



A SZÓOLVASÁSI KÉSZSÉG ONLINE MÉRÉSÉRE KIDOLGOZOTT ADAPTÍV ÉS LINEÁRIS TESZTRENSZER ÖSSZEHASONLÍTÓ HATÉKONYSÁGVIZSGÁLATA

Magyar Andrea* és Molnár Gyöngyvér**

* *Hódmezővásárhelyi Liszt Ferenc Ének-zenei Általános Iskola*

** *SZTE Neveléstudományi Intézet, Oktatásméleti Kutatócsoport*

A 20. században leginkább elfogadott és elterjedt papír alapú (PP) tesztekre alapuló mérések fejlesztése egyre több korlátba ütközött, a papír alapú tesztekre alapozott fejlesztés lehetőségei fokozatosan kimerültek (*Scheuermann és Björnsson, 2009*). A technológia rohamos fejlődése a továbblépés irányát egyértelműen a számítógépes tesztesztelésre való áttérés lehetőségében jelöli meg (*Csapó, Ainley, Bennett, Latour és Law, 2012; Molnár, 2010; Molnár és Latour, 2011; Scheuermann és Pereira, 2008*), ami számos új lehetőséget kínál a papír alapú tesztesztelés szemben. Ilyen például a motiválóbb környezet (*Thompson és Pometric, 2007*), az azonnali kiértékelés lehetősége (*Wang, 2010*), az innovatív, multimédiás elemeket is tartalmazó dinamikusan változó itemek megjelenítése (*Greiff, Wüstenberg és Funke, 2012*), illetve a személyre szabott, adaptív tesztesztelés megvalósítása (*Eggen és Straetmans, 2000*).

Az adaptív tesztesztelés (CAT – *Computerized Adaptive Testing; Weiss, 2011; Magyar, 2012*) alkalmazása során a teszt feladatai nem előre meghatározott, lineáris sorrendben követik egymást, hanem azokat egy feladatbankból választja a rendszer a tesztmegoldó képességszintjéhez igazítva, a tesztelt személy korábbi feladatokon nyújtott teljesítménye alapján. Amennyiben a tanuló meg tudja oldani a tesztesztelés során kapott feladatot, a következő lépésben egy nehezebbet kap, ha nem, akkor könnyebbet. A tesztesztelés során az előre meghatározott algoritmus biztosítja a megfelelő nehézségi szintű feladatok kiválasztását, majd a tesztesztelés végén minden tanulóhoz hozzárendel egy képességszintet. Miután a képességparaméterek és az itemek nehézségi indexei egymás által definiáltak, azaz azonos skálán jellemezhetők (*Molnár, 2013*), így a diagnosztizált képességszintnél könnyebb feladatokat nagyobb valószínűség mellett old meg helyesen a tesztelt személy, mint helytelenül. A tesztesztelés során nemcsak itemszintű lehet az elágazás, hanem részteszt szintű is, ahol a részteszten belül előre meghatározott sorrendben közvetítik ki a közel azonos nehézségű itemeket. Ezt a tesztesztelési eljárást többszakaszos adaptív tesztesztelésnek nevezzük (MST – *Multi Stage Testing; Magyar, 2013*).

A teszt típus egyesíti magában a hagyományos lineáris és az adaptív teszteszt tulajdonságait, mivel egyrészt a kérdéseket a tanuló képességszintjéhez igazítja, másrészt lehetőséget ad az itemek sorrendjének bizonyos kereteken belül történő előzetes meghatározására (*Jodoin, Zenisky és Hambleton, 2006; Zheng, 2012; Magyar, 2014a*). A modulok

előre tervezhetőek és szerkeszthetőek, így nagyobb kontrollt biztosítanak a tesztadminisztráció számára. Például kiküszöbölhetővé válik, hogy az egyik feladat megoldása információt adjon egy másik, azonos modulon belül lévő feladat megoldásához, azaz a részteszteken belül kizárható a méréselméleti mutatókat egyértelműen rontó itemfüggőség. További fontos előnyük, hogy a modulokon belül a tanulóknak lehetőségük van a visszalépésre és a javításra. Mivel adaptivitás csak a modulok között valósul meg, így ez nem veszélyezteti a teszt algoritmusát és segíti a tanulókat a minél magasabb pontszám elérésében. Az item alapú adaptív tesztekhez képest jóval kevesebb adminisztrációt és számítógépes számítást igényel ez a kiközvetítési mód, ami gördülékenyebbé teszi az adatfelvétel menetét (Hendrickson, 2007; Zheng, 2012).

Ez a típusú tesztelési mód a hagyományos, mindenki számára azonos itemeket azonos sorrendben tartalmazó, lineáris tesztekkel szemben a teljesítmények sokkal pontosabb mérését teszi lehetővé (Magyar és Molnár, 2013), illetve jelentős mértékben megnöveli a tesztelés során kinyerhető itemekre és személyekre vonatkozó információ nagysága (Molnár, 2013; Magyar és Molnár, 2013; Molnár, 2006). Elhanyagolhatóvá válik annak valószínűsége, hogy a tesztelt személyek ugyanazon feladatokat ugyanabban a sorrendben kapják, azaz növekszik a tesztelés biztonsága (Wainer, 2000). Mindez új lehetőségeket teremt a mérés-értékelés területén.

A tanulmány összegzi az adaptív tesztek lineáris teszteléssel való összevetésére vonatkozó szakirodalmat és bemutat 16, az ezredforduló óta megvalósult kutatást, melyek különböző típusú adaptív tesztek lineáris tesztekkel való hatékonyságát vizsgálják. A bemutatott kutatások főként szimulált adatbázisokon történtek, empirikus mérés csak kismintákon, egyetemi hallgatók bevonásával valósult meg.

A jelenlegi kutatás ezt a hiányt pótolja. A tanulmány második részében bemutatott nagymintás mérés fő célja 4. és 5. évfolyamos diákok szóolvasási készségének diagnosztizálása hagyományos lineáris, illetve adaptív tesztelési technika alkalmazásával, majd a kétféle tesztelési mód hatékonyságának összehasonlítása. A vizsgálat alcéljai: (1) a lineáris és az adaptív tesztek mérési pontosságának összehasonlítása; (2) a kétféle tesztkörnyezetben a becsült képességszintek évfolyam- és személyszintű összehasonlítása; (3) a különböző tesztelési módokon elért helyes válaszok arányának összevetése; (4) az adaptív tesztelés során kiközvetített itemek, illetve résztesztek nehézségi szintjének, ennek változásmintázatainak jellemzése.

Adaptív tesztelés: nemzetközi összehasonlító hatékonyságvizsgálatok

Az adaptív tesztelés hatékonyságának mérését célzó kutatások alapvetően két csoportba sorolhatóak. A kutatások egyik része médiahatás-vizsgálatot is magában foglal, mert nem azonos médián kiközvetített adaptív és lineáris tesztelést, hanem a számítógép alapú adaptív tesztelést a papír alapú lineáris teszteléssel hasonlítja össze. Miután a számítógép alapú tesztelésre való átállás sem valósult meg még a mérés-értékelés minden egyes szintjén, ezért indokoltak a médiahatás-kutatást is tartalmazó összehasonlító hatékonyságvizsgálatok (Wang és Kolen, 2001; Pásztor-Kovács és mtsai, 2013; Wan és mtsai, 2009). Ezek

biztosítják a különböző médián kiközvetített tesztek összehasonlíthatóságát, melyek egyrészt longitudinális kutatások esetén kapnak kiemelkedő szerepet, ahol szükséges a korábbi papír alapú adatfelvételek eredményeinek számítógép alapú teszteredményekkel való összehasonlíthatósága, másrészt abban az esetben, amikor a kétféle médiumon való tesztelés alternatív módon párhuzamosan folyik (Way, Davis és Fitzpatrick, 2006; Paek, 2005). A professzionális tesztstandardok (APA, 1986; AERA, APA és NCME, 1999; Wang és mtsai, 2008) is hangsúlyozzák a különböző médiumokon elért pontszámok összehasonlíthatóságának fontosságát. Az összehasonlító kutatások fő fókuszában a tesztek mérési pontosságának összehasonlítása áll, illetve annak feltárása, hogy az adaptív tesztelésre való átállás milyen hatással van a tesztelési folyamatra (idő, itemszám) és a különböző képességű egyének eredményeire.

Adaptív és PP tesztek működésének tesztelméleti összehasonlítása különösen nagy kihívás (Wang és Kolen, 2001). Mivel a vizsgázók személyre szabott tesztet kapnak, különbségek lehetnek az itemek tartalmában, az itemek elhelyezkedésében és nehézségében, valamint a pontozásban. Ezek a tényezők jelentősen befolyásolhatják az elemzéseket, melyeket a médiahatás mellett szintén javasolt figyelembe venni (Wang és Kolen, 2001; Kolen, 1999–2000). Wang és Kolen (2001) arra hívják fel a figyelmet, hogy ahhoz, hogy a CAT verzió összehasonlítható legyen a papír alapú verzióval, a CAT szempontjából nagymérvű korlátozottságot jelent, hiszen ebben az esetben az adaptívteszt-fejlesztés során nem lehet az összes számítógép adta lehetőséget kihasználni. Az ezredforduló utáni legjelentősebb, adaptív tesztelésre vonatkozó összehasonlító kutatások módszereit és főbb eredményeit összegzi az 1. táblázat. A kutatásokat minta, alkalmazott mérőeszköz, az elemzés során alkalmazott eljárás, módszer, valamint a kapott eredmény szerint csoportosítottuk.

A 16 kutatás közül 12 esetben végeztek számítógépes adaptív és a lineáris teszt összehasonlítását, melyek közül három esetben item alapú adaptív és papír alapú lineáris (Al-A'ali, 2007; Thompson és Way, 2007; Frey, Seitz és Kröhne, 2011), szintén három esetben item alapú adaptív és számítógép alapú lineáris (Olea, Revuelta, Ximenez és Abad, 2000; Kingsbury, 2004; Pyper és Lilley, 2010), egy esetben MST és papír alapú lineáris (Rotou és mtsai, 2003), két esetben item alapú adaptív, MST és számítógép alapú lineáris (Hambleton és Xing, 2006; Zheng, 2012), valamint három kutatásban MST és számítógép alapú lineáris (Jodoin, Zenisky és Hambleton, 2006; Guille és mtsai, 2011; Brossman és mtsai, 2014) tesztek működésének összehasonlítása történt. Öt esetben különböző típusú adaptív tesztek működését hasonlították össze a kutatók (Vispoel, Hendrickson és Bleiler, 2000; Rotou és mtsai, 2003; Keng, 2008; Jiban és mtsai, 2008; Crofts és mtsai, 2013).

A bemutatott 16 vizsgálat közül 11 szimulált adatbázison történt, csupán öt vizsgálat zajlott empirikus adatok felhasználásával. A minta nagysága 45 és 284 tanuló között mozgott (Vispoel, Hendrickson és Bleiler, 2000; Olea és mtsai, 2000; Al-A'ali, 2007; Jiban és mtsai, 2008; Pyper és Lilley, 2010), ebből négy esetben egyetemi hallgatók körében valósult meg a mérés (Vispoel, Hendrickson és Bleiler, 2000; Olea és mtsai, 2000; Al-A'ali, 2007; Pyper és Lilley, 2010). MST és papír alapú lineáris teszt empirikus összehasonlító vizsgálata fiatal tanulók körében nem történt.

