

SZEGEDI TUDOMÁNYEGYETEM
BÖLCSÉSZETTUDOMÁNYI KAR
NEVELÉSTUDOMÁNYI DOKTORI ISKOLA
INFORMÁCIÓS ÉS KOMMUNIKÁCIÓS TECHNOLÓGIÁK AZ OKTATÁSBAN DOKTORI PROGRAM

MAGYAR ANDREA

SZÁMÍTÓGÉP ALAPÚ ADAPTÍV ÉS LINEÁRIS TESZTEK ÖSSZEHOSONLÍTÓ
HATÉKONYSÁGVIZSGÁLATA

PHD ÉRTEKEZÉS

TÉMAVEZETŐ:

MOLNÁR GYÖNGYVÉR PHD, HABIL EGYETEMI DOCENS



SZEGED

2015

TARTALOM

BEVEZETÉS	4
I. SZÁMÍTÓGÉPES ADAPTÍV TESZTELÉS	9
I.1 TÖRTÉNETI ELŐZMÉNYEK	9
I.2 JELENLEGI ALKALMAZÁSOK	10
I.3 AZ ADAPTÍV TESZTELÉS ALGORITMUSA	11
I.4 AZ ADAPTÍV TESZTELÉS ELŐNYEI ÉS HÁTRÁNYAI	13
I.5 A VALÓSZÍNŰSÉGI TESZTELMÉLET SZEREPE AZ ADAPTÍV TESZTELÉSBEN	14
I.6 AZ ADAPTÍV TESZTKÉSZÍTÉS FOLYAMATA	22
<i>I.6.1 Megvalósíthatóság, alkalmazhatóság lehetőségeinek felmérése</i>	23
<i>I.6.2 Feladatbank létrehozása</i>	24
<i>I.6.3 Valószínűségi tesztelméleti modell kiválasztása</i>	26
<i>I.6.4 Itemek kalibrálása, skálázása</i>	27
<i>I.6.5 Kezdő item(ek) kiválasztása</i>	27
<i>I.6.6 Itemkiválasztási algoritmus meghatározása</i>	28
<i>I.6.7 Képességbecslés</i>	31
<i>I.6.8 Végződtetési kritérium</i>	32
<i>I.6.9 A teszt kiközvetítése</i>	34
I.7 A TÖBBSZAKASZOS ADAPTÍV TESZTEK	34
<i>I.7.1 A többszakaszos adaptív tesztek felépítése</i>	35
<i>I.7.2 Elágazási szabály és pontozási lehetőségek</i>	36
<i>I.7.3 A többszakaszos tesztek előnyei és hátrányai</i>	39
I.8 SZÁMÍTÓGÉPES ADAPTÍV TESZTEK ÉS LINEÁRIS TESZTEK MŰKÖDÉSÉRE VONATKOZÓ NEMZETKÖZI ÖSSZEHASONLÍTÓ KUTATÁSOK	39
II. AZ EMPIRIKUS VIZSGÁLATOK KONCEPCIÓJA	48
II.1 CÉLOK	48
II.2 KUTATÁSI KÉRDÉSEK	48
II.3 HIPOTÉZISEK	49
II.4 A VIZSGÁLATSOROZAT MENETE	49
II.5 A VIZSGÁLAT SORÁN HASZNÁLT MÉRŐESZKÖZÖK BEMUTATÁSA	50
<i>II.5.1 Induktív gondolkodást mérő teszt</i>	51
<i>II.5.2 Problémamegoldó képességet mérő teszt</i>	54
<i>II.5.3 Szóolvasó készséget mérő tesztrendszer</i>	57
II.6 A MINTÁK ÖSSZEÁLLÍTÁSÁNAK SZEMPONTJAI	59
II.7 ADATFELVÉTEL	60

II.8 AZ ADATOK ELEMZÉSE	60
III. SZÁMÍTÓGÉPES ADAPTÍV TESZTEK ÉS LINEÁRIS TESZTEK MŰKÖDÉSÉNEK ÖSSZEHASONLÍTÓ HATÉKONYSÁGVIZSGÁLATÁRA IRÁNYULÓ PILOTMÉRÉSEK.....	62
III.1 INDUKTÍV GONDOLKODÁS MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL	62
<i>III.1.1 Minta</i>	62
<i>III.1.2 Mérőeszköz</i>	62
<i>III.1.3 Adatfelvétel és eljárások</i>	65
<i>III.1.4 Eredmények és diszkusszió</i>	65
III.1.4.1 A tesztek reliabilitása	65
III.1.4.2 A becsült képességszintek összehasonlítása évfolyamonként és személyenként	67
III.1.4.3 A helyes válaszok aránya a kétféle tesztkörnyezetben	68
III.1.4.4 A lineáris és az adaptív tesztelés során kiosztott résztesztek nehézségi szintjének változásmintázata	69
III.1.4.5 A lineáris és az adaptív tesztelés során kinyert mérési hibák és kinyert információk nagyságának összehasonlítása	71
<i>III.1.5 A pilotmérés eredményeinek összefoglalása</i>	73
III.2 PROBLÉMAMEGOLDÓ KÉPESSÉG MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL.....	74
<i>III.2.1 Minta</i>	75
<i>III.2.2 Mérőeszköz</i>	75
<i>III.2.3 Adatfelvétel és eljárások</i>	77
<i>III.2.4 Eredmények és diszkusszió</i>	78
III.2.4.1 A tesztek reliabilitása	78
III.2.4.2 A becsült képességszintek évfolyamonkénti és személyszintű összehasonlítása.....	79
III.2.4.3 A helyes válaszok aránya adaptív és lineáris tesztkörnyezetben.....	81
III.2.4.4 Az adaptív tesztelés során kiosztásra került résztesztek jellemzése	82
III.2.4.5 A tesztelés során kinyert információk és mérési hibák nagyságának összehasonlítása	83
<i>III.2.5 A pilotmérés eredményeinek összefoglalása</i>	84
IV. SZÓOLVASÓ KÉSZSÉG MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL.....	87
IV.2 A SZÓOLVASÓ KÉSZSÉG KRITÉRIUMORIENTÁLT MÉRÉSÉT LEHETŐVÉ TEVŐ TESZTSOROZAT TOVÁBBFEJLESZTÉSÉNEK LEHETŐSÉGEI	87
IV.3 ADAPTÍV TESZTELÉS ALKALMAZÁSA A SZÓOLVASÓ KÉSZSÉG MÉRÉSÉRE	87
IV.4 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ FELADATBANK KIALAKÍTÁSA ÉS PARAMÉTEREZÉSE	88
<i>IV.4.1 A papíralapú feladatok digitalizálása</i>	88
<i>IV.4.2 A feladatok paraméterezéséhez szükséges lineáris tesztrendszer kialakítása</i>	90
<i>IV.4.3 A feladatok paraméterezését szolgáló kutatás mintája</i>	92
<i>IV.4.4 A feladatok paraméterezését szolgáló kutatás eljárásai és eredménye</i>	93

IV.5 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ TESZTRENSZER KIALAKÍTÁSA	95
IV.6 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ TESZTRENSZER KIPRÓBÁLÁSA	99
IV.6.1 <i>Minta, adatfelvétel</i>	99
IV.6.2 <i>Eredmények és diszkusszió</i>	100
IV.6. A PILOTMÉRÉS EREDMÉNYEINEK ÖSSZEFOGLALÁSA.....	104
V. SZÁMÍTÓGÉP ALAPÚ ADAPTÍV ÉS LINEÁRIS TESZTELÉS ÖSSZEHASONLÍTÓ HATÉKONYSÁGVIZSGÁLATÁRA IRÁNYULÓ NAGYMINTÁS MÉRÉS	105
V.1 A MÉRÉS MÓDSZEREI	105
V.1.1 <i>Minta</i>	105
V.1.2 <i>Mérőeszköz</i>	105
V.1.3 <i>Adatfelvétel és eljárások</i>	105
V.2 AZ EMPIRIKUS VIZSGÁLAT EREDMÉNYEI ÉS DISZKUSSZIÓ	106
V.2.1 <i>A tesztek reliabilitása</i>	106
V.2.2 <i>A becsült készségszintek összehasonlítása évfolyamonként és személyenként</i>	111
V.2.3 <i>A helyes válaszok aránya adaptív és lineáris tesztkörnyezetben</i>	112
V.2.4 <i>A lineáris és az adaptív tesztelés során kiosztott résztesztek nehézségi szintjének változásmintázata</i>	113
V.2.5 <i>A lineáris és az adaptív tesztelés során kinyert információ és a mérési hiba nagyságának összehasonlítása</i>	116
V.3 A SZÓOLVASÓ KÉSZSÉG VIZSGÁLATÁRA IRÁNYULÓ NAGYMINTÁS MÉRÉS EREDMÉNYEINEK ÖSSZEFOGLALÁSA	119
ÖSSZEGZÉSEK, KÖVETKEZTETÉSEK.....	121
HIPOTÉZISEK BEIGAZOLÓDÁSA	123
EREDMÉNYEK ÖSSZEFOGLALÁSA, ÉRTELMEZÉSE	126
A KUTATÁS EREDETISÉGE ÉS KORLÁTAI	127
TOVÁBBI KUTATÁSI IRÁNYVONALAK	127
KÖSZÖNETNYILVÁNÍTÁS	129
IRODALOM.....	130
ÁBRAJEGYZÉK.....	144
TÁBLÁZATOK JEGYZÉKE	146
MELLÉKLETEK	147
1. SZÁMÚ MELLÉKLET: AZ INDUKTÍV GONDOLKODÁST MÉRŐ TESZT FELADATAI	147
2. SZÁMÚ MELLÉKLET: A PROBLÉMAMEGOLDÓ GONDOLKODÁST MÉRŐ TESZT FELADATAI	152
3. SZÁMÚ MELLÉKLET: A SZÓOLVASÓ KÉSZSÉGET MÉRŐ TESZT FELADATAI	156

BEVEZETÉS

A 20. században leginkább elfogadott és elterjedt papíralapú (PP) tesztekkel történő mérések fejlesztése egyre több korlátba ütközött, a papíralapú tesztekre alapozott fejlesztés lehetőségei fokozatosan kimerültek (*Molnár és Magyar, 2015*). A továbblépéshez, a 21. században jelentkező új mérés-értékelési igények kielégítéséhez alapvető, minőségi változtatásra van szükség (*Beller, 2013; Scheuermann és Pereira, 2008*). A technológia rohamos fejlődésével a továbblépés iránya egyértelműen a számítógépes tesztelésre való átállás (*Scheuermann és Björnsson, 2009; Molnár, 2010; Csapó, Ainley, Bennett, Latour és Law, 2012*), mely számos új lehetőséget kínál a papíralapú teszteléssel szemben. Ilyen például a motiválóbbr környezet (*Thompson és Pometric, 2007*), az azonnali kiértékelés lehetősége (*Wang, 2010*), az innovatív, multimédiás elemeket is tartalmazó, dinamikusan változó itemek megjelenítése (*Greiff, Wüstenberg és Funke, 2012*), illetve a személyre szabott, adaptív tesztelés megvalósítása (*Eggen és Straetmans, 2000*).

Attól függően, hogy a technológia adta lehetőségek mekkora mértékben kerülnek kiaknázásra, a papíralapú tesztek számítógépre való konvertálása többféleképpen megtörténhet (*Csapó, Molnár és R. Tóth, 2008*). Az áttérés legelső szintjén a papíralapú tesztek változatlan formában történő digitalizálása áll. Ebben az esetben megmarad az eredeti lineáris tesztstruktúra, minden tesztelt személy ugyanazon sorrendben ugyanazon feladatokat kapja a tesztelés során, függetlenül képességszintjétől és teljesítményétől. Ezzel a technikával azonban csak egy viszonylag szűk képességtartomány mérhető a szükséges pontossággal. Ha a tesztet szélesebb képességtartomány mérésére tesszük alkalmassá, akkor minden adatfelvételben részt vevő személy számára csak a teszt néhány feladata jelent kihívást, melyek nehézségi szintjei közel állnak a tesztet megoldó személy képességszintjéhez. A teszt többi feladata, esetlegesen a teszt nagyobb része jóval kisebb mértékben járul hozzá a személy képesség- vagy tudásszintjének pontos meghatározásához. Azok az itemek, amelyek túl könnyűek vagy túl nehezek, azaz vagy nem jelentenek kihívást, vagy frusztrálóak a tesztelt személy számára, jelentős mértékben csökkenthetik a személy teszteléssel kapcsolatos érdeklődését, attitűdjét, motivációját (*Csapó, Molnár és R. Tóth, 2008*).

A számítógépes tesztelés ma ismert leginnovatívabb formája, az adaptív tesztelés (CAT – *Computerized Adaptive Testing; Weiss, 2011*) ezt a problémát küszöböli ki. Az adaptív technika alkalmazása során a teszt feladatai nem előre meghatározott fix sorrendben követik

egymást, hanem azok egy feladatbankból kerülnek kiválasztásra a tesztmegoldó képességszintjéhez igazítva, a korábbi feladaton nyújtott teljesítménye alapján. Például feladatszintű adaptivitás esetén amennyiben a tanuló meg tudja oldani a teszt egy feladatát, következőleg egy nehezebbet kap, ha nem, akkor könnyebbet. Ezen algoritmus alkalmazásával a tesztelés végén minden tanulóhoz hozzárendelhető egy képességszint, melynél könnyebb feladatokat nagyobb valószínűség mellett old meg helyesen, mint helytelenül (Weiss, 2011).

Ez a típusú feladatadás és tesztösszeállítás a hagyományos, mindenki számára azonos itemeket azonos sorrendben tartalmazó, lineáris tesztekkel szemben a teljesítmények sokkal finomabb mérését teszi lehetővé (Linacre, 2000; Magyar és Molnár, 2013). Jelentős mértékben megnő a tesztelés során kinyerhető itemekre és személyekre vonatkozó információ nagysága (Molnár, 2013; Magyar és Molnár, 2013). Elhanyagolhatóvá válik annak valószínűsége, hogy a tesztelt személyek ugyanazon feladatokat ugyanabban a sorrendben kapják, azaz növekszik a tesztelés biztonsága (Wainer, 2000a). Mindez új lehetőségeket teremt a mérés-értékelés területén. Ha nem törekszünk több információ kinyerésére, azaz megelégszünk a hagyományos tesztelés során elérhető pontossággal, akkor a kiközvetített feladatok száma, vagyis a teszt hossza (Thompson és Way, 2007), ezzel párhuzamosan a teszt megoldásához szükséges idő is jelentős mértékben rövidül, utóbbi átlagosan felére csökken (Frey és Seitz, 2009; Frey, Seitz és Kröhne, 2011).

Miután jelenleg a számítógép alapú tesztelésre való átállás fázisában vagyunk, indokoltak a médihatást tanulmányozó összehasonlító hatékonyságvizsgálatok, melyek egyrészt longitudinális kutatásokban kapnak kiemelkedő szerepet, ahol szükséges a korábbi papíralapú adatfelvételek eredményeinek számítógép alapú teszteredményekkel való összehasonlíthatósága, másrészt abban az esetben, amikor a kétféle médiumon való tesztelés alternatív módon párhuzamosan folyik.

Adaptív tesztek bevezetése során számos esetben végezték el az adaptív és a lineáris változat összehasonlító hatékonyságvizsgálatát (Al-A'Ali, 2007; Brossman és Guille, 2014; Frey, Seitz és Kröhne, 2011; Guille, Becker, Zhu, Zhang, Song és Sun, 2011; Hambleton és Xing, 2006; Jodoin, Zenisky és Hambleton, 2006; Kingsbury és Hauser, 2004; Olea, Revuelta, Ximénez és Abad, 2000; Pyper és Lilley, 2010; Rotou, Patsula, Manfred és Rizavi, 2003; Thompson és Way, 2007; Vispoel, Hendrickson és Bleiler, 2000; Zheng, 2012), az eddigi kutatások azonban főként szimulált adatbázisokkal dolgoztak. Empirikus kutatások elsősorban az egyetemista korosztály körében folytak, melyek a legtöbb esetben kis mintán végzett pilotvizsgálatok voltak.

Ezt a hiányt pótolja a disszertációban ismertetett kutatássorozat, melynek fő célja az adaptív tesztelés hatékonyságának vizsgálata az 1-8. évfolyamos tanulók körében. A kutatás elsősorban az adaptív tesztelés technikai jellemzőire fókuszál, a számítógépes adaptív tesztek kidolgozásának menetét, alkalmazását és működését veszi górcső alá, továbbá azt vizsgálja, hogy az általános iskolai korosztály körében osztálytermi környezetben milyen előnyöket jelenthet alkalmazásuk.

A disszertáció öt fő részre tagolódik. Az első részében a számítógépes adaptív tesztelés szakirodalmi háttérét összegezzük és az adaptív tesztszerkesztés megvalósításának feltételeit, lépéseit vázoljuk fel. Az adaptív tesztelés fogalma nem új, már a századforduló első éveiben felvetődött, ezért rövid történeti leírást adunk kialakulásától kezdve a napjainkban főként nagymintás méréseken történő alkalmazásáig. Ezt követően mutatjuk be a teszt működését, mely szigorú algoritmushoz kötött, ami a mérs során biztosítja, hogy minden kiközvetített item a vizsgált személy képességszintjéhez igazodjék. A képességszintek becslése a valószínűségi tesztelmélet (IRT – *Item Response Theory*) segítségével történik. Mivel a valószínűségi tesztelmélet alkalmazása a tesztelés folyamán több ponton is nélkülözhetetlen, ezért részletesen kitérünk a különböző logisztikus modellek ismertetésére, az item/teszt karakterisztikus görbék és item/teszt-információs függvények adaptív tesztelésben való alkalmazására. Ezt követően mutatjuk be az adaptív tesztkészítés folyamatát a kalibrált feladatbank létrehozásától a tesztek kiközvetítéséig. Az adaptív teszteknek számos típusa létezik. Mérésmethodikai jellemzői alapján a leggyakrabban alkalmazott típus a többszakaszos adaptív teszt, mely során nem diszjunkt itemeket, hanem rövid fix tesztek (modulokat) alkalmaznak. Mivel a kutatás során is ezt a tesztípust használtuk, részletes ismertetést adunk ennek a tesztípusnak a felépítéséről, az elágazási szabály kialakításának lehetőségeiről, a pontozási módszerekről, valamint alkalmazásának előnyeiről és hátrányairól. A számítógépes adaptív és hagyományos lineáris tesztek működésével foglalkozó összehasonlító kutatások ismertetésével zárjuk az első fejezetet.

A második fejezetben ismertetjük a vizsgálat céljait, kutatási kérdéseit, a hipotéziseket és a vizsgálatssorozat menetét. A hatékonyságvizsgálatokat pilotméréseken és nagymintás méréseken keresztül valósítottuk meg. A kutatásban három pilot és két nagymintás mérést folytatunk le, melyek három különböző képességterület mérését valósították meg. Ebben a fejezetben ismertetjük részletesen az adaptív tesztekhez felhasznált korábban papíralapon

használt készség- és képességmérő tesztek, a minták összeállításának szempontjait, az adatgyűjtés körülményeit, valamint az adatok elemzése során használt eljárásokat.

A harmadik részben a számítógépes adaptív tesztek és lineáris tesztek működésének összehasonlító hatékonyságvizsgálatára irányuló két pilotmérés eredményeit ismertetjük. Az első mérés előzőleg papír alapon végzett induktív gondolkodás mérésére irányuló teszt online adaptív és lineáris formára való konvertálásának lépéseit mutatja be, majd az 5-8. évfolyamon végzett pilotmérés eredményeit összegzi. A második mérés során a problémamegoldó képesség mérése történt meg ugyanezen korosztály körében, ugyancsak hagyományos lineáris és többszakaszos adaptív tesztek alkalmazásával.

A pilotmérések eredményei alapján nagymintás mérés során volt célunk igazolni az adaptív tesztelés megvalósíthatóságát a 4-5. évfolyamos tanulók körében, melyet a szóolvasó készség mérésére *Nagy József* által kifejlesztett papíralapú rendszer adaptív rendszerré való továbbfejlesztése során végeztünk el.

A negyedik fejezetben mutatjuk be a szóolvasási készség mérésére kifejlesztett feladatbankot és az adaptív rendszert, valamint a rendszer kismintás tesztelésével kapcsolatos eredmények elemzését. A médiahatás kiküszöbölése és a minél pontosabb mérés megvalósítása érdekében a nagymintás mérés során nem a papíralapú tesztelés során bemért itemparamétereket alkalmaztuk, hanem külön nagymintás mérés alkalmával elvégeztük a teljes itembank számítógép alapon való paraméterezését. A rendszer online formára való konvertálásakor az eredeti rendszerstruktúra megtartására törekedtünk, miszerint a szóolvasó készség mérése négy különböző szempont szerint valósult meg (címszóolvasás, toldalékos szóolvasás, szinonimaolvasás és szójeletés olvasás). A feladatok és a tanulók közös skálán történő elhelyezése és a pozíciós hatás kiküszöbölése végett a papíralapú rendszer résztesztjeiből klasztereket képeztünk, melyek horgonyként funkcionáltak a paraméterezés során. Az egyparaméteres Rasch-moddal bemért itemekből állítottuk össze a négyszakaszos adaptív tesztrendszert, mely öt különböző szintre tudja a tanulókat besorolni elért eredményük alapján. Az adaptív rendszerrel végzett kismintás mérés eredményeit is összegzi a fejezet.

Az ötödik fejezetben térünk rá a nagymintás mérés során lefolytatott összehasonlító hatékonyságvizsgálatra, mely az adaptív tesztstruktúra előnyeit vizsgálja a hagyományos lineáris teszteléssel szemben.

A disszertáció végén összegezzük és értelmezzük az elért eredményeket, kiemeljük a kutatás eredetiségét, rámutatunk a korlátaira és kijelöljük a további kutatási irányvonalakat.

