

A SZÓOLVASÁSI KÉSZSÉG ADAPTÍV MÉRÉSÉT LEHETŐVÉ TEVŐ ONLINE TESZTRENSZER KIDOLGOZÁSA

Magyar Andrea* és Molnár Gyöngyvér**

* SZTE Neveléstudományi Doktori Iskola

** SZTE Neveléstudományi Intézet, Oktatásméleti Kutatócsoport

Az iskola kezdő szakaszában meghatározó szerepet tölt be a nyelvi fejlődés segítése és az olvasási készségek fejlesztése (Blomert és Csépe, 2012). Megfelelő szövegfeldolgozó és értelmező képességek elsajátítása nélkül a tanulók nem képesek a tananyag elsajátítására, ami a későbbi tanulmányaik során súlyos problémákat okozhat (Józsa és Steklács, 2012). Biztos szövegfeldolgozó készség nélkül eleve kudarcra ítélt a tankönyvekből történő tanulás, az olvasott ismeretek megértése, feladatok, tesztek megoldása. Az olvasás megfelelő szintű elsajátítása ezért minden további tanulás előfeltétele. Ebből következően az olvasási készségek fejlesztése a kisiskoláskori oktatás során kiemelt feladat (Csapó és Csépe, 2012).

Hazánkban az olvasási képesség megfelelő szintű elsajátításával jelentős problémák vannak (pl. Nagy, 2004a; D. Molnár, Molnár és Józsa, 2012; OECD, 2014). Az első három PISA-mérésben, 2000-ben (480 pont; OECD, 2001), 2003-ban (482 pont; OECD, 2004) és 2006-ban (482 pont; OECD, 2007) a magyar tanulók közel azonos szinten, a nemzetközi átlag (500 pont) alatt teljesítettek. 2009-ben a papír alapú szövegek olvasásában szignifikáns javulás (494 pont; OECD, 2010), majd a 2012-es adatfelvétel eredményei alapján jelentős mértékű képességszint-esés (488 pont; OECD, 2014) következett be, aminek következtében 15 éves diákjaink teljesítménye ismét az OECD-átlag alattinak bizonyult. Digitálisszöveg-értés esetén még kedvezőtlenebbül alakultak a teljesítmények, a tanulók mind 2009-ben (468 pont; Balácsi és Ostorics, 2011), mind 2012-ben (450 pont; Balácsi, Ostorics, Szalay, Szepesi és Vadász, 2013) szignifikánsan az OECD-átlag (499 pont) alatt teljesítettek.

Miután az olvasási képesség egy nyelvi képesség, szintjei jellemezhetőek a nyelv strukturális szintjeivel: a fonetikai szint a betű-hang megfeleltetéssel, a morfológiai szint a szótagolással, a lexikai szint a szórutinokkal, a szintaktikai szint a mondatok olvasásával és a szöveg szintje a szövegértéssel (Csapó, Józsa, Steklács, Hódi és Csikos, 2012). Ennek következtében az olvasási készség négy komponens-készséggel jellemezhető, négy olvasástechnikai készségből áll: beszédhanghalló, betűolvasó, szóolvasó és mondatolvasó készség (Nagy, 2004b), melyek egymás előfeltételei is. Például az olvasáskészség fejlettségi szintjét meghatározza a szóolvasó készség fejlettségi szintje, mely a mondatolvasó

készség előfeltétele. A szóolvasó készség a szavak vizuális felismerésén alapul, mely aktiválja az olvasóban a szó jelentését (Nagy, 2004a). A köznyelvi szövegek közel 95–96%-a a leggyakoribb 4-5 ezer szóból variálódik (Nation és Waring, 1995), továbbá bármilyen típusú szöveg esetén az azt tartalmazó szavak 95 százalékának ismerete szükséges annak megértéséhez (Nagy, 2004a), ezért a köznyelvi szövegek sikeres, értő olvasásának nélkülözhetetlen előfeltétele a leggyakoribb, körülbelül 4000 szó ismerete és e szavak megfelelő szintű felismerése, olvasása.

A szóolvasó készség mérése papír alapú tesztrendszerrel

Nagy József (2006), felismerve a szókincs és a szóolvasó készség meghatározó szerepét az olvasási képesség fejlettsége tekintetében, kidolgozott egy szóolvasó készség vizsgálatára alkalmas kritériumorientált papír alapú tesztsorozatot. Miután a kutatás célcsoportja 2–12. évfolyamos diákok voltak, ezért az említett, a kívánatos 95%-os határt biztosító 4000 szóból álló szókészletet kibővítették – csökkentve az ismeretlen szavak arányát 4 százalékra – 5000 köznyelvi szóra.

A munka első lépéseként különböző források felhasználásával meghatározták ezt, a valószínűleg leggyakoribb 5000 magyar köznyelvi szóból álló szókészletet, amit az optimálisan fejlett olvasáskészség kritikus szókészletének tekintettek. Ezt követően a szókészletben szereplő szavak felhasználásával tíz diszjunkt, különböző szavakat tartalmazó ekvivalens tesztváltozatot készítettek, lehetővé téve tesztenként 500 különálló szó mérését. Minden tesztváltozatban azonos arányban fordultak elő a különböző gyakoriságú szavak. A tesztek felépítése azonos volt, azonos mennyiségben tartalmaztak címszóolvasást (C), toldalékoszó-olvasást (T), szinonimaolvasást (S) és szójelentés-olvasást (J) mérő feladatokat (Nagy, 2006).

A 10 címszóolvasás részeszt összesen 250 feladatot, feladatonként négy különböző szó olvasásvizsgálatát tette lehetővé, azaz a 10 tesztváltozatban összesen 1000 címszó olvastatására került sor. Mindegyik feladat négy szó elolvasását igényelte, majd mind a négy szó kapcsán külön-külön döntést kellett hozni, hogy az adott szó megfelel-e vagy sem a szavak mellett látható színes képnek. A toldalékoszó-olvasás részeszt hasonló felépítésű volt, szintén 250 feladatban összesen 1000 toldalékos szó elolvasását mérte. A szinonima részeszt 250 feladatban feladatonként négy szóról kellett eldönteni, hogy a feladatban szereplő szavak között szerepel-e szinonima (1000 szó). Végül a szójelentés-olvasás 100 feladata feladatonként öt szó és öt hozzá tartozó rövid, több szóból álló szómagyarázat olvasását és párosítását igényelte (2000 szó). Nagy József a szóolvasás kritériumorientált diagnosztikus, 850 feladatot tartalmazó tesztsorozatának kidolgozása során együtt vizsgálta azt, hogy (1) a diákoknak milyen a szókincse, (2) mennyire tudják elolvasni az általa kritikus szókészletbe tartozó leggyakoribb 5000 szót, valamint (3) milyen szókincsre, mely szavak hatékony olvasására lenne szükségük annak érdekében, hogy sikeres olvasóvá váljanak.

A papír alapú tesztrendszer kipróbálására 2–10. évfolyamon került sor (Nagy, 2004b). Az eredmények szerint a 10. évfolyam végéig sem alakult ki minden tanulóban a szóolvasó készség megfelelő szintje. Ennek egyik oka, hogy az alsó tagozatban elért képességfejlettségi szint a felső tagozatba lépés után alig fejlődött tovább. A tanulók között hatalmas szintbeli különbségeket azonosítottak. Már a második évfolyamon voltak olyan tanulók, akiknél optimális szinten működött a szóolvasó készség, vagyis számukra az olvasott szövegben legfeljebb 5%-nyi ismeretlen szó fordult elő, míg a 8. évfolyamos tanulók közel 40%-ánál nem alakult ki a folyékony szóolvasási készség az évfolyam végére sem. Az 5000 szavas alapszókincsre alapozó papír alapú tesztrendszerrel végzett hazai reprezentatív mintán történő adatfelvétel eredménye szerint a 80%-os kritériumszintet csak a 2. évfolyamos diákok 23, a 6. évfolyamos diákok 72%-a érte el vagy haladta meg (Nagy, 2004b).