1. táblázat. Adaptív teszteken végzett összehasonlító vizsgálatok

<i>Publikáció</i>	<i>Minta</i>	<i>Vizsgált mérőanyag</i>	<i>Elemzési módszer</i>	<i>Eredmények</i>
<i>Vispoel, Hendrickson és Bleiler (2000)</i>	242 egyetemi hallgató	Szókincsvizsgálat: 40 itemű különböző szerkezetű CAT	Átlagok összehasonlítása	Visszalépést engedve a teszten jelentősen javultak az átlagok.
<i>Olea, Revuelta, Ximénez és Abad (2000)</i>	184 egyetemi hallgató	Szókincsvizsgálat: 20 itemű CB és 20 itemű CAT	ANOVA, teszt információ, standard error	A CB teszteken jobb eredményeket értek el a tanulók, a mérési hiba kisebb volt az adaptív teszt esetében.
<i>Rotou, Patsula, Manfred és Rizavi (2003)</i>	10 000 fős szimulált adatbázis	55 itemű PP és 54 itemű MST 33 itemű MST és 32 itemű itemalapú CAT	Reliabilitás, standard error	Az MST mért a legpontosabban, a legkisebb mérési hibával.
<i>Kingsbury és Hauser (2004)</i>	NCLB 2003-as mérés eredményei alapján szimulált adatbázis	CAT és lineáris CB matematika és szövegértési tesztek	Teszt információ, standard error	Az adaptív teszt minden képességszinten több információt szolgáltatott, mint a lineáris teszt.
<i>Hambleton és Xing (2006)</i>	Szimulált adatbázis	MST, CAT és lineáris random kiosztású CB teszt	Reliabilitás, teszt információ	A tanulók képességszintjéhez igazított tesztek pontosabban mértek, a legpontosabban a CAT mért. Legrosszabb reliabilitás a random teszt kiosztás esetén volt.
<i>Jodoin, Zenisky és Hambleton (2006)</i>	5000 fős szimulált adatbázis	60 itemű MST, 40 itemű MST, 60 itemű lineáris CB tesztváltozat	Valódi és a kapott képességszintek közötti korrelációja, reliabilitás, teszt információ	Az MST tesztek közel azonos reliabilitásúak voltak, mindkettő pontosabban mért, mint a lineáris teszt.
<i>Thompson és Way (2007)</i>	Szimulált adatbázis	20-60 itemű CAT, 35 itemű PP teszt	Valódi és a kapott képességszintek közötti korrelációja, reliabilitás, teszt információ	Adaptív tesztekkel nagyobb mérési precizitás érhető el, és nagyban csökkenthető az itemszám (60-ról 55-re).
<i>Al-A'ali (2007)</i>	45 tanuló	14-20 itemű CAT, 25 itemű PP	Teszt információ	Az adaptív teszt esetén kevesebb item elegendő ugyanazon mérési pontosság eléréséhez.
<i>Keng (2008)</i>	10000 fős szimulált adatbázis	42 itemű tesztlet szintű CAT, item szintű tesztlet CAT, MST	Valódi és a kapott képességszintek közötti eltérés, AAD, RMSE, standard error, teszt információ	A legprecízebben az item szintű CAT mért, legrosszabb pontossággal a tesztlet szintű CAT. Alacsony képességűek esetében a CAT mért legprecízebben, átlagos képességűeknél az MST.
<i>Jiban, Ayodele, McCarthy és Christ (2008)</i>	287 kisiskolás	20 és 40 itemből álló olvasáskészséget mérő itemalapú CAT	RReliabilitás, teszt információ, standard error	20 item fölött nem változott jelentősen a teszt információ és a mérési hiba nagysága.

1. táblázat folytatása

Publikáció	Minta	Vizsgált mérőanyag	Elemzési módszer	Eredmények
<i>Pyper és Lilley</i> (2010)	180 egyetemista	20 itemű CAT, 20 itemű lineáris CB teszt	Korreláció, t-próba	A teljesítményekben magas korreláció a két teszt között.
<i>Guille, Becker, Zhu, Zhang, Song és Sun</i> (2011)	6287 fős szimulált adatbázis	MST és CB teszt	Reliabilitás, standard error	Elsősorban a magas képességű egyéneknél mérséklődött szignifikánsan a tesztelési idő.
<i>Frey, Seitz és Kröhne</i> (2011)	14624 PISA teszt eredménye alapján szimulált adatbázis	PISA 2000, 20003 és 2006-os tesztek alapján PP és CAT szimuláció	Reliabilitás	Az eredmények szerint a mérés hatékonysága (Mérési precizitás/prezentált itemek száma) 74%-kal nőtt, a szükséges itemszám a PP tesztnél szükséges 55-ről 26-ra csökkent, és a tesztelés időtartama 120 percről 57 percre csökkent.
<i>Zheng</i> (2012)	Szimulált adatbázis	600 itemű feleletválasztós kérdések nyolc különböző területről, CAT, különböző szerkezetű MST-k és lineáris CB	Teszt információ, RMSE, korreláció	Az MST mérési precizitása megegyezett a CAT precizitásával, az MST hatékonyabb volt az itembank kihasználtságát tekintve.
<i>Crotts, Zenisky, Sireci és Li</i> (2013)	Szimulált adatbázis	35 és 40 itemű olvasás-szövegértés MST	Reliabilitás, teszt információ, standard error	A tesztrövidítés nem változtatott jelentősen a teszt reliabilitásán, a tesztelés időtartama 12–24%-kal csökkent.
<i>Brossman és Guille</i> (2014)	Szimulált adatbázis	MST és lineáris CB teszt	Teszt információ, standard error, RMSE	A mérési hiba kisebb az MST esetén.

A kutatások jelentős része a különböző összeállítású, szerkezetű MST-tesztek működését, mérési hatékonyságát, pontosságának összehasonlítását vette górcső alá. Az MST-tesztrendszerek első modulja általában egy közepes nehézségű modul, amit 2–5 ágon 2–6 szakasz követ. Az eredmények szerint a szakaszok és modulok számának növelése emelte a teszt mérési precizitását, ezért egyértelműen kettőnél több szakasz alkalmazását javasolták a kutatások, ami kiküszöbölte a tanulók esetlegesen hibás szintre történő induló besorolását is. Túl sok szakasz alkalmazása azonban indokolatlanul megnövelte a teszt hosszát és az adatfelvétel idejét, miközben ez nem járt a mérési precizitás arányos növekedésével. A bemutatott kutatások esetén az 1-3 (*Rotou és mtsai*, 2003), 1-2-3-4, 1-2-4 (*Zheng*, 2012), 1-3-3 (*Keng*, 2008), 1-2-2, 1-3-3, 1-2-3, 1-3-2 (*Jodoin, Zenisky és Hambleton*, 2006), 5-5-5-5-5-5 (*Crotts és mtsai*, 2013) és 1-3-3-3-3 (*Brossman és Guille*, 2014) szerkezetű MST-tesztek fordultak elő. A kutatások egy részében, amennyiben az alkalmazott itembank mérete engedte, a tesztbiztonság növelése érdekében több ekvivalens tesztváltozatot, illetve modulát állítottak össze, melyeket random módon osztottak ki (*Brossman és Guille*, 2014;

Crotts és mtsai, 2013). A korai elemzésekben a teszteredmények összehasonlítását klaszterizációs teszelméleti módszerek alkalmazásával végezték (ANOVA-elemzések, átlagok összehasonlítása; Vispoel, Hendrickson és Bleiler, 2000; Olea és mtsai, 2000), azonban a későbbiekben általánossá vált a valószínűségi teszelméleti módszerek, eljárások alkalmazása, például az item- és tesztinformációk összehasonlítása. A tesztek mérési precizitását jellemzi a reliabilitás és a mérési hiba (SE – *standard error*). A szimuláción alapuló kutatásokban gyakran használt párhuzamos mutató a valódi és a mért képességszámok korrelációs mérőszáma, az RMSE (*Root Mean Square Error*), valamint az AAD indexek (*Average Absolute Difference*), melyek a valódi és a mért képességszámok eltéréseinek jellemzésére alkalmas mutatószámok (Keng, 2008). Jiban és munkatársai (2008) a különböző tesztelési módok esetén elért helyes válaszok arányát vizsgálta és hasonlította össze. Az eredmények szerint az adaptív tesztek alkalmazásával jelentősebb mérési precizitás volt elérhető, magasabb volt a tesztelés reliabilitása és minden képességszinten több információt szolgáltatott, mint a lineáris tesztek. Ezáltal csökkenthető volt a kiköszvetített itemek száma, ami a mérési idő csökkenését vonta maga után. Ezek az előnyök mindkét típusú (item alapú és MST) esetén is kimutathatóak voltak, azaz bármelyik típusra való átállás vonzó lehet. A mérési eredmények szerint adaptív tesztelés esetén az alacsonyabb képességtartományban jelentősen növekedett a helyes megoldások száma, viszont a magas képességű személyeknél jóval kevesebb helyes válasz született, mivel az alacsony képességű tanulók könnyebb feladatokat, a magasabb képességűek nehezebbeket kaptak (Jiban és mtsai, 2008).

A szóolvasási készség számítógép alapú lineáris és adaptív tesztelési technikával történő diagnosztizálásának nagymintás összehasonlító hatékonyságvizsgálata

Minta és mérőeszköz

A mérések megvalósításához a Nagy József és munkatársai (2004) által kidolgozott papír alapú szóolvasó készség kritériumorientált mérését lehetővé tevő tesztsorozatot használtuk fel. A kutatás első fázisában megtörtént a rendszer számítógépes alapra helyezése, mely során megtartottuk a rendszer alapvető struktúráját, vagyis a szóolvasó készség vizsgálatának online formája továbbra is tartalmazza a négy alapvető dimenzió (címszó-olvasás, toldalékoszó-olvasás, szinonimaolvasás és szójelentés-olvasás) mindegyikének mérését. A feladatokat digitalizáltuk és négyszakaszos adaptív tesztrendszerként építettünk ki belőlük (Magyar és Molnár, 2014; Magyar, 2014b; l. 7. ábra).

Az adaptív tesztrendszer kifejlesztésének lépései a következők voltak: (1) a papír alapú feladatok digitalizálása, feltöltése az eDia-rendszerbe, (2) az eredeti lineáris rendszer újrastrukturálása, horgonyrészesztek alkalmazásával, (3) az online lineáris tesztek felvétele, majd az eredmények alapján a feladatok skálázása, (4) az eredeti papír alapú rendszer köztöttségeihez leginkább igazítható, ugyanakkor a 21. században elvárható mérés-értékelési céloknak leginkább megfelelő adaptív rendszer kiválasztása és a közös nehézségi skálán

jellemzett feladatok többszakaszos adaptív tesztrendszerre alakítása, végül (5) a rendszer viselkedésének kismintán történő tesztelése.

A kismintás kutatás eredményei alapján a négyszakaszos adaptív rendszer megbízhatóan működött, alkalmasnak bizonyult az 1–5. évfolyam szóolvasási készségének mérésére. Az alacsonyabb képességszintű diákok tipikusan a könnyebb, a magasabb képességszintűek a nehezebb részteszteket kapták a teszteléskor. Az adatfelvétel során kinyert információ mennyisége fokozatosan javult, a mérési hiba mértéke elfogadható határon belül maradt.