I. SZÁMÍTÓGÉPES ADAPTÍV TESZTELÉS

I.1 TÖRTÉNETI ELŐZMÉNYEK

Az adaptív tesztelés eredete 1905-re vezethető vissza és *Alfred Binet* IQ tesztjeinek kidolgozásához kötődik (*van der Linden és Glas, 2000*). *Binet* elsősorban egyéni diagnózissal foglalkozott, és kidolgozott egy stratégiát, mellyel személyre szabottá tehetette a vizsgát; a gyerekek életkora alapján sorolta be az itemeket. A vizsgálatot a gyermek korának megfelelő kérdéssel kezdte és a tesztelés során a tanuló képességeihez mérten addig kapott egyre nehezebb kérdéseket, amíg képességszintje bemérhetővé vált (*Linacre, 2000*). Ez a módszer azonban csak egyéni tesztelést tett lehetővé. A tesztelés nagymintán való megvalósítása először a hadsereg keretein belül történt (*Office of Naval Research; Navy Personnel Research and Development Center; Air Force Human Resources Laboratory; Army Research Institute; Wainer, 2000b*), ahol hamar felismerték az adaptivitás előnyeit, és széleskörű elméleti kutatásokat indítottak, mely során számos, adaptív tesztelést megvalósító pszichometriai programot fejlesztettek ki (*ASVAB*). Az adaptív tesztelés igazi széles körű alkalmazása azonban a számítógépek elterjedésével vált lehetővé.

Frederic Lord (1971a; 1971b) a 1960-as évek végén számítógépen valósította meg az item nehézségszintjének a tanuló képességszintjével való összehangolását (*Way, Davis és Fitzpatrick, 2006*). A szóbeli vizsgáit egy közepes nehézségi szintű itemmel kezdte. Ha a vizsgált személy helyesen válaszolt, kissé nehezebb kérdést kapott, ha viszont helytelen választ adott, akkor kissé könnyebb kérdéssel folytathatta. A vizsga addig folytatódott, míg a szintje beméréséhez elegendő számú itemre tudott válaszolni (*Weiss, 2004*).

1973-ban *Weiss* dolgozott ki egy *Binet* logikáján alapuló számítógépes tesztet, melyet 'stratified adaptive', vagy 'stradaptív' tesztnek nevezett. A teszt *Binet*-jéhez hasonló itembankot használt, amely az itemeket szintenként csoportosítva tartalmazta, a különbség az itemek közvetítésében és a teszt befejezésében volt (*Weiss, 1973*).

Wainer és Kiely (1987) fejlesztették tovább az adaptív tesztelés módszertanát; itemek helyett ún. *tesztleteket* használtak. A *tesztlet* rövid, néhány itemből álló kérdések csoportja, mely rövid szöveghez, táblázathoz vagy grafikonhoz kötődik. Ezek voltak az első nem itemalapú adaptív tesztek, használatuk később igen népszerűvé vált.

Az első adaptív teszteknel a klasszikus tesztelmélet módszerét alkalmazták az itemek kalibrációjára, a valószínűségi tesztelmélet (IRT – *Item Response Theory*) elterjedése óta viszont e modellek segítségével kalibrálják be az itemek nehézségi szintjét, mely jóval pontosabb mérést tesz lehetővé (Eggen, 2004; Al-A'ali, 2007). Innen számítjuk az adaptív tesztelés újabb generációját (Weiss, 1974, 2011).

I.2 JELENLEGI ALKALMAZÁSOK

Nemzetközi viszonylatban egyre elterjedtebb az adaptív tesztelés alkalmazása (van der Linden, 2001) A világ számos egyetemén foglalkoznak kutatócsoportok adaptív tesztekre irányuló kutatásokkal. Amerikában a minnesotai és a kansasi egyetemeken folynak jelentős kutatások. Európát tekintve Hollandiában a CITO és a Twente kutatócsoportja, Németországban az Aachen Egyetem RWTH kutatócsoportja, Angliában az NFER kutatócsoportja vizsgálja az adaptív tesztek megvalósítási lehetőségeit.

Az adaptív tesztekkel foglalkozó kutatók nemzetközi szervezete az IACAT (*International Association for Computerized Adaptive Testing*) 2010-ben alakult, és célja az adaptív tesztelés pszichológia és oktatás terén történő fejlesztése és terjesztése. Adaptív tesztek használata terén jelenleg az Amerikai Egyesült Államok a meghatározó (Weiss, 2004). Az egyik legismertebb nagymintán alkalmazott teszt a *Graduate Management Admission Test* (GMAT) az üzleti iskolák számára került kifejlesztésre. Kifejlesztője és működtetője az amerikai *Graduate Management Admission Council*, de ma már a világ több mint 150 országában elérhető (Frey és Seitz, 2009; Wainer, 2000a). Szintén amerikai a COMPASS tesztkészlet, melyet az *American College Test* (ACT) program során fejlesztettek ki, és a főiskolákra jelentkezők matematika és olvasáskészségét méri fel.

További amerikai adaptív felmérések a „*Measures for Academic Progress*” (MAP), a „*Virginia Standards of Learning Tests*” (SOL), valamint az „*Oregon Assessments of Knowledge and Skills*” (OAKS) (Molnár, 2010; Kingsbury és Wise, 2011). Több nagymintás amerikai tesztnek létezik adaptív verziója is: *Graduate Record Examination* (GRE), *Computerized Placement Test* (Meijer és Nering, 1999).

Európában kisebbek az online tesztelés hagyományai, de az utóbbi években itt is egyre több felmérésben alkalmaznak adaptív tesztek (Molnár, 2010). A „*Program for International Assessment of Adult Competence*” (PIAAC) felmérése a felnőtt lakosság szövegértési és

matematika képességeit méri fel. A teszt 25 perces adaptív blokkokat tartalmaz, a számítógép a megoldás alapján adagolja a következő nehézségű blokkot. A TOEFL nyelvvizsga is adaptív rendszerű: Összesen 4-4,5 órás számítógépes adaptív teszttel mérik a tanulók nyelvtani ismeretét, szókincsét, hallás utáni szövegértését és íráskészségét (Molnár, 2010). Az ACE (*Adjustable Competence Evaluation*) teszt különböző, munkavállaláshoz szükséges logikai képességeket mér, Dániában, Angiában, Finnországban, Németországban, Norvégiában és Svédországban érhető el. A holland CITO számos, az oktatás területén alkalmazandó adaptív tesztet dolgozott ki. A tesztek között szerepelnek matematikai, számolási, szókincsfelmérési tesztek, 4 éves kortól egészen felnőttkorig.

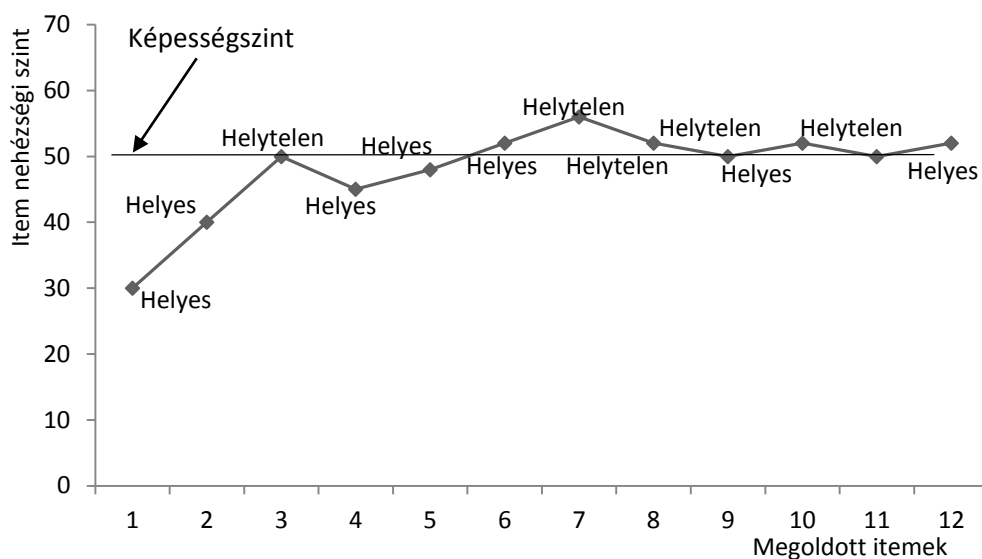
Az OECD PISA felmérés a fejlett országok legjelentősebb, nagymintás, rendszerszintű, felmérése. A 2000 óra háromévente megrendezendő adatfelvétel három fő terület (szövegértés, matematika, természettudomány), illetve egy negyedik, kiegészítő terület mérését valósítja meg. A három fő terület közül egy, ciklikusan, nagyobb hangsúllyal szerepel az adatfelvétel során. A 2015-ös mérés fő fókuszja a természettudomány lesz, mely számítógép alapú mérése adaptív blokkokat is fog tartalmazni (Frei és Seitz, 2009).

Hazánkban is léteznek adaptív modulokat alkalmazó tesztek. Ilyen a Vienna Test Sytem magyarul is elérhető értékelő eszköze, mely a pszichológiai vizsgálatok széles területén alkalmazható, úgymint a klinikai, sport-, repülés-, személyiség-, neuro-, és közlekedés-pszichológia. Az oktatás területén nagymintás vizsgálatokban is alkalmazható adaptív platform, az eDia bevezetése jelenleg a Szegedi Tudományegyetem Oktatáseleméleti Kutatócsoportja (SZTE OK) által folyamatban van (Molnár és Csapó, 2013; R. Tóth, Molnár, Latour és Csapó, 2011; Magyar, 2013).

I.3 AZ ADAPTÍV TESZTELÉS ALGORITMUSA

Az adaptív tesztek szigorú algoritmus szerint működnek (Linacre, 2000; Magyar, 2012). A kezdő item/részteszt kiválasztása többféleképpen történhet: amennyiben rendelkezésre áll előzetes információ a tesztelt személyről, akkor egy becslési algoritmus felhasználásával már a teszt kezdő iteme személyre szabott lehet, ha nem, akkor az itembankból vagy annak részhalmazából véletlenszerű kiválasztással történhet a teszt első feladatának kiválasztása. Rendszerint egy közepes nehézségű itemmel indul a tesztelés. Az adott item megoldása után

egy újabb kiválasztása következik. Ha a személy jól válaszolt, egy nehezebbet kap, amennyiben hibázott, akkor könnyebbet (1. ábra).



1. ábra: Adaptív tesztelés algoritmus (Linacre, 2000 alapján)

A program algoritmus biztosítja, hogy minden soron következő item a személy képességeihez mért legyen. A megoldott itemeket a gép algoritmus értékeli, és dönt arról, hogy szükséges-e új item kiválasztása, vagy a tesztelés véget ért. A tesztelés végén a tanuló azonnali visszajelzést kap elért eredményéről (Csapó, Molnár és R. Tóth, 2008; Eggen, 2004). Ez alapján a CAT individualizált teszt. A tanulók különböző itemekkel kezdenek és folytathatják a tesztet, és különböző lehet a megoldott itemek száma is, így a CAT dinamikus, személyre szabott, a tanuló képességszintjéhez igazodó tesztelési mód (Weiss, 2011).

I.4 AZ ADAPTÍV TESZTELÉS ELŐNYEI ÉS HÁTRÁNYAI

Az adaptív tesztek számos fontos előnnyel rendelkeznek a lineáris tesztekhez képest. Linacre (2000), Wainer (2000a), Eggen (2007) és Tian, Miao és Zhu (2007) az adaptív tesztek következő előnyeit emelik ki:

- Biztonságosabbá válik a tesztelés. Mivel a tanulók különböző feladatokat kapnak, nincs lehetőség a kérdések előzetes „betanulására” (Eggen, 2007; Wainer, 2000a; Tian, Miao és Zhu, 2007).
- A tanulók saját szintjüknek megfelelő feladatokat kapnak. Ezáltal kiküszöbölhető, hogy a tanulók túl könnyű, illetve túl nehéz kérdéseket kapjanak, és ezáltal unalmassá váljon számukra a vizsga, valamint csökken a frusztráció is (Eggen, 2007; Wainer, 2000a).

- A tesztelési idő lerövidül, és így kevésbé fárasztó a tanulók számára, és az értékelés is gyorsabbá, egyszerűbbé válik (*Linacre, 2000; Tian, Miao és Zhu, 2007*).
- Az adaptív tesztelés individualizált jellege miatt a pontosabb képességmérés lehetőségét teremti meg (*Linacre, 2000*).

Azonban a CAT hátrányokkal és korlátokkal is rendelkezik, melyek közül a legfontosabbak (*Cisar és mtsai, 2010; Frey és Hartig, 2009; Linacre, 2000; Meijer és Nering, 1999; Wainer, 2000a; Wainer és Kiely, 1987*):

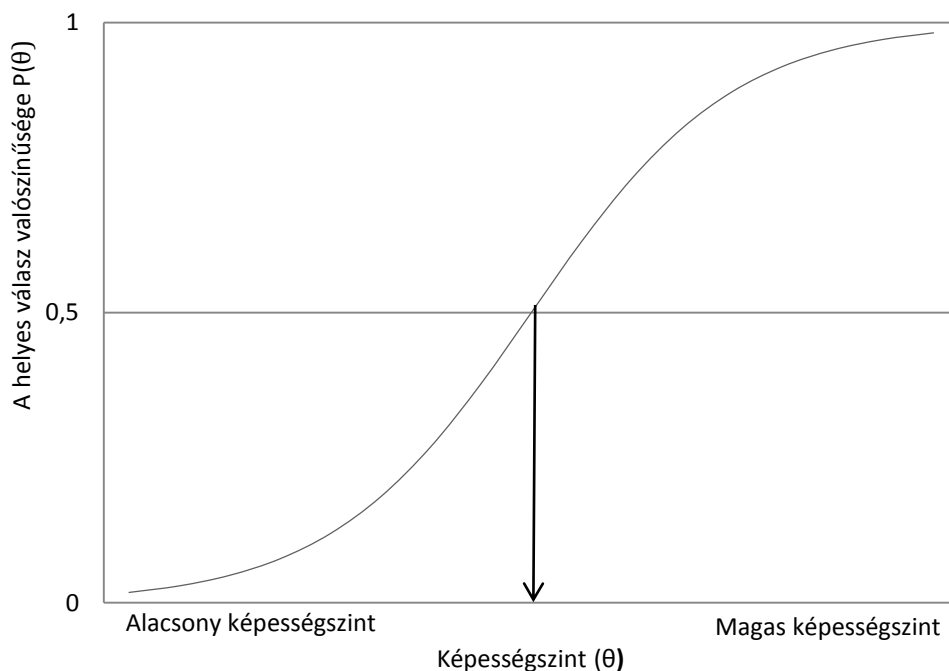
- Az adaptív tesztek előállítási költsége jóval magasabb, mint a lineáris teszteké (*Linacre, 2000; Meijer és Nering, 2000; Wainer és Kiely, 2000*).
- A legtöbb adaptív tesztben nincs lehetőség a feladatok visszamenőleges áttekintésére és javítására (*Linacre, 2000; Wainer, 2000a*).
- A lineáris teszteknel a tanulóknak lehetőségük van kérdéseket kihagyni, illetve a kérdésekre tetszőleges sorrendben válaszolni. A legtöbb adaptív teszt nem engedi ezt meg (*Cisar és mtsai, 2010; Wainer, 2000a*).
- Az itemek nagyon precíz kalibrálást igényelnek (*Cisar és mtsai, 2010; Linacre, 2000*), amennyiben az itemek nincsenek precízen paraméterezve, a tanulók képességbecslése pontatlanná válhat.
- Item pozíciós hatás: Az itemek elhelyezkedése befolyásolhatja a megoldást: ugyanaz az item a tesztben elfoglalt helyétől függően lehet könnyebben, illetve nehezebben megoldható (*Frey és Hartig, 2009; Wainer és Kiely, 1987*).
- Kereszt-információ: Mivel az itemek sorrendje nem előre meghatározott, az előző item véletlenül információt szolgáltat a következő item számára (*Wainer és Kiely, 1987*).

I.5 A VALÓSZÍNŰSÉGI TESZTELMÉLET SZEREPE AZ ADAPTÍV TESZTELÉSBEN

Az adaptív tesztek térhódítása nemcsak a technológia fejlődésének, hanem a méréselméletben végbement változásoknak is köszönhető (*Keng, 2008*). Egyszerűbb, kevés itemet tartalmazó és szűk képességtartományt vizsgáló adaptív tesztek a klasszikus tesztelmélet alkalmazásával is létrehozhatók, igazán megbízható, tágabb képességtartomány lefedését lehetővé tevő tesztek kialakításához azonban a valószínűségi tesztelmélet (IRT – *Item Response Theory*) alkalmazására van szükség.

A valószínűségi tesztelmélet két fontos tulajdonsága miatt alkalmas az adaptív tesztelésben való alkalmazáshoz: (1) Miután a tesztelés során a tanulók különböző itemeket oldanak meg, fontos, hogy eredményük összehasonlítható legyen az itembank bármely részhalmazát tekintve. (2) Az itemek nehézségi szintjei és a tanulók képességszintjei közös skálára hozhatók, ezáltal lehetővé válik a tanulók képességszintjéhez legközelebb álló itemek kiválasztása és alkalmazása. Ezek csak az IRT segítségével oldhatók meg (Eggen, 2007; Molnár, 2013; Magyar, 2013), ezért az adaptív tesztek működése ezért számos ponton a valószínűségi tesztelméletre épül.

Az IRT abból a feltevésből indul ki, hogy minden tanuló esetén meghatározható egy konkrét kérdésre adott helyes válasz valószínűsége, és ez a valószínűség nagyobb lesz a jobb képességszintű tanulóknál, mint a gyengébb képességűeknél (Gupta, Vijn és Vats, 2010; Molnár, 2003, 2006). Ez alapján a valószínűségi tesztelméletek matematikai kapcsolatot létesítenek a tesztelésben részt vevő tanuló képességszintje és az itemre adott helyes válasz valószínűsége között oly módon, hogy a tanuló meghatározott kérdésre adott helyes válaszána valószínűségét a tanuló képességszintje függvényében ábrázolják és az item karakterisztikáját jellemzik (Keng, Tusng-Han, Tzu-An és Dodd, 2000; Molnár, 2013).



2. ábra: Itemkarakterisztikus görbe (Molnár, 2013 alapján)

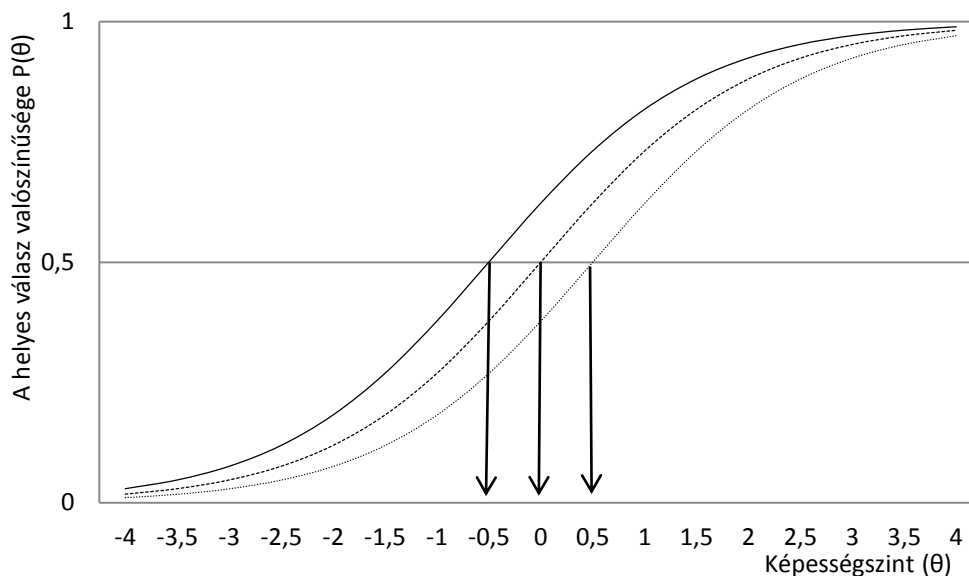
A valószínűségi tesztelméleti modellek alap építőkövei az itemkarakterisztikus görbék (2. ábra). Az itemkarakterisztikus görbe megmutatja, hogy a különböző képességszinten lévő

tanulók milyen valószínűséggel válaszolnak jól az itemre. Minél magasabb a tanuló képességszintje, annál nagyobb a helyes válasz valószínűsége. Az item nehézségét az a képességszint (θ) mutatja meg, amely az 50%-os valószínűségű helyes válaszhoz szükséges (Molnár, 2013). A teszt minden iteméhez tartozik egy karakterisztikus görbe, és minden képességszintű tanulóhoz hozzárendelhető egy valószínűségi szint, mely valószínűség szerint sikerrel oldaná meg az adott itemet, még akkor is, ha valójában nem került sor az item megoldására (Molnár, 2003; Rasch, 1960).

A 2. ábra egy átlagos nehézségű itemkarakterisztikus görbét ábrázolja. Az ábrázolt item 0 (átlagos) nehézségi szintű (a 0 nehézségi szint a közepes nehézségi szintet jelenti, ettől jobbra a magasabb képességszint, balra az alacsonyabb képességszint található). Erre az itemre a közepes képességszintű tanulók 50%-os valószínűséggel lennének képesek helyesen válaszolni. Hasonlóan, a feladatbankban szereplő mindegyik itemhez hozzárendelhető bizonyos nehézségi szint, melyet az alapján definiálásnak, hogy milyen képességszint szükséges ahhoz, hogy a helyes megoldás valószínűsége 50% legyen (Molnár, 2013). Így lehetővé válik a tanulók képességszint szerinti és az itemek közös skálán való ábrázolása, mely megkönnyíti a tanulók képességszintjéhez legközelebb álló itemek kiválasztását (Eggen, 2004).

