

8.

A DIFER tesztek online változatával végzett mérések tapasztalatai

Csapó Benő

MTA-SZTE Képességfejlődés Kutatócsoport
Szegedi Tudományegyetem Neveléstudományi Intézet

Molnár Gyöngyvér

Szegedi Tudományegyetem Neveléstudományi Intézet

Nagy József

Szegedi Tudományegyetem Neveléstudományi Intézet

Bevezetés

A sikeres iskolakezdés alapvetően meghatározza a későbbi évek tanulási eredményeit, ezért világszerte jelentős erőfeszítéseket tesznek annak érdekében, hogy minden gyermek felkészülten kezdje meg az iskolát. Ennek a törekvésnek számos összetevője van, kutatási, fejlesztési, oktatáspolitikai, szociálpolitikai és gazdasági problémák sokaságát kell megoldani a cél érdekében. A kutatási-fejlesztési feladatok egyik leglényegesebb eleme olyan könnyen alkalmazható tesztek kidolgozása, amelyekkel pontosan meg lehet állapítani, hol tart egy gyermek a sikeres iskolakezdéshez szükséges készségek fejlődése tekintetében.

A szegedi egyetemen közel négy évtizedes hagyománya van az iskolakészültség-tesztek fejlesztésének. A fejlődés természetes lépése, hogy amint rendelkezésre áll a szükséges technológiai háttér, a tesztek átkerülnek a legmodernebb informatikai környezetbe. A papír-ceruza tesztek számítógépre való átültetésének, a technológiaalapú tesztek fejlesztésének már ugyancsak több évre visszanyúló előzménye van, ugyanakkor az online iskolakészültség-tesztek fejlesztése során számos új problémát is meg kell oldani. Kisgyermekkorban alkalmazható teszteket kell készíteni, óvodások, írni-olvasni még nem tudó kisiskolások, korlátozott számítógépes tapasztalattal rendelkező gyermekek számára. Továbbá kevés kutatási előzménye van a személyes megítélésen alapuló egyéni adatfelvételt igénylő tesztek és az online tesztek összehasonlításának.

Ez a fejezet az online mérés előnyeit mutatja be a DIFER öt tesztjének hagyományos és digitalizált változatát összehasonlítva. Az öt teszt különböző pszichológiai jellemzőket mér, és online változataik különböző technológiai megoldásokat igényelnek. A mérőeszközöknek ez a változatos összetétele lesz segítségünkre annak bemutatásában, hogy a nehézségeket le lehet küzdeni, és a technológiaalapú tesztelés nem csupán kivitelezhető ilyen fiatal korban, hanem sok tekintetben még alkalmasabb is, mint a személyes felmérés vagy papíralapú tesztelés.

Az iskolakészültség mérésével kapcsolatos vizsgálatok

Egyre több kutatási és fejlesztési program foglalkozik a sikeres iskolakezdés feltételeinek megteremtésével, és ennek során különös szerepet kap az iskolaérettség felmérésére alkalmas eszközök fejlesztése és a fejlődési folyamat nyomon követése. Az eszközök széles választéka, köztük megfigyelési protokollok, tesztek és tesztbattériák állnak rendelkezésre a kognitív fejlődés különböző aspektusainak, valamint az iskolában elsajátítandó alapkészségek előfeltételeinek felmérésére. Gyakran előfordul azonban, hogy egy-egy kutatási keretek között érvényesnek és megbízhatónak bizonyult eszköz rendszeres iskolai használatra túl bonyolultnak bizonyul. Nem elég pontos, ha nem standardizált feltételek között használják, vagy nem speciálisan erre kiképzett pedagógusok alkalmazzák. Sok esetben a mérés megvalósításának és a tesztek javításának idő- és emberi munkaigénye jelenti a rendszeres használat legfőbb akadályát. Ilyen helyzetekben a technológiaalapú tesztelés nyújthat segítséget. Kisgyermekek körében még beiskolázás előtt vagy annak kezdeti szakaszában azonban komoly problémákat vethet fel a technológia használata.

Mivel bizonyos készségeknek az óvodáskori fejlődése jelentős hatással van a gyerekek későbbi iskolai sikerességére, a kutatások egyik fő célja azon készségek azonosítása, amelyek meghatározó szerepet töltenek be a későbbi fejlődésben. Annak megállapításához, hogy a gyerekek elérték-e az iskolakezdéshez szükséges fejlődési szintet, érvényes és megbízható eszközre van szükség (Snow és Van Hemel, 2008).

Az iskolakészültség-tesztek az oktatásban használt tesztek egy speciális csoportját képviselik. Fejlődésorientáltak, egy dinamikus folyamatot

mérnek, amelyben gyors és hirtelen minőségi változások következhetnek be. Formatív, gyakran diagnosztikus eszközök segítik a fejlődés követését, és ha szükséges, a megfelelő fejlesztési beavatkozások megválasztását. A legelterjedtebb eszközöket szűrésre használják: minden diákot tesztelnek, hogy meg lehessen állapítani, kinek van szüksége különös figyelemre, egyes területekre koncentráló további tesztekre vagy a jelzett fejlődési zavarra szakosodott szakember segítségére. Bár még marad nyitott kérdés az iskolaérettségi felmérés bizonyos tartalmi részleteire és az adatok felhasználásának módjaira vonatkozóan, általános az egyetértés abban, hogy az alkalmas és könnyedén használható mérőeszközök elérhetősége kulcsfontosságú a gyerekek sikeres iskolakezdésében (Snow, 2006).

A kisgyermekkorban alkalmazott tesztek sajátossága, hogy olyan gyermekeket kell velük felmérni, akik még nem tudnak olvasni. Ellentétben a hagyományos papíralapú tesztek döntő többségével, ezeket a tesztek többször egyénileg veszik fel, a feladatokat a tesztet felvevő személy (gyakran a felméréndő diákok tanára) mutatja be, ő mondja el az instrukciókat, és ő jegyzi fel a válaszokat. Ez a sajátos helyzet korlátozza a teszt-körülmények standardizálását, és nem zárja ki szubjektív hatások érvényesülését a diákok válaszáinak értelmezésében. Az iskolaérettségi felmérések sokszor a gyerekek megfigyelésén alapulnak (pl. az EDI, lásd Guhn, Janus és Hertzman, 2007). A pedagógusok megítélései gyakran elfogultnak bizonyulnak, amikor a véleményüket összevetjük objektív szintfelmérő eszközök eredményeivel (Mashburn és Henry, 2004). Minthogy az első iskolaévek fő célja az alapvető írás-olvasási és számolási készségek fejlesztése, az iskolaérettségi tesztek gyakran az írás, olvasás, számolás és következtetés alapkészségeit mérő feladatokat és az iskolai munkához szükséges társas, szociális készségek fejlettségi szintjét mérő feladatokat tartalmaznak (Konold és Pianta, 2005).