A médiahatás (*Wang és Kolen, 2001; Hülber és Molnár, 2013*) kiküszöbölése érdekében az adaptív rendszer hatékonyságvizsgálatát nem a papír alapú rendszerrel történő adatfelvétel eredményeivel összevetve végeztük el, hanem az itembank paraméterezéséhez összeállított lineáris tesztrendszer számítógép alapú nagymintás mérésének adataival. Ezen eljárás alkalmazásával küszöböltük ki az eredményeket esetlegesen befolyásoló médiahatást.

Mindkét teszt 85 feladatból állt és az eredeti rendszernek megfelelően öt különböző szinten különítette el a diákokat: előkészítő (0–59%), kezdő (60–69%), haladó (70–79%), befejező (80–89%) és optimális szint (90–100%). A tanulók különböző szintekbe való sorolása a teszten nyújtott százalékos teljesítmény függvényében történt, tehát az a tanuló volt optimális szinten, aki a szavak legalább 90%-át felismerte (*Magyar és Molnár, 2014*).

A nagymintás adatfelvétel 2014 tavaszán 4. és 5. évfolyamos diákok részvételével zajlott. Az elemzésben azon 3220 tanuló ($N_{4.évf.}=1706$, $N_{5.évf.}=1514$) eredményét használtuk fel, akik mindkét mérésben részt vettek, mivel így személyszintű összehasonlításra is lehetőség nyílt.

Adatfelvétel és eljárások

Az első adatfelvétel során a tanulók mindegyike a lineáris tesztrendszer egyik változatát oldotta meg. A második adatfelvétel egy hónap múlva történt, amikor minden tanuló az adaptív tesztrendszer feladatait kapta. Mivel a lineáris tesztnél 40 különböző tesztváltozat volt, az adaptív teszt esetén 30 különböző útvonal és útvonalanként 2–5 tesztváltozat volt elérhető, ezért elhanyagolható volt annak esélye, hogy egy tanuló ugyanazt a tesztet kapja. A diákok számára rendelkezésre álló idő mindkét esetben 45 perc volt. Az adatfelvétel az eDia-plaformon keresztül a tanulók saját iskolájában, az iskolák saját internethálózatán keresztül történt. A tanulók képességszintjének megállapítása a Rasch-modell segítségével történt, majd a logitegységben adott értékeket egy 500 pontos átlagú és 100 pontos szórású skálára transzformáltuk.

Az elemzések során a tesztek összehasonlítására használtuk a klasszikus tesztelméleti elemzéseket (átlagok, szórások összehasonlítása, a két teszt közötti korreláció meghatározása) és a valószínűségi tesztelméleti modellt (Rasch-modell). Az adatok elemzését a ConQuest program felhasználásával többdimenziós parciáliskredit-moddellel végeztük el. A többdimenziós elemzésnél a címszóolvasást, a toldalékoszó-olvasást, a szinonimaolvasást és a szójelentés-olvasást a szóolvasó készség különböző dimenzióinak tekintettük. A parciáliskredit-modell használatát az indokolta, hogy bár a szavak szintjén külön kellett

döntést hozni a diákoknak, mégis a feladatok bizonyos szinten összekötötték az egy feladaton belül szereplő szavakat.

A tesztekben kinyerhető információk nagyságának megállapítása is a Rasch-moddal történt. A tesztinformációs görbék a tesztből kinyert információ nagyságát a tesztet megoldó tanulók átlagos képességszintje és az itemek nehézségi szintje közötti különbségek segítségével jellemzik. A kinyert információ nagyságát akkor tekintettük maximálisnak, ha a feladatok nehézségi szintje és az azokat megoldó diákok képességszintje azonos. Minél távolabb volt egymástól ez a két érték, annál kisebb volt a tesztelés során kinyert információ nagysága.

Az empirikus vizsgálat eredményei

A tesztek reliabilitása

Az eredmények kiterjeszhetőségének, általánosíthatóságának körét első szinten jól jellemzi a teszt reliabilitásmutatójának értéke, aminek meghatározására a WLE személyszeparációs reliabilitásmutatót használtuk mindkét tesztváltozat esetében. Az adaptív teszt reliabilitásmutatója 0,92, ami magasabb, mint a lineáris teszt megbízhatósági mutatója (0,90). A 2. táblázat dimenziók szerinti bontásban mutatja a lineáris és az adaptív teszteken mért reliabilitásokat.

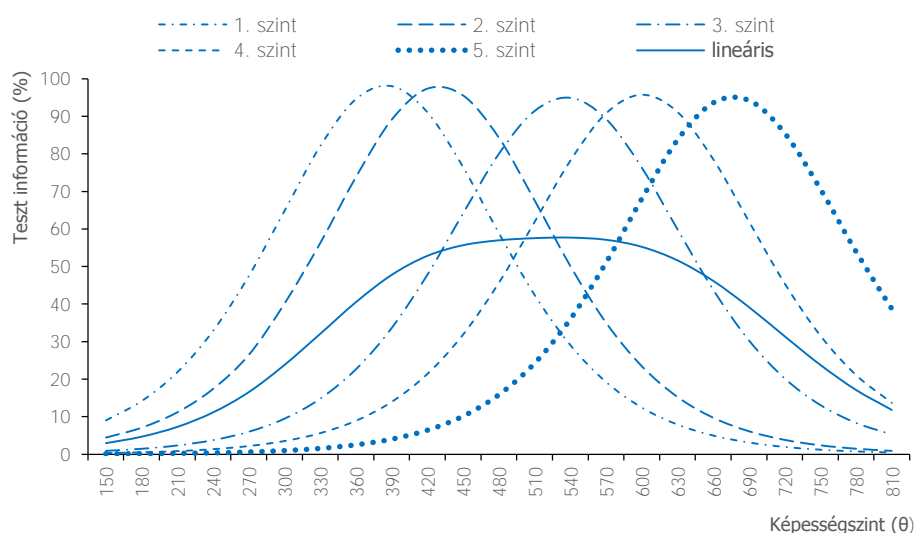
2. táblázat. A tesztek WLE személyszeparációs reliabilitása dimenziónkénti bontásban

Dimenzió	WLE személyszeparációs reliabilitás	
	Lineáris teszt	Adaptív teszt
Címszóolvasás	0,73	0,73
Toldalékoszó-olvasás	0,81	0,85
Szinonimaolvasás	0,88	0,91
Szójelentés-olvasás	0,87	0,90

A reliabilitásmutatók alapján megállapítható, hogy mindkét tesztrendszer megbízhatóságát tekintve alkalmas 4. és 5. évfolyamos diákok szóolvasó készségének mérésére. Az adaptív teszt reliabilitása három dimenzióban (toldalékoszó-olvasás, szinonimaolvasás, szójelentés-olvasás) magasabb, a címszóolvasás esetén megegyező a lineáris teszttel. Mivel a címszóolvasási feladatokat képességtől függetlenül minden tanuló megkapta az adaptív tesztnél is – hiszen a kezdő modul véletlenszerűen került kiközvetítésre, képességszinttől függetlenül, csakúgy mint a lineáris teszten –, ezért az azonos reliabilitásmutató megfelel az elvártnak. A többi dimenzióban az adaptív teszt magasabb megbízhatósággal működött.

A lineáris és az adaptív tesztelés során kinyert információ és a mérési hiba nagyságának összehasonlítása

A reliabilitás mellett a teszt mérési pontosságának egy másik mutatója a tesztelés során kinyert információ nagysága, amit a teszteken nyújtott teljesítmények alapján számított tesztinformációk segítségével jellemzünk. Az 1. ábra görbéi grafikusan szemléltetik, hogy már akár a kizárólagosan az 1., 2., 3., 4. és 5. szintű modulok résztesztjeiből összeállított tesztek (a 30 tesztváltozathból csak 5) is több információt szolgáltatott a tesztelés során, mint az egyetlen, sokféle nehézségű feladatot tartalmazó lineáris teszt. A görbék alapján minden képességszinten több információt tudtunk kinyerni adaptív tesztek alkalmazásával (a görbék minden esetben a lineáris teszt által adott információs függvény felett futnak az érintett képességtartományokban).



1. ábra

A lineáris és az adaptív technikával összeállított, azonos nehézségű modulokat tartalmazó tesztek információs függvényei

A mérés során minden tanuló megoldotta az adaptív és a lineáris verziót is, így lehetőség volt az eredmények személyszintű összehasonlítására. A tesztek mindkét változatát megíró 3220 tanuló közül 1927 esetben, vagyis közel a tanulók kétharmadánál volt az adaptív tesztből kinyert információ magasabb, mint a lineáris teszt esetében (1293 tanuló). A különbség átlagosan 2,9% volt, nagysága képességszintenként különbözött. A 3. táblázat mutatja a kinyert információk mennyiségét képességszintenkénti bontásban. Minden képességszinten több információ volt a tesztből kinyerhető adaptív teszt esetén, mint a lineáris tesztnél.

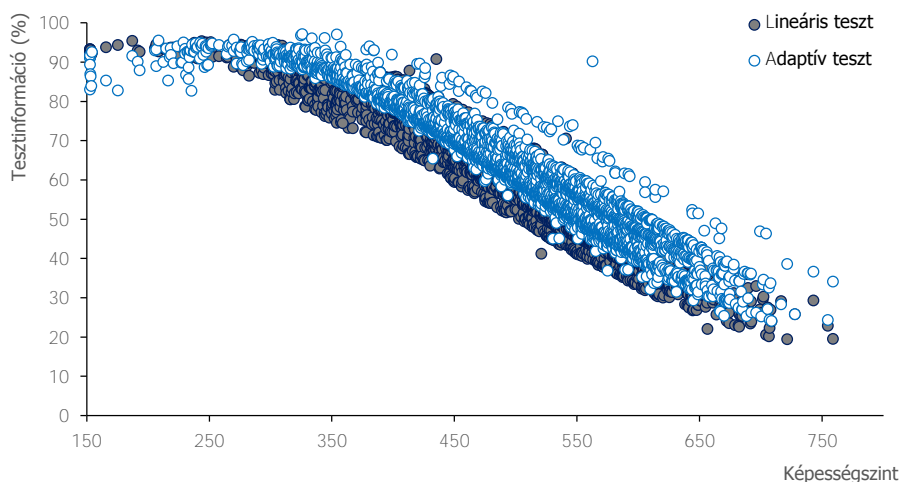
Crotts és munkatársai (2013) többféle adaptív teszt esetén vizsgálták a kinyerhető információ mennyiségét és átlagosan 36–62% információt nyertek ki 40 itemű MST-tesztek

esetén. Lineáris és adaptív tesztek alkalmazása során kinyert információ mennyiségét hasonlította össze *Rotou* és *mtsai* (2003), illetve *Thomson* (2007) kutatásában. *Rotou* és munkatársai a magas képességtartományban, míg *Thomson* mind az alacsony, mind a magas képességtartományokban statisztikailag jelentősebb mértékű információnövekedést detektált adaptív tesztelés során.

