó képesség online mérésének kiemelkedő jelentősége van a mai kor elvárásainak való megfelelés szempontjából. A felgyorsult gazdasági és társadalmi közeg újszerű, eredeti gondolatokkal rendelkező egyéneket keres, akik váratlan helyzetekben is feltalálják magukat, és képesek másokkal együttműködve bármilyen felmerülő nehézség esetén megoldást találni. A képességterület online mérésével egyrészt átlépjük az egyéni tesztelés adta korlátokat, másrészt megvalósítjuk a képességterület objektív, minden csoporttag számára azonos feltételeket biztosító vizsgálatát.

A bemutatott új generációs technológia alapú tesztelési lehetőségek alkalmazása jelentős mértékű változást eredményezhet a pedagógiai mérés-értékelésben, és hozzájárulhat az új technológiák egyre szélesebb körben való elterjedéséhez.

A kutatást a TÁMOP 3.1.9/11 kutatási program támogatta.

Irodalom

- Ainley, J. F., Gebhardt, E. és Schulz, W. (2012): *ACARA National Assessment Program – ICT Literacy Years 6 & 10 Report*. Australian Curriculum, Assessment and Reporting Authority, Sydney.
- Anderson, J. (2010): *ICT Transforming Education. A Regional Guide*. UNESCO, Bangkok.
- Barkóczi I. és Zétényi T. (1981): *A kreativitás vizsgálata*. OPI Kiadó, Budapest.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Martin, R., Miller-Ricci, M. és Rumble, M. (2012): *Defining Twenty-First Century Skills*. In Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and Teaching of 21st Century Skills*. Springer, New York. 17–66.
- 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, Luxemburg. 24–29.
- Cheung, P. C. és Lau, S. (2010): *Gender differences in the creativity of Hong Kong school children: Comparison by using the new electronic Wallach-Kogan creativity tests*. *Creativity Research Journal*, 2. 2. sz. 194–199.
- Chiu, M. M. (2000): *Group Problem-Solving Processes: Social Interactions and Individual Actions*. *Journal for the Theory of Social Behaviour*, 30. 1. sz. 26–49.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): *A papír alapú tesztek a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái*. *Iskolakultúra*, 3-4. sz. 3–16.
- Csapó Benő, Molnár Gyöngyvér, Pap-Szigei Róbert és R. Tóth Krisztina (2009): *A mérés-értékelés új tendenciái: a papír és számítógép alapú tesztelés összehasonlító vizsgálatai általános iskolás, illetve főiskolás diákok körében*. In: Perjés István és Kozma Tamás (szerk.): *Új kutatások a neveléstudományokban. Hatékony tudomány, pedagógiai kultúra, sikeres iskola*. Magyar Tudományos Akadémia, Budapest. 99–108.
- Csapó, B., Ainley, J., Bennett, R., Latour, T., és Law, N. (2012): *Technological issues of computer-based assessment of 21st century skills*. In Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and Teaching of 21st Century Skills*. Springer, New York. 143–230.
- Csapó, B., Lőrincz, A. és Molnár, G. (2012): *Innovative assessment technologies in educational games designed for young students*. In D. Ifenthaler, D. Eseryel, X. Ge (szerk.): *Assessment in game-based learning: foundations, innovations, and perspectives*. Springer, New York. 235–254.
- Csapó Benő és Molnár Gyöngyvér (2012): *Gondolkodási készségek és képességek*. In Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 407–440.
- CTB/McGraw-Hill (2003): *The Computer-Based or Online Administration of paper-Pencil Tests*. *Journal of Technology, Learning, and Assessment*, 2. 1. sz.
- Gallagher, A., Bridgeman, B. és Cahalan, C. (2000): *The effect of computerbased tests on racial/ethnic, gender, and language groups* (GRE Board Professional Report No. 96–21P). Educational Testing Service, Princeton, NJ.
- Greiff, S. (2012): *From interactive to collaborative problem solving: Current issues in the Programme for*