A pedagógiai értékelés szinte minden területén intenzív kutató-fejlesztő munka folyik a technológiaalapú tesztek kidolgozása érdekében (Csapó, Ainley, Bennett, Latour és Law, 2012). A PISA keretében 2006-ban kezdődtek az előkészületek a természettudomány számítógépes felmérésével (OECD, 2010), majd 2009-ben sor került a digitális szövegértés tesztelésére (OECD, 2011). A 2012-es mérésben már két fő területen (szövegértés és matematika) a papíralapú tesztekkel párhuzamosan számítógépen is futott a tesztelés, és emellett a dinamikus problémamegoldás volt az innovatív területe, amely szimulációra, interakcióra épülő feladatokat is tartalmazott,

bevonva a PISA-mérésekbe a számítógépes tesztelés legújabb lehetőségeit (OECD, 2014). A 2015-ös ciklussal kezdődően azonban már csak technológiaalapú adatfelvételre kerül sor.

Számos előnye miatt a számítógépes módszer az iskolakészültség vizsgálatában is fokozatosan felváltja a papíralapút, mivel lehetővé teszi a tanulók egyéni jellemzőihez igazított mérést (például adaptív tesztelés), az automatikus kiértékelést (akár az ígéretesen fejlődő beszédhang-felismerést is; lásd Price és mtsai., 2009), továbbá az azonnali visszajelzést. Különböző innovatív feladatformátumokat lehet alkalmazni (pl. multimédiás elemek, szimulációt tartalmazó feladatok, dinamikus itemek), a teszttingerek prezentációja pontosan szabályozható, ki lehet küszöbölni az adatrögzítést, és a tesztek adminisztrációs költségeit is jelentősen csökkenteni lehet.

Az utóbbi időben számos vizsgálatot végeztek annak elemzésére, hogy a tesztelés módja befolyásolja-e a teszteredményeket (Clariana és Wallace, 2002; Kingston, 2009; Wang, Jiao, Young, Brooks és Olson, 2008). A papíralapú és a számítógépes tesztek eredményei közötti különbségeket mind az érvényesség és megbízhatóság, előnyök és hátrányok, mind pedig a háttérváltozók (nem, etnikum, technológiával kapcsolatos tényezők, mint például számítógépes jártasság; Gallagher, Bridgeman és Cahalan, 2000; Csapó, Molnár és R. Tóth, 2009) szempontjából vizsgálták már, azonban alig néhány tanulmány foglalkozik a nagyon fiatal tanulók technológiaalapú környezetben való tesztelésével (Choi és Tinkler, 2002; Carson, Gillon és Boustead, 2011).

A fejezet több átfogó projekt tapasztalataira épít. A DIFER kifejlesztésére irányuló kutatásokon és az online diagnosztikai feladatbank létrehozását szolgáló munkálatokon túl felhasználja a *Szegedi Iskolai Longitudinális Program* (Csapó, 2007, 2014) eredményeit is, ahol az iskolaérettségi teszt-csomag biztosította az első mérési pontot (Józsa, 2004). A longitudinális mérések online megvalósításához nélkülözhetetlen a megbízható eszközök kifejlesztése. A legtöbb eddigi médiahatás-vizsgálat csupán a papíralapú és a számítógépes tesztek hasonlította össze. Ebben a fejezetben viszont az egyéni személyes tesztelés és az online tesztelés közötti különbségekkel foglalkozunk, aminek nincsenek még kutatási előzményei.

A DIFER teszt online változatának kidolgozása

A fejezet középpontjában álló iskolaérettségi teszt kifejlesztése az 1970-es években kezdődött, amikor egy nagymintás felmérés számos olyan készséget vizsgált, amely szükséges a sikeres iskolakezdéshez. Hozzávetőleg 10 ezer gyermeket mértek fel ugyanazzal a tesztsorozattal az óvodától kezdve az első iskolás évekig. E munka eredménye (Nagy, 1980) képezte az alapját a PREFER tesztsomag kifejlesztésének, amely lefedte a sikeres iskolakezdéshez szükséges legfontosabb alapkészségeket (Nagy, 1987). Ezt az eszközt több mint egy évtizedig használták az oktatási gyakorlatban, majd az új kutatási eredmények alapján megújították. Az új, DIFER elnevezésű tesztsomagot standardizálták, és statisztikai optimalizálással egy rövidített változatot is szerkesztettek belőle (Nagy, Józsa, Vidákovich és Fazekasné Fenyvesi, 2004a). A tesztsorozat megbízhatóságát és előrejelző validitását több nagy mintán alapuló felmérés igazolta már, az egyik ezek közül lefedte az adott évben iskolába lépő teljes gyermekpopulációt (Nagy, Józsa, Vidákovich és Fazekasné Fenyvesi, 2004b).

A DIFER-t úgy tervezték, hogy a használata ne igényeljen különleges szakértelmet; óvónők és általános iskolai tanítók is felvehetik. A tesztsorozat egyik fő hátránya, hogy csak személyesen és egyénileg végezhető el, így egy tanuló teljes körű felmérése akár több órát is igénybe vehet. Ez különösen az iskolában probléma, ahol nehéz a tanárok és a diákok óraarendjébe illeszteni a tesztelési alkalmakat. Mivel az óvodai foglalkozások rugalmasabbak, ott gyakrabban is használják a DIFER-t. Egy másik kérdés a tesztek felvételének objektivitása. Előfordulhat, hogy a tanár kicsit másképpen olvassa fel az instrukciókat az egyes tanulóknak, és a válaszok pontozása is kismértékben változó lehet. A feladatok online közvetítésével, melynek része az előre rögzített hanginstrukció (a szövegeket képzett szakemberek olvassák fel), és a válaszok automatikus pontozásával ezek a problémák kiküszöbölhetők. Mivel a DIFER hét tesztje közül kettő (szociális készségek és írásmozgás-koordináció) jelenleg nem digitalizálható, a jelen tanulmányban a fennmaradó öt tesztet (beszédhanghallás, relációszőkincs, elemi számlálás, tapasztalati összefüggés-megértés, tapasztalati következtetés) használjuk.

Beszédhanghallástereszt

A teszt 60 fonémakontraszt-észlelést mérő itemet tartalmazott, melyek segítségével megállapítható, hogy a diákoknak jó-e a beszédhanghallásuk, azaz meg tudják-e különböztetni egymástól az egyes kritikus fonémapárok tagjait, mint például /v/ – /f/ és /b/ – /p/. A teszt első résztesztje váltóhangos szópárokat (pl. *fél-vél*) tartalmazott (22 item). A diákok képi illusztráció mellett meghallgatták a szópárok mindkét szavát egy-egy mondatba ágyazva, majd a két szó közül csak az egyiket hallották, és ki kellett választaniuk az elhangzott szót ábrázoló képet. A második részteszt (8 item) feladatai azt vizsgálták, hogy a diákok a folyamatos beszédben mennyire hallják tisztán a beszédhangokat. A diákoknak a hallásbeli pontos különbségtétel alapján kellett döntést hozniuk, miközben a szó jelentését egy rajz ábrázolta. A hallott szópár egyike minden esetben helyes ejtésű szó, a másik helyes vagy torzított ejtésű szó volt (pl. *citom-citrom*). Végül a harmadik és a negyedik részteszt valós (17 item) és álszavakat (13 item), szópárokat tartalmazott képi segítség nélkül, melyek egymástól legfeljebb egy fonémában különböztek. A diákoknak el kellett dönteniük, hogy a szópárok két tagja azonos volt-e, vagy sem.