3. táblázat. A tesztekől kinyert információk mennyisége képességszintenkénti bontásban

Képesség-pont	Lineáris teszt			Adaptív teszt		
	N	Tesztinformáció (%)	Szórás (%)	N	Tesztinformáció (%)	Szórás (%)
300 alatt	105	92,0	2,0	116	92,8	2,8
301–400	450	82,6	5,4	397	87,1	4,1
401–500	928	67,7	7,3	904	71,3	6,4
501–600	1235	50,9	7,2	1352	53,8	6,9
601–700	482	37,0	6,0	435	43,0	6,0
700 felett	20	25,0	4,0	16	30,0	7,0

A jelenlegi kutatásunkban az adaptív teszt esetében átlagosan 63% volt a kinyert információ mennyisége, a lineáris teszt esetében 59% – a különbség szignifikáns ($t=12,61$, $p<0,001$). A 2. ábra a tanulók képességszintjének függvényében ábrázolja a tanulók lineáris, illetve az adaptív teszten kinyert információk mennyiségét. Mindkét teszt esetében a magasabb képességszint felé haladva csökkent a kinyert információ nagysága, azonban az adaptív teszt esetén mindegyik képességszint tekintetében magasabb volt.



2. ábra

A lineáris és az adaptív teszten kinyert információk a tanulók képességszintjének függvényében

A kinyert információ csökkenésével a teszt mérési precizitása is csökken. *Jodoin* és munkatársai (2006) eredményei szerint az információ mértékének 25%-os esése a tanulók 1–1,5%-ának téves szintre való besorolását eredményezheti, további 50%-os esés 3%-nyi tanuló esetén is jelenthet hibás minősítést. A tesztből kinyerhető információ a tanuló képességszintjéhez közel álló nehézségi indexű itemek kiközvetítésével növelhető, ami adaptív tesztelés esetén valósítható meg nagyobb mértékben, hiszen itt érhető el, hogy a tanulók képességszintjükhöz illeszkedő itemeket kapjanak, ellentétben a lineáris tesztekkel, ahol képességszinttől függetlenül többféle, különböző nehézségű itemeket kapnak a tanulók.

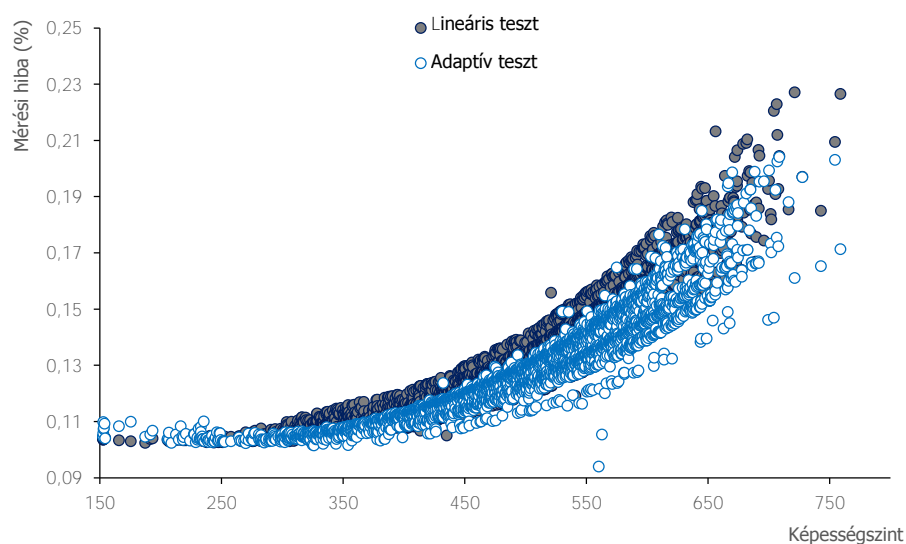
A kutatásban használt tesztek esetében mindkét típusú tesztnél csökkent a kinyerhető információ mennyisége az egyre magasabb képességszinteken, ami annak tulajdonítható, hogy a tesztelés elején minden tanuló címszóolvasási feladatot kapott, mely itemek elsősorban az alsó képességtartományban lévő tanulók esetén tudtak magasabb mennyiségű információt szolgáltatni (l. 4.a és 4.b ábra). A tesztelés későbbi szakaszában, a lineáris teszt megoldása során minden tanuló egyformán kapott mindegyik dimenzióból itemeket, így többféle nehézségű feladatot kellett megoldania, melyek csak egy kis része volt a képességszintjéhez illeszkedő, jelentős mértékben csökkentve a kinyerhető információt. Az adaptív tesztelés során a gyengébb képességűek végig címszóolvasási és toldalékoszóolvasási feladatokat oldottak meg, melyek minden iteme nehézségi indexében közel állt a képességszintjükhöz. Ezért volt a gyenge tanulók esetén a kinyert információk mennyisége igen magas. Ám a magasabb képességszintű tanulók a tesztelés első felében címszó- és toldalékoszó-olvasási feladatokat kaptak, melyek alacsonyabb nehézségi indexűek voltak, mint a képességszintjük, és csak a tesztelés második fele biztosította számukra a képességszintjükhöz közel álló feladatokat. Ez okozhatta a magasabb képességszinten a kinyerhető információ mértékének csökkenését, ami még így is magasabb volt, mint a lineáris teszt esetén kinyert információ.

A kinyert tesztinformáció nagyságához hasonlóan a mérési hibák is összehasonlíthatóak diákonkénti bontásban. A nemzetközi kutatásokban a hiba mértéke fontos mutatója a teszt mérési precizitásának. *Crotts* és munkatársai (2013) 0,16–0,24 közötti értékeket kaptak. *Olea* és munkatársai (2000) az adaptív teszten 0,25, a lineáris teszten 0,31 SE-értékeket mértek. *Rotou* és munkatársai (2003) az adaptív teszten 0,78–2,90, a lineáris teszten 0,98–3,10 közötti értékeket kaptak, és az adaptív teszten mért hiba minden képességtartományban kisebb volt, mint a PP tesztnél. *Thompson* (2007) elsősorban az alacsony és a magas képességtartományokban mért jelentős eltérést. *Kingsbury* (2004) 0,3 alatti SE-értékeket tart elfogadhatónak. Vizsgálatunkban a kétféle tesztkörnyezetben történt képességszint-becslés során elkövetett hiba nagyságát összehasonlítva megállapítható, hogy a lineáris formátumú teszt alapján történt képességszint-becslés hibáinak nagysága diákszinten átlagosan nagyobb ($t=13,52$, $p<0,01$; $SE_{\text{lineáris}}=0,14$), mint ugyanazon diákok adaptív tesztkörnyezetben történt képességszint becslésének hibája ($SE_{\text{adaptív}}=0,13$). A teljes mintát tekintve mindegyik képességszinten pontosabban, kisebb mérési hibával történt adaptív tesztkörnyezetben a diákok képességszintjének becslése (4. táblázat).

4. táblázat. A teszteken mért mérési hibák nagysága képességszintenkénti bontásban

Képességpont	Lineáris teszt			Adaptív teszt		
	<i>N</i>	<i>SE</i>	<i>Szórás</i>	<i>N</i>	<i>SE</i>	<i>Szórás</i>
300 alatt	105	0,10	0,01	116	0,10	0,01
301–400	450	0,11	0,01	397	0,10	0,01
401–500	928	0,12	0,01	904	0,11	0,01
501–600	1235	0,14	0,01	1352	0,13	0,01
601–700	482	0,16	0,01	435	0,15	0,01
700 felett	20	0,20	0,01	16	0,18	0,01

Összehasonlítva a lineáris és az adaptív teszten elért eredmények sztenderd hibáinak változását (3. ábra), mindegyik képességtartományban nagyobb hibával mért a lineáris teszt, mint az adaptív tesztváltozat.



3. ábra

Az adaptív és a lineáris teszt standard hibáinak alakulása a tanulók képességszintjének függvényében

A feladatok nehézségi indexeinek eloszlása a képességskálákon

A tanulók személy-ítem térképét mutatja a 4. és az 5. ábra. A minta képességeloszlása mindkét teszten, mind a négy dimenzió vonatkozásában közelíti a normál eloszlást. Mind a lineáris, mind az adaptív teszt esetében a minta képességeloszlása a -2 – $+2$ képességsávban helyezkedik el mind a négy dimenzióban.

A szóolvasási készség online mérésére kidolgozott adaptív és lineáris teszrendszer összehasonlító ...

	Címszó	Szín	Szój	Told
				540.5 617.4 697.4 730.4
3				708.4
				751.4
				119.4 636.4
				332.2 520.5 614.4
				754.4
2				513.5 514.5 647.4 811.4
				551.5 722.4
				375.2 554.5 565.5 759.4 794.4
		X		543.5 550.5 587.5 621.4 644.4
		X		405.2 501.5 504.5 510.5 533.5
		X		54.4 268.2 520.4 537.5 544.5
		X		522.5 549.5 558.5 574.5 635.4
	X	X	X	553.5 561.5 571.5 585.5 595.5
	X	XX	X	X 69.4 441.2 503.5 507.5 508.5
	XX	XX	XX	X 57.4 188.4 282.2 293.2 502.5
	XX	XXX	XX	X 120.4 343.2 385.2 407.2 511.5
1	XXX	XXXX	XXX	XX 27.4 217.4 277.2 326.2 337.2
	XXXX	XXXX	XXX	XXX 82.4 127.4 195.4 203.4 213.4
	XXXXX	XXXXX	XXXX	XXXX 22.4 221.4 245.4 261.2 290.2
	XXXXX	XXXXX	XXXXXXX	XXXXXXX 84.4 114.4 137.4 164.4 252.2
	XXXXXXX	XXXXXXX	XXXXXXX	XXXXXXX 89.4 151.4 157.4 211.4 230.4
	XXXXXXX	XXXXXX	XXXXXXX	XXXXXXX 81.4 115.4 146.4 264.2 284.2
	XXXXXXX	XXXXXXX	XXXXXXX	XXXXXXX 9.4 206.4 286.2 294.2 339.2
	XXXXXXX	XXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX 105.4 119.3 154.4 201.4 204.4
	XXXXXXXX	XXXXXX	XXXXXXXXXX	XXXXXXXXXX 87.4 117.4 143.4 172.4 258.2
0	XXXXXXXX	XXXXX	XXXXXXXXXX	XXXXXXXXXX 35.4 173.4 184.4 196.4 219.4
	XXXXXXXX	XXXXXX	XXXXXXXXXX	XXXXXXXXXX 39.4 111.4 132.4 182.4 185.4
	XXXXXXXX	XXXXXX	XXXXXX	XXXXXX 10.4 30.4 62.4 67.4 131.4 136.4
	XXXXXX	XXXXXX	XXXXXXX	XXXXXXX 6.4 24.4 41.4 78.4 158.4 186.4
	XXXXXX	XXXXX	XXXXXX	XXXXXX 1.4 45.4 70.4 108.4 121.4 127.3
	XXXXXX	XXXXX	XXXXX	XXXXX 27.3 54.3 55.4 63.4 65.4 103.4
	XXX	XXXX	XXXX	XXX 47.4 91.4 96.4 109.4 150.4 168.4
	XX	XXXX	XXXX	XX 16.4 34.4 42.4 99.4 120.3 138.4
	XX	XXX	XXX	XX 23.4 36.4 50.4 92.4 102.4 126.4
-1	XX	XX	XXX	XX 22.3 38.4 89.3 100.4 107.4 110.4
	X	XXX	XXX	X 3.4 4.4 40.4 53.4 60.4 77.4 86.4
	X	XX	XX	X 2.4 27.2 58.4 66.4 74.4 75.4
	X	XX	X	X 5.4 28.4 32.4 35.3 51.4 56.4
		XX	X	X 25.4 59.4 64.4 68.4 76.4 79.4
		X		12.4 13.4 15.4 21.4 36.3 41.3
	X	X		1.3 37.4 52.4 53.3 62.3 70.3
	X	X	X	X 7.4 19.4 26.4 31.4 43.4 44.4
				8.4 10.3 29.4 33.4 34.3 38.3
		X		6.3 11.4 21.3 24.3 28.3 33.3
-2				2.3 3.3 14.4 18.4 33.2 42.3 46.4
				8.3 12.3 18.3 28.2 29.3 30.3
				4.3 9.3 17.4 18.2 20.4 21.2 25.3
				7.3 10.2 15.3 16.3 19.3 23.3
				3.2 5.3 11.3 14.3 26.3 28.1 33.1
				2.2 16.2 17.3 23.2 31.3 37.2
				1.2 17.2 20.3 25.2 38.2 41.1
				9.2 11.2 12.2 13.3 14.2 17.1
				4.2 7.2 15.2 19.2 22.2 24.2 26.2
-3				6.2 8.1 8.2 9.1 10.1 11.1 12.1
				1.1 2.1 3.1 4.1 5.1 5.2 6.1 7.1