A valószínűségi tesztelméleti modellek többféleképpen csoportosíthatóak. Az egyik legelterjedtebb besorolás az itemparaméterek száma szerinti osztályozás, mely szerint egy-, két-, és háromparaméteres modelleket különböztethetünk meg (Gupta, Vijn és Vats, 2010; Molnár, 2003, 2013).

Az egyparaméteres logisztikus modell (pl. Rasch-modell; 1960) a személyparaméter mellett egy paramétert tartalmaz, az itemnehézségi mutatót, azaz a modell feltételezi, hogy az itemek csak nehézségi szintjükben különböznek egymástól. Az itemeket grafikonon ábrázolva az itemek karakterisztikus görbéi párhuzamosan futnak. A 3. ábrán a folytonos vonallal rajzolt görbe a legkönnyebb, a szaggatott az átlagos, a pontozott vonallal ábrázolt a legnehezebb item helyes megoldásának valószínűségi görbét ábrázolja. A példában szereplő itemek nehézségi mutatói: -0,5, 0 és 0,5. A helyes válasz valószínűsége mindegyik képességszinten a legkönnyebb item (folytonos vonallal jelölt) esetben a legnagyobb.



3. ábra: Az egyparaméteres logisztikus modell itemkarakterisztikus görbéje

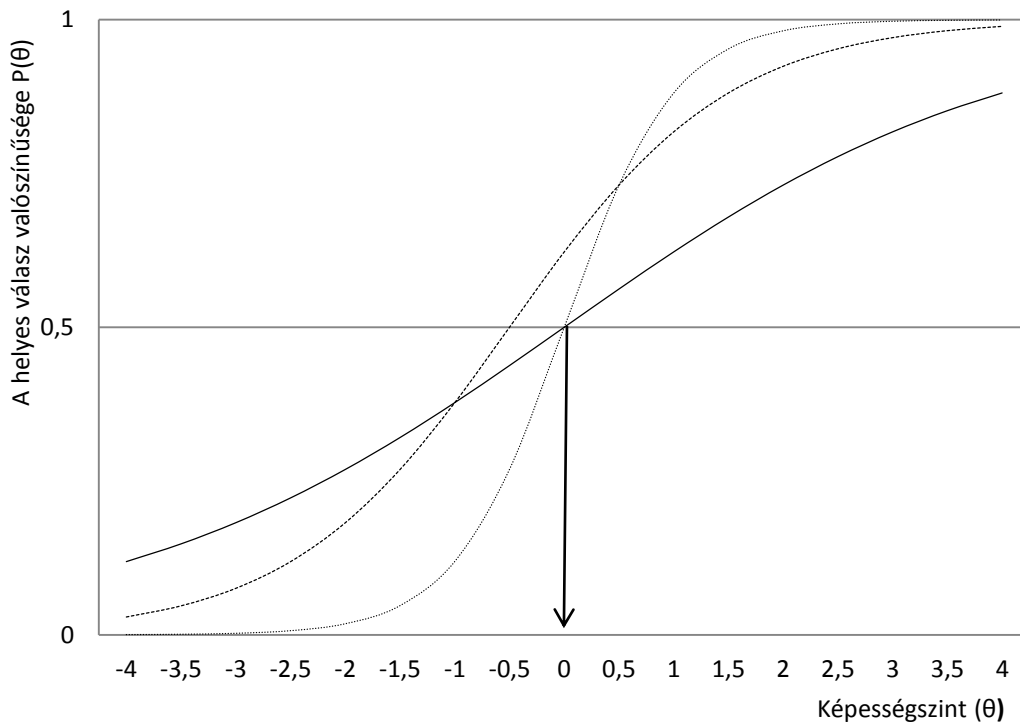
Matematikai formulával:

$$P_i(\theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}$$

ahol θ tanuló képességparamétere a vizsgált képességskálán,

b_i az item nehézségi mutató ugyanazon a skálán.

A kétparaméteres modellek az itemnehézségi mutató mellett még egy paramétert tartalmaznak, a diszkriminációs indexet. A 4. ábra három különböző megkülönböztető erővel bíró itemet ábrázol. A pontozott vonallal jelölt görbének a legnagyobb a diszkrimináló ereje, tehát ez az item képes leginkább különbséget tenni a tanulók között, a folytonos vonallal jelölt pedig legkevésbé. Minél meredekebb a görbe, annál nagyobb a megkülönböztető ereje. Egy teszt akkor jó, ha minél több magas diszkrimináló erővel rendelkező itemeket tartalmaz (Baker, 2001). A grafikonon ábrázolt görbék közül a pontozott és a folytonos vonallal jelölt itemek azonos nehézségi szintűek, de különböző diszkriminációs indexekkel rendelkeznek. Ennek következtében a helyes válasz valószínűsége a 0 képességszint alatti tanulóknál a folytonos görbével ábrázolt itemnél nagyobb, vagyis ez a könnyebb item számukra, míg a 0 képességszint feletti tanulók esetén a pontozott vonallal ábrázolt görbénél nagyobb a helyes válasz valószínűsége, tehát ez az item a könnyebb számukra (Yu, Jannasch-Pennel és DiGangi, 2008).



4. ábra: A kétparaméteres logisztikus modell itemkarakterisztikus görbéje

Kétparaméteres modellnél az alábbi matematikai formula használatos:

$$P_i(\theta) = \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]}$$

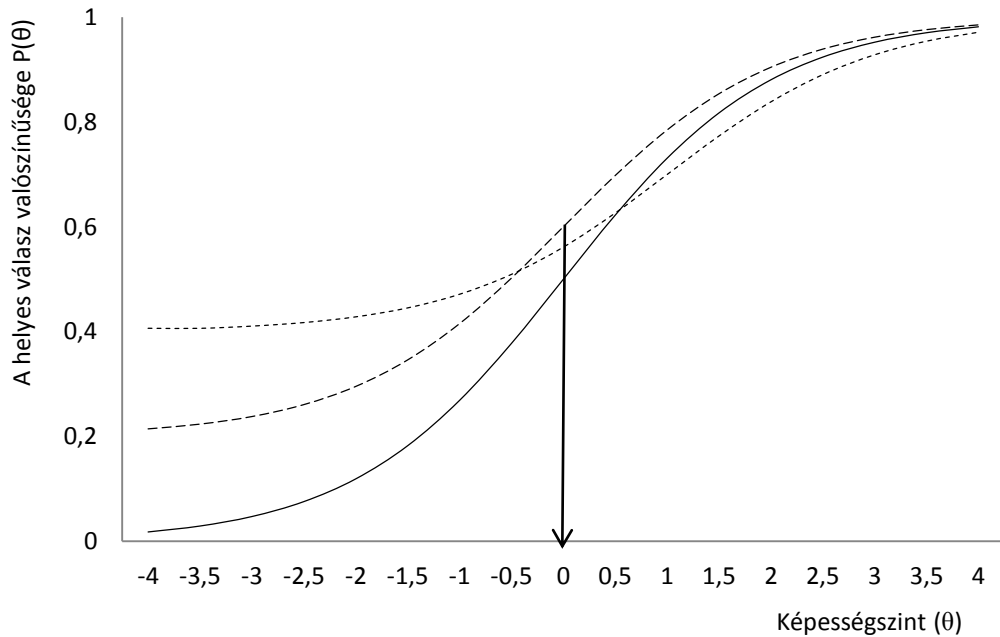
ahol θ tanuló képességparamétere a vizsgált képességskálán,

b_i az item nehézségi mutató ugyanazon a skálán,

a_i a diszkriminációs mutató ugyanazon a skálán.

Az egyparaméteres modellnél $a_i=1$.

A háromparaméteres modellek az előbbi mutatókon kívül az eredményes találgatás valószínűségét is figyelembe veszik. Grafikonon ábrázolva az itemkarakterisztikus görbéje az alacsony képesszintű egyéneknél nem a nullához konvergál. Az 5. ábrán a két szaggatott vonallal jelölt itemnél a sikeres találgatás valószínűsége nagyobb, mint 0, a görbék aszimptotái 0,2, illetve 0,4-hez tartanak, ugyanakkor a helyes válasz valószínűsége is magasabb, mint 50% (Partchev, 2004; Molnár, 2013).



5. ábra: A háromparaméteres logisztikus modell itemkarakterisztikus görbéje

A használatos matematikai képlet:

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]}$$

ahol θ tanuló képességszintje a vizsgált képességskálán,

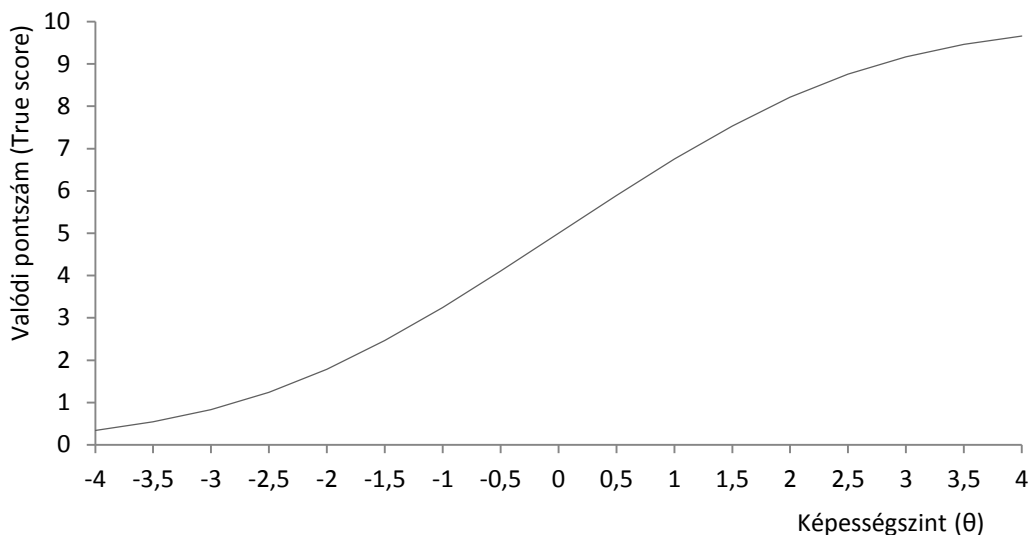
b_i az item nehézségi mutató ugyanazon a skálán,

a_i a diszkriminációs mutató ugyanazon a skálán,

c_i a találgatási paraméter.

Az egyparaméteres modellnél $a_i=1$ és $c_i=0$.

Az itemek karakterisztikus görbéjének pontonkénti összegzésével előállítható a teszt karakterisztikus görbéje, melynek elsődleges szerepe a becsült képességpontok valódi pontszámokká való konvertálásában van (Baker, 2001). A 6. ábrán látható görbe egy tíz itemből álló teszt karakterisztikus görbéjét ábrázolja. Az item jelleggörbéhez hasonlóan leolvasható, hogy adott képességszintű tanulók hány pontot érnének el az adott teszten. Ezen az ábrán például egy 3-as képességszintű tanuló 9 pontot érne el, míg egy 1-es képességszintű 7 pontot. Az itemkarakterisztikus görbéhez hasonlóan a teszt nehézségét az a képességszint (θ) mutatja meg, amely az 50%-os megoldáshoz szükséges, vagyis a valódi pontszámhoz tartozó képességszintnél olvasható le, a jelen esetben 0 az értéke (Baker, 2001).



6. ábra: Teszt karakterisztikus görbe

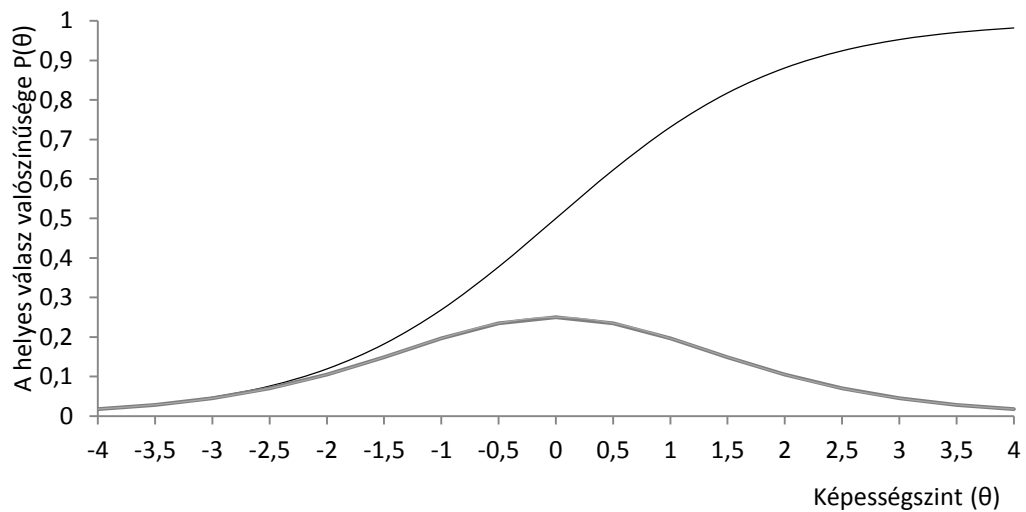
A tesztelés során információhoz jutunk arról, hogy egy adott itemet egy bizonyos képességszintű tanuló milyen valószínűséggel oldana meg. Az eddigi egyenletek alapján a sikeres válaszadás valószínűségét az adott személy képességparaméterének és az item nehézségi indexének különbsége határozza meg ($\theta - b_i$). A tesztelés során tehát akkor jutunk leginkább megbízható információhoz, ha ez a különbség minél kisebb, tehát a személy képességparamétere és az item nehézségi indexe minél közelebb van egymáshoz (Molnár, 2013). Az item információt a következő képlettel definiálhatjuk (egyparaméteres modell esetén):

$$I_i(\theta) = P_i(\theta)Q_i(\theta),$$

ahol $Q_i(\theta) = 1 - P_i(\theta)$ (Baker, 2001).

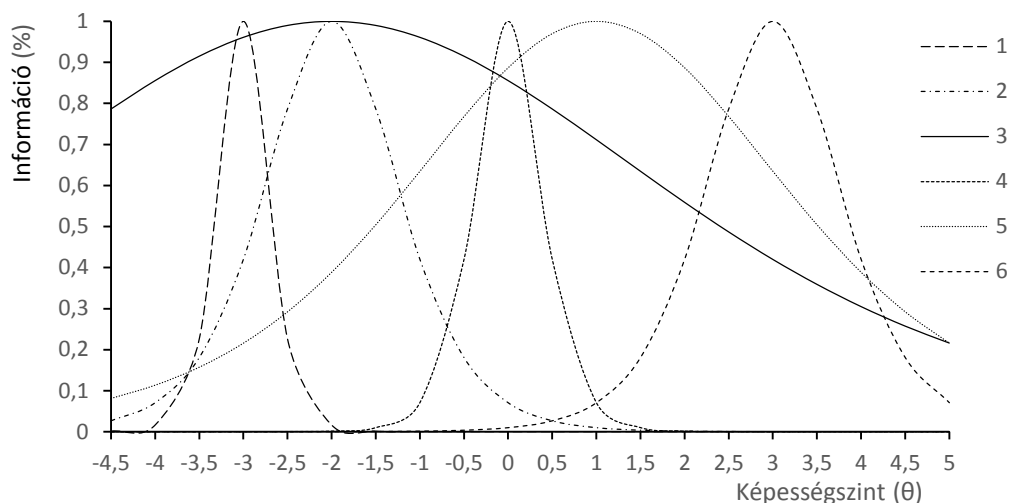
A 7. ábra egy egyparaméteres itemkarakterisztikus és információs görbét ábrázolja. A grafikonról leolvasható, hogy az item a különböző képességszinteken mekkora információt szolgáltat. Az item információ értékei szimmetrikusak a nehézségi paraméterre. A függvény maximuma 0,25, melyet ott ér el, ahol a helyes és a helytelen válaszok valószínűsége egyaránt 0,5. Ez akkor történik meg, amikor az item nehézségi indexe megegyezik a személy képességparaméterével (jelen esetben a 0 képességszintnél). Mindkét oldalra távolodva ettől a ponttól a kinyerhető információ nagysága csökken. Tehát az item akkor mér pontosabban, ha minél több információt szolgáltat az adott személy képességszintjéről, ez pedig akkor valósul meg, ha nehézségi szintben minél közelebb áll a vizsgált személy képességszintjéhez

(Molnár, 2013). Az adaptív tesztelés során tehát az algoritmusnak ilyen itemek kiválasztását kell biztosítania.



7. ábra: Egyparaméteres itemkarakterisztikus és információs görbéje (Molnár, 2013 alapján)

A 8. ábra hat különböző item item információs görbéjét ábrázolja. A függvények középpontja mutatja az item nehézségét, ennek megfelelően az 1-es item a legkönnyebb, a 6-os a legnehezebb. A görbék meredeksége a megkülönböztető erejükre utal, minél meredekebb a görbe, annál nagyobb a megkülönböztető ereje, azaz jelen esetben az 1-es item különbözteti meg a legjobban a tanulókat, a 3-as a legkevésbé. A 3-as képességtartományban a 6-os item szolgáltatja a legtöbb információt. A 3-as item a legszélesebb sávon szolgáltat információt, de a szolgáltatott információ precizitása jóval kisebb, mint a többi item esetében. A tesztelés során amennyiben átlagos képességű tanuló számára ($\theta=0$) szeretnénk itemet kiválasztani, akkor a 4-es számú item a legmegfelelőbb, mivel ez szolgáltat a legtöbb információt erről a képességszintről (Weiss, 2011). Az item-információs függvénynek az item-kiválasztási algoritmus meghatározásánál van kiemelt jelentősége.



8. ábra: Hat különböző item item-információs függvénye

Az itemkarakterisztikus függvényhez hasonlóan a teljes teszt szintjén képezhetjük a teszt információs függvényt, mely az előzőhöz hasonlóan az itemek információs függvényeinek összege az adott képességszinten (*Baker, 2001; Molnár, 2013*). Ennek megfelelően, a teszt szintjére kivetítve, akkor jutunk a tanulókról képességszintjét illetően legtöbb információhoz, ha olyan tesztet alkalmazunk, melynek nehézségi szintje minél közelebb van a tanulóéhoz.

Minél több információ nyerhető ki a tesztből, annál pontosabban képes a tanulók képességszintjét becsülni, vagyis annál kisebb hibával mér. A teszt standard hibája (SE) a kapott információ reciprokának a négyzetgyökével számítható:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Az adaptív tesztelés folyamata során a valószínűségi tesztelmélet alkalmazása a kalibrált itembank létrehozásánál, a személyek képességbecslésénél, az item-kiválasztási algoritmusoknál, a végződtetési kritérium meghatározásánál és a végső képességbecslésnél kap kiemelt szerepet (lásd később).

I.6 AZ ADAPTÍV TESZTKÉSZÍTÉS FOLYAMATA

Az adaptív tesztek számos változata létezik (*van der Linden, 2008*), az itemalapú tesztetől a lineáris alteszteket alkalmazó többszakaszos tesztekig, alapelvüket tekintve azonban mindegyik adaptív teszt hasonlóan épül fel. A fejezet célja, hogy áttekintést adjunk

az adaptív tesztkészítés folyamatáról. Bemutatjuk az adaptív tesztek fő komponenseit, a kalibrált feladatbank létrehozásának folyamatát, az item-kiválasztási algoritmusokat, a tesztvégződtetési kritériumokat, valamint a teszt kiközvetítésére használható platformokat (*van der Linden és Glas, 2010*).

Az adaptív tesztek működése szigorú algoritmushoz kötött (*Linacre, 2000*). Ezen algoritmus biztosítását a tesztek szerkesztési folyamata biztosítja, mely a következő fő komponensekből tevődik össze (*Green, Bock, Humphreys, Linn és Reckase, 1984; Weiss és Kingsbury, 1984; Thompson és Prometric, 2007; Thomson és Weiss, 2011; Sereci, 2003*):

- 1) megvalósíthatóság, alkalmazhatóság lehetőségeinek felmérése,
- 2) feladatbank létrehozása,
- 3) valószínűségi tesztelméleti modell (IRT – Item Response Theory) kiválasztása,
- 4) itemek előtesztelése, kalibrálása, skálázása,
- 5) kezdő item(ek) kiválasztása,
- 6) itemkiválasztási algoritmus meghatározása,
- 7) képességbecslés,
- 8) végződtetési kritérium,
- 9) a teszt kiközvetítése.

A továbbiakban ezen komponensek mentén ismertetjük a tesztkészítés folyamatát.

I.6.1 Megvalósíthatóság, alkalmazhatóság lehetőségeinek felmérése

Adaptív tesztelésre való átállásnál számos gyakorlati és üzleti kérdés merül fel, melyek befolyásolhatják a döntést. Egyrészt nem minden hagyományos teszt konvertálható adaptívvá (*Čisar és mtsai, 2010; Linacre, 2000*), másrészt meggondolandó, hogy az adaptív tesztre való átállás eredményezi-e az elvárt mérészetodikai javulást, vagyis a teszt hosszának és a tesztelési időnek a rövidülését, valamint a precizitás és a tesztbiztonság növekedését (*Thompson és Weiss, 2011*).

Az adaptív tesztek előállítása jelentős anyagi befektetéssel jár. Egyrészt alapfeltétel a több száz itemből álló kalibrált feladatbank létrehozása, melynek kifejlesztése szakembereket igényel, másrészt a számítógép alapú kiközvetítéshez speciális szoftverekre van szükség, melyek szintén szakembereket és jelentős anyagi forrásokat kívánnak (*Thompson és Weiss, 2011; van der Linden, 2008*).