A teszt eredeti egyéni adatfelvételt igénylő változatában a tesztet felvevő személy például az első részteszt esetén felolvasta a szópárt tartalmazó mondatot, miközben rámutatott a szavakhoz tartozó képekre. Ezután felolvasta a szópár egyikét. A diákok válaszadása a szóhoz tartozó képre való mutatóval valósult meg, amit a tesztvezető értékel, pontozott és feljegyzett a válaszlapon. A számítógép-alapú változatban az instrukciókat a gyerekek fejhallgatón keresztül hallgatták meg, a hangfájl lejátszása közben a program piros keretet rajzolt az eredeti változatban mutatóval jelzett kép köré. Ez, miután előre időzítéssel történt, mindenki számára egységes volt. A diákok a megfelelő képre kattintással adtak választ. Az egérrel való kattintás egy nagy piros pöttyöt hagyott a kiválasztott képen. A 8.1. ábrán látható egy illusztráció erről a feladattípusról.

Relációsószókincs-teszt

Minden nyelvben találunk relációkat kifejező szavakat, amelyek különböző dolgokat összekötő téri viszonyokra (pl. *alatt*, *fölött*), jellemzőkre és folyamatokra utalnak. Az ilyen relációsószavak ismerete és megfelelő hasz-

nálata az iskolai tanulás előfeltétele. A DIFER programcsomag a relációszókincs-teszt négy ekvivalens változatát tartalmazza; jelen fejezet mind a négy változat számítógépes és papíralapú kivitelezését is felhasználja.



8.1. ábra. Minta feladat a beszédhanghallásteresztből egyéni adatfelvétel és számítógép-alapú adatfelvétel esetén

A négy relációszókincs-teszt mindegyike 24 itemet tartalmazott, ezek térbeli (8 item; pl. belül, között), időbeli (4 item; pl. délután, tavasz), mennyiségi (4 item; pl. páratlan, kevés) vagy hasonlósági (4 item; legrövidebb, azonos hosszúságú) viszonyokat fejeznek ki, továbbá az igekötők relációs szerepét (4 item; pl. rálép, belép) vizsgálják. A feladatok egy részének több mint egy helyes megoldása volt. A 8.2. ábra egy számítógépes feladatot mutat be, melynek ábrája megegyezik a papíralapú változat ábrájával.

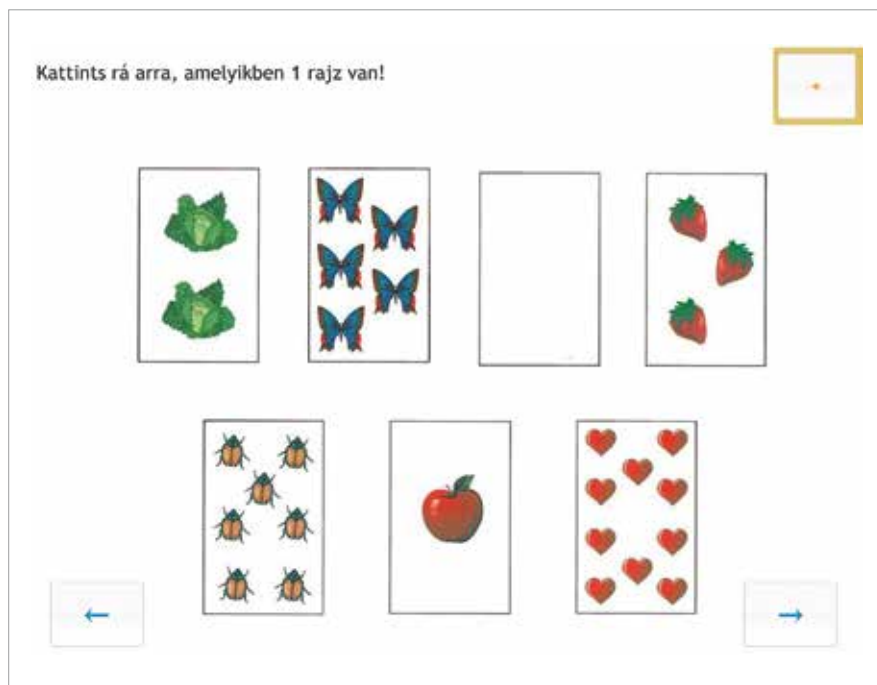
Az egyéni adatfelvétel során a tesztet felvevő személy olvasta fel az utasításokat és értékelte a diákok válaszait, amelyeket a gyerekek a megfelelő kép(ek)re mutatással adtak meg. Számítógépes helyzetben az instrukciókat fülhallgatón keresztül hallgathatták meg, majd a megfelelő kép(ek)re való kattintással adhatták meg válaszaikat. Ez utóbbi esetben a kiértékelés automatikus volt; a program a többszörös jó megoldásokat is tudta kezelni.



8.2. ábra. Példafeladat a relációszókinsz-teszt számítógépes változatából

Elemi számolási készség tesztje

Az eredeti egyéni felvételre tervezett teszt hat részesztből állt, melyek a számossági viszonyokat és a számérzékenységet mérték: 21-ig számolás (egy item), 501-ig számolás (öt item), visszafelé számolás 501 és 0 között (nyolc item), húszas számkörben manipulatív számolás (tárgyakkal végzett műveletek; tizenegy item), tízes számkörbeli számképfelismerés (kilenc item), százas számkörbeli számolvasás (számok jelének felismerése; négy item). Mivel az online platform még nem tud szóbeli válaszokat kezelni, nem volt lehetőségünk minden elemi számolási készséget mérő feladat digitalizálására, a szóbeli választ igénylő feladatokat elhagytuk a számítógépes verzióban. A 8.3. ábra illusztrálja a teszt számítógépes változatának feladatait. A személyes és számítógépes adatgyűjtés menete megegyezett a beszédhanghallástereszténél leírtakkal.



8.3. ábra. Példafeladat az elemi számolási készség tesztjéből

Tapasztalati összefüggés-megértés

A tapasztalati összefüggés-megértést vizsgáló teszt eredeti, személyes adatfelvételre készült változata 32, kontextusba ágyazott feleletalkotó feladatot tartalmazott. Ugyanezen képességterületet mérő teszt számítógépes változata 32 többszörös választásos feladathból állt, lehetővé téve az azonnali kiértékelés megvalósítását. Minden feladat két premisszával (állítás-sal) kezdődött, majd a tanulóknak meg kellett fogalmazniuk a megfelelő következtetést.

Az egyéni adatfelvétel során a tesztet felvevő egyszer, ismétlés nélkül felolvasta az instrukciókat, és a tanulóknak szóban kellett válaszolniuk a feltett kérdésre, amit az adatfelvételt végző személy pontozott és feljegyzett egy válaszlapon. Számítógépes módban a tanulók tetszés szerinti alkalommal meghallgathatták az instrukciókat és a lehetséges következtetéseket, mielőtt a megfelelő rádiógombra való kattintással kijelölték válaszukat.

Tapasztalatikövetkeztetés-teszt

A teszt 32 mindennapi szituációba ágyazott többszörös választásos feladata a legfontosabb kijelentés- és predikátumlogikai sémákat fedte le. Minden egyes feladat két premisszát és egy konklúziót tartalmazott. A diákoknak a két állítás alapján be kellett fejezniük a megkezdett harmadik állítást, azaz a konklúziót (l. *Nagy, Józsa, Vidákovich és Fazekasné Fenyvesi, 2004*).