4. ábra

A nagymintás lineáris adatfelvétel alapján kirajzolható többdimenziós személy-Item térkép [címszóolvasás (Címszó), szinonima- (Szin), szójelentés- (Szój) és toldalékoszó-olvasás (Told)]

Magyar Andrea és Molnár Gyöngyvér

	Címszó	Szin	Szój	Told
				540.5
				375.2
				520.4 670.4
				697.4 811.4
3				343.2
				119.4
				405.2
				514.5
				501.5 520.3 537.5 550.5 551.5
				544.5 617.4 647.4 722.4 730.4
2				293.2 522.5 587.5 621.4 711.4
				337.2 511.5 533.5 554.5 574.5
		X	X	282.2 469.2 510.5 513.5 543.5
		X	X	424.2 503.5 517.5 531.5 561.5
		X	X	342.2 385.2 410.2 435.2 502.5
		XX	XX	277.2 326.2 507.5 508.5 523.5
		XX	XX	54.4 151.4 501.4 504.5 506.5
		XXX	XX	22.4 69.4 120.4 286.2 327.2
1	X	XXXX	XXX	X 39.4 127.4 268.2 290.2 341.2
	XX	XXXX	XXX	X 84.4 332 464.2 515.5 516.5 520.2
	XX	XXX	XXX	XX 9.4 164.4 245.4 284.2 303.2
	XXXX	XXXX	XXX	XXXX 188.4 217.4 221.4 252.2 261.2
	XXXXX	XXXX	XXXX	XXXXX 27.4 115.4 157.4 203.4 218.4
	XXXXXXX	XXXX	XXXX	XXXXXXX 57.4 81.4 82.4 114.4 137.4 195.4
	XXXXXXXX	XXXX	XXXX	XXXXXXXX 45.4 96.4 146.4 211.4 213.4
	XXXXXXXX	XXXX	XXXXXXXX	XXXXXXXXXXXXXXXXXXXX 35.4 89.4 99.4 105.4 117.4 119.3
0	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 136.4 172.4 173.4 196.4 219.4
	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 24.4 75.4 87.4 135.4 154.4 170.4
	XXXXXXXX	XXX	XXX	XXXXXXXX 10.4 91.4 94.4 101.4 111.4 131.4
	XXXXX	XXX	XXX	XXXXX 26.4 41.4 65.4 70.4 78.4 85.4
	XXXXX	XXX	XXX	XXXX 6.4 13.4 28.4 47.4 63.4 67.4
	XXX	XX	XXX	XXX 1.4 15.4 22.3 36.4 54.3 74.4
	XXX	XX	XXX	XX 12.4 16.4 30.4 34.4 40.4 42.4
	X	XX	XX	XX 5.4 32.4 43.4 56.4 57.3 62.4
-1	X	XX	XX	X 3.4 7.4 23.4 27.3 28.3 53.4 55.4
		X	XX	X 2.4 4.4 14.4 18.4 31.4 38.4 49.4
		XX	XX	8.4 14.3 20.4 21.4 33.4 47.3
		X	X	7.2 7.3 15.3 17.4 19.4 26.3 28.2
		X	X	15.2 17.3 25.4 26.2 27.2 29.4
		X	X	5.3 10.3 12.3 13.3 18.3 20.3
		X	X	1.3 9.3 12.2 18.2 21.3 31.3 33.3
		X	X	6.3 11.4 24.3 33.2 35.3 36.3
-2		X		30.3 32.3 39.2 39.3 41.2 43.2
			X	2.3 3.3 5.1 5.2 9.2 11.3 21.2
			X	10.2 11.1 11.2 29.2 29.3 32.2
				4.3 8.2 8.3 22.2 36.2 37.3 38.3
				2.2 3.2 19.3 25.3 41.1 44.1 46.2
				4.2 6.2 20.2 23.3 24.2 37.2 42.2
				38.2 40.3 49.2 50.2 54.2 64.3
				52.1 59.1 70.2 74.3 79.1 83.1
-3				2.1 3.1 20.1 49.1 50.1 55.2 78.1
				13.1 13.2 17.2 24.1 25.2 69.1
				23.2 54.1 62.1 75.1 82.2 87.1
				16.3 39.1 53.2 85.1 90.2 111.2
				12.1 14.1 14.2 15.1 16.1 16.2
				1.1 1.2 4.1 6.1 7.1 8.1 9.1 10.1

5. ábra

A nagymintás adaptív adatfelvétel alapján kirajzolható többdimenziós személy-
 térkép [címszóolvasás (Címszó), szinonima- (Szin), szójelentés- (Szó) és
 toldalékosszó-olvasás (Told)]

A lineáris teszt esetében mind a négy készségterületen hasonlóan alakult a minta képességeloszlása, mivel a gyengébb és a magasabb képességtartományba eső tanulók ugyanazokat a feladatokat kapták ugyanolyan arányban. Ugyanakkor az adaptív teszt esetében csak a címszóolvasási és a toldalékoszó-olvasási feladatokat kapta meg minden tanuló, a szinonima és a szójelentés feladatokat viszont csak a közepes és a magasabb képességsávba tartozó tanulók részére történtek kiközvetítésre, és őket ezek alapján rangsorolta a rendszer.

A feladatok nehézségi indexeit tekintve mindkét tesztelési mód esetén a címszóolvasási feladatok estek a képességsála aljára, ezt követték a toldalékoszó-olvasási feladatok, majd a szinonima és a szójelentés feladatok bizonyultak a legnehezebbeknek. A feladatok nehézségi indexei lefedték a teljes képességskálát, tehát alkalmasak voltak a tanulók képességmérésére. Az alacsonyabb képességszintűek mérésére elsősorban a címszó- és a szóolvasási feladatok voltak alkalmasak, a magasabb képességtartományokat a szinonima- és a szójelentés-olvasási feladatok tudták – átlagosan magasabb nehézségi szintjük miatt – precízebben mérni. A mérésben a 4. és 5. évfolyamos korosztály vett részt, de mivel a feladatok jóval nagyobb tartományban szóródnak, ezért alacsonyabb és magasabb évfolyamok és képességgel rendelkező tanulók mérésére is alkalmas lehet a rendszer.

Az egyes dimenziók közötti összefüggések erősségét mutatja az 5. és a 6. táblázat. Mind a lineáris, mind az adaptív tesztek esetében szoros összefüggés volt a különböző dimenziókon elért eredmények között. A legszorosabb összefüggés a szinonima- és a szójelentés-olvasás közötti, majd ezt követte a címszóolvasás és a toldalékoszó-olvasás dimenzióban elért eredmény, a leggyengébben a címszóolvasás és a szinonimaolvasás függött össze. A dimenziók között magas korrelációk jelzik, hogy releváns volt az alkalmazott adaptív tesztrendszer alkalmazása, a mért dimenziók szorosan összefüggnek, ezért annak ellenére, hogy a tanulók a feladatbank egy bizonyos részalmazát oldották meg, vagyis képességüktől függően csak bizonyos dimenziókból kaptak feladatokat, a megoldott itemekből nagy valószínűséggel következtethetünk arra is, hogy a tanuló a többi dimenzióhoz tartozó feladatokon milyen eredményt ért volna el.

Az egyes dimenziók közötti összefüggések erősségét mutatja az 5. és a 6. táblázat. Mind a lineáris, mind az adaptív tesztek esetében szoros összefüggés volt a különböző dimenziókon elért eredmények között. A legszorosabb összefüggés a szinonima- és a szójelentés-olvasás közötti, majd ezt követte a címszóolvasás és a toldalékoszó-olvasás dimenzióban elért eredmény, a leggyengébben a címszóolvasás és a szinonimaolvasás függött össze. A dimenziók között magas korrelációk jelzik, hogy releváns volt az alkalmazott adaptív tesztrendszer alkalmazása, a mért dimenziók szorosan összefüggnek, ezért annak ellenére, hogy a tanulók a feladatbank egy bizonyos részalmazát oldották meg, vagyis képességüktől függően csak bizonyos dimenziókból kaptak feladatokat, a megoldott itemekből nagy valószínűséggel következtethetünk arra is, hogy a tanuló a többi dimenzióhoz tartozó feladatokon milyen eredményt ért volna el.

A diákok lineáris teszten nyújtott teljesítménye (átlag=0,04, szórás=0,55) és az adaptív teszt alapján számolt képességszintje (átlag=0,06 szórás=0,61) magasan korrelált egymással ($r=0,74$, $p<0,01$). Hasonló eredményt kapott *Al-A'ali* (2007) kutatásában is ($r=0,63$, $p<0,01$). A páros t-próba eredménye szerint nem volt szignifikáns különbség a tanulók két teszten elért eredményei között ($t=-1,29$, $p=0,20$). Az adaptív rendszerré való kifejlesztés

során célunk az eredeti struktúra megtartása volt, ami a mutatók szerint sikeres, mivel a tanulók különböző képességszintekre sorolása hasonlóan történt mindkét tesztrendszer esetében, és az elért eredmények sem különböztek jelentősen a kétféle tesztkörnyezetben.