A szóolvasó készség kritériumorientált mérését lehetővé tevő tesztsorozat továbbfejlesztésének lehetőségei

A szóolvasó készség kritériumorientált mérését lehetővé tevő rendszer továbbfejlesztése több szempont alapján is megvalósítható. A számítógépek elterjedésével megteremtődtek a feltételek a számítógép alapú tesztek adta lehetőségek minél szélesebb körben való kihasználására (Molnár, 2011; Thompson és Prometric, 2007). A szóolvasó készség mérését megvalósító feladatok kapcsán motiválóbbr környezet, azonnali visszajelzés lehetősége, a személy képességszintjéhez illesztett adaptív tesztelési technika alkalmazásával pontosabb és jóval rövidebb idő alatt történő képességszint meghatározás valósítható meg (Eggen, 2007; Thompson és Weiss, 2011; Magyar, 2014). Utóbbi feltétele a feladatok közös nehézségi skálára hozása, ami az eredeti rendszer struktúrájának újragondolását vonja maga után.

Adaptív tesztelés alkalmazása a szóolvasó készség mérésére

Az adaptív tesztelési technika alkalmazása során a tesztelt személyek attól függően kapják a teszt következő feladatait, hogy miképpen oldották meg az előzőeket. Ez a típusú feladatadás és tesztösszeállítás a hagyományos, rögzített formátumú, azonos hosszúságú tesztekkel szemben a teljesítmények sokkal finomabb mérését teszi lehetővé (Weiss, 2004, 2011a,b; Molnár, 2013), mivel jelentős mértékben megnő a tesztelés során kinyerhető itemekre és személyekre vonatkozó információ nagysága (Magyar és Molnár, 2013). Másik előnye, hogy elhanyagolhatóvá válik annak valószínűsége, hogy a tesztelt személyek ugyanazon feladatokat ugyanabban a sorrendben kapják, azaz növekedik a tesztelés biztonsága (van der Linden és Glas, 2010). Mindezen túl a kiközvetített feladatok száma, azaz a teszt hossza (Thompson és Prometric, 2007) és ezzel párhuzamosan a teszt megoldásához szükséges idő is jelentős mértékben rövidül, utóbbi átlagosan felére csökken (Frey és Seitz, 2009; Frey, Seitz és Kröhne, 2011). Az adaptív tesztelés során nő a helyes

válaszok száma, arányát tekintve mindegyik tanuló közelítően azonos helyes választ produkál, mivel a feladatok közel állnak a tanulók képességszintjéhez. Az átlagosnál alacsonyabb képességszintű tanulók számára ez nagyobb sikerélményt, míg a magasabb képességszintű diákok számára nagyobb kihívást jelent (*Jiban, Ayodele, McCarthy és Christ, 2008*).

Az adaptív teszteknek számos típusa létezik az item alapú tesztekől a többszakaszos adaptív tesztekig (*Magyar, 2012; Magyar és Molnár, 2013*). Item alapú tesztelés során a rendszer minden kiköszvetített itemet követően képességszintet számol, és ennek megfelelően választja ki a következő kiköszvetítendő itemet. Azonban lehetséges olyan elrendezés is, amikor itemek helyett rövidebb-hosszabb rögzített formátumú tesztek közvetítenek ki több szakaszban, ezt nevezik többszakaszos adaptív tesztelésnek (*Zenisky, Hambleton és Luecht, 2010*). Ennek legegyszerűbb esete a kétszakaszos teszt: egy bevezető tesztet követően a tanulók a teszten elért teljesítményük függvényében kapnak könnyebb vagy nehezebb tesztet a második részben (*Magyar, 2012*). A többszakaszos tesztek igen gyakran alkalmazott adaptív tesztek, mivel egyesítik magukban a rögzített formátumú, lineáris tesztek és az item alapú adaptív tesztek tulajdonságait, vagyis egyrészt a kérdéseket a tanuló képességszintjéhez igazítják, másrészt lehetőséget adnak az itemek sorrendjének előzetes meghatározására (*Amstrong, Jones, Koppel és Pashley, 2004; Molnár, 2013*).

A tanulmányban bemutatott kutatás célja a korábban papír alapon alkalmazott és jól működő rendszer továbbfejlesztése volt, aminek keretein belül a valószínűségi tesztelmélet adta skálázási eszközrendszer és a számítógép által kínált lehetőségeket minél szélesebb körben kihasználtuk. A rendszer átstrukturálásánál alapvető szempont volt, hogy ne változzon az alap gondolat, miszerint négy különböző szempont szerint történjék a szóolvasó készség vizsgálata (címszóolvasás, toldalékos szó olvasása, szinonimaolvasás és szójelentés-olvasás). Az azonnali visszacsatolást biztosító kritériumorientált diagnosztikus és online adaptív szóolvasó készség fejlettségi szintjét mérő tesztrendszer fejlesztésének nélkülözhetetlen feltétele volt, hogy az adaptivitás megvalósítása mellett továbbra is elegendő információt szolgáltatson mind a négy területen a tesztelt diák esetében. Ebből adódóan az item alapú adaptivitás nem biztosította volna a négy szempont együttes tesztben történő megjelenését, ezért a többszakaszos adaptív rendszerek használata mellett döntöttünk. Miután az eredeti rendszer egyes tesztváltozatai egymástól teljes mértékben különböző feladatokat tartalmaztak, így a rendszer eredeti struktúrájában nem volt alkalmas arra, hogy a feladatokat egy közös nehézségi skálán tudjuk elhelyezni. Ennek következtében a skálázás során nem alapozhattunk a korábbi papír alapú eredményekre, nélkülözhetetlen volt a feladatok ismételt, rögzített formátumú, de már számítógép alapú bemérése.

A kutatás során a következő kutatási kérdésekre kerestük a választ: (1) összeállítható-e és milyen változtatásokkal *Nagy József* szóolvasási készséget mérő papír alapú teszt-sorozatának 850 feladatából egy osztálytermi környezetben alkalmazható online adaptív diagnosztikus tesztrendszer; (2) milyen megbízhatósággal és hatékonysággal alkalmazható az online tesztrendszer a szóolvasási képesség kritériumorientált diagnosztikus mérésére 1–5. évfolyamon; (3) a tesztrendszer viselkedését befolyásolja-e a célcsoport életkora, azaz megfelelő mértékben diszkriminálja-e a rendszer a különböző életkorú diákokat?

Az adaptív tesztrendszer fejlesztésének lépései a következők voltak: (1) a papír alapú feladatok digitalizálása, feltöltése az eDia-rendszerbe (elektronikus diagnosztikus mérési rendszer), (2) az eredeti rögzített formátumú rendszer újrastrukturálása, horgonyrészesztek alkalmazásával, (3) az online rögzített formátumú tesztek felvétele, majd az eredmények alapján a feladatok skálázása, (4) az eredeti papír alapú rendszer kötöttségeihez leginkább igazítható, ugyanakkor a 21. században elvárható mérés-értékelési céloknak leginkább megfelelő adaptív rendszer kiválasztása, (5) a közös nehézségi skálán jellemzett feladatok többszakaszos adaptív tesztrendszerré alakítása, valamint (6) a rendszer viselkedésének kismintán történő tesztelése.

A szóolvasó készség online adaptív mérését előkészítő lineáris tesztrendszer kialakítása

Az eDia-rendszer

A feladatok számítógépre való konvertálása, illetve a rendszer adaptív tesztrendszerré alakítása a Szegedi Tudományegyetem Oktatásméleti Kutatócsoportja által kifejlesztett elektronikus diagnosztikus mérési rendszer (eDia; *Molnár és Csapó, 2013*) alkalmazásával történt. A platform számos feladattípus számítógépes formában való megjelenítését teszi lehetővé. A rendszer használatához elegendő egy böngésző és internetkapcsolat, ezért a tanulók a saját iskolájukban, a saját hálózatukon keresztül elérhetik és használhatják a rendszert. A tanulók a teszten nyújtott teljesítményükről azonnali visszajelzést kapnak, ezáltal a rendszer használata gyors és a papír alapúnál pontosabb visszajelzést tesz lehetővé.