A számítógépre való átültetés korlátai

A tesztek digitalizálása során az egyik fő célunk a feladatok lehető legtöbb részletének megőrzése volt annak érdekében, hogy a két tesztelési mód összehasonlítható legyen. A hét DIFER tesztből kettőt, a tanulók viselkedésének megfigyelésére épülő szocialitáskészség-tesztet, illetve a finommozgásokat mérő írástesztet nem lehetett megvalósítani az online platformon. A fennmaradó öt DIFER tesztet sikeresen konvertáltuk számítógépes formátumba, bár néhány feladatot ki kellett hagynunk, és a nyitott kérdéseket átalakítottuk többszörös választásos feladatokká az automatikus pontozás érdekében. A számítógépes tesztelés általános előnyeinek túl két jelentős jellemzővel bővült az online kivitel, amely a személyes felvételi DIFER tesztekben nem szerepelt: az instrukciókat standard előre rögzített hang adta, és a pontozás automatikus volt. Az előbbi a tesztek megbízhatóságát javította, az utóbbi pedig a használhatóságot növeli, és a tanulók motivációját erősítheti az azonnali visszajelzésnek köszönhetően.

A felmérések módszerei

A minta

A mintát első osztályos (6–7 éves) diákok alkották. Az öt teszt digitális változatát más-más diákok oldották meg. A számítógépes adatfelvételben tesztenként kb. 400 diák vett részt (lásd 8.1. táblázat). A beszédhanghallás teszt esetében a személyes és a számítógépes adatfelvétel mintája megegyezett, a többi esetben nem.

8.1. táblázat. A vizsgálat mintái

Teszt	N (egyéni)	N (számítógépes)
Beszédhanghallás	364	364
Relációszőkincs 1	1892	426
Relációszőkincs 2	1204	409
Relációszőkincs 3	1328	456
Relációszőkincs 4	999	449
Számlálás	1895	435
Tapasztalati összefüggés-megértés	424	402
Tapasztalati következtetés	1094	416

A relációszőkincs, az elemi számolás és a tapasztalati következtetés tesztjeinek online eredményeit összevetettük egy egyéni adatfelvételen alapuló országos reprezentatív mintán felvett felmérés eredményeivel. A relációszőkincs-teszt mind a négy változatát használtuk, ezért mindegyiket a teljes minta egynegyedén vettük fel.

A mérések lebonyolítása

A személyes adatfelvétel során a tanulókat egyenként vizsgáltuk. A feladatok instrukcióit a tesztet felvevő pedagógusok olvasták fel, majd egy értékelőlapon ők pontozták a válaszokat. Az adatokat központilag vittük számítógépre, a tanulók nem kaptak visszajelzést az adatgyűjtés után.

Az online adatfelvétel az eDia elektronikus diagnosztikus platformon történt interneten keresztül. A számítógépes környezetben az előre felvett instrukciókat fejhallgatón keresztül hallgatták meg a diákok. Az adatfelvétel a részt vevő iskolák számítógépes termeiben zajlott. A tesztfelvétel 20-45 percet vett igénybe az adott tesztől függően. A válaszokat automatikusan pontozta a program, és a tanulók azonnali visszajelzést kaptak (helyes válaszok aránya) a teszt végén.

Az online tesztelés alkalmazhatósága az iskolakészültség felmérésében és a kisiskolások fejlődésének követésében – eredmények

Az online tesztek iskolai alkalmazásának tapasztalatai

Míg a DIFER személyes egyéni felvétele tanulónként hozzávetőlegesen egy órát vett igénybe, a számítógépes változathoz tipikusan két iskolai tanóra volt szükség. A csoportos tesztelést megnehezíthette, ha az iskolának nem volt teljes osztályokat befogadó számítógépterme. Az iskolák többségében azonban van ilyen terem, és más intézményekben a tanulókat kisebb csoportokban vagy egyénileg lehet tesztelni.

A számítógépes tesztelés további korlátja a beiskolázás kezdetén a tanulók nem megfelelő számítógépes ismerete lehet. A tesztek fejlesztésénél különös figyelmet fordítottunk arra, hogy ne legyen szükség speciális számítógépes készségekre a teszt megoldásához, és a gyakorlatban a tanulók eltérő informatikai műveltsége nem okozott problémát. A számítógépek, operációs rendszerek és böngészők sokféleségének hatásait részletesebben nem vizsgáltuk, de ez a körülmény nem okozott megfigyelhető problémát, és nem volt érezhető hatással az eredményekre.

Az online tesztek megbízhatósága

A tesztek reliabilitása (belső konzisztenciája, Chronbach- α) megfelelőnek bizonyult mind az egyéni, mind a számítógép-alapú adatfelvétel során. Előbbi esetben a reliabilitásmutatók értékei 0,71 és 0,89 között mozogtak, utóbbi esetben pedig 0,74 és 0,94 között. Általában az online adatfelvétel mutatói magasabbnak bizonyultak, mint a szemtől szemben, egyénileg végzett adatfelvétel eredményei alapján számolt mutatók értékei (8.2. táblázat).

A legmagasabb reliabilitásmutató a beszédhanghallásteszt esetében volt tapasztalható, ahol a személyes adatfelvételen alapuló eleve magas megbízhatósági mutató (Chronbach- $\alpha = 0,89$) még magasabbra emelkedett (Chronbach- $\alpha = 0,94$). A javulást a standardizált beszédhangingereknek tulajdoníthatjuk. Az egyéni adatfelvétel során viszonylag alacsony megbízhatósági értéket mértünk a relációszőkincs-, a tapasztalatkövetkeztetés- és a tapaszt-

latati összefüggés-megértési teszteken, amelyek közül az első kettő jelentős, a harmadik pedig némi javulást mutatott a számítógépes változatban.

8.2. táblázat. Az egyéni és a számítógép-alapú adatfelvétel eredményeinek megbízhatósági mutatói

Teszt	Itemek száma	Chronbach- α	
		Egyéni adatfelvétel	Számítógépes adatfelvétel
Beszédhanghallás	60	0,887	0,938
Relációszőkincs 1	24	0,796	0,844
Relációszőkincs 2	24	0,706	0,824
Relációszőkincs 3	24	0,749	0,809
Relációszőkincs 4	24	0,768	0,851
Tapasztalati összefüggés-megértés	32	0,743	0,831
Tapasztalati következtetés	32	0,885	0,904
Elemi számolás	13	0,812	0,739

Az elemi számolási készség tesztje esetében ezzel szemben a megbízhatósági mutató jelentős romlását figyelhetjük meg. Miután a hagyományos verzió több feladatát is ki kellett hagynunk a számítógépes változatból, mert szóbeli választ igényeltek, amit nem lehetett megvalósítani az online tesztkörnyezetben, a két különböző módon felvett tesztet nem tekinthetjük ugyanazon konstruktum mérőeszközének. A 13 feladatból álló rövidített teszt egyéni felvétele is viszonylag magas megbízhatóságot mutatott (0,81), azaz az alacsony elemszám nem ad magyarázatot a számítógépes változat alacsonyabb megbízhatóságára.