5. táblázat. Korrelációs mátrix (lineáris teszt)

Dimenziók	Dimenziók			
	Címszó	Szin	Szój	Told
Címszó	1,00			
Szin	0,49	1,00		
Szój	0,50	0,97	1,00	
Told	0,78	0,65	0,58	1,00

Megjegyzés: címszóolvasás (Címszó), szinonima- (Szin), szójelentés- (Szój) és toldalékosszó-olvasás (Told)

6. táblázat. Korrelációs mátrix (adaptív teszt)

Dimenziók	Dimenziók			
	Címszó	Szin	Szój	Told
Címszó	1,00			
Szin	0,51	1,00		
Szój	0,54	0,92	1,00	
Told	0,76	0,64	0,68	1,00

Megjegyzés: címszóolvasás (Címszó), szinonima- (Szin), szójelentés- (Szój) és toldalékosszó-olvasás (Told)

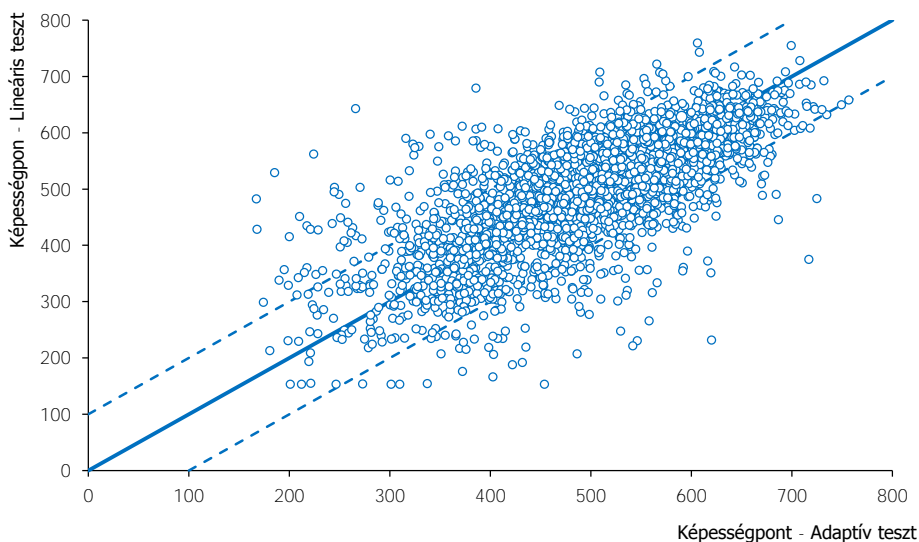
A becsült képességszintek összehasonlítása évfolyamonként és személyenként

A kétféle teszten elért képességszintek között a korábbi kutatások szerint jelentős eltérések lehetnek. *Olea* és munkatársai (2000) eredményei szerint a tanulók a lineáris teszten átlagosan 0,95, az adaptív teszten 0,58 képességpontokat értek el, azaz jelentős mértékű eltérés realizálódott a két tesztelési eljárással kivitelezett adatfelvétel során. A diákok lineáris, illetve adaptív tesztkörnyezetben mutatott teljesítményének alapstatisztikai mutatóit – évfolyamonkénti bontásban – a 7. táblázat mutatja. Sem a negyedik évfolyamon, sem az ötödik évfolyamon nem volt szignifikáns különbség a lineáris, illetve az adaptív tesztkörnyezetben becsült képességszintek között. Az alacsonyabb képességű diákokat segítette a rendszer a jobb eredmény elérésében, a magasabb képességszinten nehezebb volt jobb eredményt elérni az adaptív tesztkörnyezetben, mint a lineáris teszt esetében. Évfolyamonkénti összehasonlítást nézve sem különböztek az eredmények szignifikánsan egyik testtenszem ($t_{\text{lineáris}}=-0,59$, $p=0,55$; $t_{\text{adaptív}}=-1,82$, $p=0,07$, $F=0,67$, $p=0,41$), ami azt jelzi, hogy nem fordult elő olyan eset, hogy valamelyik tanuló egyik teszten kiugróan jó eredményt ért volna el, a másikon pedig gyengén teljesített volna, ami mindkét tesztrendszer esetében a jó működést mutatja.

7. táblázat. A lineáris és adaptív tesztkörnyezetben mutatott teljesítmények klasszikus tesztelméleti mutatói évfolyamonkénti bontásban

Évfolyam	N	Lineáris teszt (pont)				Adaptív teszt (pont)				t	p	szign.
		Átlag	Szórás	Min.	Max.	Átlag	Szórás	Min.	Max.			
4.	1706	0,01 (492)	0,53 (96)	-1,78 (167)	1,27 (723)	0,04 (497)	0,61 (101)	-2,05 (152)	1,91 (804)	1,15	0,25	{4}={5}
5.	1514	0,09 (508)	0,57 (103)	-1,77 (167)	1,45 (757)	0,08 (503)	0,60 (99)	-2,05 (152)	1,63 (758)	-1,24	0,21	

A 6. ábra a két tesztkörnyezetben nyújtott teljesítmények diákszintű összehasonlítását ábrázolja. Ha a diák képességszintje tesztkörnyezettől függetlenül számszerűen ugyanannak bizonyult, akkor a diákot reprezentáló alakzat a folytonos vonalon helyezkedik el. Amennyiben megállapított képességszintje nem különbözött egymástól szignifikánsan lineáris és adaptív környezetben, az őt reprezentáló jel a szaggatott vonalakon belül található. A szaggatott vonalak által képzett sávon kívül elhelyezkedő diákok esetében az adaptív tesztkörnyezet szignifikánsan különböző képességszintet állapított meg, mint a fix tesztekkel diagnosztizált képességszint.



6. ábra

A lineáris és az adaptív teszten nyújtott teljesítmények összehasonlítása diákonkénti bontásban

A mintában szereplő tanulók 7%-ánál magasabb, 6%-ánál alacsonyabb képességszintet diagnosztizáltunk adaptív környezetben, azaz a diákok 13%-ánál alapvetően más képességszint lett meghatározva. Jellemzően az adaptív teszten alacsonyabb képességtartományba sorolt diákok esetén fordult inkább elő, hogy a lineáris teszt magasabb képességszintet mutatott, valamint az alacsony és átlagos képességszintű diákok körében fordult ennek fordítottja is elő. Ha az értelmezés során figyelembe vesszük az adaptív tesztkörnyezetben tapasztalt méréselméleti mutatók (mint a reliabilitás növekedése vagy a mérési hiba csökkenése) javulását, akkor megállapítható, hogy a diákok e 13%-a számára volt igazán meghatározó az adaptív környezet. A többiek képességszintje nem különbözött jelentős mértékben egymástól lineáris és adaptív környezetben.

A helyes válaszok aránya adaptív és lineáris tesztkörnyezetben

A vonatkozó nemzetközi kutatások szerint jelentős különbségek lehetnek a helyes válaszok számát illetően a kétféle tesztelési mód alkalmazása során. *Olea* és munkatársai (2000) szerint az adaptív teszten 13,39 volt a helyes válaszok száma, a lineáris teszten 11,72. Az adaptív tesztrendszer helyes működését jelzi, ha a tanulók teszten nyújtott teljesítménye a képességszinttől függően azonos (a rendszer beállításakor meghatározott, jelen esetben 80% körüli) szinten mozog. Ebben az esetben ugyanis mindenki a képességszintjéhez leginkább közel álló feladatokat kapta a tesztben, az alacsony képességszintűek könnyebb, a magasabb képességszintűek nehezebb feladatot. A helyes megoldások aránya így minden esetben közel azonos annak ellenére, hogy az egyes százalékos teljesítmények más-más képességfejlettségi szintet jeleznek. A helyes válaszok aránya 4. évfolyamon 81,2%, míg 5. évfolyamon 81,6% volt, azaz előzetes feltételezésünknek megfelelően működött a tesztrendszer.

A kétféle tesztkörnyezetben elért eredmények képességszint szerinti eloszlását ábrázolja a 8. táblázat. Általánosságban mindkét teszten a képességszint növekedésével párhuzamosan emelkedett a teszteken elért helyes válaszok aránya is, bár a növekedés mértéke különbözőnek bizonyult.

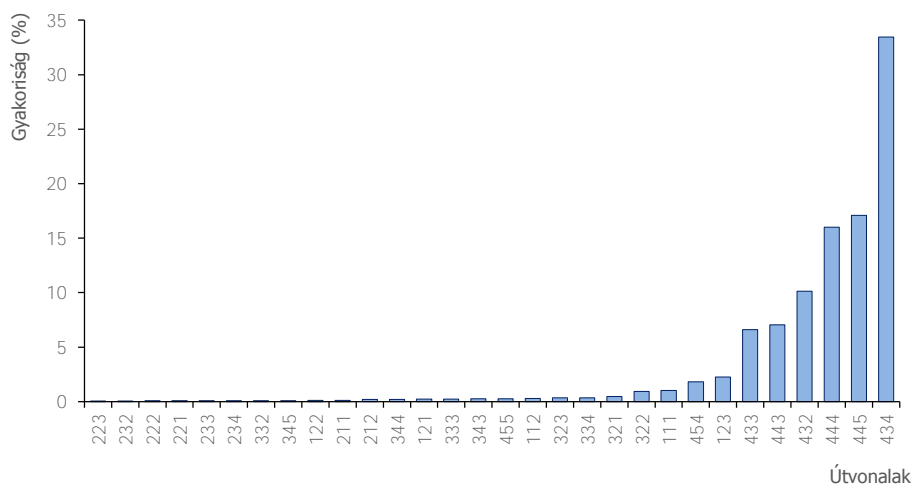
8. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak eloszlása képességszint szerinti bontásban

Képességszint	Helyes válaszok eloszlása					
	Lineáris teszt			Adaptív teszt		
	N	Átlag (%)	Szórás (%)	N	Átlag (%)	Szórás (%)
300 alatt	105	55,9	7,3	116	71,8	12,5
301–400	450	71,1	6,2	397	77,4	7,4
401–500	928	79,2	4,8	904	80,3	5,3
501–600	1235	85,9	3,4	1352	82,9	5,8
601–700	482	90,7	2,6	435	84,8	6,6
700 felett	20	94,0	1,7	16	88,1	8,9

Adaptív teszt esetén az átlag alatti tanulóknál jóval magasabb arányban fordultak elő helyes válaszok, mint a lineáris teszt esetén, átlag feletti tanulók esetében viszont fordítva, kisebb arányban fordultak elő helyes válaszok. Ez arra enged következtetni, hogy az adaptív teszt az alacsony képességszintűek esetén több sikerélményt hozott a tanulóknak (kevesebb számukra túl nehéz, nagy valószínűség szerint megoldhatatlan feladatot osztott a rendszer), a magas képességszintűek számára viszont nagyobb kihívást jelentett (kevesebb könnyebb feladatot kiköszvetítve), mint a lineáris teszt. Az eredmények összecsengnek *Jiban* és munkatársai (2008) kutatási eredményeivel, akik ugyanilyen eredményre jutottak kisiskolások adaptív teszten elért eredményeit vizsgálva.