A papír alapú feladatok számítógépesítése


A rendszer összes feladatát megtartva igyekeztünk azokat minél kevesebb változtatással számítógépesíteni, azaz az eDia-rendszerbe történő feltöltéssel alapvetően nem változtattuk meg a feladatok típusát. A papír és a számítógép alapú feladatok megjelenítését szemlélteti két részeszt alapján az 1–2. ábra (a bal oldali ábra a papír, a jobb oldali ábra ugyanazon feladat elektronikus formában történő megjelenítését mutatja). A papír alapú teszten karikázással és áthúzással kellett megadni a diákoknak a választ, azaz minden szó esetében döntést kellett hozniuk, hogy az adott szóval – címszóolvasás esetén például – jellemezhető-e a mutatott kép vagy sem.

Miután az eDia-rendszer alapaxiómái közé tartozik, hogy minden feladat kapcsán a válaszmezőket üresen hagyva 0%-os, a helyes megoldást megadva 100%-os teljesítményt regisztráljon a rendszer, ezért az 1. és a 2. ábrán ábrázolt feladatokat itemenkénti (szavankénti) értékelést alkalmazva nem lehetett jelölőnégyzetes feladatként digitalizálni. (Jelölőnégyzetes feladatként a diákok mind a két példa esetében 3-3 üresen hagyott válaszmezőre is kaptak volna pontot, azaz a feladatokat nem megoldó diák teljesítménye jelentősen különbözött volna a 0%-tól.) Miután a feladat megoldása során – a papír alapú formával


analog módon – mindegyik szóról döntést kellett hoznia a diáknak, ezért az igen-nem-es megoldás mellett döntöttünk.

a) épület
b) kertés
c) ház
d) ól

1.



Döntsd el, hogy a szó jelentése illik-e a képre!
Ha illik, kattints a szó után lévő "Igen" feliratú gombra,
ha nem illik, kattints a "Nem" feliratú gombra!



épület Igen Nem

kertés Igen Nem

ház Igen Nem


ól Igen Nem

1. ábra

Címszóolvasás-feladat (papír alapú és számítógépes formátumban)

a) összegyűlik
b) fordították
c) szórakozik
d) alázatosan

14.



Döntsd el, hogy a szó jelentése illik-e a képre!
Ha illik, kattints a szó után lévő "Igen" feliratú gombra,
ha nem illik, kattints a "Nem" feliratú gombra!



összegyűjtik Igen Nem

fordították Igen Nem

szórakozik Igen Nem

alázatosan Igen Nem

2. ábra

Toldalékosszó-olvasás feladat (papír alapú és számítógépes formátumban)

Egy másik jelentős változtatás a teszt egészének megjelenítésében volt. Míg papír alapon a diákok egyszerre több feladatot láttak egy tesztlapon (résztesztől függően 10–29 darabot), addig számítógépes formában egyszerre mindig csak egyet. Ennek előnye az volt, hogy a környező itemek nem befolyásolhatták a tanulót döntéshozatalában. A feladatok közötti teljes mértékű navigáció, azaz a visszalépés lehetősége csak részteszten belül volt engedélyezve.

A teszt pontozása a papír alapú rendszer szókészletmutatójával azonosan történt. Első körben azt vizsgáltuk, hogy a tesztben szereplő szavak hány százalékánál adott helyes választ a tanuló. Minden szóról meghozott döntés nulla vagy egy pontot ért, azaz a címszó-olvasás és a toldalékos szavak feladatai négy pontos feladatok, a szinonimaolvasás feladatok két pontos, a szójelentés olvasás feladatok maximum öt pontos feladatok voltak annak függvényében, hány kifejezésről hozott a diák helyes döntést.

A feladatok paraméterezése, a feladatok paraméterezéséhez szükséges lineáris tesztrendszer szerkezete, kialakítása

Az adaptív rendszer felépítésének alapvető feltétele, hogy a rendszerben lévő minden feladat viselkedése, nehézségi, illetve diszkriminációs indexe a többi feladat fényében ismert legyen. A feladatok mutatóinak meghatározásakor, azaz a feladatok paraméterezése során egyrészt a horgonyfeladatok hiánya, másrészt az esetleges médiahatás (l. *Hülber és Molnár*, 2013) megléte miatt nem alapozhattunk a korábbi papír alapú adatfelvétel eredményeire, újabb, nagymintás, rögzített formátumú, azaz lineáris tesztekkel történő számítógép alapú adatfelvételre volt szükség, ahol az egyes tesztek között horgonyfeladatok biztosították a feladatok közös skálára történő konvertálását.

A 850 feladat rögzített formátumú, lineáris teszté alakítása során a következő szempontokat vettük figyelembe: (1) a lehetőségekhez mérten a rendszer eredeti felépítésének megtartása, minél kisebb mértékű változtatás alkalmazása; (2) a feladatok és diákok közös skálán történő jellemzéséhez megfelelő mennyiségű horgonyitemek rendszerbe történő illesztése (*Yousfi és Böhme*, 2012); (3) a feladat tesztben elfoglalt pozíciójából (item position effect; *Hartig és Buchholz*, 2012) és elhelyezkedésétől, környezetéből (carry-over effect; *Yousfi és Böhme*, 2012) adódó eltérések kiküszöbölése. Mindennek megvalósítása céljából a kiegyenlített nem teljes blokk design (*balanced incomplete block design=BIBD*) latin négyzetten alapuló elrendezése volt a legmegfelelőbb (*Frey és Hartig*, 2009). A papír alapú rendszer résztesztjeiből klasztereket képeztünk, a klaszterekből bookleteket (tesztváltozatokat) oly módon, hogy egy booklet négy klasztert tartalmazott. A rendszer felépítése során a következő szempontokat vettük figyelembe:

- 1) Egy klaszteren belül minél inkább az eredeti rendszer felépítésének követése.
- 2) Egy bookletben egy klaszter csak egyszer fordult elő.
- 3) A teljes rendszerben mindegyik klaszter pontosan négyszer fordult elő.
- 4) A bookletekben lévő klaszterpárok pontosan egyszer fordultak elő, azaz ugyanaz a klaszterpár nem szerepelhetett két bookletben is.
- 5) Mindegyik klaszter mind a négy (a teszt első, második, harmadik és negyedik negyedében) pozícióban szerepelt, méghozzá pontosan egyszer.

A feltételeknek megfelelően 40 bookletet alakítottunk ki, melynek összefűzését, egymáshoz való viszonyát a 3. ábra mutatja.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	j1	j2	j3	j4	j5	j6	j7	j8	j9	j10
j7	j8	j9	j10	j6	c2	c3	c4	c5	c1	c7	c8	c9	c10	c6	t2	t3	t4	t5	t1	t7	t8	t9	t10	t6	s2	s3	s4	s5	s1	s7	s8	s9	s10	s6	j2	j3	j4	j5	j1
s8	s9	s10	s6	s7	j3	j4	j5	j1	j2	j8	j9	j10	j6	j7	c3	c4	c5	c1	c2	c8	c9	c10	c6	c7	t3	t4	t5	t1	t2	t8	t9	t10	t6	t7	s3	s4	s5	s1	s2
s5	s1	s2	s3	s4	s10	s6	s7	s8	s9	j5	j1	j2	j3	j4	j10	j6	j7	j8	j9	c5	c1	c2	c3	c4	c10	c6	c7	c8	c9	t5	t1	t2	t3	t4	t10	t6	t7	t8	t9

3. ábra

A BIBD design.