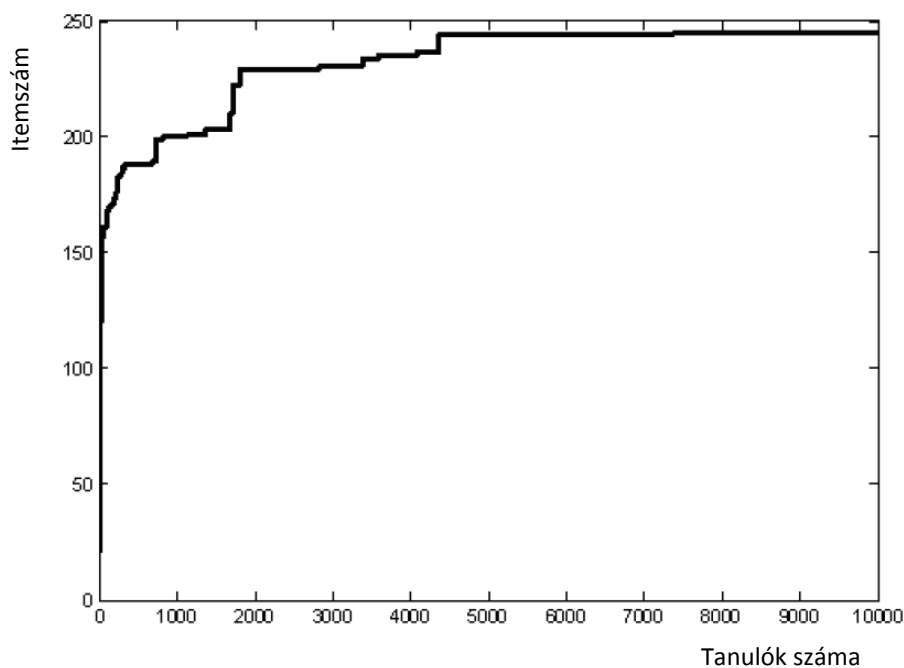
I.6.2 Feladatbank létrehozása

Az adaptív tesztek alapvető feltétele a megfelelően kalibrált feladatbank. A feladatbank *Milman* és *Arter* (1984, 315. o.) definíciója szerint „könnyen hozzáférhető tesztkérdések viszonylag nagy gyűjteménye”. A „viszonylag nagy” azt jelenti, hogy az itemek száma többszöröse a tesztnél előforduló itemek számának, a „könnyen hozzáférhetőn” pedig azt érti, hogy az itemek indexeltek, paraméterekkel ellátottak, hogy a tesztelés folyamán minél könnyebben elérhetőek legyenek. A feladatbank az itemeken túl azok jellemzőit is tartalmazza, mint például a megoldásokat, különböző tartalmi információkat, és az itemek paramétereit (*Burghof*, 2001; *Linacre*, 2000). *Vidákovich Tibor* (1993, 9. o.) megfogalmazásában azokat a „feladathalmazokat, amelyek a feladatgyűjteményekben is szereplő jellemzőkön, paramétereken kívül legalább a feladatok egy jóságmutatóját (többnyire a reliabilitásukat, megbízhatóságukat) és egy-két viszonyítási adatot (országos átlagot, szórást) is tartalmazznak, *feladatbankoknak* nevezzük.” A feladatbankban szereplő itemparaméterek lehetnek előzetes bemérés alapján meghatározottak (pl. reliabilitás, nehézségi mutató, differenciáló erő, magyarázóérték, a megoldáshoz szükséges átlagos idő), illetve előzetes mérés nélkül is megállapíthatóak (pl. tartalom, értékelési kritérium, feladattípus, a megoldáshoz használható eszközök) (*Vidákovich*, 1993).

Feladatbankok létrehozásánál számos kérdés felmerül. Először is fontos a megfelelő itemszám. A kezdeti adaptív teszteknel 100-120 itemből álló bankok már elérték a lineáris tesztek pontosságát, nagymintás mérésnél azonban ez a szám kevés. *Wise* és *Kingsbury* (2000) három fő faktort említ, melyet a feladatbank méretének meghatározásánál figyelembe kell venni: A hagyományos lineáris tesztekkel is nagy pontosságú mérések végezhetőek, az itemkiválasztási folyamatnál alkalmazott korlátozások nagyobb itemszámot követelnek meg, valamint a magas tétet képviselő tesztek esetén a tesztbiztonság veszélybe kerülhet, ha a feladatbank túl kicsi (*Wise* és *Kingsbury*, 2000; *Csapó, Molnár* és *R. Tóth*, 2008). Ezért a teszt tétjétől és a felhasználás gyakoriságától függően a szükséges itemek száma több száz is lehet. A feladatbankok kifejlesztésének költsége igen magas, ezért ezt a tényezőt is ajánlott figyelembe venni (*Thompson* és *Weiss*, 2011). *Revueita* és *Ponsoda* (1998) rámutatnak, hogy amennyiben túl nagy a létrehozott feladatbank, az algoritmustól függően az itemek bizonyos százaléka csak ritkán választódik ki. Ezért olyan méretű feladatbank létrehozása javasolt,

ahonnan minden itemet kiválasztanak bizonyos százalékban, és elkerülhető, hogy bizonyos itemek túlzottan sokszor szerepeljenek.

Reckase (2007) az egyparaméteres Rasch-modellt felhasználva próbált módszert kidolgozni az optimális feladatbank méretének meghatározására. *Segall*-hoz (2004) hasonlóan, ő is hangsúlyozta, hogy a feladatbank mérete szoros összefüggésben van a tesztelésbe bevont kohorsz képességeloszlásával. A 9. ábra azt mutatja, hogy kisebb mintás méréseknél (100-200 fő) 150-200 item elegendő, nagyobb minta esetén azonban minimum 250-es itemszám szükséges.



9. ábra: A feladatbankhoz szükséges itemszám a tanulók számának függvényében (*Reckase*, 2007 alapján)

A feladatbankba a feladatok megfelelően kalibrálva kerülnek. Az item kalibrálása a választott valószínűségi tesztelméleti modell segítségével az adott item paramétereinek becslését jelenti (*Eggen*, 2007; *Weiss*, 2013). A legtöbb itembank eredetileg papíralapú tesztekhez készült, és paramétereiket ez alapján állapították meg. Több kutatás szerint a legtöbb item a papír alapon bemért paramétereivel használható adaptív tesztekhez is (*Linacre*, 2000; *Segall*, 2004). Azonban a papíralapú tesztekhez használt itembankok általában nem elég nagyok, és az itemek paramétereik nem fedik le a teljes képességskálát, mivel hagyományos tesztelésnél többségben vannak az átlagos képességszinthez illeszkedő itemek. Adaptív

tesztelés esetén azonban az az ideális, ha a feladatbank a képességskála minden szintjére vonatkoztatva tartalmaz elegendő feladatot, és többségben vannak a magas megkülönböztető erejű, alacsony találgatási indexszel rendelkező itemek (Segall, 2004; Stocking, 1994). Ennek biztosítására az adaptív tesztelésre való átállás során az itembankok bővítése szükséges, mely az IRT segítségével történhet az új itemek meglévő itemekhez való hozzákálázásával. Az összeskálázást az itemek kalibrálása során alkalmazott közös itemek, ún. horgonyitemek biztosítják (Eggen, 2004). A horgonyzásnak több módszere használatos, adaptív teszteknel a leggyakrabban az ún. kövér horgonyt (fat anchor) alkalmazzák (Molnár, 2013). A különböző csoportok különböző teszteknek kapnak, oly módon, hogy a tesztek egy része a már kalibrált horgonyitemekből áll, a másik részben pedig a kalibrálatlan új itemek szerepelnek. A közös horgonyitemek biztosítják, hogy a tesztek további itemei egymáshoz skálázhatóak legyenek (Lee, 2011).

Az itembankok használata során fontos, hogy a nem jól működő itemeket törölni szükséges, mivel súlyos pontatlanságokat okozhatnak a képességbecslés során (Wise és Kingsbury, 2000). Ilyen itemek lehetnek például a magas találgatási indexszel rendelkező itemek vagy a részmintafüggő itemek (Molnár, 2013). Mivel az adaptív tesztek viszonylag rövidek, lényeges, hogy a felhasznált itemek jól működjenek. A nem jól működő itemek kiszűrésének többféle módja létezik. Az egyik gyakran használt eljárás, hogy a vizsgált item empirikus itemkarakterisztikus görbáját az elméleti görbével vetjük össze, és az illeszkedés mértékét leíró mutató (MNSQ) értékét vizsgáljuk (Molnár, 2013). Amennyiben ez túlzottan eltér az elvárt értéktől, célszerű az adott itemet törölni.

I.6.3 Valószínűségi tesztelméleti modell kiválasztása

Az adaptív tesztek a valószínűségi tesztelmélet (IRT – *Item Response Theory*) felhasználásával készülnek (Thompson és Weiss, 2011). Ennek az az oka, hogy az IRT segítségével a különböző teszteken elért eredmények összehasonlíthatóak lesznek annak ellenére, hogy a tanulók különböző teszteknek oldanak meg (Eggen, 2007; Molnár, 2006). Az IRT így megkönnyíti annak meghatározását, hogy adott képességszintű tanuló milyen valószínűséggel teljesítene adott feladatbankban szereplő feladatok megoldásán, még akkor is, ha csak a feladatbankban szereplő itemek egy bizonyos részét oldja meg (Molnár, 2013). Míg a klasszikus tesztelméleti mutatók csak az adott tanulócsoporthoz érvényesek, a

valószínűségi tesztelmélet lehetővé teszi a mintafüggetlen, illetve tesztfüggetlen összehasonlítást (Csapó, Molnár és R. Tóth, 2008; Reckase, 2009).

Adaptív teszteknel leggyakrabban a többparaméteres modelleket használják (Harris, 1989), ezek közül is sokáig a háromparaméteres modellek voltak a leginkább elterjedtek. Ennek az az oka, hogy a legtöbb nagymintás tesztben a domináns feladattípus a többválasztásos feladat (*multiple choice*), ahol a találgatás valószínűsége 25% is lehet (Han, 2012). Azonban a *c*-paraméter nem becsülhető egyértelműen, máshogy viselkedik gyenge és magas képességű tanulóknál, valamint kisebb és nagyobb minták esetén. Ezért az utóbbi időben az egy- és kétparaméteres modellek, illetve a háromparaméteres modell fix találgatási indexszel alkalmazott változata terjedt el (Han, 2012).

I.6.4 Itemek kalibrálása, skálázása

Bármelyik modellt választjuk, mindenképpen szükséges az itemeket empirikus tesztelési eljárásokon keresztül vinni. Kétféle eset lehetséges attól függően, hogy a létrehozott feladatbank teljesen újonnan fejlesztett, vagy az előzetesen használt lineáris tesztből átvett itemeket bővítették újabakkal (Thompson és Weiss, 2011). Mivel az adaptív tesztek több száz itemet tartalmazó feladatbank felhasználásával működnek megfelelően, fizikailag lehetetlen, hogy a kalibrálás során minden diák mindegyik itemet megkapja. Ezért célszerű a feladatbankot részalmazaira bontani oly módon, hogy minden részalmaz tartalmazzon közös elemeket, ún. horgony itemeket (Molnár, 2013). Ezen horgony itemek segítségével végezhető el aztán az itemek közös skálán való elhelyezése, és paramétereik meghatározása (Lee, 2011). Amennyiben meglévő feladatbank kerül kibővítésre újonnan fejlesztett itemekkel, az újonnan fejlesztett itemek előtesztelése és a feladatbankba illesztése szintén a horgony itemek segítségével történhet (Thompson és Weiss, 2011).

I.6.5 Kezdő item(ek) kiválasztása

Adaptív tesztelés során fontos kérdés a kezdő item kiválasztása, melyre többféle lehetőség van. Amennyiben nem állnak rendelkezésre előzetes információk a tanulóról, ajánlatos átlagos nehézségű kezdő itemmel kezdeni a tesztet. Mivel a továbbiakban a kérdések szintje egyre közelebb kerül a tanuló szintjéhez, a rosszul kalibrált kezdő itemnek nincs nagy jelentősége, viszont a megfelelően választott kezdő item lerövidítheti a tesztelés időtartamát (Weiss és Kingsbury, 1984).

A teszt biztonsága és az itemek kiválasztásának szabályozása érdekében célszerű különböző kezdő itemekkel indítani a tesztet. Az egyenlő feltételek biztosítása érdekében a kezdő itemet a -0,5 és +0,5 közötti skálatartományba eső itemek közül választják ki (*Thompson és Weiss, 2011*). Amennyiben előzetes információk rendelkezésre állnak a tanuló képességszintjét illetően, például előző teszteredmények, akkor ezek felhasználása is lehetőség lehet a kezdő item nehézségi szintjének a meghatározására (*Thompson és Way, 2007; Weiss, 2011*).

I.6.6 Itemkiválasztási algoritmus meghatározása

Itemalapú tesztelés folyamán minden megoldott item után egy újabbat választanak ki a feladatbankból. Azt, hogy melyik item kerüljön a következő lépésben a tanuló elé, az alkalmazott algoritmus dönti el, melyre különböző szabályzókat építenek be. A szabályozás célja kettős, egyrészt kiküszöbölni, hogy bizonyos itemek túlzottan sokszor kerüljenek kiközvetítésre, másrészt biztosítsa, hogy minden item egyenletesen szerepeljen a tesztelés során, ne legyenek olyan itemek, melyek folyton szerepelnek, és olyanok, amelyek egyszer sem kerülnek kiközvetítésre. Amennyiben bizonyos itemek túlzottan gyakran választódnak ki (ezek általában az átlagos nehézségi szinten a magas információval rendelkező és magas diszkriminációjú itemek), gyorsan „elhasználódnak”, azaz a gyakori használat következtében ismertté válnak, ami a tesztbiztonság csökkenéséhez vezet, veszélyeztetve a teszt validitását. Az itembank egyenletes kihasználtsága érdekében fontos, hogy minden item használatban legyen, hiszen így sokkal többféle tesztváltozat összeállítására nyílik lehetőség (*Revuelta és Ponsoda, 1998*).

Három tipikus terület van, melyre szabályozást építenek be, az item kiválasztási gyakoriság, a tartalmi és az egymásnak információt szolgáltató itemek feletti szabályozás (*Wise és Kingsbury, 2000*). Az item kiválasztási gyakoriságának szabályozása során az algoritmus azt vezérli, hogy nehogya bizonyos itemek túl sokszor szerepeljenek, ami veszélyeztetné a tesztbiztonságot. A tartalmi szabályozás a különböző tartalmi elemek azonos mértékű kiválasztását kontrollálja (*van der Linden, 2005*). Az egymásnak információt szolgáltató itemeken alkalmazott szabályozás kiküszöböli a nagyon hasonló, vagy egymásnak információt szolgáltatatható elemek kiválasztását (*Weiss, 2013*).

Számtalan itemkiválasztási algoritmust dolgoztak ki az adaptív tesztekkel foglalkozó kutatók, melyek közül a leggyakrabban a legnagyobb valószínűség módszerét (MLE –

maximum likelihood estimation), az egyparaméteres módszert (*1P method*), a véletlen módszert (RA – *random method*), a *Sympson-Hetter* módszert, a súlyozott eltérés modellt (WDM – *weighted deviation model*) és az árnyék teszt megközelítést (STA – *Shadow Test Approach*) alkalmazzák.

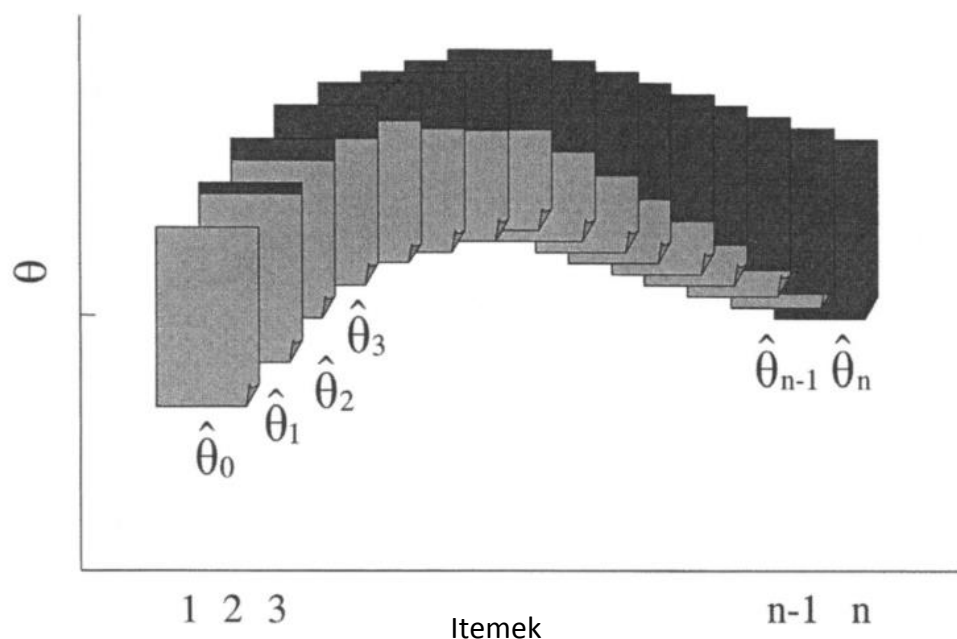
Az egyik legelterjedtebb módszer a legnagyobb valószínűség módszere, melynél mindig a legmagasabb információtartalommal rendelkező item választódik ki (*Kingsbury és Zara, 1989; Weiss, 2013; Schnipke és Green, 1995*). A 8. ábra egy hat itemből álló itembank információs függvényeit ábrázolja példaként. A tanuló képességbecslése az előző lépésben $\theta=0,0$ volt. Ezen a képességszinten három item közvetít információt; a 3-as, a 4-es és az 5-ös. Mivel a 4-es item szolgáltatja a legnagyobb információt, ezért a következő lépésben ezt az itemet választják ki (*Weiss, 2013*). Egyparaméteres modellnél az item információ helyett elég az itemek nehézségi paraméterét figyelembe venni, és a következő lépésben azt az itemet kiközvetíteni, amelynek nehézségi paramétere a legközelebb van a becsült képességszinthez (egyparaméteres módszer; *Davis, 2005; Eggen és Verschoor, 2006*). A véletlen módszernél az algoritmus előre meghatározott mennyiségű, a legnagyobb információkat szolgáltató item közül választ, véletlenszerűen (*Revuelta és Ponsoda, 1998*). A módszerek előnye, hogy könnyen alkalmazhatóak, nincs szükség komplikált algoritmusok beépítésére, viszont hátrányuk, hogy a legnagyobb információval rendelkező itemek használata kerül túlsúlyba. Másik gondot jelenthet főleg a tesztelés elején, amikor még a személy képességparamétere nagyobb hibával van becsülve, hogy ehhez, a kevésbé pontos képességszinthez igazítja az algoritmus a kiválasztódó itemeket, így előállhat az a paradoxon, hogy eredetileg legjobban mérő itemek rosszabbul fognak a tesztelés során mérni (*Keng, 2000*).

A *Sympson-Hetter* módszer ennek kiküszöbölésére törekszik, és limitálja egy-egy item tekintetében a kiválasztási lehetőség számát. Amennyiben bizonyos item eléri a megengedett kiválasztási számot, a továbbiakban nem kerülhet kiközvetítésre (*Revuelta és Ponsoda, 1998; Sympson és Hetter, 1985*). A tesztelés során ez úgy nyilvánul meg, hogy a leggyakrabban kiválasztódó itemek száma egy idő után lecsökken, majd amikor a többi item is kiválasztódik bizonyos százalékban, újra elérhetővé válnak, ezáltal biztosítva az egyenletes itemkiválasztást (*Keng, 2000*).

A súlyozott eltérés modellnél (*Swanson és Stocking, 1993*) az itemek karakterisztikája súlyozottan szerepel. Ez biztosítja, hogy az adott itembankból generált tesztek a lehetőségekhez mérten minél hasonlóbbak legyenek tartalmi, itemformátum és

itemparaméterek tekintetében. Ennek biztosítása érdekében nemcsak a legnagyobb információval rendelkező item választódik ki, hanem amely megfelel a különböző tartalmi szabályzóknak is (Stocking, 1996).

További elterjedt módszer az árnyékteszt megközelítés (STA – *shadow test approach*; van der Linden és Glas, 2000), ami filozófiáját tekintve a WDM-modellhez áll legközelebb, azonban ennél a modellnél nemcsak a következő itemről dönt az algoritmus, hanem egy teljes, úgynevezett „árnyéktesztet” állít elő, mely a rendelkezésre álló legkedvezőbb karakterisztikájú itemekből áll össze. Így minden megoldott itemet követően képez egy-egy új lineáris tesztet, amely a becsült képességpontnál a legnagyobb információt szolgáltat, mindegyik tartalmi szabályozásnak megfelel és tartalmazza az összes eddig ki nem választott itemet (10. ábra). Ennek a módszernek az az előnye a súlyozott eltérés modellhez képest, hogy a teszt egészére optimális itemkiválasztást képes előre tervezni.



10. ábra: Az árnyékteszt megközelítés grafikus ábrázolása (a sötét terület mutatja a kiközvetített itemeket, a világosabb terület az egy-egy item megoldását követően újrastrukturált lineáris tesztet; van der Linden, Ariel és Veldkamp alapján, 2006)

A bemutatott módszerek közös vonása, hogy bizonyos algoritmusok szerint döntenek arról, hogy a következő lépésben melyik item kerüljön a tanuló elé. Ez a folyamat addig ismétlődik, amíg az ismétlődő képességbecslés eléri a végződtetési kritérium szintjét, és a becsült képességpont véglegessé válik. A kiválasztási szabályozókról további részletek

olvashatók Swanson és Stocking (1993), van der Linden és Glas (2000), Veerkamp és Berger (1997), Rijn, Eggen, Hemker és Sanders (2000), Chang és Ansley (2003), van der Linden (2005), Barrada, Olea, Ponsoda és Abad (2009), Luecht és Sireci (2011) Revuelta és Ponsoda (1998), Georgiadou, Triantafyllou és Economides (2007), valamint Davis és Dodd (2003) tanulmányaiban.