A lineáris és az adaptív tesztelés során kiosztott résztesztek nehézségi szintjének változásmintázata

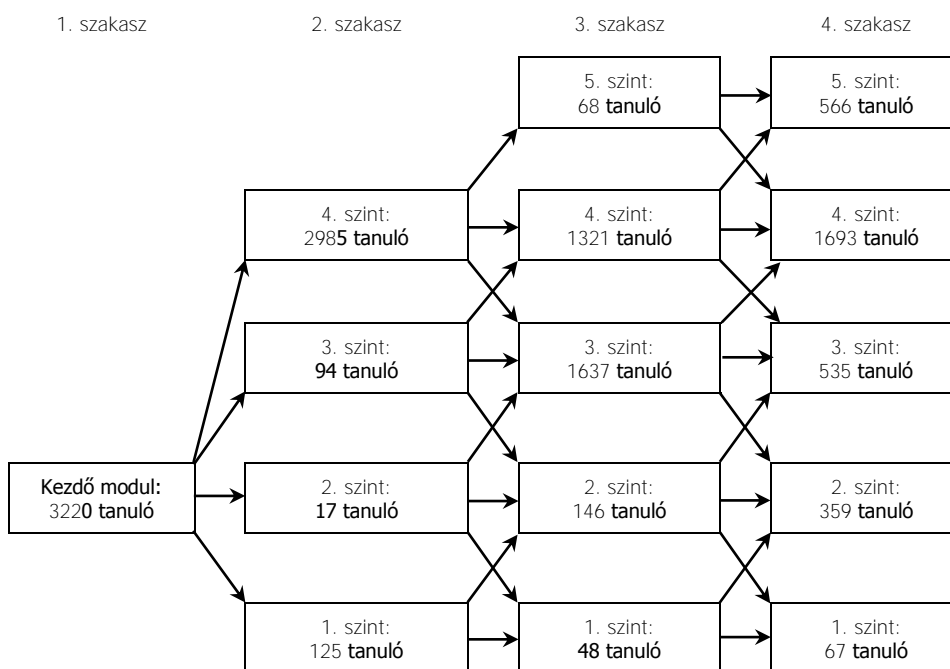
Az adatfelvétel során – a többszakaszos adaptív teszt esetében – az öt szakaszból összeállítható 30 különböző teszt mindegyike kiosztásra került (7. ábra). A kiosztott résztesztekkel illetően a legnagyobb részben az átlagos nehézségű, a 3. és a 4. szinten szereplő feladatokat közvetítette ki a rendszer. A hat leggyakoribb útvonal a 4. és a 3. szinteken haladt át. Az esetek 20%-ában jelentek meg a legnehezebb, az 5. szint feladatai, 5%-ban kizárólagosan könnyű feladatokból álló tesztekkel oldottak meg a tanulók. A kezdő részteszt után a 4. szinten folytatták a tanulók, majd innen a 3., illetve a 4. szintű modul felé haladtak. A negyedik szakaszban a tanulók közel fele maradt a 4. szinten, harmaduk a 3. és az 5. szinten végzett. A tanulók 10%-a a könnyebb, 2-es szint felé haladt a 4. szakaszban. Az esetek 75%-ában a tanulók a 3. szakasz után még szintet váltottak, ami indokolja a 4. szakasz szükségességét.



7. ábra

Az adaptív teszrendszeren belül a második, a harmadik és a negyedik szakaszban kiosztott útvonalak gyakorisága

A szakaszokon belül a tanulók modulonkénti eloszlását képességszint és gyakoriság szerinti bontásban a 8. ábra szemlélteti. A kezdő modulon mutatott teljesítmény alapján a tanulók legnagyobb része a teszt második szakaszában a 4. szintű feladatokat kapta, majd a teszt 3. és 4. szakaszában egyre egyenletesebben oszlottak el a 2–5. szint között. A teszt utolsó szakaszában a tanulók közel fele a 4. szinten végzett, a többiek közel azonos arányban oszlottak el a 2., a 3. és az 5. szinteken. A legelső szinten viszonylag kevés, 67 tanuló végzett, ők 90%-ban a teszt kezdetétől az 1-es szinten haladtak végig.



8. ábra

A tanulók gyakorisági és képességszint szerinti eloszlása a szakaszokon és a modulokon belül

A papír alapú rendszerrel összehangban a lineáris teszt esetén is besorolhatóak voltak a tanulók az öt szint valamelyikébe. A 8. táblázat a tanulók lineáris teszten és az adaptív teszt 4. szakaszában történt szintenkénti besorolását összegzi. A két rendszer hasonló arányban végezte el a tanulók besorolását, azonban az adaptív teszt esetében több tanuló jutott el az 5. szintre, és a második szinten is többen végeztek, mint a lineáris teszt esetében, tehát a kiemelkedő és a gyengébb képességű tanulók jobb eredményeket értek el az adaptív teszt kiosztásnál. Itt mutatkozik meg a pontosabb képességmérés jelentősége, hiszen a tanulók végső besorolását befolyásolhatja, hogy mennyire működik pontosan a rendszer.

9. táblázat. A tanulók lineáris és adaptív teszten való szintenkénti besorolása

Szint	Lineáris teszt (N)	Adaptív teszt (N)
5. szint (91–100%)	321	566
4. szint (81–90%)	1703	1693
3. szint (71–80%)	824	535
2. szint (61–70%)	259	359
1. szint 60% alatt	113	67

Mindezt *Thompson és Way* (2007) kutatásai is megerősítik, eredményeik szerint elsősorban az alacsony és a magas képességtartományokban volt jelentős a tanulók adaptív teszt általi precízebb besorolása. *Crotts és munkatársainak* (2013) eredménye szerint 9113 esetből 91–137 esetben változott a tanulók besorolása, azaz a tesztelésben résztvevők 1-1,5%-ának eredményeire volt hatással a mérés pontossága.

Összegzés

A 21. században jelentkező mérési-értékelési igények a számítógépes tesztelés felé jelölik ki a fejlődés irányvonalát. A számítógépes tesztek számos új lehetőséget kínálnak a készség- és képességmérésre, segítségükkel lehetővé válik az azonnali kiértékelés, új, innovatív itemtípusokat dolgozhatnak ki, új, eddig nem vagy csak nehezen mérhető képességtérületek pontos és hatékony mérésére adnak lehetőséget. A papír alapú tesztek számítógépes formára való konvertálása több szinten megvalósulhat. A ma létező leginnovatívabb forma a számítógépes adaptív tesztelés. Adaptív tesztelési technika esetén az itemeket, illetve részteszteket egy pontosan bemért, paraméterezett itemeket tartalmazó itembankból közvetítik ki, és minden tanuló a képességszintjének legmegfelelőbb itemeket, illetve részteszteket kapja. Ez a tesztelési mód a hagyományos, lineáris tesztelési technikához képest a képességek sokkal pontosabb és hatékonyabb mérését teszi lehetővé. Mivel a tanulók saját képességszintjükhöz illeszkedő feladatokat kapnak, a teszt feladatai egyformán kihívást jelentenek számukra, ezáltal a teszt minden iteme egyforma mértékben járul hozzá a személy képességszintjének meghatározásához, így sokkal pontosabb képességszint meghatározásra nyílik lehetőség.

A papír alapú tesztekre az adaptív tesztelésre való átállás során megvizsgálandó kérdés, hogy az átállás biztosítja-e az elvárt szintű mérésmethodikai javulást és a hatékonyabb képességmérést. A vonatkozó nemzetközi vizsgálatok szerint az adaptív tesztek reliabilitása magasabb a lineáris tesztekénél, a kinyerhető információk mennyisége szintén több, a mérési hiba viszont alacsonyabb, ezáltal a lineáris tesztekénél jóval pontosabb mérést tesznek lehetővé. Azonban a nemzetközi kutatásokon elért eredmények elsősorban szimulált adatbázisokra alapozottak, néhány pilot mérés történt empirikus adatok felhasználásával, egyetemista hallgatók bevonásával.

Kutatásunk célja a papír alapú tesztelestől adaptív tesztelest való átállás feltételeinek vizsgálata volt empirikus kutatás keretében általános iskolai korosztály körében. A kutatás során a szóolvasó készség mérésére kidolgozott adaptív és lineáris tesztváltozatok esetében történt az adaptív tesztek hatékonyságának, mérési precizitásának összehasonlítása a hagyományos, lineáris tesztkörnyezetben folyó tesztelessel. Az elemzések során – a vonatkozó nemzetközi kutatásoknak megfelelően – összehasonlítottuk a tesztek jószágmutatóit, a kinyert információkat és a mérési hibák nagyságát az adaptív és lineáris tesztkörnyezetben, megvizsgáltuk a kétféle tesztelest során kiosztott résztesztek nehézségi szintjének változásmintázatát és a helyes válaszok arányát mindkét tesztelestési mód tekintetében.

Az eredmények alapján az adaptív tesztek pontosabb képességbecslést tettek lehetővé a vizsgált évfolyamok tekintetében. A feladatok nehézségi indexei lefedték a vizsgált korosztály képességszintjét, tehát alkalmasak voltak a korosztály képességszintjének becslésére. A tesztben szereplő dimenziók szorosan korreláltak egymással, ugyanígy a kétféle tesztkörnyezetben elért eredmények is erős összefüggést mutattak, vagyis nem különbözött jelentős mértékben a tanulók kétféle tesztkörnyezetben elért eredménye. A teszteken elért helyes válaszok arányát összehasonlítva: az adaptív teszt az alacsonyabb képességszintű tanulók esetén magasabb helyes válaszokat mutatott, a magas képesség-tartományban viszont fordítva, viszonylag kevesebb jó válasz született adaptív teszt-kiosztással, mint a lineáris teszttel. Ez azt mutatja, hogy az adaptív kiosztás nagyobb sikerélményt jelentett a gyengébb képességű tanulók számára és kihívást a magas képességű tanulók részére.

A tanulók jelentős része átlagos képességű, ennek megfelelően a közepes nehézségű résztesztek szerepeltek a legnagyobb gyakorisággal az adaptív teszt-kiosztás esetén, azonban a harmadik, illetve még a negyedik szakaszban is sok tanuló esetén módosult a szint, ami indokolja az öt különböző nehézségi szintű modul alkalmazásának szükségességét. A mérés során összehasonlítottuk a tesztinformációkat, illetve a mérési hibák nagyságát, és mindkét esetben az adaptív tesztnél mértünk magasabb tesztinformációt és kisebb mérési hibát a teljes képességskála tekintetében.

A kutatás egyedisége, hogy az adaptív tesztelest hatékonyságát vizsgáló legtöbb kutatással szemben nem szimulált adatbázison, hanem empirikus adatok segítségével hasonlította össze a lineáris és az adaptív tesztkörnyezetben becsült képességszintek alakulását, továbbá az azonos minta alkalmazása lehetővé tette a diákszintű összehasonlítást is. A mérésekben 10 és 11 évesek vettek részt, ami szintén egyedinek mondható az adaptív tesztekkel foglalkozó kutatások között. Az eredmények alátámasztották a szimulációs kísérletekben is tapasztaltakat, miszerint jelentős mértékű mérési precizitás érhető el adaptív-teszt-algoritmus alkalmazásával a hagyományos lineáris tesztekhez képest.

Kutatási eredményeink általánosíthatóságának korlátja, hogy az adatfelvétel során egy készség, a szóolvasó készség tekintetében vizsgáltuk az adaptív tesztelest való átállás lehetőségét. Az alkalmazott ítebankok mérete, az ítemek típusai befolyásolhatják a képességszint-becslés során kinyert és kinyerhető információ mennyiségét, ezért a kinyert információ mértékének pontosabb meghatározásához további kutatások szükségesek a különböző méretű és tartalmi lefedésű ítebankok felhasználásával.

A tanulmány megírását TÁMOP 3.1.9/11 kutatási program és az Oktatásméleti Kutatócsoport támogatta.