- International Student Assessment. *Review of Psychology*, **19**. 2. sz. 111–121.
- Greiff, S., Wustenberg, S., Holt, D. V., Goldhammer, F. és Funke, F. (2013): Computer-based assessment of Complex Problem Solving: concept, implementation, and application. *Educational Technology Research and Development*, **61**. 407–421.
- Griffin, P. és Care, E. (2012): Challenges in internet-based CPS Assessment. Előadás: Conference of the International Testing Commission. Amszterdam, 2012. július 5.
- Guilford, J. P. (1967): *The nature of human intelligence*. McGraw-Hill, New York.
- Gunawardena, C. N., Lowe, C. A. és Anderson, T. (1997): Analysis of Global Online Debate and the Development of an Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing. *Journal of Educational Computing Research*, **17**. 4. sz. 397–431.
- Henri, F. (1992): Computer Conferencing and Content Analysis. In Kaye, A. R. (szerk.): *Collaborative Learning through Computer Conferencing*. Springer Verlag, Berlin. 117–136.
- Hsieh, I.-L. és O’Neil, H. F., Jr. (2002): Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior*, **18**. 1. sz. 699–715. The International Test Commission (2006): International Guidelines on Computer-Based and Internet-Delivered Testing. *International Journal of Testing*, **6**. 2. sz.
- 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.
- Kárpáti Andrea, Molnár Gyöngyvér, Tóth Péter és Főző Attila (2008, szerk.): *A 21. század iskolája*. Nemzeti Tankönyvkiadó, Budapest.
- Katz, I. R. és Macklin, A. S. (2007): Information and Communication Technology (ICT) Literacy: Integration and Assessment in Higher Education. *Systemics, Cybernetics and Informatics*. **5**. 4. sz. 50–55.
- Keng, L. (2008): *A Comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas, Austin.
- Kozma, R. (2009): Assessing and teaching 21st century skills: A call to action. In F. Schueremann és J. Bjornsson (szerk.): The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment. European Communities, Brussels. 13–23.
- Kwon, M., Goetz, E. T. és Zellner, R. D. (1998): Developing a computer-based TTCT: Promises and problems. *Journal of Creative Behavior*, **32**. 2. sz. 96–106.
- Lau, S., és Cheung, P. C. (2010): Creativity assessment: Comparability of the electronic and paper-and-pencil versions of the Wallach–Kogan Creativity Tests. *Thinking Skills and Creativity*, **5**. 3. sz. 101–107.
- Law, N., Lee, Y. és Yuen, H. K. (2010): The impact of ICT in Education policies on teacher practices and student outcomes in Hong Kong. In Scheuermann, F. és Pedró, F. (szerk.): *Assessing the effects of ICT in Education – indicators, criteria and benchmarks for international comparisons*. Publications Office of the European Union/OECD, Luxembourg. 143–164.
- Luecht, R. M. és Sireci, S. G. (2011): *A Review of models for computer-based testing*. Research Report 2011–12. College Board.
- Magyar Andrea (2012): Számítógépes adaptív tesztelés. *Iskolakultúra*, **6**. sz. 52–60.
- Magyar Andrea (2013a): Többszakaszos adaptív tesztek. *Oktatás-Informatika*. megjelenés alatt.
- Magyar Andrea (2013b): Adaptív tesztek készítésének folyamata. *Iskolakultúra*. megjelenés alatt.
- Magyar Andrea és Molnár Gyöngyvér (2013): Adaptív és rögzített formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*. **113**. 3. sz. 181–193.
- Mayrath M. C., Clarke-Midura J. és Robinson D. H. (2012): Introduction to Technology-Based assessments for 21st Century Skills. In Mayrath, M. C., Clarke-Midura, J. és Robinson, D. H. (szerk.): *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research* (Current Perspectives on Cognition, Learning, and Instruction). Information Age Publishing, USA. 1–12.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, **7–8**. sz. 22–34.
- Molnár Gyöngyvér (2011): Az információs-kommunikációs technológiák hatása a tanulásra és oktatásra. *Magyar Tudomány*, **9**. sz. 1038–1047.
- Molnár Gyöngyvér és Kárpáti Andrea (2012): Informatikai műveltség. In: Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 441–476.
- Molnár Gyöngyvér (2013): *A Rasch-modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér és Csapó Benő (2013): Az eDia online diagnosztikus mérési rendszer. XI. Pedagógiai Értékelési Konferencia. Szeged, 2013. április 11–13. 82.
- Molnár, Gy., Greiff, S., Csapó, B. (2013): Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity*, **9**. 8. sz. 35–45.