A két tesztelési módból adódó különbségek

A tesztelés módja számos tekintetben befolyásolhatja az eredményeket. Másutt, részletesebb elemzésekkel megmutattuk, hogy amennyiben pontosan ugyanazokból a feladatokból állítjuk össze a teszteket a kétféle médiumon, a mérési skálák invariánsak maradtak a papíralapúról a technológiai alapúra való átültetés során, vagyis a két mérőeszköz lényegében ugyanazt méri (Csapó, Molnár és Nagy, 2014). Ebben a fejezetben a felmérés tartal-

mának, a feladatok típusának és a vizsgált személyek jellemzőinek hatásait elemezzük. Az eltérő módokon felvett tesztek lehetővé teszik a személyes és számítógépes változatok páronkénti összehasonlítását. Mivel különböző kutatási elrendezést alkalmaztunk a tesztek felvételekor, más és más elemzéseket végezhattünk a különböző tesztekre vonatkozóan.

Az egyéni és az online tesztelési módból adódó különbségek

Megvizsgáltuk a tesztelés megváltozott módjának hatását a diákok teljesítményére (8.3. táblázat). A DIFER kritériumorientált tesztekben áll, ezért az első osztályos tanulók többsége már elérte az iskolakezdéshez szükséges fejlettségi szintet.

8.3. táblázat. Az egyéni és számítógépes adatfelvétel eredményei közötti különbségek

Teszt	Egyéni adatfelvétel (%)		Számítógép-alapú adatfelvétel (%)		d	t	p<
	Átlag	Szórás	Átlag	Szórás			
Beszédhanghallás	91,35	9,88	82,61	16,80	0,59	-8,79	0,001
Relációszőkincs 1	80,86	15,44	78,03	17,86	0,17	-3,32	0,01
Relációszőkincs 2	83,88	12,06	73,06	18,23	0,72	-13,62	0,001
Relációszőkincs 3	84,07	13,39	72,86	18,16	0,71	-13,98	0,001
Relációszőkincs 4	84,84	13,62	81,55	17,59	0,21	-3,86	0,001
Tap. összefügg.-m.	70,69	14,34	63,59	20,87	0,39	-5,72	0,001
Tapasztalati köv.	69,64	20,66	55,93	23,64	0,58	-8,61	0,001
Elemi számolás	87,35	18,46	88,90	14,84	-0,09	1,52	n. sz.

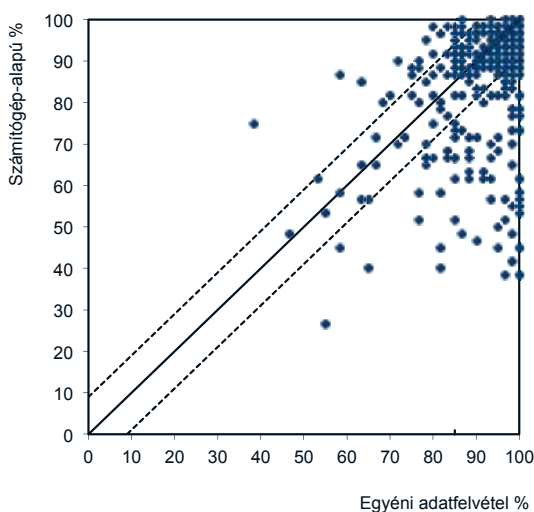
Megjegyzés: n. sz.: nem szignifikáns

Az eredmények szerint nyolcból hét esetben a számítógépes teszteken gyengébb eredményt értek el a tanulók, mint a személyes adatfelvételen alapuló teszteken. Ez arra utal, hogy a tesztet felvevő pedagógusok gyakran magasabb pontszámokat adnak a gyerekeknek, mint a számítógép automatikusan (és objektíven) kiszámolt pontszámai. A különbség másik magyarázata lehet, hogy a tanulóknak nehézségeik voltak a számítógép kezelésével, és ez rontotta a teljesítményüket. Miután nem találtunk jelentős különbségeket az elemi számolás (magas pontszám mindkét verzióban) esetén, a számítógép-kezelési nehézség nem lehet kielégítő magyarázat a teljesítménykülönbségekre. Valószínűbb, hogy a pedagógusok elnézőbbek voltak a gyerekek válaszaival szemben.

A beszédhanghallástereszt eredményeiben számottevő esés volt megfigyelhető: az igen magas 91,4 százalékról a szintén magas, de lényegesen alacsonyabb szintű 82,6 százalékos teljesítményre. A relációszókincs-teszt négy változatában kétféle változás jelentkezett: jelentős teljesítménysökkenés két esetben, kismértékű teljesítményváltozás a másik kettőben. Az elemi számolás esetében nem észleltünk számottevő változást.

Az egyéni és a számítógépes adatfelvétel eredményeinek különbségei fejlettségi szint függvényében

A tesztelési mód teljesítménybefolyásoló hatásainak pontosabb megértése érdekében megvizsgáltuk, hogy mely képességfejlettségi szinteken volt a legerősebb a megváltozott közvetítő eszköz hatása. A beszédhanghallásteresztet ugyanazon a mintán vettük fel a hagyományos és számítógépes tesztelés során, ami lehetőséget adott az eredmények tanulónkénti összehasonlítására. Amint korábban láttuk, a pedagógusoktól kapott átlagos pontszámok sokkal magasabban voltak, mint a számítógépen elért pontszámok. A tanulók eredményei a 8.4. ábrán láthatók, ahol a vízszintes tengely a személyes módban, a függőleges tengely pedig a számítógépes módban elért eredményeket mutatja tanulónként.



8.4. ábra. A tanulók hagyományos, illetve számítógépes módban nyújtott teljesítményei

A tanulók hagyományos és számítógépes környezetben mutatott teljesítménye között szignifikáns, de viszonylag alacsony volt a korreláció együttható ($r = 0,346$, $p < 0,001$). Sok esetben a tanulók hagyományos tesztelési módban elért teljesítménye 100 százalékos vagy 90 százalék fölötti, míg ugyanezen diákok számítógépes környezetben nyújtott teljesítménye messze elmarad ettől a szinttől. Ez arra utal, hogy az adatfelvételt végző pedagógusok feltételezhetően hajlamosak akkor is elfogadni a tanulók választát, amikor a gyerekek nem a helyes választ adják meg, vagy a kérdés feltevésekor egyéb eszközökkel segítenek a diákoknak.

Lehetséges ugyanakkor, hogy a pedagógus megítélése pontosabb, és a tanulók valamilyen oknál fogva nem tudják kezelni a számítógépet, ezért az nem helyesen méri a képességeiket, és alulbecsüli a teljesítményüket. Az a tény azonban, hogy néhány másik teszt esetében a tanulók nem ütköznek hasonló akadályokba, ellentmond ennek a feltételezésnek. A másik lehetőség szerint a számítógépes teszt eredményei helyesek, és tanítók túlbecsülik a tanulók valódi képességeit. Több tényező is közrejátszhat ebben az eredményben: a tanulókat segítheti a helyes kép kiválasztásában az instrukciók felolvasásának módja, a helyes és a helytelen válaszlehetőség artikulálása (szemben a semleges előre rögzített hanggal), a tanítók arckifejezései és mozdulatai, vagy a tanítók esetleg elfogadnak kétértelmű válaszokat helyesnek. A pontok eloszlását tekintve a 8.4. ábrán ez utóbbi értelmezés valószínűbbnek tűnik, és a teszt számítógépes felvétele feltehetően objektívebb módszer. További kutatás (pl. a két tesztelési mód előrejelző validitásának összehasonlítása) szükséges azonban ahhoz, hogy pontosabban tudjuk megítélni a két tesztformátum objektivitását.