I.6.7 Képességbecslés

Adaptív tesztnél nem használható a hagyományos pontozási módszer (Kolen, 1999-2000, 2010), mert a diákok különböző feladatokat kapnak, és nem mindegy, hogy azonos nyerspontot elérő tanulók milyen nehézségű feladatokat oldottak meg. A képességbecslés többféle módon történhet (Diao és Reckase 2009; Hoe, Kiong, Sam és Usop, 2009; Molnár, 2013), a leggyakrabban alkalmazott eljárás a legnagyobb valószínűség melletti közelítés (MLE – *Maximum Likelihood Estimates*). Itemalapú adaptív teszteknel a tanulók képességbecslése minden megoldott itemet követően megtörténik (Weiss, 2011; Diao és Reckase, 2009). Amennyiben a tanuló válasza helyes volt, nő a képességindex, ha helytelen, akkor csökken. Ez a procedúra minden megválaszolt item után megismétlődik, és ez alapján kapja a tanuló a következő itemet. Az MLE módszer nagy előnye, hogy a tanuló összes addigi válaszában mért információt figyelembe veszi az itemek nehézségi szintjeivel összefüggésben, vagyis ha egy nehezebb kérdésre jól válaszol, akkor a képességszintje jobban nő, mintha egy könnyű itemre válaszolt volna helyesen. Hasonlóan, ha egy könnyű itemre válaszol helytelenül, képességszintje alacsonyabbra kerül, mintha egy nehezebb itemre válaszolt volna helytelenül. Ez alapján különböző képességpontot kap az a tanuló, akinek '0011' volt a válaszmintázata, mint akinek '1100' (Weiss, 2011).

A képességbecslés tehát ismétlődő folyamat. A tesztelés kezdetekor egy *a priori* értékről indul, mely általában egy átlagos érték és a következő lépésben az előző értékből számítható a következő (egyparaméteres modellnél):

$$\theta_{s+1} = \theta_s + \frac{\sum_{i=1}^N -[u_i - P_i(\theta_s)]}{\sum_{i=1}^N P_i(\theta_s) Q_i(\theta_s)}$$

ahol θ_s a személy képességbecslése a s -edik lépésnél,

u_i az i -edik itemre adott válasz, értéke 1, ha a válasz helyes, 0, ha a válasz helytelen,

$P_i(\theta_s)$ a θ nehézségi szinten az i -edik itemre adott helyes válasz valószínűsége az s -edik lépésben,

$Q_i(\theta_s)$ a θ nehézségi szinten az i -edik itemre adott helytelen válasz valószínűsége az s -edik lépésben.

A képlet magyarázata az, hogy a képességbecslés során mindig az előző becsült értékhez ad hozzá egy korrekciós értéket. A korrekciós érték számlálójában az $u_i - P_i(\theta_s)$ a tanuló válasza és a helyes válasz valószínűségének a különbsége. Ahogy a képességbecslés egyre pontosabbá válik, ezeknek a különbségeknek az összegei egyre kisebbek lesznek. A nevezőben található képlet pedig a standard hiba képlete, mely szintén egyre kisebb lesz, ahogy a képességbecslés egyre pontosabbá válik. Tehát a képességbecslés során a korrekciós érték egyre kisebb lesz, és amikor ez az érték átlép egy bizonyos szintet, θ_{s+1} értéke lesz a végső becsült képességpont (*Baker, 2001; Gupta, Vijn és Vats, 2010*).

I.6.8 Végződtetési kritérium

Az adaptív tesztek lehetnek fix hosszúságúak (a tanulók egyenlő számú tesztkérdést kapnak, de az itemek személyre szabottan választódnak a feladatbankból), valamint lehetnek változó hosszúságúak, amikor nemcsak az itemek választódnak adaptívan, hanem az itemek száma is. Ez esetben a teszt végződésének meghatározására különböző módszerek léteznek attól függően, hogy a teszt végződésének meghatározásához a tanuló képességszintjét, a standard hibát, a feladatbankban lévő itemeket vagy időkorlátot vesznek alapul (*Thompson és Weiss, 2011; Luecht és Sireci, 2011*).

A tanuló képességszintjét figyelembe vevő módszernél a tesztelés folyamata akkor fejeződik be, amikor az újabb és újabb itemekre adott válaszok alapján a kiközvetített itemek paraméterei már csak minimális mértékben különböznek a tanuló képességszintjétől, vagyis az itemparaméterek közötti eltérés egy előre meghatározott sávon belülre kerül. A standard hibát figyelembe vevő módszer hasonló elv szerint működik, csak itt a mérési hiba mértékét számolják, amely nagyobb a tesztelés elején, és bizonyos számú item után minimálisra csökken (*Thompson és Weiss, 2011*). A harmadik megközelítés a tanuló képességszintje helyett a feladatbankot veszi alapul. Ilyen például a minimum információs kritérium, ahol a tesztelés akkor ér véget, mikor a feladatbankban már nem marad olyan item, mely újabb, az előzőeknél pontosabb információt szolgáltatna a vizsgált személy képességszintjéről (*Thompson és Weiss, 2011*). A negyedik lehetőség az időkorlát, amikor bizonyos idő elteltével a tesztelés véget ér. A módszer részletes leírása *Weiss (2004), Thompson és Prometric (2007)*, valamint *Čisar és munkatársai (2010)* tanulmányában található.

Az itemek kiválasztásához hasonlóan a végződtetési kritérium is lehet szabályozott. A minimális és maximális teszthosszúság tipikus szabályozók, melyek biztosítják, hogy a tanulók biztosan megkapjanak bizonyos számú itemet, de a tesztelés se legyen túlságosan hosszú (Thompson és Weiss, 2011).

A teszt céljától függően is különböző algoritmusok szabályozhatják a teszt végződtetését. Egyrészt lehet a tanulók képességbecslése, és közös skálán való elhelyezése, mely az oktatásban használt tesztekénél a leggyakrabban fordul elő. Másik cél lehet a tanulók kritériumorientált tesztelése, ahol egy bizonyos kritérium értéken felül elfogadott a teljesítmény, ez alatt viszont nem, ilyenek például a vizsgatesztek.

Amennyiben a mérés célja minden tanuló képességszintjének hasonló pontossággal való becslése, a mérés akkor fejeződik be, amikor az újabb és újabb itemekre adott válaszok alapján a becsült képességszintek már csak minimális mértékben különböznek a tanuló képességszintjétől, vagyis a becsült képességszintek közötti különbségek egy előre meghatározott értéken belülre kerülnek. Ehhez hasonlóan gyakran a mérési hiba értékét veszik figyelembe, melynek határa általában 0,25 (Thompson és Weiss, 2011; Luecht és Sireci, 2011).

Kritériumorientált tesztek esetében más végződtetési kritérium használatos. Először a képességskálán a ponthatár és a megfelelő pontosságú becslést biztosító hibasáv kijelölése történik. A tesztelés során minden képességértékhez kiszámítanak egy standard hibasávot. A tesztelés akkor ér véget, amikor a becsült képességpont a hibasávval együtt átlépi a ponthatár felső vagy alsó hibasávját.

A tesztelés végén kerül sor a tanuló végső képességszintjének meghatározására, mely általában az utolsó megoldott item utáni becsült képességpont. A könnyebb áttekinthetőség érdekében, ezt az értéket általában átkonvertálják egy itemstandard skálára, mint pl. 500-as átlagú, 100-as szórású skálára. A teszten elért eredmény meghatározható a becsült képességpont valódi pontszámra (true score) konvertált értékeként is (12. ábra; Stocking, 1996). Ez az eljárás közelebb áll a hagyományos pontozáshoz, ezért könnyebben érthető a hétköznapi emberek számára (Yi, Wang és Ban, 2001). Ennek a pontozási módnak az előnye, hogy a személyeket és az itemeket közös képességskálára hozza, és így összehasonlíthatóvá válnak a tanulói teljesítmények. A másik fontos előnye, hogy nagy valószínűséggel megállapítható, hogy a tanulók hogyan teljesítettek volna bizonyos itemeken még akkor is, ha nem került számukra kiközvetítésre az a bizonyos item (Molnár, 2013).

I.6.9 A teszt kiközvetítése

Az adatfelvétel utolsó lépése a tesztek kiközvetítése. Ez történhet különböző, kereskedelmi forgalomban kapható, adaptív tesztek kezelésére alkalmas programok segítségével (FastTestWeb, FastTEST Professional Testing System, Pearson VUE, Prometric, McCann), de saját kutatócsoportok által kifejlesztett szoftvereken, platformokon keresztül is végbemehet (MATE, eDIA). Hazánkban a Szegedi Tudományegyetem Oktatásméleti Kutatócsoportja által kifejlesztett eDia rendszer biztosít online adaptív tesztelésre alkalmas platformot (R. Tóth és mtsai, 2011; Molnár és Latour, 2011; Molnár és Csapó, 2013). Mivel a tesztet használó intézmények különböző infrastrukturális feltételekkel rendelkeznek, az adaptív tesztet közvetítő szoftvernek megfelelően rugalmasnak kell lennie, hogy tudja ezeket a különbségeket kezelni (Way, Davis és Fitzpatrik, 2006). A tesztek leggyakrabban online platformokon keresztül működnek, és folyamatos kapcsolatot tartanak a központi szerverrel. A tanulók előre generált belépési kóddal kezdenek meg a tesztelést és azonnali visszajelzést kapnak elért eredményükről. A kapott eredmények a tanulók képességszintjét mutatják, melyek alapján az őket tanító tanárok képet kapnak arról, hogy az adott diák a képességskála mely részén helyezkedik el. Ez kiindulópont lehet a további fejlesztések megtervezéséhez, mivel a képességskála azt is megmutatja, hogy melyek azok a hiányterületek, amiben a tanulónak még fejlődnie kell.

I.7 A TÖBBSZAKASZOS ADAPTÍV TESZTEK

Az I.4 fejezetben részleteztük az adaptív tesztek előnyeit és hátrányait. Az itemalapú adaptív tesztek legfontosabb hátrányaként említik, hogy a legtöbb itemalapú teszt nem ad lehetőséget a visszalépésre, a feladatok utólagos áttekintésére és javítására, valamint a véletlenszerű itemkiosztás miatt az előző item információt szolgáltat a következő item számára. Az itemek elhelyezkedése is befolyásolhatja a megoldást; ugyanaz az item attól függően, hogy a tesztelés elején vagy a végén helyezkedik el, lehet könnyebb, vagy nehezebb (Frey és Hartig, 2009; Linacre, 2000; Wainer, 2000a; Wainer és Kiely, 1987).

Ezek közül több probléma kiküszöbölését oldja meg a többszakaszos adaptív teszt (MST – *Multi Stage Test*; Magyar, 2013), mely egyesíti magában a hagyományos lineáris és az adaptív tesztek tulajdonságait, egyrészt a kérdéseket a tanuló képességszintjéhez igazítja,

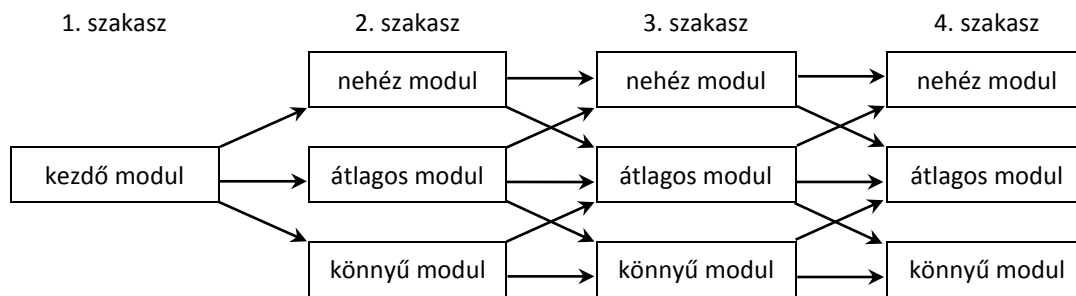
másrészt lehetőséget ad az itemek sorrendjének előzetes meghatározására (*Amstrong, Jones, Koppel és Pashley, 2004; Molnár, 2013*).

I.7.1 A többszakaszos adaptív tesztek felépítése

Szerkezetüket tekintve a többszakaszos tesztek a hagyományos lineáris és az itemalapú adaptív tesztek között félúton helyezkednek el (*Jodoin, Zenisky és Hambleton, 2006; Patsula, 1999; Zheng, 2012*). A tesztek több szakaszban itemek helyett modulokat tartalmaznak, melyek tulajdonképpen különböző nehézségi szintű rövid fix tesztek. Az alapvető különbség abban van, hogy a modulok csak egy-egy szűk képességtartomány mérését célozzák, ezért hasonló nehézségű feladatokból épülnek fel, ellentétben a hagyományos lineáris tesztekkel, melyek vegyesen tartalmaznak könnyebb és nehezebb feladatokat, ezáltal széles képességtartomány mérésére alkalmasak. Egy teszt minimum két szakaszból áll. Egy szakaszon belül két vagy több modul lehet, melyek nehézségi szintjükben különböznek egymástól. Miután a tanuló végez egy-egy modullal, képességszintje becslésre kerül, és ez alapján kap a következő szakaszban újabb nehézségi szintűt (*Zenisky, Hambleton és Luecht, 2010*).

A többszakaszos teszteknek többféle változata létezik (*Yan, von Davier és Lewis, 2014*), ennek megfelelően szerkesztésük is különböző. A legegyszerűbb szerkezetűek a kétszakaszos ('two-stage') tesztek, melyeknél a tanulók először egy kezdő modult (*routing test*) kapnak, mely különböző nehézségű itemeket tartalmaz. Ezen teszten elért eredményük függvényében kapják a tanulók saját képességszintjéhez illeszkedő nehézségű felmérő teszt (*measurement test*) a második részben (*Adema, 1990; Puhán és Gierl, 2003*). A felmérő tesztek különböző nehézségi szintű lineáris tesztek, melyek nagyjából azonos nehézségi szintű kalibrált itemeket tartalmaznak. A felmérő teszt megoldásával és értékelésével be is fejeződik a tesztelés. Több szakaszból álló tesztek esetében több felmérő teszt követi egymást, nehézségi szintjük szerint modulokba rendezve (*Hendrickson, 2007*).

Az MST nagy változatosságot enged a szakaszok, a szakaszokon belül a modulok, és a modulokon belül az itemek számát illetően (*Davis, 2005; Yan, von Davier és Lewis, 2014*). A 11. ábra egy négy szakaszból álló tesztet ábrázol, szintenként három különböző nehézségű modullal. A szintekről való továbblépés a teljesítmény függvényében egy-egy szinttel változhat.



11. ábra: 1-3-3-3 szerkezetű négyszakaszos teszt (Molnár, 2013 alapján)

A szakaszok száma nagyban befolyásolja a teszt összetettségét. Minél több szakaszból áll a teszt, annál több a lehetséges útvonal száma és annál többféle nehézségi szintű teszt állítható elő. (A 11. ábrán összesen 17 különböző útvonal lehet.) A szakaszok számának növelésével azonban egyre összetettebbé válik a teszt adminisztráció a mérési precizitás arányos növekedése nélkül (Amstrong és mtsai, 2004; Hendrickson, 2007). Ezen megfontolások alapján a leggyakrabban 2-4 szakaszos tesztek alkalmazása terjedt el (Zenisky, Hambleton és Luecht, 2010).

A szakaszokon belül a modulok is változó számúak lehetnek. Általában a teszt egy modullal kezdődik, és azt követően háromra, esetenként ötre nő az egy szinten lévő modulszám (Patsula, 1999). A megfelelő pontosság eléréséhez azonban általában három, maximum négy modul elegendő szakaszonként (Amstrong és mtsai, 2004).

A teszteken belül az itemek számára vonatkozóan oszlanak meg leginkább a vélemények. Hendrickson (2007) tanulmánya szerint a különböző kutatásokban az itemek száma egy-egy modulon belül 1-90 közötti, azonban átlagosan öt itemből álló modulok fordulnak elő a leggyakrabban. További kutatási kérdés, hogy a kezdő, vagy a következő modulok legyenek-e hosszabbak. Davis és Dodd (2003), valamint Kim és Plake (1993) a bevezető mérés pontosságának kiemelt jelentőségével indokolja a hosszabb kezdő teszt alkalmazását (Keng, 2008).

1.7.2 Elágazási szabály és pontozási lehetőségek

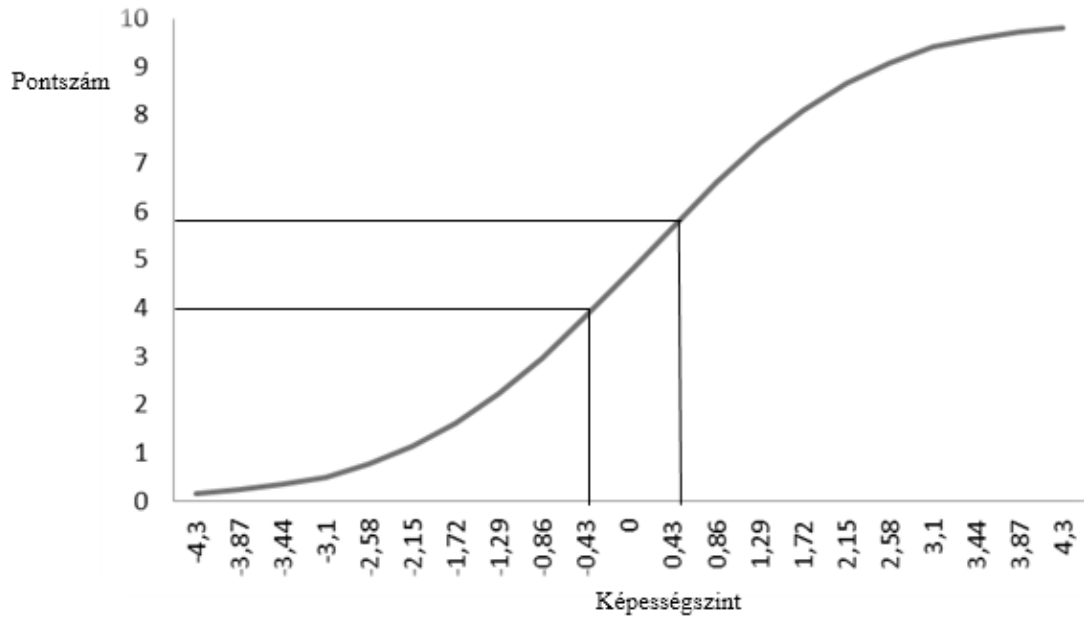
A tesztelés során a tanulók az előző modulon elért pontszámtól függően kapnak könnyebb vagy nehezebb modult a következő lépésben. Az elágazásoknál alkalmazott algoritmus alapvetően meghatározó a tesztelés során, mivel ennél a pontnál történik a tanulók hozzárendelése a különböző nehézségű modulokhoz (Zenisky és Hambleton, 2004; Amstrong,

2002). *Zenisky és Hambleton (2004)* négy leggyakrabban használatos módszert említene: a DPI, az AMI, a NC és a random módszer.

A DPI módszer (DPI – *Defined Population Interval*) a tanulókat egyenlően osztja szét a következő szakaszban következő modulokba mindegyik tanulóhoz hozzárendelve az adott modul alapján becsült képességszintjét. Ez alapján egy meghatározott képességszint alatti tanulókat a könnyű modulba, az e felettieket a közepes modulba sorolja. Ha több modulra ágazik a következő szakasz, akkor több osztópont alapján dönt az algoritmus. Az osztópont standard normál eloszlású mintánál ($N(0,1)$) a tanulók három modulba történő sorolásánál -0,43 és +0,43-nál lesz; két modul alkalmazása esetén, ha a mintát felezni szeretnénk, akkor 0,0-nál.

Az AMI módszernél (AMI – *Approximate Maximum Information*) az osztópont a soron következő teszt teszt információs függvénye (*Baker, 2001; Molnár, 2013*) alapján kerül kiszámításra, és amelyik modul a legtöbb információt szolgáltatja, azaz értéke a legközelebb áll a tanuló eredményéhez, az választódik ki az elágazásnál. Mivel ez a technika áll a legközelebb az itemalapú adaptív teszteknel használatos módszerekhez (*Keng, 2008*), ettől várható a legnagyobb mérési pontosság.

Az NC módszer (NC – *Number Correct*) az első és második modul közötti elágazásnál a kezdő modul teszt-karakterisztikus görbéje alapján osztja a tanulókat három hozzávetőlegesen egyenlő részre. Ennek megfelelően, ha két modulból áll a második szakasz, akkor 0,0 képességponthoz tartozó pontszámánál lesz az osztópont, három modul esetén pedig -0,43 és +0,43 által meghatározott értékeknél sorolják be a tanulókat a modulon elért pontszámuk alapján (12. ábra). A 0,0 képességpont az átlagos képességet jelenti, ettől balra helyezkednek el az alacsonyabb képességsávok, amelyeket a negatív számok jellemeznek, jobbra pedig az átlag felettiek, melyeket a pozitív számok szemléltetnek. A következő elágazásoknál hasonlóan határozzák meg az osztópontokat, az addig megtett útvonal alapján. Az NC módszer előnye, hogy alkalmazásával elkerülhető a tanulók minden egyes modul utáni képességbecslése, és hozzávetőlegesen egyenlő számú tanuló sorolható minden modulhoz (*Zenisky és Hambleton, 2004*).