A világosszürke: címszó (c), sötétszürke: szójelentés (j), pontozott: toldalékos (t), vonalazott: szinonima (s)

A feladatok paraméterezését szolgáló kutatás mintája

A rendszer komplexitása miatt a feladatok paraméterezéséhez szükséges minta meghatározása során több szempont figyelembe vételére is szükség volt. Egy feladat biztos paraméterezése csak akkor valósítható meg, ha legalább 300 adat áll rendelkezésünkre a feladat kapcsán, illetve a korábban említett szempontok szerinti elemzés megvalósításának feltétele, hogy szempontként és feladatonként is legalább 100 adatunk legyen. Ennek megfelelően a 40 booklet beméréséhez minimum 4000 fős mintára volt szükség, ahol minden egyes bookletet legalább 100 diák megold. Ebben az esetben, miután minden klaszter négy bookletben szerepelt, minden klaszter vonatkozásában minimum 400 diák válaszra állt rendelkezésünkre.

A magyar iskolarendszer jelentős mértékű szelektivitása (Csapó, Molnár és Kinyó, 2009) miatt fennállt az adatok erős klasztereződésének veszélye, ezért a tesztek kiközvetítését egyrészt egy, a magyar iskolarendszert reprezentatívan leképező 400 iskolából álló mintából véletlenszerűen alakítottuk ki, másrészt a kutatásban részt vevő 4. és 5. évfolyamos osztályokban a diákok véletlenszerűen kapták a 40 booklet egyikét. A 4. és az 5. évfolyam kiválasztását egyrészt az indokolta, hogy a megfelelő szintű szóolvasási készségnek az alsó tagozatból felső tagozatba való lépésnél van kiemelt jelentősége, mivel a felső tagozatba lépéssel megszűnik az iskolai olvasásóra. Ugyanakkor a felső tagozaton számos olyan tantárgy lép be (irodalom, történelem), mely hosszabb szövegek elolvasását és megértését követeli meg a tanulóktól. Másrészt épp a korábbiakból adódóan – a papír alapú mérések alapján – mind 4., mind 5. évfolyamon vannak már diákok, akik képességének fejlettségi szintje vetekszik a magasabb képességszintű középiskolásokéval, és vannak olyan diákok is, akik az általános iskola első évfolyamos átlagos képességfejlettségi szintet sem érik el. A tág képességszintbeli különbségek a feladatok széles nehézségi skálán való paraméterezését teszi lehetővé. Ennek következtében a 850 feladat vonatkozásában létrejövő, 4. és 5. évfolyamos diákok teljesítményén alapuló adatbázis (N=4480) megfelelő a rendszerben lévő feladatok skálázásához.

A feladatok paraméterezését szolgáló kutatás eljárásai és eredménye

A tesztek kiközvetítése az eDia-rendszeren keresztül, az iskolák saját internethálózatát használva történt. A teszt megoldására minden diáknak egy tanítási óra (45 perc) állt rendelkezésére. Az adatok elemzése és az itemek paraméterezése a ConQuest-program felhasználásával, többdimenziós parciáliskredit-moddellel történt. Miután az alkalmazott modell a Rasch-modell egy továbbfejlesztett változata, ami egy paraméter, a nehézségi paraméter szerint rendezi az elemeket a képességskálára, így a rendszerben lévő bármely feladat nehézségi indexe összehasonlíthatóvá vált bármely másik rendszerben szereplő feladatával (Molnár, 2013).

A többdimenziós elemzést az tette szükségessé, hogy bár a rendszer a szóolvasó készség mint egy egységes konstruktum mérését valósítja meg, ám az eredeti rendszerben kialakított eszközök, résztesztek kifejlesztése különböző mérési szempontok mentén történt. Ennek megfelelően a címszóolvasást, a toldalékoszó-olvasást, a szinonimaolvasást és a szójelentés-olvasást úgy kezeltük, mint a szóolvasó készség különböző dimenzióit.

A dichotóm adatok elemzésére alkalmas Rasch-modell helyett annak továbbfejlesztett változatát, a parciáliskredit-modellt alkalmaztuk az elemzések során. Ennek oka, hogy bár a szavak szintjén külön kellett döntést hozni a diákoknak, mégis, a feladatok bizonyos szintjén összekötötték az egy feladaton belül szereplő szavakat. A teszt EAP/PV reliabilitása 0,91. Az egyes dimenziókon mért reliabilitásmutatókat az 1. táblázat mutatja.

1. táblázat. A teszt EAP/PV-reliabilitása dimenziónkénti bontásban

<i>Részteszt</i>	<i>EAP/PV-reliabilitás</i>
Címszóolvasás	0,74
Toldalékoszó-olvasás	0,89
Szinonimaolvasás	0,88
Szójelentés-olvasás	0,80

Átlagosan a címszóolvasás feladatok voltak a legkönnyebbek, majd ezt követték a toldalékoszó-olvasás feladatai, a szinonimaolvasás és a szójelentés-olvasás feladatok hasonló nehézségűeknek bizonyultak. A feladatok a nehézségi index tekintetében a diákok teljes képességskáláját lefedték, tehát a 850-ből álló feladatbank alkalmasnak bizonyult a különböző képességszintű tanulók szóolvasó készségének mérésére (4. ábra).

Magyar Andrea és Molnár Gyöngyvér

	Címszó	Szin	Szój	Told
				540.5 697.4 730.4
3				617.4
				614.4 708.4 751.4
				119.4
				332.2 636.4
				514.5 520.5 811.4
				647.4 754.4
2				554.5 565.5 722.4
		X		513.5 551.5 621.4 794.4
				375.2 501.5 543.5 550.5 574.5
		X		54.4 268.2 405.2 504.5 510.5
		X		522.5 533.5 537.5 544.5 587.5
		X	X	69.4 520.4 553.5 558.5 571.5
	X	XX	X	282.2 502.5 507.5 549.5 561.5
	X	XX	XX	441.2 503.5 508.5 511.5 517.5
	XX	XXX	XX	X 57.4 120.4 188.4 293.2 343.2
1	XXX	XXXX	XXX	XX 213.4 277.2 326.2 337.2 380.2
	XXXX	XXXX	XXX	XX 27.4 217.4 221.4 245.4 290.2
	XXXX	XXXX	XXXX	22.4 127.4 195.4 203.4 252.2
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 82.4 84.4 115.4 137.4 157.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 81.4 114.4 164.4 211.4 230.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 89.4 151.4 261.2 264.2 284.2
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 9.4 146.4 201.4 273.2 286.2
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 105.4 154.4 204.4 206.4 258.2
0	XXXXXX	XXXXXX	XXXXXX	XXXXXX 117.4 119.3 143.4 196.4 219.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 35.4 39.4 87.4 172.4 173.4 184.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 10.4 30.4 111.4 131.4 132.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 6.4 24.4 41.4 62.4 67.4 136.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 108.4 121.4 127.3 150.4 182.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 1.4 45.4 54.3 63.4 70.4 78.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 27.3 47.4 55.4 65.4 96.4 103.4
	XXXXXX	XXXXXX	XXXXXX	XXXXXX 16.4 42.4 91.4 120.3 152.4 157.3
	XX	XXXX	XXXX	XXXX 23.4 34.4 99.4 102.4 109.4 138.4
	XX	XXXX	XXXX	XXXX 22.3 50.4 75.4 92.4 107.4 110.4
-1	XX	XXXX	XXXX	XX 4.4 36.4 38.4 40.4 60.4 74.4
	XX	XXXX	XX	X 3.4 53.4 59.4 64.4 66.4 77.4
	X	XX	X	X 2.4 5.4 27.2 28.4 35.3 56.4 58.4
	X	XXXX	X	X 21.4 25.4 32.4 51.4 68.4 72.4
		XX		12.4 13.4 15.4 43.4 52.4 62.3
		X		19.4 31.4 37.4 41.3 46.4 47.3
	X	X	X	X 1.3 7.4 10.3 29.4 36.3 42.3 44.4
		X		3.3 6.3 11.4 21.3 24.3 26.4 38.3
-2	X	X		8.4 9.3 18.4 28.3 33.4 34.3 42.2
	X			4.3 12.3 14.4 18.3 30.3 33.3
				2.3 7.3 15.3 17.4 18.2 20.4 21.2
				8.3 10.2 23.3 24.2 25.3 28.2
				3.2 9.2 11.3 12.2 14.3 15.2 19.3
				4.2 7.2 16.3 17.3 24.1 26.3 31.3
				5.3 6.2 37.2 38.2 43.2 43.3 49.3
				13.3 20.3 22.2 32.3 33.1 47.2
				17.2 23.2 28.1 31.2 41.1 44.2
-3				2.2 14.2 16.2 46.2 55.1 56.2
				1.2 17.1 25.2 34.2 39.2 40.2