Irodalom

- American Educational Research Association, American Psychological Association, és National Council on Measurement in Education (1999): *Standards for educational and psychological testing*. American Educational Research Association, Washington, DC.
- Al-A'ali, M. (2007): Implementation of an improved adaptive testing theory. *Educational Technology & Society*, **10**. 4. sz. 80–94.
- American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment (1986): *Guidelines for computer-based tests and interpretations*. Washington, DC.
- Brossman, B. G. és Guille, R. A. (2014): A comparison of multi-stage and linear test designs for medium-size licensure and certification examinations. *Journal of Computerized Adaptive Testing*, **2**. 2. sz. 18–36. DOI: [10.7333%2Fjcat.v2i0.31](https://doi.org/10.7333%2Fjcat.v2i0.31)
- Csapó, B., Ainley, J., Bennett, R. E., Latour, T. és Law, N. (2012): Technological issues for computer-based assessment. In: Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and teaching of 21st century skills*. Springer, New York. 143–230. DOI: [10.1007/978-94-007-2324-5_4](https://doi.org/10.1007/978-94-007-2324-5_4)
- Crotts, K. M., Zenisky, A. L., Sireci, S. G. és Li, X. (2013): Estimating measurement precision in reduced-length multi-stage adaptive testing. *Journal of Computerized Adaptive Testing*, **1**. 4. sz. 67–87. DOI: [10.7333%2Fjcat.v1i0.19](https://doi.org/10.7333%2Fjcat.v1i0.19)
- Eggen, T. J. H. M. és Straemans, G. J. J. M. (2000): Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, **60**. 5. sz. 713–734. DOI: [10.1177/00131640021970862](https://doi.org/10.1177/00131640021970862)
- Frey, A., Seitz, N. N. és Kröhne, U. (2011): Reporting differentiated literacy results in PISA by using multidimensional adaptive testing. In: Prenzel, M., Kobarg, M., Schöps, K. és Rönnebeck, S. (szerk.): *Research in the context of the Programme for International Student Assessment*. Springer, Berlin. 103–133. DOI: [10.1007/978-94-007-4458-5_7](https://doi.org/10.1007/978-94-007-4458-5_7)
- Greiff, S., Wüstenberg, S. és Funke, J. (2012). Complex Problem Solving. More than reasoning? *Intelligence*, **40**. 1–14. DOI: [10.1016/j.intell.2011.11.003](https://doi.org/10.1016/j.intell.2011.11.003)
- Guille, R. A., Becker, K. A., Zhu, R. X., Zhang, Y., Song, H. és Sun, L. (2011): Comparison of asymmetric early termination MST with linear testing. Előadás. National Council on Measurement in Education, New Orleans, LA.
- Hambleton, R. K. és Xing, D. (2006): Optimal and nonoptimal computer-based test designs for making pass–fail decisions. *Applied Measurement in Education*, **19**. sz. 221–239. DOI: [10.1207/s15324818ame1903_4](https://doi.org/10.1207/s15324818ame1903_4)
- Hendrickson, A. (2007): An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, **26**. 2. sz. 44–52. DOI: [10.1111/j.1745-3992.2007.00093.x](https://doi.org/10.1111/j.1745-3992.2007.00093.x)
- Hülber László és Molnár Gyöngyvér (2013): Papír és számítógép alapú tesztelés nagymintás összehasonlító vizsgálata matematika területén, 1-6. évfolyamon. *Magyar Pedagógia*, **113**. 4. sz. 243–263.
- Jiban, C. L., Ayodele, A., McCarthy, A. és Christ, T. (2008): *CBAS-R fall screening pilot: Technical report on psychometric and practical feasibility*. University of Minnesota, Minnesota.
- Jodoin, M., Zenisky A. és Hambleton, R. K. (2006): Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, **19**. 3. sz. 203–220. DOI: [10.1207/s15324818ame1903_3](https://doi.org/10.1207/s15324818ame1903_3)
- Keng, L. (2008): *A comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas, Austin.
- Kingsbury, G. G. és Hauser, C. (2004): Computerized adaptive testing and the No Child Left Behind. Előadás. Annual Meeting of the American Educational Research Association. San Diego, CA.

- Kolen, M. J. (1999–2000): Threats to score comparability with applications to performance assessments and computerized adaptive tests. *Educational Assessment*, 6. 2. sz. 73–96.
DOI: [10.1207/S15326977EA0602_01](https://doi.org/10.1207/S15326977EA0602_01)
- Magyar Andrea (2012): Számítógépes adaptív tesztelés. *Iskolakultúra*, 22. 6. sz. 52–60.
- Magyar Andrea (2013): Többszakaszos adaptív tesztek felépítése, működése. *Oktatás-Informatika*, 1–2. sz. <http://www.oktatas-informatika.hu/2013/11/magyar-andrea-tobbszakaszos-adaptiv-tesztek-felepitesemukodese>. Letöltés ideje: 2015. március 2.
- Magyar Andrea és Molnár Gyöngyvér (2013): Adaptív és lineáris formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*, 113. 3. sz. 181–193.
- Magyar Andrea (2014a): Adaptív tesztek készítésének folyamata. *Iskolakultúra*, 24. 4. sz. 26–33.
- Magyar Andrea (2014b): Szóolvasási készséget mérő adaptív tesztelésre alkalmas feladatbank fejlesztése. VI. Oktatás-informatika Konferencia tanulmánykötete. 404–412. http://www.eltereader.hu/media/2014/03/VI_OKTINF_Tanulmanykotet_READER.pdf. Letöltés ideje: 2015. január 5.
- Magyar Andrea és Molnár Gyöngyvér (2014): A szóolvasási készség adaptív mérését lehetővé tevő online tesztrendszer kidolgozása. *Magyar Pedagógia*, 114. 4. sz. 259–279.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, 16. 12. sz. 99–113
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, 20. 7–8. sz. 22–34.
- Molnár Gyöngyvér és Thibaud Latour (2011): Online tesztelés: lehetőségek és kihívások. Előadás. IX. Pedagógiai Értékelési Konferencia. Szeged, 2011. április 29–30. 63.
- Molnár Gyöngyvér (2013): *A Rasch-modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Nagy József (2004): A szóolvasó készség fejlődésének kritériumorientált diagnosztikus feltérképezése. *Magyar Pedagógia*, 104. 2. sz. 123–142.
- Olea, J., Revuelta, J., Ximénez, M. C. és Abad, F. J. (2000): Psychometric and psychological effects of review on computerized fixed and adaptive tests. *Psicológica*, 21. sz. 157–173. DOI: [10.1177/0013164403251282](https://doi.org/10.1177/0013164403251282)
- Paek, P. (2005): *Recent trends in comparability studies*. PEM Research Report 05-05.
- Pásztor-Kovács Anita, Magyar Andrea, Hülber László, Pásztor Attila és Tongori Ágota (2013): Áttérés online tesztelésre – a mérés-értékelés új dimenziói. *Iskolakultúra*, 23. 11. sz. 86–100.
- Pyper, A. és Lilley, M. (2010): A comparison between the flexilevel and conventional approaches to objective testing. Előadás. CAA Konferencia, University of Hertfordshire.
- Rotou, O., Patsula, L., Manfred, S. és Rizavi, S. (2003): Comparison of multi-stage tests with computerized adaptive and paper and pencil tests. Előadás. American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME). Chicago, IL.
- Scheuermann, F. és Björnsson, J. (2009, 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. DOI: [10.2788/60083](https://doi.org/10.2788/60083)
- Scheuermann, F. és Pereira, G. A. (2008, szerk.): *Towards a research agenda on computer-based assessment*. Office for Official Publications of the European Communities, Luxembourg.
- Thompson, N. A. és Prometric, T. (2007): A practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research and Evaluation*, 12. 1. sz. 1–13.
- Thompson, T. és Way, D. (2007): Investigating CAT designs to achieve comparability with a paper test. Applications and Issues Paper Session. Pearson. <http://publicdocs.iacat.org/cat2010/cat07thompson.pdf>. Letöltés ideje: 2015. április 12.

- Vispoel, W. P., Hendrickson, A. B. és Bleiler, T. (2000): Limiting answer review and change on computerized adaptive vocabulary tests. Psychometric and attitudinal results. *Journal of Educational Measurement*, **37**. 1. sz. 21–38. DOI: [10.1111/j.1745-3984.2000.tb01074.x](https://doi.org/10.1111/j.1745-3984.2000.tb01074.x)
- Wainer, H. (2000): CATs: Whither and whence. *Psicológica*, 21. sz. 121–133.
DOI: [10.1002/j.2333-8504.2000.tb01835.x](https://doi.org/10.1002/j.2333-8504.2000.tb01835.x)
- Wan, L., Keng, L., McClarty, K. és Davis, L. (2009): Methods of comparability studies for computerized and paper-based tests. *Test, measurements and research services bulletin*, 10. sz. 1–4.
- Wang, T. és Kolen, M. J. (2001): Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement*, **38**. 1. sz. 19–49.
DOI: [10.1111/j.1745-3984.2001.tb01115.x](https://doi.org/10.1111/j.1745-3984.2001.tb01115.x)
- Wang, S., Jiao, H., Young, M. J., Brooks, T. és Olson, J. (2008): Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments. *Educational and Psychological Measurement*, **68**. 1. sz. 5–24. DOI: [10.1177/0013164407305592](https://doi.org/10.1177/0013164407305592)
- Wang, H. (2010): Comparability of computerized adaptive and paper-pencil tests. *Test, measurements and research services bulletin*, 13. sz. 1–7.
- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): *Practical questions in introducing computerized adaptive testing for K-12 assessments*. Pearson, San Antonio.
- Weiss, D. J. (2011): Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, **2**. 1. sz. 1–27.
- Zheng, Y. (2012): Multistage adaptive testing for a large-scale classification test: design, heuristic assembly, and comparison with other testing modes, ACT research report series.
http://www.act.org/research/researchers/reports/pdf/ACT_RR2012-6.pdf. Letöltés ideje: 2015. április 15.

Magyar Andrea és Molnár Gyöngyvér

ABSTRACT

ADAPTIVE MEASUREMENT OF READING COMPETENCE: AN EFFECTIVENESS STUDY

Andrea Magyar and Gyöngyvér Molnár

During the first years of primary education, language development plays a significant role, as children are unable to perform well in school without suitable reading ability. A paper-and-pencil diagnostic test system was developed by József Nagy in the 1990s to diagnose students' word reading ability. The spread of computers and the use of computer-based assessment have provided us with a number of new opportunities and made it possible to renew the paper-and-pencil test system by using various advantages of online testing, such as objective administration and scoring, rapid response time, immediate feedback, cheaper data-collection and adaptive testing. The aim of this paper was to explore and quantify the advantages of electronic testing and study the media effect by making detailed comparisons of test results delivered by different media. In this paper, we: (1) outline the original diagnostic assessment system for word reading ability; (2) compare 1st–5th-grade students' achievement in PP and CB mode; (3) outline the potential for improvement in the system; and (4) present the renewed online adaptive assessment system for word reading abilities. The original paper-and-pencil test system was converted into a computer-based format, the online test system was delivered via the eDia platform, and the data collection was carried out via internet using computer facilities available at schools. The pilot sample involved 154 primary school children between grades 1 and 5. A partial credit model was used to scale the items. According to the results, the online version of the test system was suitable for assessing the children's word reading ability. The person separation reliability of the test was 0.88. The difficulty parameters of the items were spread over a wide range (349–855), making the test suitable for a wide range of abilities. The research represented a promising step towards more precise educational assessment in using computerised adaptive testing among young children.

Magyar Pedagógia, 115(4). 403–428. (2015)
DOI: 10.17670/MPed.2015.4.403

Levelezési cím / Address for correspondence:

Magyar Andrea, Hódmezővásárhelyi Liszt Ferenc Ének-zenei Általános Iskola, H-6800
Hódmezővásárhely, Szent István tér 2.

Molnár Gyöngyvér, SZTE Neveléstudományi Intézet, Oktatásméleti Kutatócsoport, H-6722
Petőfi Sándor sgt. 30–34.