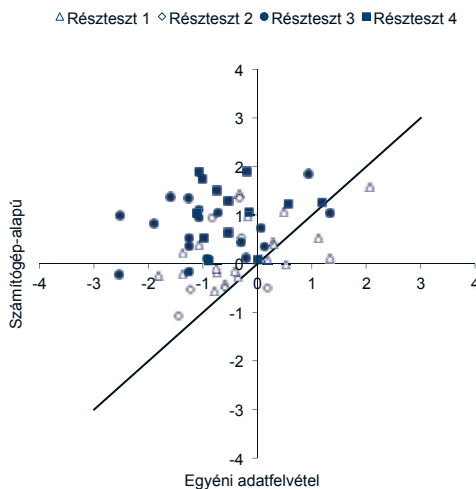
Itemszintű különbségek

A számítógépes tesztelés jellemzőit az itemek szintjén is tanulmányozhatjuk, összehasonlítva, miként viselkednek a két különböző tesztelési médiumon. Ezzel a módszerrel megvizsgálhatjuk, hogy melyek a feladatoknak azok a sajátosságai, amelyek inkább vagy kevésbé érzékenyek arra, hogy milyen médiumon valósul meg a tesztelés. Itemszintű elemzéseket a klasszikus és a valószínűségi tesztelmélet (IRT) eszközeit alkalmazva egyaránt végezhetünk. A legnyilvánvalóbb összehasonlítás annak tanulmányozása, hogy mennyire bizonyultak nehéznek az egyes itemek a különböző médiumokon. Ehhez az IRT modelleknel az itemek

nehézségi indexét, a klasszikus tesztelméleti megközelítésben a megoldottság arányát (p érték) használhatjuk fel.

Ismét kihasználjuk a beszédhanghallásterzt felvétele kapcsán alkalmazott kutatási elrendezést, azt, hogy mindkét tesztelési módban ugyanazon diákok oldották meg a feladatokat. Ez lehetőséget biztosít az IRT-elemzések elvégzésére és az itemek nehézségi szintjének közös skálán való jellemzésére, azok összehasonlítására.

A 8.5. ábra minden egyes alakzata egy itemet mutat. Az item vízszintes (x) koordinátája a hagyományos adatfelvétel alapján számolt nehézségi indexe, függőleges (y) koordinátája a számítógép-alapú adatfelvételen nyújtott teljesítmény alapján számolt indexe. A könnyebb feladatokhoz kisebb, a nehezebb feladatokhoz nagyobb nehézségi index tartozik. Ennek megfelelően azok a feladatok, amelyek könnyebbnek bizonyultak az egyéni adatfelvétel során, az ábrán a folytonos egyenessel jelzett vonal fölött helyezkednek el.



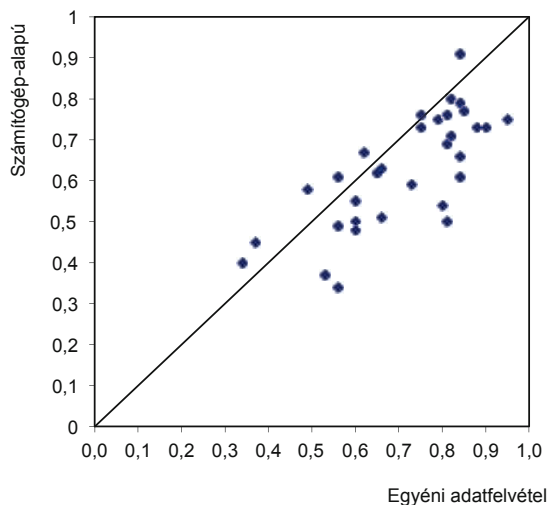
8.5. ábra. Az IRT modellel számított itemparaméterek összehasonlítása a beszédhanghallásterzt egyéni és számítógépes adatfelvételi módjában

A tartalmi elemzés szerint a legnagyobb eltérés az álszavakra építő feladatok esetén van. Ugyanez az effektus figyelhető meg az olyan szavak esetében, ahol a kontextus hiányát részben helyettesíthette a tanító segí-

tó magatartása, míg a teljes mondatokban szereplő szavaknál a személyes tesztelés hatása kevésbé nyilvánvaló. (Csak ez utóbbi esetben volt szignifikáns a korreláció az itemnehézségi paraméterek között: $r = 0,55$, $p < 0,01$.) Ezek az eredmények szintén azt a hipotézist igazolják, hogy bizonyos esetekben a tesztfeltevő kevésbé objektív, mint a gép, főként, ha az ingert kiemeljük a szokásos szövegkörnyezetéből.

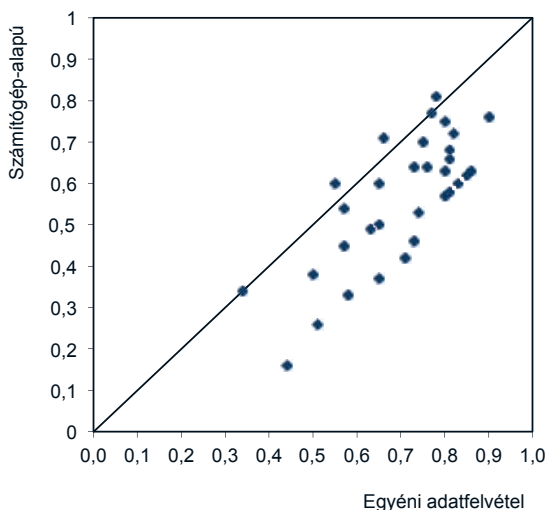
Ahol különböző mintákon vettük fel a két különböző médiumra készült tesztek, a teljes mérőeszközökre vonatkozó összehasonlítás pontossága a két minta összehasonlíthatóságán is múlik. Ugyanakkor az itemek egymáshoz viszonyított helyzete hasznos információkat szolgáltat a médiumok különbségeit illetően is.

A 8.6. ábra a tapasztalati összefüggés-megértési teszt itemeinek viselkedését mutatja be a két médiumon. Itt a nehezebb itemekhez kisebb megoldottsági arány tartozik. A vonal alatt elhelyezkedő itemek könnyebbek voltak a hagyományos tesztfelvétel esetén. Az eredmények szerint a tapasztalati összefüggés-megértés tesztjében ezek az itemek voltak többségben.



8.6. ábra. A tapasztalati összefüggés-megértési teszt itemeinek viselkedése a hagyományos és a számítógép-alapú adatfelvétel során

A tapasztalati következtetés tesztjének itemeire vonatkozó hasonló elemzés eredményét a 8.7. ábrán szemléltetjük. A 8.6. és 8.7. ábrák összehasonlításából kitűnik, hogy ezek a tesztek, illetve tesztfeladatok a két médiumon nagyon hasonlóan viselkedtek.



8.7. ábra. A tapasztalati következtetés-teszt itemeinek viselkedése a hagyományos és a számítógép-alapú adatfelvétel során

Mindkét tesztben a vonal alatt elhelyezkedő itemek voltak többségben, vagyis azok, amelyek a hagyományos tesztfelvételi módban könnyebbek. Az itemnehezégi mutatók korrelációja mindkét esetben magasnak bizonyult (tapasztalati összefüggés-megértés $r = 0,750$, $p < 0,01$; tapasztalati következtetés $r = 0,763$, $p < 0,01$).