4. ábra

A feladatok nehézségi szintje a diákok képességszintje fényében – többdimenziós személy-ítem térkép [címszóolvasás (Címszó), szinonimaolvasás (Szin), szójelentés-olvasás (Szój) és toldalékosszó-olvasás (Told) dimenziókban]

A többdimenziós parciáliskredit-modellel becült feladatparamétereket 500-as átlagú és 100-as szórású skálára konvertáltuk. A feladatok résztesztenkénti átlagos nehézségi indexét és azok szórását dimenzióként mutatja a 2. táblázat.

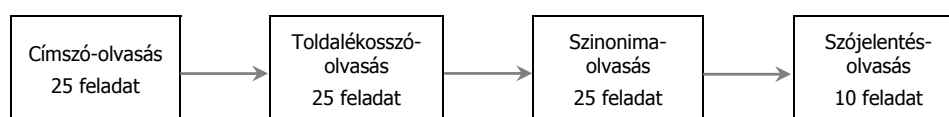
2. táblázat. Az itemek átlagos nehézségi indexei a különböző dimenziókban

Dimenziók	Átlag (képességpont)	Szórás (képességpont)
Címszóolvasás	397	34
Toldalékoszó-olvasás	473	66
Szinonimaolvasás	588	65
Szójelentés-olvasás	603	46

A szóolvasó készség online adaptív mérését megvalósító tesztrendszer kialakítása és tesztelése

A négyszakaszos adaptív tesztrendszer összeállítása

Az eredeti, papír alapú tesztsorozat 10 különböző tesztváltozatot tartalmazott, minden tesztváltozatban 85 feladat szerepelt. A feladatok dimenziónkénti megoszlását az 5. ábra mutatja.



5. ábra

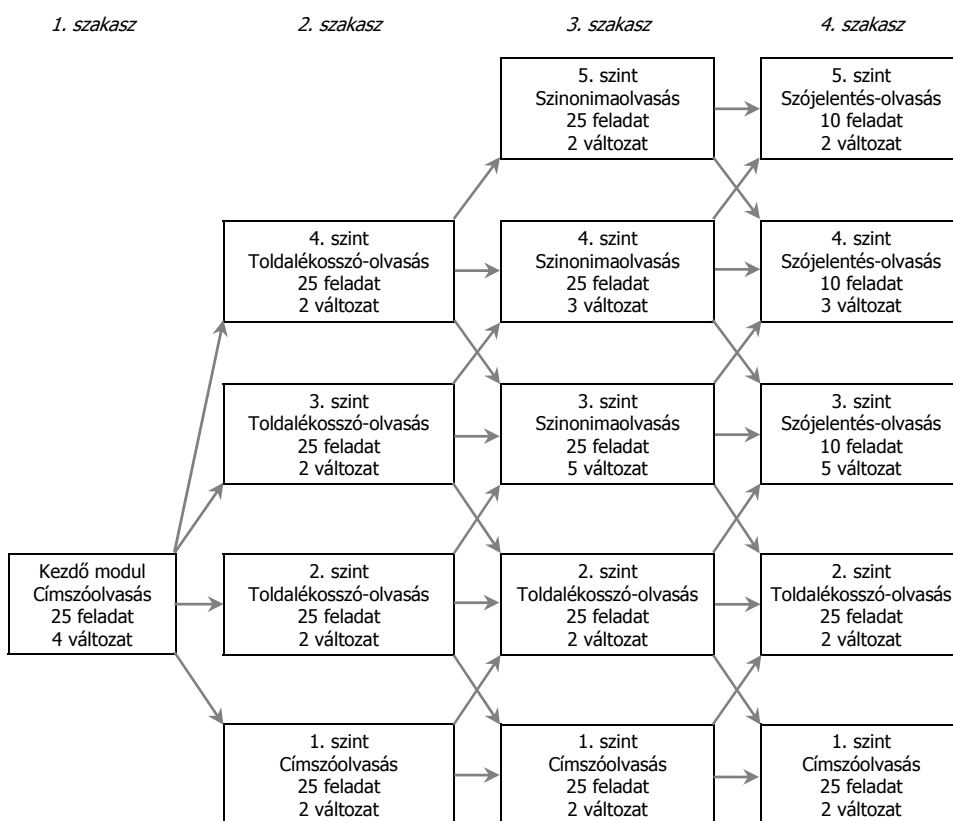
Az eredeti, lineáris tesztek felépítése

Az adaptív tesztrendszer összeállításánál fő szempontként szerepelt az eredeti teszt-szerkezet megtartása, ezért a többféle adaptív elrendezés közül a négyszakaszos, öt különböző szintet megkülönböztető tesztrendszer kidolgozását választottuk. Nagy József (2004b) a szóolvasás fejlettségét tekintve öt különböző szintet definiált: 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.

Ezt a struktúrát megtartva, az adaptív rendszer öt különböző képességfejlettségi szinten különíti el egymástól a diákokat, illetve ezzel párhuzamosan az eltérő nehézségi szintű feladatokat. A tesztrendszer kezdő, mindenki által megoldott klasztere egy közepes nehézségű címszóolvasás feladatokat tartalmazó klaszter. Ez a klaszter négy, azonosan felépített, azonos viselkedésű klaszterből kerül ki véletlenszerűen, ezzel kiküszöbölve a rendszer többszöri használatából adódó gyors elévülést, minimalizálódik a tesztre történő em-

lékezés. E klaszter feladatain nyújtott teljesítmény függvényében lépnek a tanulók nehezebb vagy könnyebb feladatokat tartalmazó klaszterekre. A címszóolvasást, a toldalékoszó-olvasást és a szinonimaolvasást tartalmazó klaszterek 25-25 feladatot tartalmaznak, a szójelentés-felismerés klaszterei 10 feladatot.

Az adaptív rendszer elágazási szabályát (6. ábra) a Nagy József-i rendszer megtartásával, azzal összhangban alakítottuk ki. Az első szintről legalább 60%-os teljesítmény elérése volt szükséges a második szintre való lépéshez, a második szintről 70%-os eredménnyel lehetett a harmadik szintre lépni, a harmadik szinten 80%-os teljesítmény volt az elvárt követelmény, illetve a negyedik szintről 90% teljesítése után lehetett az ötödik szintet elérni.



6. ábra

A szóolvasó készség mérésére alkalmas négyszakaszos adaptív tesztrendszer szerkezete

Azzal, hogy a különböző szinteken különböző nehézségű feladatok szerepelnek, a rendszer egyrészt segíti a gyengébb tanulókat, mivel számukra könnyebb a feljebbjutás,

másrészt kihívást állít a magasabb szinten lévők elé, mert a magasabb szinteken nehezednek a feladatok. Miután az eredeti tesztrendszer összes feladatát integráltuk, ezért a tesztrendszer minden szakaszának minden szintjén több azonos felépítésű és nehézségű klaszter szerepel, melyek közül véletlenszerűen történik annak kiválasztása, hogy melyiket kell a diáknak az adott szakasz adott szintjén megoldani (6. ábra).