12. ábra: Osztópontok meghatározása az NC módszer szerint: 10 itemből álló modulnál 4 és 6 pontnál helyezkednek el az osztópontok

A random módszernél véletlenszerűen történik a tanulók modulokhoz való hozzárendelésre, az egyetlen szempont, hogy a minta egyenlő arányban kerüljön elosztásra (Zenisky és Hambleton, 2004). A felsorolt módszereknek számos kombinációja létezik (Zenisky, Hambleton és Luecht, 2010).

Az elágazási szabályokhoz szorosan kapcsolódik a modulok és a teljes teszt pontozása. Amint az elágazási szabályoknál látható volt, a tanulók képességi szintje egy-egy szakasz végeztével becslésre kerül. Gyakran a becsült képességpontokat NC (number-correct) pontszámokká konvertálja az algoritmus (Keng, 2008; Yan, von Davier és Lewis, 2014). Azonban, míg a modulok pontozására elegendő az NC pontozás, a teljes teszt pontozására nem megfelelő, mivel a tanulók statisztikailag különböző itemeket kapnak (Zenisky, Hambleton és Luecht, 2010). Ezért a teljes teszt pontozására az item-alapú adaptív teszteknel használatos módszerek alkalmazhatók a többszakaszos teszteknel is, a megfelelő IRT modellt alkalmazva (Keng, 2008).

I.7.3 A többszakaszos tesztek előnyei és hátrányai

Az itemalapú adaptív tesztekkel összehasonlítva a többszakaszos tesztek számos előnnyel rendelkeznek (Magyar és Molnár, 2013). A modulok előre tervezhetőek és szerkeszthetőek, így nagyobb kontrollt biztosítanak a teszt adminisztráció számára. Ezáltal kiküszöbölhetővé válik, hogy az itemek egymásnak információt szolgáltatassanak (Hendrickson, 2007). Különösen előnyös alkalmazásuk a tartalmi korlátozások esetében (Hendrickson, 2007). További fontos előnyük, hogy a modulokon belül a tanulóknak lehetőségük van a visszalépésre és javításra (Zheng, 2012). Mivel adaptivitás csak a modulok között valósul meg, így ez nem veszélyezteti a teszt algoritmusát és segíti a tanulókat a minél magasabb pontszám elérésében (Vispoel, Hendrickson és Bleiler, 2000). Az itemalapú adaptív tesztekhez képest jóval kevesebb adminisztrációt és számítógépes számításokat igényelnek (Hendrickson, 2007; Zheng, 2012).

Előnyei mellett Hendrickson (2007) hangsúlyozza, hogy bizonyos hátrányokkal is rendelkezhetnek a többszakaszos tesztek. Általában több itemre van szükség azonos precizitás eléréséhez. A teszt szerkesztőknek több munkába kerül előállításuk, mivel az itemeken túl azok egymásra hatását is ellenőrizniük kell. A kétszakaszos teszteknel könnyen előfordulhat, hogy a kezdő teszt nagyobb hibával méri a tanulók képességszintjét. További hátránya, hogy a teszt csak az adott modul végén érhet véget, így a teszt hossza kevésbé flexibilis az itemalapú adaptív tesztekhez képest (Zheng, 2012). Ezen hátrányok ellenére, az MST mégis egyensúlyt képvisel a pontosság, adaptivitás, gyakorlati használhatóság és az itemek feletti kontroll tekintetében (Zenisky, Hambleton és Luecht, 2010).

I.8 SZÁMÍTÓGÉPES ADAPTÍV TESZTEK ÉS LINEÁRIS TESZTEK MŰKÖDÉSÉRE VONATKOZÓ NEMZETKÖZI ÖSSZEHASONLÍTÓ KUTATÁSOK

Az adaptív tesztelés hatékonyságának mérését célzó kutatások alapvetően két csoportba sorolhatóak. A kutatások egyik része médiahatásvizsgálatot is magában foglal, mert nem azonos médián kiközvetített adaptív és lineáris tesztelést, hanem a számítógép alapú adaptív tesztelést a papíralapú lineáris teszteléssel hasonlítja össze. Miután a számítógép alapú tesztelésre való átállás sem valósult meg még a mérés-értékelés minden egyes szintjén, ezért indokoltak a médiahatáskutatást is tartalmazó összehasonlító hatékonyságvizsgálatok (Wang és Kolen, 2001; Pásztor-Kovács, Magyar, Hülber, Pásztor és Tongori, 2013; Wan, Keng, McCarty és Davis, 2009). Ezek biztosítják a különböző médián kiközvetített tesztek

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

Adaptív és PP tesztek működésének tesztelméleti összehasonlítása különösen nagy kihívás (Wang és Kolen, 2001). Mivel a tanulók személyre szabott tesztet kapnak, különbségek lehetnek az itemek tartalmában, az itemek elhelyezkedésében és nehézségében, valamint a pontozásban. Ezek a tényezők jelentősen befolyásolhatják az elemzéseket, melyeket a médiahatás mellett szintén javasolt figyelembe venni (Wang és Kolen, 2001; Kolen, 1999-2000). Wang és Kolen (2001) arra hívják fel a figyelmet, hogy ahhoz, hogy a CAT verzió összehasonlítható legyen a papíralapú verzióval, a CAT szempontjából nagymérvű korlátozottságot jelent, hiszen ez esetben az adaptív tesztfejlesztés során nem lehet az összes, a számítógép adta lehetőséget kihasználni. Az ezredforduló utáni legjelentősebb, adaptív tesztelésre vonatkozó összehasonlító kutatásokat összegzi az 1. táblázat a bevont minta, a vizsgált mérőanyag jellege, az elemzés során alkalmazott eljárások, módszerek, valamint az eredmények szerint. A táblázat célja, hogy a vizsgált különböző szempontok (minta mérete, mérőanyag, elemzési módszerek, eredmények) szerint összegezze az ismertetett kutatásokat.

1. táblázat: Adaptív teszteken végzett összehasonlító vizsgálatok

Publikáció	Minta	Vizsgált mérőanyag	Elemzési módszer	Eredmények
Vispoel, Hendrickson és Bleiler (2000)	242 egyetemi hallgató	szókincsvizsgálat: 40 itemű CAT és PP teszt	átlagok összehasonlítása	A PP teszt átlagai magasabbak voltak, mint a CAT-on elért átlagok.
Olea, Revuelta, Ximénez és Abad (2000)	184 egyetemi hallgató	szókincsvizsgálat: 20 itemű PP és 20 itemű CAT	ANOVA, teszt információ, standard error	A PP teszteken jobb eredményeket értek el a tanulók, a mérési hiba kisebb volt az adaptív teszt esetében
Rotou, Patsula,	10 000 fős szimulált adatbázis	55 itemű PP és 54 itemű MST	reliabilitás, standard error	Az MST mért a legpontosabban, a legkisebb mérési hibával.

Az 1. táblázat folytatása

<i>Manfred és Rizavi (2003)</i>		33 itemű MST és 32 itemű itemalapú CAT		
<i>Kingsbury és Hauser (2004)</i>	NCLB 2003-as mérés eredményei alapján szimulált adatbázis	CAT és lineáris CB matematika és szövegértési tesztek	teszt információ, standard error	Az adaptív teszt minden képességszinten több információt szolgáltatott, mind a lineáris teszt
<i>Hambleton és Xing (2006)</i>	szimulált adatbázis	MST, CAT és lineáris random kiosztású CB teszt	reliabilitás, teszt információ	A tanulók képességszintjéhez igazított tesztek pontosabban mértek, a legpontosabban a CAT mért. Legrosszabb reliabilitás a random teszt kiosztás esetén volt.
<i>Jodoin, Zenisky és Hambleton (2006)</i>	5000 fős szimulált adatbázis	60 itemű MST, 40 itemű MST, 60 itemű lineáris CB tesztváltozat	valódi és a kapott képességpontok korrelációja, reliabilitás, teszt információ	Az MST tesztek közel azonos reliabilitásúak voltak, mindkettő pontosabban mért, mint a lineáris teszt.
<i>Thompson és Way (2007)</i>	szimulált adatbázis	20-60 itemű CAT, 35 itemű PP teszt	valódi és a kapott képességpontok korrelációja, reliabilitás, teszt információ	Adaptív tesztekkel nagyobb mérési precizitás érhető el, és nagyban csökkenthető az itemszám (60-ról 55-re).
<i>Al-A'ali (2007)</i>	45 tanuló	14-20 itemű CAT, 25 itemű PP	teszt információ	Az adaptív teszt esetén kevesebb item elegendő ugyanazon mérési pontosság eléréséhez.
<i>Keng (2008)</i>	10000 fős szimulált adatbázis	42 itemű tesztlet szintű CAT, item szintű tesztlet CAT, MST	valódi és kapott képességpontok közötti eltérés, AAD, RMSE, standard error, teszt információ	A legprecízebben az item szintű CAT mért, legrosszabb pontossággal a tesztlet szintű CAT. Alacsony képességűek esetében a CAT mért legprecízebben, átlagos képességűeknél az MST.
<i>Jiban, Ayodele, McCarthy, és Christ (2008)</i>	287 kisiskolás	20 és 40 itemből álló olvasáskészséget mérő itemalapú CAT	reliabilitás, teszt információ, standard error	20 item fölött nem változott jelentősen a teszt információ és a mérési hiba nagysága.
<i>Pyper és Lilley (2010)</i>	180 egyetemista	20 itemű CAT, 20 itemű lineáris BC teszt	korreláció, t-próba	A teljesítményekben magas korreláció a két teszt között.
<i>Guille, Becker, Zhu, Zhang, Song és Sun (2011)</i>	6287 fős szimulált adatbázis	MST és CB teszt	reliability, standard error	Elsősorban a magas képességű egyéneknél mérséklődött szignifikánsan a tesztelési idő
<i>Frey, Seitz és Kröhne (2011)</i>	14624 PISA teszt eredménye alapján szimulált adatbázis	PISA 2000, 20003 és 2006-os tesztek alapján PP és CAT szimuláció	reliabilitás	Az eredmények szerint a mérés hatékonysága (Mérési precizitás/prezentált itemek száma) 74%-kal nőtt, a szükséges itemszám a PP tesztnél szükséges 55-ről 26-ra csökkent, és a tesztelés időtartama 120 percről 57 percre csökkent.
<i>Zheng (2012)</i>	szimulált adatbázis	600 itemű feleltválasztós	teszt információ, RMSE, korreláció	Az MST mérési precizitása megegyezett a CAT

		kérdések nyolc különböző területről, CAT, különböző szerkezetű MST-k és lineáris CB		precizitásával, az MST hatékonyabb volt az itembank kihasználtságát tekintve.
<i>Crotts, Zenisky, Sireci és Li (2013)</i>	szimulált adatbázis	35 és 40 itemű olvasás-szövegértés MST	reliabilitás, teszt információ, standard error	A tesztrövidítés nem változtatott jelentősen a teszt reliabilitásán, a tesztelés időtartama 12-24%-kal csökkent
<i>Brossman és Guille (2014)</i>	szimulált adatbázis	MST és lineáris CB teszt	teszt információ, standard error, RMSE	A mérési hiba kisebb az MST esetén

A 16 kutatás közül 13 esetben végezték számítógépes adaptív és lineáris teszt összehasonlítását, melyek közül öt esetben itemalapú adaptív és papíralapú lineáris (*Vispoel, Hendrickson és Bleiler, 2000; Olea és mtsai, 2000; Al-A'ali, 2007; Thompson és Way, 2007; Frey, Seitz és Kröhne, 2011*), két esetben itemalapú adaptív és számítógép alapú lineáris (*Kingsbury és Hauser, 2004; Pyper és Lilley, 2010*), egy esetben MST és papíralapú lineáris (*Rotou és mtsai, 2003*), két esetben itemalapú adaptív, MST és számítógép alapú lineáris (*Hambleton és Xing, 2006; Zheng, 2012*) és három esetben MST és számítógép alapú lineáris (*Jodoin, Zenisky és Hambleton, 2006; Guille és mtsai, 2011; Brossman és Guille, 2014*) tesztek működésének összehasonlítása történt. Négy esetben különböző típusú adaptív tesztek működését hasonlították össze a kutatók (*Rotou és mtsai, 2003; Keng, 2008; Jiban és mtsai, 2008; Crotts és mtsai, 2013*).

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

A kutatások jelentős része a különböző összeállítású, szerkezetű MST tesztek működését, mérési hatékonyságát, pontosságának összehasonlítását vette górcső alá. Az MST tesztrendszerek első modulja általában egy közepes nehézségű modul, amit 2-5 ágon 2-6 szakasz követ. Az eredmények szerint a szakaszok és modulok számának növelése növelte a teszt mérési precizitását, ezért egyértelműen kettőnél több szakaszt javasoltak a kutatások,

ami kiküszöbölte a tanulók esetlegesen hibás szintre történő besorolását is. Túl sok szakasz alkalmazása azonban indokolatlanul megnövelte a teszt hosszát és az adatfelvétel idejét, miközben ez nem járt a mérési precizitás arányos növekedésével. A bemutatott kutatásokban az 1-3 (Rotou és mtsai, 2003), 1-2-3-4, 1-2-4 (Zheng, 2012), 1-3-3 (Keng, 2008), 1-2-2, 1-3-3, 1-2-3, 1-3-2 (Jodoin, Zenisky és Hambleton, 2006), 5-5-5-5-5-5 (Crotts és mtsai, 2013) és 1-3-3-3-3 (Brossman és Guille, 2014) szerkezetű MST tesztek fordultak elő. A kutatások egy részében, amennyiben az alkalmazott itembank mérete engedte, a tesztbiztonság növelése érdekében több ekvivalens tesztváltozatot, illetve modult állítottak össze, melyeket random módon osztottak ki (Brossman és Guille, 2014; Crotts és mtsai, 2013).

A korai elemzésekben a teszteredmények összehasonlítását klasszikus tesztelméleti módszerek alkalmazásával végezték (ANOVA elemzések, átlagok összehasonlítása; Vispoel, Hendrickson és Bleiler, 2000; Olea és mtsai, 2000) a későbbiekben azonban általánossá vált a valószínűségi tesztelméleti módszerek, eljárások alkalmazása, mint például az item- és tesztinformációk összehasonlítása. A tesztek mérési precizitását jellemzi a reliabilitás és a mérési hiba (SE - *standard error*). A szimuláción alapuló kutatásokban gyakran használt mutató a valódi és a mért képességpontok korrelációs mérőszáma, valamint az RMSE (*Root Mean Square Error*) indexek, illetve AAD (*Average Absolute Difference*), melyek a valódi és a mért képességpontok eltéréseinek jellemzésére alkalmas mutatószámok (Keng, 2008). Jiban és munkatársai (2008) a különböző tesztelési módokban elért helyes válaszok arányát vizsgálták és hasonlították össze.

A kutatási eredmények szerint a kétféle teszten elért képességszintek között a korábbi kutatások szerint jelentős eltérések lehetnek. Olea és mtsai (2000) kutatási eredményei szerint a tanulók a lineáris teszten átlagosan 0,95, az adaptív teszten 0,58 képességpontokat értek el, azaz jelentős mértékű eltérés realizálódott a két tesztelési eljárással kivitelezett adatfelvétel során. Az adaptív tesztek alkalmazásával jelentősebb mérési precizitás volt elérhető, magasabb volt a tesztelés reliabilitása és minden egyes képességszinten több információt szolgáltatott, mint a lineáris tesztek. Jodoin és mtsai (2006) eredményei szerint az információ mértékének 25%-os esése a tanulók 1-1,5% ának téves szintre való besorolását eredményezheti, további 50%-os esés 3%-nyi tanuló esetén is jelenthet hibás minősítést. Thompson és Way (2007) kutatásai is megerősítik ezt, eredményeik szerint elsősorban az alacsony és a magas képességtartományokban volt jelentős a tanulók adaptív teszt általi precízebb besorolása. Crotts és mtsai (2013) eredményei szerint 9113 tanuló közül 91-137

esetben változott a tanulók besorolása, azaz a tesztelésben résztvevők 1-1,5%-ának eredményeire volt hatással a mérés pontossága.

Crotts és munkatársai (2013) többféle adaptív teszten vizsgálták a kinyerhető információ mennyiségét, és átlagosan 36-62% információ nyertek ki 40 itemű MST tesztekéből. *Rotou* és munkatársai (2003) szerint a magas képességtartományban volt nagyobb a kinyert információ adaptív teszten, mint a PP teszten, *Thompson* (2007) szerint az alacsony és a magas képességtartományokban volt jelentős az eltérés.

A tesztinformáció mellett a mérési precizitás másik fontos mutatója a becsült hiba mértéke. *Crotts* és munkatársai (2013) adaptív teszten a becsült hiba értékére 0,16-0,24 közötti értékeket kaptak. *Olea* és munkatársai (2000) az adaptív teszten 0,25, a lineáris teszten 0,31 SE értékeket mértek. *Rotou* és mtsai (2003) az adaptív teszten 0,78-2,90, a lineáris teszten 0,98-3,10 közötti értékeket kapott, és az adaptív teszten mért hiba minden képességtartományban kisebb volt, mint a PP tesztnél. *Thompson* (2007) elsősorban az alacsony és a magas képességtartományokban mért jelentős eltérést. *Kingsbury* és *Hauser* (2004) 0,3 alatti SE értékeket tartott elfogadhatónak.

Az adaptív tesztek alkalmazásával csökkenthető volt a kiközvetített itemek száma (*Keng*, 2008), ami a mérési idő csökkenését vonta maga után. Ezek az előnyök mindkét típusú (itemalapú és MST) esetén kimutathatóak voltak, azaz bármelyik típusra való átállás vonzó lehet. A mérési eredmények szerint adaptív tesztelésnél az alacsonyabb képességtartományban jelentősen növekedett a helyes megoldások száma, a magas képességű személyeknél viszont jóval kevesebb helyes válasz született, mivel az alacsony képességű tanulók könnyebb feladatokat, a magasabb képességűek viszont nehezebbeket kaptak (*Jiban* és mtsai, 2008). *Olea* és mtsai (2000) kutatása szerint az adaptív teszten átlagban 13,39 volt a helyes válaszok száma, a lineáris teszten 11,72.

A továbbiakban részletesebben kitérünk azon négy kutatás bemutatására, amelyek a disszertációban bemutatott kutatásokhoz hasonlóan az adaptív tesztek különböző típusait lineáris tesztekkel hasonlították össze. *Rotou* és munkatársai (2003) az MCAT verbális érvelés tesztjét használták fel vizsgálatukban, összesen 440 item (64 feladat) alkotta az itembankot. 32 itemű CAT és 33 itemű MST, valamint 55 itemű PP és 54 itemű MST eredményét hasonlították össze, mindegyik tesztet egyenként 1, 2 és 3 paraméteres modellel kalibrálták. Mindkét MST változatnál az 1-2 szerkezetet használták, vagyis egy átlagos nehézségű kezdő

modult követően háromfelé ágazott a rendszer, könnyű, közepes vagy nehéz modul következett. A tesztbiztonság érdekében kétféle tesztváltozat készült, vagyis összesen nyolc modul alkotta a tesztet. A választott tesztkonstrukciót az itembank kis mérete (64 feladat) indokolta. A mérési pontosságot tekintve az MST kissé jobban mért az 1 és 2 paraméteres modellek esetében, mint a CAT és a PP teszténél, és a hosszabb MST reliabilitása is magasabb volt, mint a rövid, 33 iteműé. A helyes és helytelen válaszok közötti eltérések hasonlóak voltak mindegyik teszt esetében. Az itemek kiválasztását tekintve az MST esetén az itemek 51%-a kevesebb, mint 10%-ban választódott ki, és 24%-a több mint 50%-ban, míg a CAT esetén az itemek 70%-ának a kiválasztódási aránya 10% alatti volt, és nem volt olyan item, mely 30% fölött szerepelt volna. A kiválasztási arány az átlagos itemek körében volt a legmagasabb. A mérési eredményeket összevetve *Rotou* és munkatársai (2003) szerint a 33 itemű, 2 paraméteres modellel kalibrált MST a leginkább ajánlott CAT típus.

Keng (2008) szintén MST és itemalapú CAT tesztek viselkedését vizsgálta. Tanulmányában tesztlet alapú CAT, item-alapú CAT és 1-3-3 struktúrájú MST tesztek vizsgált szimulációs módszerrel. Az itemek kalibrációja 3 paraméteres modellel történt. A kísérlet során a tesztek hosszát, az itembank méretét és a képességszintet változtatták. Mérési pontosságát tekintve mindhárom teszt hasonlóan magas precizitással mért, de az item szintű CAT a legjobban, a legkisebb standard hibával. A teszt rövidítése mindhárom teszténél csökkentette a precizitást, de az item-szintű CAT alkalmával volt a legkisebb hatása, ez a teszt mért továbbra is a legpontosabban. Az itembank méretének csökkentése nem okozott jelentős eltérést egyik teszt esetében sem. A képességszint változtatása mindhárom tesztre negatív hatással volt, de az MST-nél csökkent legkevésbé. Az itembank felhasználását illetően a teszt rövidítése pozitív hatással volt a tesztlet szintű CAT-re, jelentősen romlott az item-szintű CAT esetén, és nem mutatott lényeges eltérést az MST-nél.

Zheng (2012) különböző típusú MST tesztek hatékonyságát vizsgálta itemalapú adaptív és lineáris CB tesztekkel összevetve. A vizsgálat során 600 itemből álló feleletválasztós itemeket tartalmazó itemeket használt fel, melyet háromparaméteres modellel kalibráltak. Az itemek nyolc különböző területet fedtek le három nehézségi szinten, a kategóriák közötti nehézségi eloszlásuk egyenletes volt. A vizsgált MST tesztek 1-2-3-4, illetve 1-2-4 szerkezetűek voltak és összesen 21 itemet tartalmaztak. A lineáris teszt 30 itemű, az itemalapú CAT 21 itemű volt. A szimulációs kísérlet 5000 fős adatbázison történt. A mérés során az RMSE értékeket, a teszt információkat és a különböző tesztváltozatok közötti korrelációkat hasonlították össze.