Ezek az itemszintű elemzések is azt mutatják, hogy az online mérés jól helyettesítheti a hagyományos, személyes adatfelvételt. Ahol a két mérés között eltéréseket tapasztaltunk, ott a rendelkezésre álló információk arra utalnak, hogy az online mérések objektívebb eredményei állnak az eltérések mögött.

Következtetések, további kutatási és fejlesztési feladatok

Az iskolakészültség-tesztek digitalizálásának lehetőségei, az új médium használatának általános konklúziói

A tanulmány az iskolakészültségi tesztek új médiumra való átalakításának lehetőségeit és korlátait mutatta be. Nem lehetett digitalizálni egy magas megbízhatóságú és jó előrejelző képességű, egyéni adatfelvételt igénylő tesztet, a szociális készségek tesztjét. A szociális készségek számítógépes mérésének elvi korlátai vannak, a gépi környezet mindig csak korlátozottan helyettesítheti a személyes interakciót. Mindemellett érdemes megjegyezni, hogy a szociális készségnek az a dimenziója, amely az iskolakészültség szempontjából meghatározó jelentőségű, elsősorban az utasítások követésére való hajlandóságot, feladatokra való folyamatos odafigyelést és a feladattartást jelenti. Az ilyen jellegű tulajdonságok mérésére már lehet online teszteket készíteni.

Szintén elmaradt az írásmozgás-koordináció tesztjének digitalizálása, ugyanakkor egy ahhoz közeli konstruktum (a finom kézmozgás) könnyen mérhető más technológiák felhasználásával, és egy alternatív konstruktum (billentyűzet- és egérkezelés) szintén jól mérhető. Az idevágó technológia rendelkezésünkre áll kutatási célokra, de az kérdéses, hogy az iskolák számára széles körben elérhető-e. A teszt viszonylag alacsony előrejelző validitása szintén felveti a kérdést, vajon megéri-e az energiabefektetést, amikor az új technológiákhoz szükséges hasonló pszichomotoros készségek mérése ígéretesebbnek tűnik.

Kimaradt a digitalizálásból sok lényeges feladat az elemi számolási készség tesztjének esetében, mivel a tanulók szóbeli válaszait nem tudtuk rögzíteni. Ennek a veszteségnek az ára a gyengébb validitás és alacsonyabb megbízhatóság volt. A digitalizált változat tehát nem tekinthető az eredeti (magas megbízhatóságú és jó előrejelző érvényességű) teszt egyenértékű helyettesítésének, ezért további munka szükséges az elemi számolási képesség megfelelő számítógépes tesztjének kidolgozása érdekében.

A pozitívumokat tekintve a tanulmány alátámasztotta, hogy a technológiaalapú tesztek már a beiskolázás legkorábbi pontján is alkalmazhatók több lényeges kompetenciaterületen a rendszeres oktatási gyakorlatban is. Ezek a mérések gyakorlatilag bármikor elvégezhetők nagyon alacsony

költséggel, és a tanárok részéről kevés többlet-időráfordítást igényelnek. Miközben a számítógépes tesztelés néhány lényeges hátrányát is feltártuk, számos előnyös oldal is megmutatkozott, melyek felhasználhatók a számítógépes tesztek továbbfejlesztése során.

A jelenlegi vizsgálat korlátai

Az adatgyűjtést az eDia platform korábbi verziójával végeztük el, ami még csak a leggyakoribb itemformátumokat kezelte, de kevesebb lehetőséget adott új itemtípusok létrehozására. A korábbi nagymintás felmérésekhez képest a tanulmányban bemutatott kutatáshoz kisebb minta állt rendelkezésünkre, mivel még fejlesztés alatt levő rendszert használtunk. Az elemzési lehetőségeket szintén korlátozza a háttéradatok hiánya. A kutatás jelenlegi fázisában nincs adatunk annak a lehetőségéről, hogy azonos eszközzel végzett ismételt teszteléssel kísérjük figyelemmel a fejlődést. (A DIFER korábbi személyes tesztelési módszerét rendszeresen használják erre a célra minden probléma nélkül.) Nincs adat továbbá a számítógépes tesztelés előrejelzési validitásáról sem.

További korlátja a jelenlegi elemzésnek, hogy az eredeti DIFER teszteket az óvodából az iskolába átlépő gyerekek mérésére tervezték. Így a már iskolás tanulók rendre a maximum körül teljesítenek ezeken a teszteken, és az ő adataik nem igazán alkalmasak a teszt működésének vizsgálatára. Ez a hiányosság esetleg kiküszöbölhető azzal, hogy a jövőben a kutatást kiterjesztjük az óvodai populációra, bár a számítógép-használat ebben a korban még további megvalósíthatósági tanulmányokat igényel.

Az új iskolaérettségi tesztek fejlesztésének további lehetőségei

A jelen tanulmányban az iskolaérettségi felmérések online médiára való konvertálását vizsgáló kutatást mutattunk be, megalapozva ezzel egy olyan új tesztesomag létrehozásának a lehetőségét, amely a legnagyobb mértékben kihasználja a számítógépes technológia nyújtotta előnyöket. Azonban ennek az eszköznek a létrehozása és felhasználása még további kutatásokat igényel, legalább három irányban: (1) a jelenleg hagyományos módszerekkel mért területeket a számítógépes technológia által mérhetővé kell tenni,

és olyan új konstruktumokat is kell mérni, amelyek különösen jól illenek a számítógépes technológiához; (2) tovább kell fejleszteni az online felmérési technológiát olyan funkciókkal, amelyeket az infotechnológia más területein már használnak; (3) meg kell találni annak a módját, hogy a rendszeres kisgyermekkorai felmérések beépíthetők legyenek az oktatási folyamatba.

A jelenleg számítógéppel még nem mérhető területek között a szociális készségek számítógépes mérésének megoldása különösen fontos feladat. Intenzív kutatási munka szükséges ahhoz, hogy olyan alternatívákat és itemhelyettesítéseket találjunk, amelyekkel kiválthatóak a jelenleg használatos egyénileg felvett tesztek. Szintén további erőfeszítések szükségesek olyan új jelenségek mérésére, mint például az olvasás előfeltételei, a fonológiai tudatosság vagy a megfelelő olvasási fejlődés indikátorai, mint például a gyors automatikus megnevezés.

Számos technológiai megoldás létezik más területeken, amelyek beépíthetők az online mérési rendszerekbe, azok lehetőségeit bővítve. Már most használható az interakció, szimuláció, a tárgyak irányítása a képernyőn, az ingerek új típusai, mint például a film és animáció. Az ingerek időzítése és az információ kontrollált megjelenítésének más módszerei szintén lehetségesek. Szintén egyszerűen megoldható a reakcióidő mérése, a billentyűzet gombnyomásának vagy az egér mozgásának rögzítése, bár további vizsgálatok szükségesek ahhoz, hogyan tudnak ezek az eszközök hozzájárulni a korai mérések gyakorlati problémáinak megoldásához. Egy példa a gyakorlati problémákra, ahol a létező technológiai megoldások kulcsfontosságúak lehetnek, épp a jelen vizsgálatok során merült fel: hangfelismerési technológia szükséges ahhoz, hogy a számolási tesztek online is végrehajthatók legyenek.