- Molnár, Gy., Greiff, S., Wüstenberg, S. és Fischer, A. (2014): Empirical study of computer based assessment of domain-general dynamic problem solving skills. In Funke, J. és Csapó, B. (szerk.): *The Nature of Problem Solving*. OECD, Paris. Benyújtott kézirat.
- NAGB (é. n.): Technology and Engineering Literacy Framework for the 2014 National Assessment of Educational Progress. Pre-Publication Edition. WestEd. 2013. 10. 30-i megtekintés, http://www.edgateway.net/cs/naepsci/download/lib/249/prepub_naep_tel_framework.pdf?x-r=pcfile_d.
- OECD (2013): PISA 2015 Draft Collaborative Problem Solving Framework. 2013. 10. 30-i megtekintés, <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>
- O'Neil, H. F., Chuang, S. és Chung, G. K. W. K. (2003): Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, 10. sz. 361-373.
- Palaniappan, A. K. (2012): Web-based Creativity Assessment System. *International Journal of Information and Education Technology*, 2. 3. sz. 255-258.
- Partnership for 21st Century Skills (2008): Overview. Framework for 21st Century Learning. Partnership for 21st Century Skills. 2013. 10. 30-i megtekintés, <http://www.p21.org/overview>.
- Pearson Educational Measurement (2003): Virginia standards of learning web-based assessments comparability study report – Spring 2002 administration: Online & paper tests. Austin.
- Peak, P. (2005): Recent Trends in Comparability Studies. Pearson Educational Measurement. 2013. 10. 30-i megtekintés, http://www.pearsonassessments.com/NR/rdonlyres/5FC04F5A-E79D-45FE-8484-07AACAE2DA75/0/TrendsCompStudies_rr0505.pdf.
- Pearson Educational Measurement (2003): Virginia standards of learning web-based assessments comparability study report – Spring 2002 administration: Online & paper tests. Austin.
- Piirto, J. (2011): *Creativity for 21st century skills. How to embed creativity into the curriculum*. Sense Publisher, Rotterdam.
- Plucker, J. A. és Renzulli, J. S. (1999): Psychometric approaches to the Study of Human Creativity. In R. J. Sternberg (szerk.). *Handbook of Creativity*. Cambridge University Press, London. 35–62.
- Pretz, J. E. és Link, J. A. (2008): The Creative Task Creator: A tool for the generation of customized, Web-based creativity tasks. *Behavior research methods*, 40. 4. sz. 1129–1133.
- Psychological Testing Centre (2006): Using online assessment tools for recruitment. The British Psychological Society, Leicester.
- Reckase, M. D. (2009): Computerized adaptive testing using MIRT. In *Multidimensional item response theory*. Springer, Dordrecht. 311–339.
- Rosen, Y. és Tager, M. (2013): Computer-based Assessment of Collaborative Problem Solving Skills: Human-to-Agent versus Human-to-Human Approach. Pearson Education. 2013. 10. 30-i megtekintés, <http://researchnetwork.pearson.com/wp-content/uploads/CollaborativeProblemSolvingResearchReport.pdf>
- 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.
- Schroeders, U. és Wilhelm, O. (2011): Equivalence of Reading and Listening Comprehension Across Test Media. *Educational and Psychological Measurement*, 71. 5. sz. 849–869.
- Tian J., Miao D., ZhuXia, G. J. (2007): An Introduction to the computerized adaptive testing. *US-China Education Review*, 4. 1. sz. 72–81.
- Tongori Ágota (2012): Az IKT műveltség fogalmi keretének változása. *Iskolakultúra*, 22. 11. sz. 34–47.
- Tongori Ágota (2013): Az információs és kommunikációs technológiai műveltség mérési lehetőségei. *Új Kép*, 1–2. sz. 12–19.
- Torrance, E. P. (1966): *Torrance Tests of Creative Thinking*. IL: Scholastic Testing Service, Bensenville.
- Wallach, M. A. és Kogan, N. (1965): *Modes of thinking in young children: A study of the creativity-intelligence distinction*. Holt, Rinehart and Winston, New York.
- Wang, S. és Shin, C. D. (2009): Comparability of Computerized Adaptive and Paper-Pencil Tests. Test, Measurement & Research Service. *Bulletin*, 13. sz. 1–7.
- Zelman, M., Shmis, T., Avdeeva, S., Vasiliev, K and Froumin, I. (2011): *International Comparison of Information Literacy in Digital Environments*. Conference papers. International Association for Educational Assessment (IAEA) Manila 37th annual conference.. 2013. 10. 30-i megtekintés, http://www.iaea.info/documents/paper_30e43f54.pdf
- Zenisky, A., Hambleton, R. K. és Luecht, R. M. (2010): Multistage testing: Issues, designs and research. In: der Linden, W. J. és Glas, C. A. W. (szerk.): *Elements of adaptive testing*. Springer, New York. 355–372.
- Zétényi Tamás (1989): *A kreativitás-tesztek teszt-könyve I. és II.* Munkalélektani Koordinációs Tanács Módszertani Sorozata 22. sz. kötet, Munkaügyi Kutatóintézet, Budapest.