A lineáris teszthez képest bármelyik adaptív változat megbízhatóbb mérési eredményeket szolgáltatott annak ellenére, hogy a lineáris teszt 30%-kal hosszabb volt. A legkisebb RMSE érték az itemalapú tesztnél fordult elő. Összegezve, a legjobb mutatókat az itemalapú adaptív teszten mérték, viszont az MST esetén a 30%-os rövidülés ellenére is nagyobb mérési precizitás volt elérhető, mint a lineáris változatnál.

Az OECD a PISA mérések terén 2015-től tervezi az adaptív modulok bevezetését, mely számos előnnyel járhat. Egyrészt csökken a mérés időtartama, az eredmények azonnal elérhetővé válnak, és nő a tanulóktól begyűjthető információk mennyisége. Az átállást azonban több tényező nehezíti (*Frey, Seitz és Kröhne, 2011*):

- A PISA itembank nem optimális adaptív tesztelésre: kevés az itemszám és nagyrészt közepes nehézségű itemeket tartalmaz. Az itemek 49%-a nyílt és rövid válasz formátumú, mely alkalmatlan közvetlen számítógépes értékelésre.
- Az itemek 54%-a hogonyitem, melyek már szerepeltek előző PISA vizsgálatokban, és ezek teszik lehetővé az előző mérésekkel való összehasonlítást. Amennyiben ezeket az itemeket mellőzni, vagy átalakítani szükséges, úgy veszélybe kerülnek a trendjelentések.
- Az itemeknek csak 13%-a egyedülálló, a többi csoportokat, tesztleteket alkot, melyek csak összefüggően prezentálhatóak.
- A tesztben meghatározott százalékban szerepelnek a különböző területekhez (matematika, természettudomány, szövegértés) tartozó itemek.

Frey, Seitz és Kröhne (2011) szimulációs kísérletben próbáltak megoldást találni a felsorolt problémákra, és javaslatot tenni adaptív tesztelés valamilyen típusának bevezetésére. Az eredmények szerint adaptív teszteléssel a mérés hatékonyságának (mérési precizitás/prezentált itemek száma) akár 40%-os növekedését el lehetne érni, a szükséges itemszám a PP tesztnél szükséges 55-ről 26-ra csökkenhetne, és a tesztelés időtartama is 120 percről 57 percre csökkenne. Ezen eredmények alapján a PISA Governing Board a többszakaszos adaptív tesztek részleges bevezetését javasolta a 2015-ös méréstől kezdődően.

Összességében a nemzetközi kutatások eredményei szerint az adaptív tesztek reliabilitása a lineáris tesztekénél magasabb. A kinyerhető információk mértéke szintén magasabb, a mérési hiba viszont alacsonyabb, és ezáltal a lineáris tesztekénél jóval pontosabb

mérést tesznek lehetővé. A legtöbb vizsgálat eredménye szerint a többszakaszos teszt precizitása valamivel elmaradt az item, illetve a tesztlet alapú tesztekétől, viszont a tesztfejlesztés folyamán alkalmazható nagyobb adminisztratív kontroll előnyössé teszi ezt a tesztípust, ezért papíralapú tesztelésről való átállás esetén a legpreferáltabb típusú adaptív tesztkonstrukció.

II. AZ EMPIRIKUS VIZSGÁLATOK KONCEPCIÓJA

II.1 CÉLOK

A kutatás fő célja az adaptív tesztelés hatékonyságának vizsgálata hagyományos, lineáris tesztelési móddal való összehasonlítása során az 1-8. évfolyamos tanulók körében.

A vizsgálat alcéljai:

- 1) a korábban papír alapon alkalmazott tesztek online formára konvertálása;
- 2) a számítógép alapú alapú itemekből osztálytermi környezetben használható lineáris, illetve adaptív tesztrendszerek kialakítása;
- 3) az adaptív és lineáris tesztek mérési pontosságának összehasonlítása;
- 4) a becsült képességszintek évfolyam és személyszintű összehasonlítása;
- 5) a kétféle tesztkörnyezetben elért helyes válaszok arányának összehasonlítása;
- 6) az adaptív tesztelés során kiosztott itemek, illetve résztesztek nehézségi szintjének, ennek változásmintázatainak jellemzése;
- 7) a lineáris és az adaptív tesztelés során kinyert információ és a mérési hiba nagyságának összehasonlítása képességszint szerinti bontásban.

II.2 KUTATÁSI KÉRDÉSEK

A felvázolt célok alapján a következő kutatási kérdésekre keressük a választ:

- 1) A korábban papír alapon alkalmazott tesztek átkonvertálhatóak-e online tesztformátumra?
- 2) A számítógép alapú alapú itemekből kialakítható-e osztálytermi környezetben megbízhatóan alkalmazható adaptív tesztrendszer?
- 3) Az adaptív tesztrendszerek pontosabb mérést tesznek-e lehetővé, mint a lineáris tesztek?
- 4) Van-e különbség az adaptív, illetve lineáris tesztekkel becsült képességszintek között évfolyam, illetve személyszinten?
- 5) Hogyan alakul a kétféle tesztkörnyezetben a tanulók által elért helyes válaszok aránya?

- 6) Mely nehézségi szintű itemek/résztesztek szerepelnek leggyakrabban az adaptív teszt kiosztás során?
- 7) Hogyan alakul a kinyert információk és mérési hibák mértéke a kétféle tesztkörnyezetben?

II.3 HIPOTÉZISEK

- 1) a korábban papír alapon alkalmazott tesztekből kifejleszhető többszakaszos adaptív tesztelésre alkalmas tesztrendszer;
- 2) az adaptív tesztrendszerek hatékonyan és megbízhatóan alkalmazhatóak az 1-8. évfolyamos korosztály diagnosztikus mérésére;
- 3) az adaptív rendszer a képességek pontosabb mérését teszi lehetővé;
- 4) a becsült képességszintek nem különböznek jelentős mértékben online adaptív és lineáris tesztelés esetén;
- 5) adaptív tesztelés esetén az alacsonyabb képességtartományban magasabb a helyes válaszok aránya, mint lineáris tesztkörnyezetben, átlag feletti tanulóknál viszont fordítva, kisebb arányban fordulnak elő helyes válaszok, ezért az adaptív tesztelés több sikerélményt nyújt az alacsonyabb képességű tanulóknak, és nagyobb kihívást a magas képességszintű tanulók számára;
- 6) a tanulók többsége átlagos képességszintű, ezért adaptív teszt kiosztás esetében az átlagos nehézségű itemek/résztesztek szerepelnek leggyakrabban;
- 7) a tesztek közül kinyert információ minden képességszinten szignifikánsan magasabb, a mérési hiba viszont szignifikánsan alacsonyabb az adaptív tesztrendszer esetében, mint lineáris tesztekénél.

II.4 A VIZSGÁLATSOROZAT MENETE

A hatékonyságvizsgálatokat pilotmérések és nagymintás mérések alkalmazásával valósítottuk meg 2012 és 2014 között több részletben. Három pilotvizsgálatot és két nagymintás vizsgálatot folytattunk le, melyeket különböző képességeket mérő, korábban papír alapon alkalmazott tesztek itemeinek felhasználásával állítottuk össze. Az első pilotvizsgálat induktív gondolkodást, a második problémamegoldó képességet mérő tesztek

felhasználásával történt. A harmadik mérés alkalmával szóolvasás készséget mérő tesztrendszert konvertáltunk online formára, melynek működését először kismintás, majd ezt követően nagymintás méréssel teszteltük.

A két pilotmérés (induktív gondolkodás és problémamegoldás) alkalmával a papír alapon bemért paramétereket használtuk az adaptív tesztváltozatok összeállításánál, a szóolvasás mérése esetén viszont külön nagymintás méréssel paramétereztük az itemeket, ezt követően állítottuk össze az adaptív tesztrendszert és valósítottuk meg a hatékonyságmérő nagymintás mérést. A mérések során az adaptív teszt mellett ugyanabból az itembankból összeállított lineáris tesztváltozat biztosította a mérések összehasonlíthatóságát. Az alkalmazott adaptív tesztek minden esetben többszakaszos tesztek voltak, az itembank méretétől és összetételétől függően különböző szerkezeti struktúrában.

II.5 A VIZSGÁLAT SORÁN HASZNÁLT MÉRŐESZKÖZÖK BEMUTATÁSA

A vizsgálathoz felhasznált mérőeszközök olyan képesség-, illetve készségterületek mérésére alkalmasak, melyek kiemelt szerepet játszanak az általános iskolai korosztály készség, illetve képességfejlődésében. Mindhárom mérőeszköz eredetileg papír alapon került kidolgozásra, és több alkalommal használták őket nagymintás mérések keretében a vizsgált korosztály készség, illetve képességmérésére, melyek során a tesztek nagyon megbízhatóan működtek, mindegyik teszt megbízhatósági mutatója 0,80 Cronbach-alpha felett volt. A papíralapú mérések alapján a tesztek minden iteme paraméterezésre került, a nehézségi paraméterek lefedték a vizsgált korosztály képességskáláját, ezért ezeket a paramétereket használtuk a kismintás mérések alkalmával is. A felhasznált tesztek kiválasztásának további indoka, hogy ezek a tesztek tartalmaztak elég itemet az adaptív teszteléshez megfelelő méretű itembank kialakításához. További fontos szempont volt, hogy az itemek számítógépes formátumra való konvertálása csekély változtatással megoldható volt, ezáltal a médiahatás nem befolyásolta jelentős mértékben a tesztek validitását.

Az induktív gondolkodást és a problémamegoldó képességet mérő tesztek online adaptív formára történő konvertálásakor az eredeti, papír alapon bemért paraméterek alapján állítottuk össze a tesztek. A harmadik mérés során a mérési precizitás növelése érdekében online paraméterezést hajtottunk végre az itembank összes itemére vonatkozóan, és az így kapott paraméterek alapján állítottuk össze a pilot, illetve nagymintás mérés során használt

adaptív tesztrendszer. Az induktív gondolkodást mérő teszt esetében 1-3-3-3 szerkezetű MST tesztkonstrukciót alkalmaztunk, a problémamegoldó képesség esetében az 1-2-3 szerkezetű tesztekre esett a választásunk, mivel az előzetes szakirodalmi kutatások során kisebb itembankok esetén ezek megvalósítása tűnt a legmegfelelőbbnek. A harmadik kutatásnál a szóolvasó készség mérésekor az itembank nagy mérete és az eredeti tesztstruktúra komplikáltabb szerkezet alkalmazását indokolta, ezért ebben az esetben az 1-4-5-5 szerkezetű MST mellett döntöttünk.

II.5.1 Induktív gondolkodást mérő teszt

Az induktív gondolkodás az iskolai tanulás kognitív folyamatai között kiemelkedő helyet foglal el (Csapó, 1994; Csapó és Molnár, 2012). A megismerés szinte minden szintjén jelen van, a gondolkodási mechanizmusokban vezető szerepet tölt be (Csapó, 2001). Fejlettségi szintje jellemzi az új tudás megszerzésének módját, ami által megmagyarázhatóvá válik, hogy hogyan juthatunk el olyan új következtetésekre, amelyek túlmutatnak a közvetlen tapasztalaton (Csapó, 2002). A pszichológiai megközelítések szerint az indukció a megismerés egyik alapvető módja, az új tudás megszerzésének eszköze (Csapó, 2002). Mivel jól jellemzi a tanulók intellektuális fejlettségét, szerepe ezért kimagasló az általános iskolai korosztály gondolkodási folyamatainak feltárása terén, és mérése az utóbbi évtizedek kiemelkedő fontosságú kutatási témája (Molnár és Csapó, 2011).

Az induktív gondolkodásnak számos összetevője van, melyek közül a legtöbb kutató kisebb-nagyobb eltérésekkel ugyanazokat hangsúlyozza. A legrészletesebben kidolgozott rendszer Klauertől (1990) származik (Csapó és Molnár, 2012). Az induktív gondolkodást szabályszerűségek és rendellenességek megtalálásának tekintette olyan módon, hogy tulajdonságokat és relációkat összehasonlítva hasonlóságokat, különbségeket, továbbá együttesen megjelenő hasonlóságokat és különbségeket különböztetett meg. Így az induktív folyamatokat két nagy területre osztotta fel: a tulajdonságokkal és a relációkkal kapcsolatos indukciók, melyeket további három-három induktív folyamatra bontott. A tulajdonságokhoz az általánosítás (generalizáció), a megkülönböztetés (diszkrimináció) és az osztályozás (klasszifikáció) tartozott, a relációkkal kapcsolatos indukciók a kapcsolatok felismerése, kapcsolatok megkülönböztetése és a rendszeralkotás voltak. Ezáltal rendszerében hat alapvető induktív folyamatot azonosított, melyekhez számos fejlesztő tréninget dolgoztak ki különböző életkorú tanulók számára (Klauer, 1990; Molnár, 2006a).

Az induktív gondolkodás mérésére elsősorban a feleltválasztók technikákat alkalmazzák, analógiák, sorozatok, mátrixok alapján kell a tanulóknak a szabályosságokat felfedezni és az alapján a hiányzó elemet azonosítani. A szakirodalomban fellelhető tesztekre alapozva *Csapó Benő* széles életkori intervallumot átfogó, csoportos adatfelvételre alkalmas teszteket dolgozott ki. A feladatok tartalmául számokat, szavakat és betűket választott. A hat tesztből álló rendszer összesen 104 feladatból állt. A feladatok szám- és szóanalógiákat, átkódolást, számsorozatokat, betűsorozatokat és kizárás teszt itemeket tartalmaztak (13. ábra; *Csapó, 2002*). A szóanalógiák és a kizárás feladatok zárt típusúak voltak, a tanulók öt, illetve hat válaszlehetőség közül választották ki a helyes megoldást, a számanalógiák, a számsorok, a betűsorok és az átkódolás tesztek pedig nyílt végű feladatokat tartalmaztak, a tanulóknak kellett a helyes választ beírniuk.

Számsorok

3 6 11 14 19 22 _____

Számok analógiája

20 → 32 8 → 20 11 → _____

Szóbeli analógiák

SZÉK : BÚTOR = KUTYA : ?

a MACSKA b ÁLLAT c TACSKÓ d ASZTAL e KUTYAÓL

Betűsorok

a c e g i k m _____

Átkódolás

Minta: hétfő + szerda = csütörtök; kedd + csütörtök = szombat

Feladat: szerda + csütörtök = _____

Kizárás

a SÁL b CIPŐ c KALAP d SZÉK e TRIKÓ f KESZTYŰ

13. ábra: Mintafeladatok az induktív gondolkodást mérő tesztből (Csapó, 2002 alapján)

A számsorozatok teszt feladataiban a tanulóknak egy megkezdett számsort kellett kiegészíteni két további taggal. A sorozatok különböző nehézségűek voltak, a nehezebb feladatokban komplikáltabb szabályok is előfordultak, mint például a számok közötti különbség két növekvő hatványa, vagy két különböző szabály szerint összekapcsolt számsor.

A számanalógiás feladatokban két számpár volt valamilyen összefüggés szerint összekötve, és ennek megfelelően kellett a harmadik szám párját képezni. A könnyebb feladatokban a számpár második tagja egyszerűen valamennyivel növelt értéke az első számnak, a nehezebb itemeknél bonyolultabb lineáris összefüggésekre kellett rájönni és alkalmazni a harmadik számpár megalkotása során.

A szóbeli analógiák esetében a számanalógiákhoz hasonlóan valamilyen összefüggést kellett felfedezni az első pár két szava között, és ezt alkalmazva kellett kiválasztani a megadott öt szó közül a megfelelőt. Az összefüggés lehetett az egy halmazba tartozás, a rész-egész viszony, az időrend, az ok-okozat kapcsolat, a szinonima, az ellentét, a tulajdonság, a funkció, az átalakulás (valamiből valami lesz) stb.

A betűsorok feladatok sorrendiségi szabályokon alapultak, például minden második betű, vagy a nehezebb feladatoknál több, egymásba ágyazott sorozat. Az átkódolás-teszt feladatainál a hét napjait kellett számoknak megfeleltetni, és ezt az összefüggésrendszert alkalmazni.

A kizárás teszt feladatainál egy hat elemből álló halmazból kellett az oda nem illő elemet kiválasztani. A halmazok között szerepeltek konkrét (pl. ruhadarab, folyadék) és absztrakt halmazképző koncepciók (pl. emberi kapcsolatok, tulajdonságok) is.

Az alsóbb évfolyamokra *Molnár Gyöngyvér* (Molnár, 2006, 2008, 2009) dolgozta ki a *Klauer* rendszerén alapuló hat résztesztet. A feladatok képeket, rajzokat ábrázoltak, minimalizálva az olvasás szükségességét (14. ábra).

Karikázd be azt a három alakzatot, amiben van valami közös és különbözik a többitől!



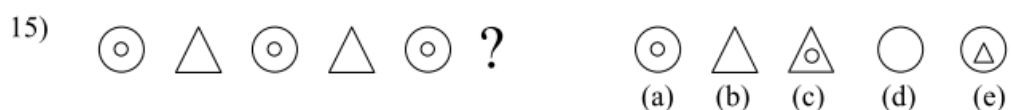
Karikázd be azt az alakzatot, amelyik leginkább illik az előtte lévőkhöz!



Karikázd be, melyik a kakukktojás?



Folytasd a sorozatot! Karikázd be, amelyik a kérdőjel helyére illik!



14. ábra: Mintafeladatok az induktív gondolkodást mérő tesztből (Molnár, 2006 alapján)

A kifejlesztett tesztek 1993-2008 között több mérés alkalmával vizsgálták országos nagymintás reprezentatív mintán az 1-11. évfolyamos tanulók körében, melyet a papíralapú adatfelvétel mellett számítógépes alapú online adatfelvétel is kiegészített (Molnár és Csapó, 2011). A mérések alkalmával a tesztek Cronbach alpha reliabilitása igen magas volt (0,91-0,97; Csapó, 2001), vagyis megbízhatóan alkalmazhatóak a vizsgált korosztály mérésére. A valószínűségi tesztelméleti elemzésekkel közös képességskálára hozott személy-item térkép szerint az itemek nehézségi szintje közel lefedte az 1-11. évfolyamos minta átlagos képességszintjét. A nehézségi indexek -2,31 és 3,73 logit között voltak, a tanulók átlagos személyparaméterei évfolyamonkénti bontásban -1,16 (sd=1,19 – 1. évfolyam) és 1,37 (sd=1,17 – 11. évfolyam) között mozogtak. Az itemek nemcsak a teszt szintjén, hanem az item szintjén is jól diszkrimináltak a tanulókat, a 104 item diszkriminációs indexének átlaga 0,48 (sd=0,10) volt (Molnár és Csapó, 2011).

II.5.2 Problémamegoldó képességet mérő teszt

A problémamegoldó gondolkodás a 21. századi kulcsképeségek között vezető helyet foglal el (Molnár, 2006c; Csapó és Molnár, 2012). Fejlesztése és mérése az utóbbi évtizedekben az egyik leggyakrabban előforduló kutatási téma a mérés-értékeléssel foglalkozó nemzetközi mutatókban (pl. OECD PISA mérések; OECD, 2007; NAEP kutatások; Bennett, Persky, Weiss

és Jenkins, 2007). A 2003-as PISA mérésben már külön területként szerepelt a komplex problémamegoldás (Csapó és Molnár, 2012; Molnár, 2006c).

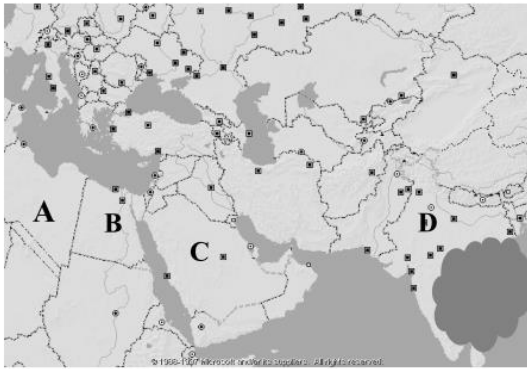
A szakirodalomban számos definíció létezik a problémamegoldó képesség leírására, sok kutatás kapcsolatba hozza az intelligenciával (Molnár, 2012). A PISA megközelítés szerint

„a problémamegoldás az egyén képessége arra, hogy kognitív eljárásokat használjon olyan reális, diszciplinákat átmetsző (cross-disciplinary) helyzetekben, amikor a megoldáshoz vezető út nem válik azonnal nyilvánvalóvá, és amikor a műveltségi területek vagy tantervi tartalmak, amelyek esetleg felhasználhatók, nem találhatók meg a matematika, az olvasás vagy a természettudomány egyetlen területén belül” (OECD, 2003, 156. o. idézi Csapó és Molnár, 2012, 416.o.).