További kutatások szükségesek ahhoz is, hogy megismerjük az online mérések alkalmazhatóságát az oktatásban. Az ökológiai érvényesség egyike azon kérdéseknek, amelyek alapos mérlegelést igényelnek. Az előrejelzési validitás kritikus fontosságú azoknál a teszteknel, amelyek a későbbi tanulás feltételeit mérik, és a későbbi problémák korai indikátorainak felismerésére használhatók. Az itt tárgyalt online tesztek vizsgálata a jelen tanulmány továbblépéseként már meg is kezdődött. Az ismételt tesztelés hatásainak feltárása szintén megkezdődött, de az elemzésekhez szükséges adatmennyiség összegyűjtése mindkét esetben hosszabb időt igényel. Mivel a tanulók felméréseinek adatait az eDia online felmérési rendszerben tároljuk, nemcsak a tesztelési adatok, hanem a kérdőívekkel gyűjtött infor-

mációk is egy helyre gyűjthetők, és a növekvő információs adatbázis lehetővé teszi azoknak a kérdéseknek a későbbi vizsgálatát, amelyekre most nem találtunk válaszokat.

A DIFER kidolgozása során meghatározó szempont volt, hogy azt az óvoda-iskola átmenet során többször fel lehessen venni, azzal a tanuló fejlődését nyomon lehessen követni. Az online diagnosztikus mérési rendszer iskolakezdéstől átveszi a fejlődés követésének feladatait, ezért egy új iskolakészültségi tesztrendszernek az iskolát megelőző években és az iskolakezdés időpontjában alkalmazandó, az induló szintet feltérképező, az alapvető szűrési feladatokat ellátó készségmérésre kell koncentrálnia.

Az itt bemutatott vizsgálatban arra törekedtünk, hogy egy hosszú fejlesztési múlttal rendelkező, a gyakorlatban is jól bevált mérőeszközt a lehető legpontosabban átültessünk az új technológiai alapra. Ez a törekvés ugyanakkor azzal is járt, hogy nem használtuk ki a technológia által kínált összes lehetőséget. Ezek az elemzések nem csupán bizonyították a technológia használhatóságát egy ilyen különleges mérési helyzetben, hanem megmutatták annak sokféle előnyét is, amit egy új generációs eszközrendszer kifejlesztése során figyelembe lehet venni.

Irodalom

- Carson, K., Gillon, G. és Boustead, T. (2011): Computer-administrated versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language and Hearing*, **4**. 2. sz. 85–101.
- Choi, S. W. és Tinkler, T. (2002): Evaluating comparability of paper and computer based assessment in a K-12 setting. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.
- Clariana R., és Wallace P. (2002): Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, **33**. 5. sz. 593–602.
- Csapó Benő (2007): Hosszsmetszeti felmérések iskolai kontextusban – az első átfogó magyar iskolai longitudinális kutatási program elméleti és módszertani keretei. *Magyar Pedagógia*, **107**. 4. sz. 321–355.
- Csapó Benő (2014): A szegedi iskolai longitudinális program. In: Pál József és Vajda Zoltán (szerk.): *Szegedi Egyetemi Tudástár 7. Bölcsész- és társadalomtudományok*. Szegedi Egyetemi Kiadó, Szeged. 117–166.

- Csapó, B., Ainley, J., Bennett, R., Latour, T. és Law, N. (2012): Technological issues of computer-based assessment of 21st century skills. In: McGaw, B., Griffin, P. és Care, E. (szerk.): *Assessment and teaching of 21st century skills*. Springer, New York. 143–230.
- Csapó, B., Molnár, Gy. és Nagy, J. (2014): Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, **106**. 2. sz. 639–650.
- Csapó, B., Molnár, Gy. és R. Tóth, K. (2009): Comparing paper-and-pencil and online assessment of reasoning skills: A pilot study for introducing TAO in large-scale assessment in Hungary. In: Scheuermann F. és Björnsson, J. (szerk.): *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 113–118.
- Gallagher, A., Bridgeman, B. és Cahalan, C. (2000): *The effect of computer-based tests on racial/ethnic, gender, and language groups (GRE Board Professional Report No. 96–21P)*. Education Testing Service, Princeton.
- Guhn, M., Janus M. és Hertzman, C. (2007): The Early Development Instrument: Translating school readiness assessment into community actions and policy planning. *Early Education & Development*, **18**. 3. sz. 369–374.
- Józsa, K. (2004): Az első osztályos tanulók elemi alapkészségeinek fejlettsége. Egy longitudinális kutatás első mérési pontja. *Iskolakultúra*, **14**. 11. sz. 3–16.
- Kingston N. M. (2009): Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. *Applied Measurement in Education*, **22**. 1. sz. 22–37.
- Konold, T. R. és Pianta, R. C. (2005): Empirically-derived, person-oriented patterns of school readiness in typically-developing children: Description and prediction to first-grade achievement. *Applied Developmental Science*, **9**. 4. sz. 174–187.
- Nagy József (1980): 5-6 éves gyermekeink iskolakészültsége. Akadémiai Kiadó, Budapest.
- Nagy József (1987): Prefer: Preventív fejlettségvizsgáló rendszer 4–7 éves gyermekek számára. Akadémiai Kiadó, Budapest.
- Nagy József, Józsa Krisztián, Vidákovich Tibor és Fazekasné Fenyvesi Margit (2004a): *Diagnosztikus fejlődésvizsgáló és kritériumorientált fejlesztő rendszer 4–8 évesek számára: DIFER programcsomag*. Mozaik Kiadó, Szeged.
- Nagy József, Józsa Krisztián, Vidákovich Tibor és Fazekasné Fenyvesi Margit (2004b): *Az elemi alapképességek fejlődése 4-8 éves életkorban. Az eredményes iskolakezdés hét kritikus alapkészségének országos helyzetképe és a pedagógiai tanulságok*. Mozaik Kiadó, Szeged.
- Snow, K. L. (2006): Measuring school readiness: Conceptual and practical considerations. *Early Education and Development*, **17**. 1. sz. 7–41.

- Snow, C. E., és Van Hemel, S. B. (szerk., 2008): *Early childhood assessment: Why, what, and how*. The National Academies Press, Washington, DC.
- Mashburn, A. J. és Henry, G. T. (2004): Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice*, **23**. 4. sz. 16–30.
- OECD (2010): *PISA Computer-based assessment of student skills in science*. OECD, Paris.
- OECD (2011): *PISA 2009 results: Students on line digital technologies and performance (Volume VI)*. OECD, Paris.
- OECD (2014): *PISA 2012 results: Creative Problem Solving. Students' skills in tackling real-life problems. (Volume V)*. OECD, Paris.
- Price, P., Tepperman, J., Iseli, M., Duong, T., Black, M., Wang, S., Boscardin, C. K., Heritage, M., Pearson, P. D., Narayanan, S. és Alwan, A. (2009): Assessment of emerging reading skills in young native speakers and language learners. *Speech Communication*, **51**. 968–984.
- Wang, S., Jiao, H., Young, M. J., Brooks, T. E. és Olson, J. (2008): Comparability of computer-based and paper-and-pencil testing in K-12 assessment: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, **68**. 5–24.