Az azonos szinten lévő modulok átlagos nehézségi szintjében nincs szignifikáns különbség. Az egyes szinteket jellemző átlagos képességszinteket mutatja a 3. táblázat.

3. táblázat. A különböző szinteken megjelenő feladatok átlagos nehézségi indexei (átlag: 500, szórás: 100)

Szint (pont)	Átlagos nehézség (pont)	Szórás (pont)
1. (350–386)	377	12
2. (398–474)	430	26
3. (477–582)	545	27
4. (530–635)	602	22
5. (636–855)	680	35

A rendszer alkalmazásával közös képességskálán jellemezhetőek mind a feladatok, mind a tanulók képességszintjük függvényben (Baker, 2001). Minden tanuló képességskálán elfoglalt helye megmutatja, hogy melyek azok a szavak, amelyeket nagy valószínűséggel ismer, és melyek azok, amelyeket nagy valószínűséggel nem még akkor is, ha a tanuló az általa megoldott tesztben azzal a konkrét szóval nem is találkozott (Molnár, 2013).

A szóolvasó készség online adaptív mérését megvalósító tesztrendszer kipróbálása

Az adaptív rendszer megfelelő működését pilotmérés keretében térképeztük fel. A kismintás adatfelvétel 2014 tavaszán, 154 általános iskolás tanuló részvételével zajlott. A tanulók évfolyam szerinti eloszlását a 4. táblázat tartalmazza. Az adatfelvétel a tanulók saját iskolájában, saját internethálózatukon keresztül az eDia-rendszer segítségével történt. A feladatok megoldására 45 perc állt a tanulók rendelkezésére. A rendszerbe egyéni mérési azonosítójukkal léptek be a tanulók, és a tesztelés végén azonnali visszajelzést kaptak teljesítményükről.

4. táblázat. A minta évfolyamonkénti eloszlása

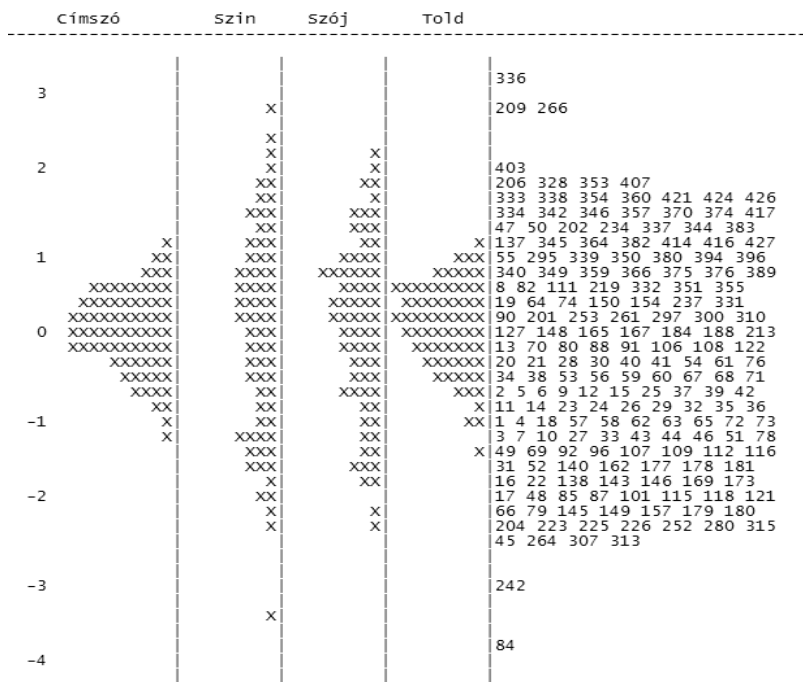
Évfolyam	Fő
1.	16
2.	28
3.	38
4.	42
5.	30

A kutatás során megbízhatónak bizonyult a rendszer, a személyseparációs reliabilitásmutató mind a teljes rendszer szintjén (0,88), mind az egyes dimenziók szintjén (5. táblázat) megfelelő volt.

5. táblázat. A teszt EAP/PV reliabilitása dimenziókénti bontásban

Dimenzió	EAP/PV-reliabilitás
Címszóolvasás	0,75
Toldalékoszó-olvasás	0,89
Szinonimaolvasás	0,87
Szójelentés-olvasás	0,88

A tanulók személy-ítem térképét mutatja a 7. ábra. Mind a négy dimenzió vonatkozásában a minta képességeloszlása közelíti a normál eloszlást. Az ábra jobb oldalán az ítemek lefedik a minta képességtartományát, tehát a feladatok megfelelőek voltak a vizsgált korosztály képességszintjének.



7. ábra

A pilot adaptív adatfelvétel alapján kirajzolható többdimenziós személy-ítem térkép [címszóolvasás (Címszó), szinonimaolvasás (Szin), szójelentés-olvasás (Szój) és toldalékoszó-olvasás (Told)]

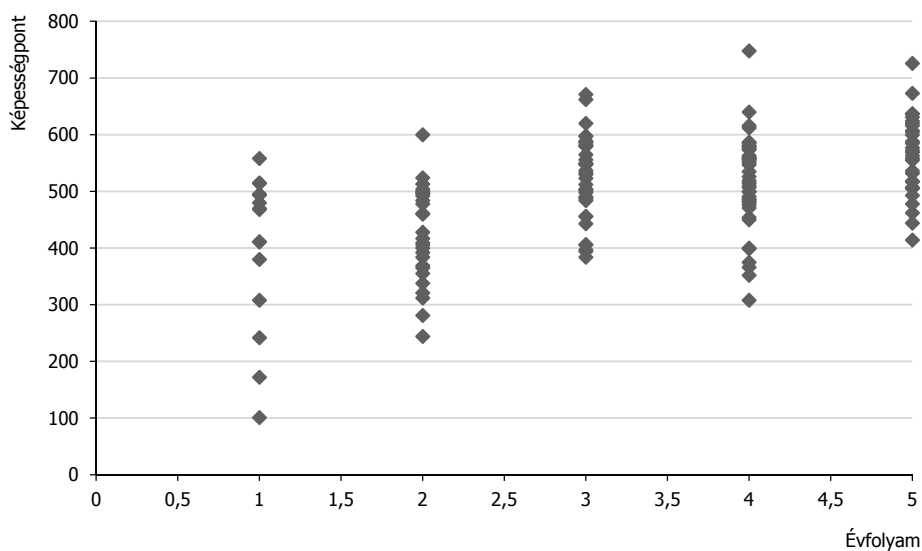
A tesztrendszer a diákok szóolvasási készségfejlettségének diagnosztizálására is alkalmasnak bizonyult. A becsült képességszintek képességpont és százalékos teljesítményének átlagát évfolyamonkénti összehasonlításban mutatja a 6. táblázat. Az első és második évfolyam átlagos képességszintjében nem volt szignifikáns különbség, tőlük szignifikánsan jobban teljesítettek a 3–5. évfolyamos diákok.

A rendszer helyes működését jellemzi, ha a diákok százalékos teljesítménye megegyezik a szintenként előre meghatározott százalékos teljesítménnyel, ami arra utal, hogy mindenki a képességszintjéhez leginkább közel álló feladatokat kapta a tesztelés során, azokat a feladatokat, amelyek megoldási sikeressége a legtöbb információval szolgál a diák képességszintjéről (Molnár, 2013). Ebben az esetben ez az elméleti érték a második szinten (átlagos teljesítmény 430 pont, 2. és 3. táblázat) 70% volt, amit igazoltak az empirikus adatok is (az első és a második évfolyamos diákok átlagos teljesítménye 70% volt). A 3–5. évfolyamos diákok átlagosan a harmadik szintnek megfelelő teljesítményt mutattak, azaz átlagos teljesítményük a rendszer kialakítása során meghatározott 80% körülnek kellett lennie, amit alátámasztanak az empirikus adatok.

6. táblázat. A teszten elért képességszintek átlaga és szórása évfolyamonkénti bontásban

<i>Évfolyam</i>	<i>N</i>	<i>Min. (pont)</i>	<i>Max. (pont)</i>	<i>Átlag (pont)</i>	<i>Szórás (pont)</i>	<i>Átlag (%)</i>	<i>Szign.</i>
1.	16	100	558	407	134	70	
2.	28	280	599	426	84	70	
3.	38	384	671	523	72	80	{1, 2} < {3, 4} < {5}
4.	42	308	748	517	84	78	
5.	30	462	726	565	70	82	

A rendszer mindamelllett, hogy alkalmasnak bizonyult az évfolyamok között meglévő átlagos különbségek detektálására, az évfolyamokon belül megjelenő különbségek diagnosztizálását is lehetővé tette. Mind a legalacsonyabb képességszintű első és második évfolyamos diákok, mint a legmagasabb képességszintű 5. évfolyamos tanulók képességszint szerinti elkülönítésére alkalmas volt (7. ábra). A kismintás mérés eredményei alapján a tesztrendszer tág intervallumban, megfelelő információ kinyerése mellett megvalósította a diákok képességszintjéhez illesztett adaptív tesztelést. Az évfolyamokon belül megjelenő képességszintbeli különbségek, illetve az egyes évfolyamok átlagos teljesítményei alapján megállapítható, hogy a feladatok skálázásához kijelölt 4. és 5. évfolyam és az ott tanuló diákok által lefedett tág képességszint-intervallum megfelelő volt a feladatok skálázásához.



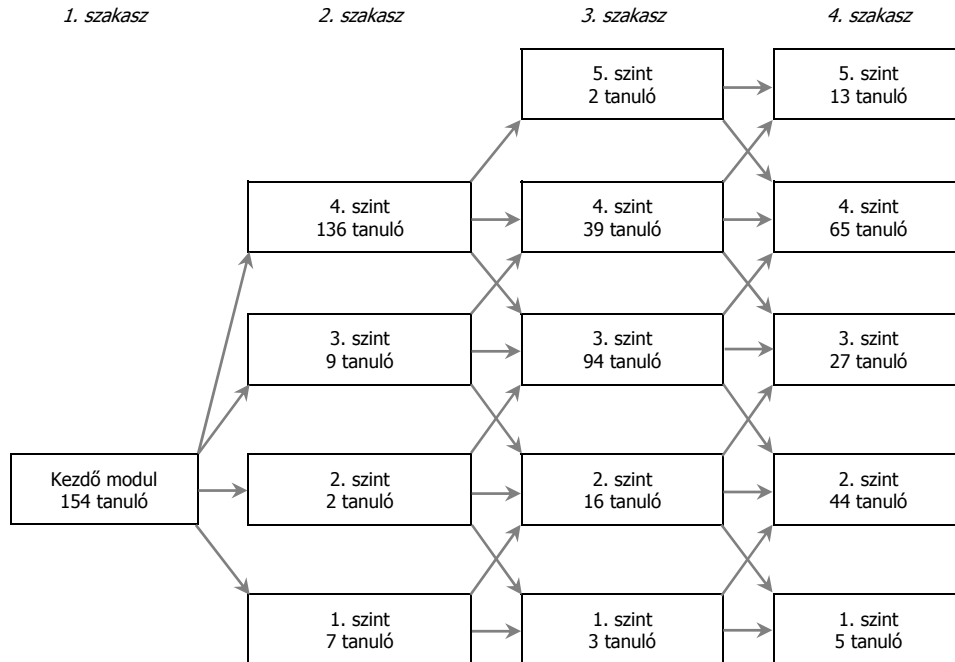
7. ábra

A tanulók képességszint szerinti eloszlása évfolyamonkénti bontásban

A diákszintű eredmények is alátámasztották a rendszer helyes és tervezett működését. Az adaptív teszt összesen 40 különböző klasztert tartalmazott, ezekből a kismintás adatfelvétel során 38-at osztott ki a rendszer? A szakaszokon belül a tanulók azonos tulajdonságokkal jellemezhető klaszterenkénti eloszlását mutatja gyakoriság szerinti bontásban a 8. ábra. A kezdő modulon nyújtott teljesítménye alapján a tanulók legnagyobb része a második szakaszban 4. szintű modult kapott, majd a harmadik szakaszban a tanulók döntő többsége a 3., illetve a 4. szinten folytatta tovább. A negyedik szakaszban a legtöbb tanuló a 4. szinten, illetve a 2. szinten helyezkedett el.

A kismintás kutatás eredményei alapján a rendszer mind évfolyamonkénti, mind diákonkénti bontásban helyesen működött. Az alacsonyabb képességszintű diákok tipikusan a könnyebb, a magasabb képességszintűek a nehezebb klasztereket kapták a tesztelés során, amivel az adatfelvételkor kinyert információ mennyisége javult, miután minden diák a képességszintjéhez relatív közel álló feladatokat kapott a teszt utolsó moduljában. Az utolsó két modul tekintetében 31 tanuló esetén nem változott a kapott modul szintje a harmadikról a negyedik szakaszba való lépésnél, ami a tanulók ötödét jelenti, tehát mindenképpen indokolt volt a négy szakasz alkalmazása.

A szóolvasási készség adaptív mérését lehetővé tevő online tesztrendszer kidolgozása



8. ábra

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

Összefoglalás

Kutatásunk célja olyan online adaptív tesztrendszer kidolgozása volt, mely alkalmas általános iskolás diákok szóolvasási készségének mérésére. Az iskolai sikeresség, az önálló tanulás elengedhetetlen feltétele a megfelelő szintű olvasási képesség birtoklása, aminek közvetlen alapfeltétele a szóolvasó készség működésének optimális szintje. A szóolvasó készség mérésére a 2006-ban kidolgozott papír alapú, 10 különböző tesztből álló tesztsoportot dolgoztuk át és alakítottuk egy egységes adaptív tesztrendszeré. A feladatok közös skálán történő jellemzését, az esetleges médiahatás kiküszöbölését, a szükséges paraméterek meghatározását egy 4500 fős mintán megvalósított kutatás keretein belül végeztük el. A minta és a kutatás felépítésének meghatározása során figyelembe vettük, többek között, a magyar iskolarendszer szelektivitásából adódó esetleges problémákat, a közös képességskála kialakításához szükséges ismérveket (pl. horgonyitemek alkalmazását), az eredeti, papír alapú rendszer filozófiáját és kialakítását, illetve több újabb vizsgált, a

mérési eredményeket befolyásoló tényezőket (pl. a feladatok elhelyezkedésének és környezetének befolyásoló hatása). A feladatok skálázása többdimenziós parciáliskredit-moddal történt. Az itemek nehézségi paraméterei széles skálán helyezkedtek el (349–855), ami a rendszer tág életkori és ezzel párhuzamosan tág képességszint szerinti intervallumban történő alkalmazását teszi lehetővé. Az adaptív tesztrendszer a papír alapú rendszer alkalmazásának eredményeire alapozva öt képességszinten különíti el egymástól a diákokat. A nagymintás empirikus adatokon alapuló skálázott tesztrendszer kipróbálása 154 1–5. évfolyamos tanuló bevonásával pilot mérés keretein belül történt. Az elemzések szerint a tesztrendszer megbízhatóan alkalmazható általános iskolás diákok szóolvasási képessége diagnosztizálására.

A kutatás jelentőségét az adja, hogy a rendszer adaptív módon, azaz a diákok képességszintjéhez illeszkedően tudja mérni a tanulók szóolvasói képességszintjét és a közös skálára konvertálás következtében minden diákról megadható, hogy a rendszerben előforduló 5000 szó olvasási képességében hol áll, még akkor is, ha az adott szó nem fordult elő tesztjében. A tesztrendszer nemcsak annak mérésre alkalmas, hogy a tanuló hány szót ismer, hanem annak meghatározására is, milyen nehézségű szavak ismeretének birtokában van, melyek azok a szavak, amelyeket nagy valószínűséggel ismer – a közös nehézségi és képességszkála következtében még akkor is, ha a számára kiközvetített tesztben azok a szavak nem szerepeltek –, és melyek azok, amelyeket nem, vagyis ahol fejlesztésre szorul.

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

Irodalom

- Armstrong, R. D., Jones, D. H., Koppel, N. B. és Pashley, P. J. (2004): Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement*, 28. sz. 147–164.
- Baker, F. B. (2001): *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation. University of Maryland, College Park, MD.
- Balázs Ildikó és Ostorics László (2011): *PISA2009 Digitális szövegértés. Olvasás a világhálón*. Oktatási Hivatal, Budapest.
- Balázs Ildikó, Ostorics László, Szalay Balázs, Szepesi Ildikó és Vadász Csaba (2013): *PISA 2012 Összefoglaló jelentés*. Oktatási Hivatal, Budapest.
- Blomert, L. és Csépe Valéria: Az olvasástanulás és –mérés pszichológiai alapjai. In: Csapó Benő és Csépe Valéria (szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 17–86.
- Csapó Benő és Csépe Valéria (2012): Bevezetés. In: Csapó Benő és Csépe Valéria (szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 9–16.
- Csapó Benő, Józsa Krisztián, Steklács János, Hódi Ágnes és Csíkos Csaba (2012): A diagnosztikus olvasás felmérések részletes tartalmi kereteinek kidolgozása: elméleti háttér és gyakorlati kérdések. In: Csapó Benő és Csépe Valéria (szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 189–218.

A szóolvasási készség adaptív mérését lehetővé tevő online tesztrendszer kidolgozása

- Csapó Benő, Molnár Gyöngyvér és Kinyó László (2009): A magyar oktatási rendszer szelektivitása a nemzetközi összehasonlító vizsgálatok eredményeinek tükrében. *Iskolakultúra*, **19**. 3–4. sz. 3–13.
- D. Molnár Éva, Molnár Edit Katalin és Józsa Krisztián (2012): Az olvasásvizsgálatok eredményei. In: Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 17–82.
- Eggen, T. J. H. M. (2007): Choices in CAT models in the context of educational testing. In: Weiss, D. J. (szerk.): *Proceedings of the 2007 GMAC conference on computerized adaptive testing*. <http://publicdocs.iacat.org/cat2010/cat07eggen.pdf>. Letöltés ideje: 2014. május 3.
- Frey, A. és Hartig, J. (2009): An NCME instructional module on booklet design sin large-scale assessments of student achievement: Theory and practice. *Educational Measurement*, **28**. 3. sz. 39–53.
- Frey, A. és Seitz, N. N. (2009): Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, **35**. 2–3. sz. 89–94.
- 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. 103–120.
- Hartig, J. és Buchholz, J. (2012): A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling*, **54**. 4. sz. 418–431.
- 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.
- Józsa Krisztián és Steklács János (2012): Az olvasás tanításának tartalmi és tantervi szempontjai. In: Csapó Benő és Csépe Valéria (szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 137–188.
- Magyar Andrea (2012): Számítógépes adaptív tesztelés. *Iskolakultúra*, **22**. 6. sz. 52–60.
- Magyar Andrea (2014): Adaptív tesztek készítésének folyamata. *Iskolakultúra*, **24**. 4. sz. 13–22.
- Magyar Andrea és Molnár Gyöngyvér (2013) Adaptív és rögzített formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*, **113**. 3. sz. 181–193.
- Molnár Gyöngyvér (2011): Az információs-kommunikációs technológiák hatása a tanulásra és oktatásra. *Magyar Tudomány*, **172**. 9. sz. 1038–1047.
- 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, 2012. április 11–13. 82. o.
- Nagy József (2004a): Olvasástanítás: a megoldás stratégiai kérdései. *Iskolakultúra*, **14**. 3. sz. 3–26.
- Nagy József (2004b): A szóolvasó készség fejlődésének kritériumorientált diagnosztikus feltérképezése. *Magyar Pedagógia*, **104**. 2. sz. 123–142.
- Nagy József (2006): A szóolvasó készség fejlődésének kritériumorientált diagnosztikus feltérképezése. In: Józsa Krisztián (szerk.): *Az olvasási képesség fejlődése és fejlesztése*. Dinasztia Tankönyvkiadó, Budapest. 91–106.
- Nation, P. és Waring, R. (1995): Vocabulary size, text coverage and word lists. <http://www.fltr.ucl.ac.be/fltr/germ/etan/bibs/vocab/cup.html>. Letöltés ideje: 2013. június 4.
- OECD (2001): *Knowledge and skills for life. First results from the OECD Program for International Students Assessment (PISA) 2000*. OECD, Párizs.
- OECD (2004): *Learning for tomorrow's world-first results from PISA 2003*. OECD, Paris.
- OECD (2007): *PISA 2006: Science competencies for tomorrow's world. Executive summary*. OECD, Paris.

- OECD (2010): *PISA 2009 results: Executive summary*. OECD, Párizs.
- OECD (2014): *PISA 2012 results: What students know and can do student performance in mathematics, reading and science*. OECD, Paris.
- 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, N. A. és Weiss, D. A. (2011): A framework for the development of computerized adaptive tests. *Practical Assessment Research and Evaluation*, **16**. 1. sz. 1–9.
- van der Linden, W. J. és Glas, C. A. W. (2010): *Elements of adaptive testing*. Springer, New York.
- Weiss, D. J. (2004): Computerized adaptive testing for effective and efficient measurement in counselling and education. *Measurement and Evaluation in Counselling and Development*, **37**. 2. sz. 70–84.
- Weiss, D. J. (2011a): Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, **2**. 1. sz. 1–27.
- Weiss, D. J. (2011b): Item banking, test development, and test delivery. In: Geisinger, Kurt F. (2013, szerk.). *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. APA handbooks in psychology. DC, US: American Psychological Association, Washington, 185–200.
- Yousfi, S. és Böhme, H. F. (2012): Principles and procedures of considering item sequence effects in the development of calibrated item pools: Conceptual analysis and empirical illustration. *Psychological Test and Assessment Modelling*, **54**. 4. sz. 366–396.
- 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.

ABSTRACT

ANDREA MAGYAR AND GYÖNGYVÉR MOLNÁR: DEVELOPING AN ONLINE ADAPTIVE TESTING SYSTEM FOR WORD READING ABILITY

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 (PP) diagnostic test system was developed by *József Nagy* to diagnose children's word reading ability in the 1990s. With the spread of computers and applications of computer-based (CB) assessment, a number of new opportunities were introduced, making it possible to renew the paper-and-pencil test system by incorporating a number of 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 the achievement of children in Years 1 to 5 in PP and CB modes; (3) outline options for improving the system; and (4) represent the renewed online adaptive assessment system for word reading ability. The original PP test system was converted to a CB format. The online test system was delivered via the eDia platform, and data collection was carried out via the Internet using computer facilities available at schools. In the pilot sample 154 primary school children were involved between Years 1–5. 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 covered a large scale (349–855), which made the test appropriate for a wide range of abilities. The research represented a promising step towards more precise educational assessment using computerised adaptive testing among young children.

Magyar Pedagógia, **114**. Number 4. 259–279. (2014)

Levelezési cím / Adress for correspondence:

Magyar Andrea, SZTE Neveléstudományi Doktori Iskola, H–6722 Szeged, Petőfi S. sgt. 30–34.

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