A PISA és a korábbi komplex problémamegoldó képesség fejlettségét mérő kutatásokra alapozva (Csapó és Molnár, 2012) fejlesztett ki Molnár Gyöngyvér több olyan komplex problémamegoldó gondolkodást vizsgáló feladatlapot, melyek tág életkori határokon belül (3-11. évfolyam) alkalmazható a hétköznapi életben felmerülő problémák megoldásának képességmérésére. A történetben felmerülő problémák azt vizsgálták, hogy hogyan tudják a tanulók az iskolában, elsősorban matematika órákon megszerzett tudást életszerű helyzetekben alkalmazni, valamint mennyire tudják az iskolai tudást transzferálni gyakorlati, iskolán kívüli helyzetekre (Molnár, 2006b). A feladatlapok kerettörténetbe ágyazottan jelenítették meg a problémákat, melyet egy megfelelő korú testvér mesélt el. A tanulóknak a történetek elolvasása során kellett a felmerülő problémákat megoldaniuk, illetve a felmerülő kérdésekre válaszolniuk. A feladatlapokon szereplő problémák három csoportra oszthatók: az első csoportba tartoztak azok a feladatok, amikor a megoldáshoz szükséges minden információt tartalmazott a feladat. A második csoportba tartozó feladatok esetében nem volt minden információ megtalálható a feladatlapon, viszont a hiányzó információ az adott korosztály matematika tananyagában szerepelt. A harmadik típusú feladatok esetében sem a feladat, sem a tananyag nem nyújtott biztos támpontot a megoldáshoz, viszont az elvárt háttérinformációkkal a tanulók nap, mint nap találkozhattak a hétköznapi életben (Molnár, 2006c; Csapó és Molnár, 2012).

A feladatlapok oldalait függőlegesen kettébontották, a lap bal oldalán jelentek meg az információk realiztikus formában (pl. kép, hirdetés, étlap, árfolyam), a jobb oldalán pedig a történet volt olvasható, a megfelelő problémákkal beágyazottan (15. ábra; Molnár, 2006b).

úgy kezdődött, hogy...



Teljesült az álmom...

Szülinapomra kaptam egy bőrkötésű üres könyvecskét, ez lett később a titkos naplóm...

Este a család (apa, anya, nővérem bátyám és én) úgy döntött, hogy július 2-án elutazunk Egyiptomba... Nagyon örültem neki, rögtön meg is kerestem, hogy hol van Egyiptom. Sajnos a suliban a térképemet leöntöttek kakaóval, azért nem tudtam pontosan kiolvasni, hogy melyik lehet Egyiptom. Szerinted?

A B C D

15. ábra: A komplex problémamegoldó feladatlap egy mintafeladata

A papíralapú tesztekkel több nagymintás adatfelvétel történt 2003 és 2011 között. Az első felmérés 3-11. évfolyamos tanulók bevonásával, reprezentatív mintán zajlott, ezt követően 3-8. évfolyamos diákok vettek részt a mérésben, majd 7. és 11. évfolyamokon történtek újabb mérések longitudinális vizsgálatok keretében. A mérésekben résztvevő tanulók száma meghaladta a 25000-et (Csapó és Molnár, 2012).

A tesztek Cronbach-alpha reliabilitásmutatója 0,77-0,83 volt, mely elfogadható ilyen belső konzisztenciájú feladatlap estében (Molnár, 2003; Molnár, 2006c). A valószínűségi tesztelméleti elemzésekkel a közös képességskálán jellemzett nehézségi szintek szerint az itemek nehézség szerint jól eloszlottak, alkalmasak voltak a tág életkori határokon történő mérésekre, sőt még a kiemelkedően magas képességszintű tanulók mérésére is (Molnár, 2003, 2006b). Az itemek modellhez való illeszkedése megfelelő volt, mindegyik teszt esetében az itemek infit paraméterei az elfogadható 0,7 és 1,3-as tartományon belül estek (Molnár, 2003). Az itemek diszkriminációját tekintve az iskolai kontextushoz közelebb álló, egyszerűbb típusú feladatok elkülönítésmutatója magasabbnak bizonyult, mint az életszerűbb, komplexebb, több háttértudást elváró problémáké (Molnár, 2006c). Az életkor emelkedésével egyre magasabb eredményeket értek el a tanulók, mely elsősorban a matematikai természetű problémamegoldó képességgel kapcsolatban volt kiugró. Ettől eltekintve az alsóbb korosztályokon is voltak olyan tanulók, akiknek teljesítménye a középiskolás diákokéval vetekedett. Ez indokolja mindegyik évfolyam esetében a tág intervallumskálán való itemek használatát (Molnár, 2003).

II.5.3 Szóolvasó készséget mérő tesztrendszer

A szóolvasó készség mérését megvalósító tesztrendszer választását több tényező indokolta. Egyrészt 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 nélkül a tanulók nem képesek a tananyag elsajátítására, mely 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 olvasásképesség megfelelő szintű elsajátításával jelentős problémák vannak (*D. Molnár, Molnár és Józsa, 2012; Nagy, 2004a; OECD, 2014*). Az első három PISA mérésben, azaz 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 papíralapú szövegek olvasásában szignifikáns javulás (494 pont; *OECD, 2010*), majd 2012-es adatfelvétel eredményei alapján jelentős mértékű képességszint esés (488 pont; *OECD, 2014*) következett be, aminek következtében teljesítményünk ismét az OECD átlag alattinak bizonyult. Digitális szövegértésnél még kedvezőtlenebben alakult helyzetünk, 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 a magyar tanulók.

Miután az olvasásképesség egy nyelvi képesség, szintjei jellemezhetőek a nyelv strukturális szintjeivel: a fonetikai szinten a betű-hang megfeleltetéssel, a morfológiai szinten a szótagolással, a lexikai szinten a szórutinokkal, a szintaktikai szinten a mondatok olvasásával és a szöveg szintjén a szövegértéssel (*Csapó, Józsa, Steklács, Hódi és Csíkos, 2012*). Ennek következtében az olvasáskészség négy komponens-képességgel jellemezhető, négy olvasástechnikai készségből áll: beszédhanghalló készség, betűolvasó készség, szóolvasó készség 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). Miután a köznyelvi szövegek közel 95-96 százaléka 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 (Nagy, 2004a) megértéséhez, megállapítható, hogy a köznyelvi szövegek sikeres, értő olvasásának nélkülözhetetlen előfeltétele a leggyakoribb kb. 4000 szó ismerete és e szavak megfelelő szintű felismerése, olvasása.

Nagy József (2004a, 2004b) felismerve a szókincs és a szóolvasó készség meghatározó szerepét az olvasási képesség fejlettségénél, munkatársaival kidolgozott egy szóolvasó készség vizsgálatára alkalmas kritériumorientált papíralapú teszt-sorozatot. Miután a kutatás célcsoportja 2-12. évfolyamos diákok voltak, ezért a fent 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, melyet az optimálisan fejlett olvasáskészség kritikus szókészletének neveztek el. Ezt követően a szókészletben szereplő szavak felhasználásával tíz diszjunkt, különböző szavakat tartalmazó ekvivalens tesztváltozatot készítettek, amely tesztenként 500 különálló szó mérését lehetővé téve. Minden egyes tesztváltozatban azonos arányban fordultak elő a különböző gyakoriságú szavak. A tesztek felépítése azonos volt, azonos mennyiségben tartalmaztak címszóolvasást (C), toldalékos szóolvasást (T), szinonimaolvasást (S) és szójelentés olvasást (J) mérő feladatokat (Nagy, 2004b).

A 10 címszóolvasás részteszt összesen 250 feladatot, feladatonként négy különböző szó olvasásvizsgálatát tette lehetővé, azaz a tíz tesztváltozatban összesen 1000 címszó olvastatására került sor. Minden egyes feladat négy szó elolvasását igényelte, majd mind a négy szó esetén külön-külön döntést kellett hozni, hogy az adott szó megfelel-e a szavak mellett prezentált színes képnek, vagy sem. A toldalékos szóolvasás részteszt hasonló felépítésű volt, szintén 250 feladatban összesen 1000 toldalékos szó elolvasását mérte. A szinonima részteszt 250 feladatában 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 szavas 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 szókincs, (2) mennyire tudják elolvasni az általa kritikus szókészletbe tartozó leggyakoribb 5000 szót, valamint (3) milyen szókincsre, mely szavak hatékony olvasására lenne szükségük, hogy sikeres olvasóvá váljanak.

A papíralapú tesztrendszer kipróbálására 2-10. évfolyamon került sor (Nagy, 2004b). Az eredmények szerint a tizedik 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 jelentős képességszintbeli különbségek voltak. Már a második évfolyamon voltak olyan tanulók, akiknél az optimális szinten működött a szóolvasó készség, vagyis számukra az olvasott szövegben legfeljebb 5%-nyi ismeretlen szó fordul elő, míg a 8. évfolyamos tanulók közel 40 százalékánál viszont nem alakult ki a folyékony szóolvasási készség az évfolyam végére sem. Az 5000 szavas alapszókincre alapozó papíralapú 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 teszt Cronbach-alpha reliabilitása igen magas, 0,95 volt, ami azt jelenti, hogy a diagnosztikus tesztrendszer igen megbízhatóan alkalmazható tág életkori határokon belül a tanulók szóolvasási készségének mérésére.

II.6 A MINTÁK ÖSSZEÁLLÍTÁSÁNAK SZEMPONTJAI

A kutatás célja szerint az általános iskolai korosztály körében vizsgáltuk az adaptív tesztelés bevezethetőségeinek lehetőségeit. Ezért a mérések során törekedtünk arra, hogy a rendelkezésre álló tesztek lehetőségeit kihasználva, minél nagyobb mértékben átfogjuk az általános iskolai korosztályt. Mivel a bemutatott tesztek alkalmasnak bizonyultak tág életkori határokon belül történő képességmérésekre, ezért ki tuduk terjeszteni a méréseket az alsó és a felső korosztály körére is. A mérésekben az általános iskolák 1-8. évfolyamos tanulói közül összesen 8165-en vettek részt a 2. táblázat szerint részletezett megoszlásban. Az induktív és a problémamegoldó gondolkodás mérésére kidolgozott tesztekkel a felső tagozatosok vizsgálatát céloztuk meg. A szóolvasás teszt feladatainak megoldásába elsősorban az alsó tagozatos korosztályt vontuk be. A nagymintás hatékonyságvizsgálathoz szűkebb életkori intervallumot, a 4-5. évfolyamot választottuk, az életkori fejlettségből eltérő jellemzők kiküszöbölése végett.

2. táblázat: A minták évfolyamok szerinti megoszlása

Mérés	Évfolyamok								Összesen (fő)
	1	2	3	4	5	6	7	8	
Induktív gondolkodás pilot					22	44	51	41	158
Problémamegoldó képesség pilot					33	41	31	48	153
Szóolvasó készség pilot	16	28	38	42	30				154
Szóolvasó készség lineáris nagy mintás				2309	2171				4480
Szóolvasó készség nagy mintás adaptív				1706	1514				3220
Összesen	16	28	38	4057	3770	85	82	89	8165

II.7 ADATFELVÉTEL

A tesztek minden esetben az eDia rendszer segítségével készültek és kerültek kiközvetítésre. A tanulók online módon, a saját iskolájukban, saját internethálózatukon keresztül oldották meg a feladatokat. A mérésekre minden esetben egy tanítási óra (45 perc) állt a diákok rendelkezésére. A segítő pedagógusok részletes mérési útmutatót kaptak az adatfelvétel lebonyolításának részletes leírásával. Mindegyik tesztet két változatban oldották meg a tanulók, az első fázisban véletlenül kaptak adaptív vagy lineáris tesztet, a második fázisban fordítva történt a teszt kiosztás, aki az első fázisban lineáris tesztet oldott meg, az adaptívat kapott, és fordítva. A két fázis között minimum két, maximum négy hét eltérés volt. A mérés végén a rendszer visszajelzést adott a tanulók teljesítményéről. A lineáris teszteknel a teszten elért százalékos teljesítményt jelezte vissza a rendszer, az adaptív teszteknel viszont képességpontot számolt a program.

II.8 AZ ADATOK ELEMZÉSE

A kapott adatok elemzése során elsősorban a tesztek technikai működését vizsgáltuk, a kutatás fő célja az itemjellemzők és az itemek viselkedésének összehasonlítása volt a kétféle tesztkörnyezetben. Az elemzések elvégzése egyrészt klasszikus tesztelméleti módszerekkel, másrészt a valószínűségi tesztelmélet alkalmazásával történt. A klasszikus tesztelméleti elemzéseket az SPSS program, a valószínűségi tesztelméleti elemzéseket a ConQuest program segítségével végeztük. A valószínűségi tesztelmélet lehetővé tette az összes feladat egy közös nehézségi skálán történő elhelyezését és az egyes itemek valószínűségi alapon történő,

populációfüggetlen elemzését. Az elemzéseket parciális kredit modellel végeztük. Az itemek paraméterezése az egyparaméteres Rasch-modell segítségével történt (*Rasch, 1960*). A logitegységben kapott diákokra vonatkozó képességpontokat 500 pontos átlagú és 100 pontos szórású skálára transzformáltuk. A mérés során készségek és képességek mérése történt, azonban a valószínűségi tesztelmélet képességskálaként értelmezi a kapott értékeket, ezért nem változtattunk technikailag az elterjedt kifejezésen.

A lineáris tesztek mérési pontosságának meghatározásához a teszt megbízhatósági mutatója, a Cronbach-alpha reliabilitásmutató szolgál. A Cronbach-alpha viszont csak olyan esetben számítható, amikor a tesztelésben résztvevő minden személy minden feladatot megold, azaz nincs hiányzó adat. Adaptív teszteknel viszont a tanulók az itembanknak csak egy részhalmazát oldják meg, így a Cronbach-alpha az itembank szintjén nem számítható. Ezért az adaptív tesztek megbízhatóságának jellemzésére a valószínűségi tesztelmélettel számítható WLE személy-szeparációs reliabilitásmutatót alkalmaztuk, mely mindig alacsonyabb értéket ad, mint a Cronbach-alpha (*Linacre, 1997; Clauser és Linacre, 1999*). A mérési pontosság további mutatójaként a teszt információt és a standard hiba mértékét használtuk (*Weiss, 2013*), melyeket szintén a Rasch-modell segítségével számítottunk. A tesztinformációs görbék a tesztből kinyert információ nagyságát a tesztet megoldó tanulók átlagos képességszintje és az itemek nehézségi szintje közötti különbségek segítségével jellemzik. A kinyert információ nagyságát akkor tekintettük maximálisnak, ha a feladatok nehézségi szintje és az azokat megoldó diákok képesség-szintje azonos. Minél távolabb volt egymástól ez a két érték, annál kisebb volt a tesztelés során kinyert információ nagysága. Az item- és tesztinformációt, valamint a mérési hiba nagyságát az 1.5. fejezetben leírtak szerint számoltuk.

Mivel mindkét tesztváltozatot minden tanuló megoldotta, a becsült képességszinteket és a helyes válaszok arányát illetően személyszintű összehasonlításra is lehetőség nyílt. A változók közötti kapcsolatokat korrelációkkal vizsgáltuk. A különbségek szignifikanciaszintjének meghatározását egymintás t-próbával, illetve varianciaanalízissel (ANOVA) végeztük. A két teszten mért különbségek jellemzésére a különbség mértékét szóráségségben kifejező mutatót, a Cohen d-t használtuk (*Cohen, 1988*) használtuk.

III. SZÁMÍTÓGÉPES ADAPTÍV TESZTEK ÉS LINEÁRIS TESZTEK MŰKÖDÉSÉNEK ÖSSZEHAONLÍTÓ HATÉKONYSÁGVIZSGÁLATÁRA IRÁNYULÓ PILOTMÉRÉSEK

Annak vizsgálatára, hogy papíralapú tesztek számítógépes formára konvertálhatóak-e, és az így kifejlesztett adaptív tesztek osztálytermi környezetben használhatóak-e az 5-8. évfolyamos tanulók körében, két empirikus pilotvizsgálatot végeztünk. Az első vizsgálat az induktív gondolkodást mérő teszt, a második pedig a problémamegoldó képességet mérő teszt adaptív és lineáris tesztváltozatainak összehasonlításával történt.

III.1 INDUKTÍV GONDOLKODÁS MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL

III.1.1 Minta

Az adatfelvétel 2012 őszén 158 5–8. évfolyamos diák részvételével zajlott. A minta évfolyam szerinti eloszlását a 3. táblázat tartalmazza.

3. táblázat. A minta évfolyam szerinti eloszlása





















Évfolyam	N (fő)
5.	22
6.	44
7.	51
8.	41

III.1.2 Mérőeszköz

Az induktív gondolkodás teszt számítógépes változatában az itemek a papíralapú feladatokkal azonos formában jelenítettük meg. A legfontosabb különbség az volt, hogy a papír alapon itemcsoportokban megjelenített feladatok helyett a képernyőn az itemek egyesével jelentek meg (16.a-b ábra). A zárt típusú feladatoknál a tanulóknak kattintással kellett a helyes választ megadniuk, a nyílt végű feladatoknál pedig a beviteli mezőbe kellett beírniuk a megfelelő karaktert. A számítógépes verziónál a rendszer csak akkor engedte tovább a tanulót, ha megoldotta a feladatot, tehát nem volt engedélyezve az itemkihagyás lehetősége. Egy-egy modulon belül viszont volt lehetőség a visszalépésre és a javításra.

Papír alapon:

Karikázd be azt a három alakzatot, amiben van valami közös és különbözik a többitől!

- 1)     
- 2)     
- 3)     
- 4)     

Számítógép alapon:

Kattints rá arra a három alakzatra, amelyekben van valami közös és különböznek a többitől!



16.a ábra: Zárt típusú feladat megjelenítése papír alapon és számítógép alapon

Papír alapon:

Folytasd a következő betűsorokat! Írd mindegyik sor végén a vonalra azt a két betűt, amelyik a sor folytatásaként legjobban illik oda! A sorokban csak az abc következő betűi fordulnak elő:

a b c d e f g h i j k l m n o p q r s t u x y v w z.

1. a c e g i k m ___ ___
2. z y t q n k h ___ ___
3. b e g j l o q ___ ___

Számítógép alapon:

Folytasd a következő betűsorokat! Írd mindegyik sor végén a vonalra azt a két betűt, amelyik a sor folytatásaként legjobban illik oda! A sorokban csak az abc következő betűi fordulnak elő:

a b c d e f g h i j k l m n o p q r s t u x y v w z.

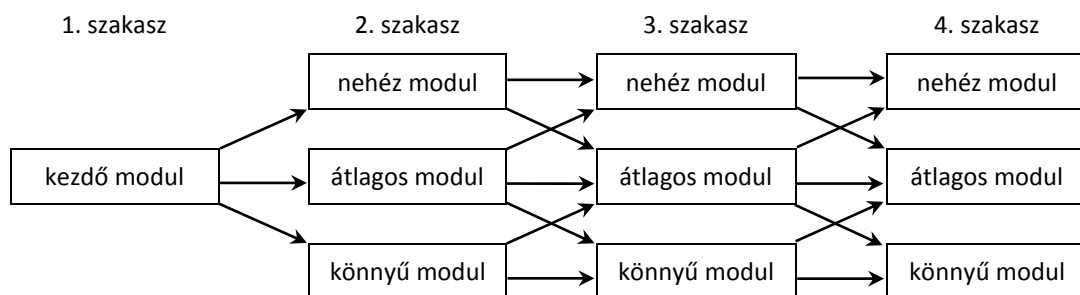
a c e g i k m

16.b ábra: Nyílt végű feladat megjelenítése papír alapon és számítógép alapon

Az itembank létrehozásánál az itemek egy paraméterét, a nehézségi paraméterét tekintettük, mely a papíralapú mérések alkalmazásával került meghatározásra a Rasch-modell segítségével. Az itemek paraméterei $-4,3$ és $+4,3$ logitegységszint között mozogtak.

A lineáris teszt összeállításánál törekedtünk a széles képességtartomány megbízható vizsgálatára, ezért a teszt összeállítása során különböző nehézségű itemeket alkalmaztunk. A lineáris tesztváltozat 28 itemből állt és vegyesen tartalmazott mindegyik feladattípusból itemeket.

A többszakaszos adaptív teszt kezdő résztesztje (1. szakasz) 10, szélesebb itemnehézségi skálán mozgó itemet tartalmazott. A 2–4. szakasz résztesztjei 6-6 itemből álltak, és minden szakaszon belül három különböző nehézségi szintű résztesztet tartalmaztak. Összesen 10 különböző részteszt (modul) kialakítása történt, melyből 17 különböző tesztváltozat összeállítására volt lehetőség. A részteszt (modulok) egymáshoz való viszonyát és az egész tesztelés során elfoglalt helyét a 17. ábra szemlélteti.



17. ábra: A többszakaszos adaptív teszt szerkezete

A modulok közötti elágazási szabály meghatározása az NC-módszer (*Number Correct*) segítségével történt (Zenisky és Hambleton, 2004). Minden feladatban a helyes válaszáért 1, a helytelen vagy hiányzó válaszáért 0 pontot adott a rendszer. Az első és a második modul közötti elágazásnál az osztópontot a kezdő modul tesztkarakterisztikus görbéje (Molnár, 2013) alapján számítottuk ki a korábbi adatfelvételek eredményeire alapozva, így a 4-nél kevesebb pontot elért tanulók a könnyű modult kapták, a 7-nél több pontot elérők a nehezet (457 és 543 képességszintek által meghatározott értékek). A második és a harmadik elágazásnál is hasonló módon határoztuk meg az osztópontokat: a könnyű modulnál 0–4 pont elérésekor a könnyű modul felé, 5–6 pont esetén a közepesen nehéz modul felé ágazott el a teszt. A legnehezebb részteszten 0–2 pontot elérő diákok a tesztelés következő szakaszában a közepes

nehézségű modulot kapták, míg a 3–6 pontot teljesítők maradtak a legnehezebb itemeket tartalmazó tartományban. A közepes (átlagos) nehézségű résztesztet megoldók három irányban léphettek tovább. A 0–2 pontot elérők a legkönnyebb, a 3–4 pontot teljesítők a közepes nehézségű, az 5–6 pontot kapott diákok a legnehezebb feladatokat tartalmazó részteszt felé léptek tovább. Ezzel a módszerrel, előzetes hipotézisünk szerint, a negyedik szakasz végére három egyenlő részre osztottuk a diákokat képességszintjük szerint.

III.1.3 Adatfelvétel és eljárások

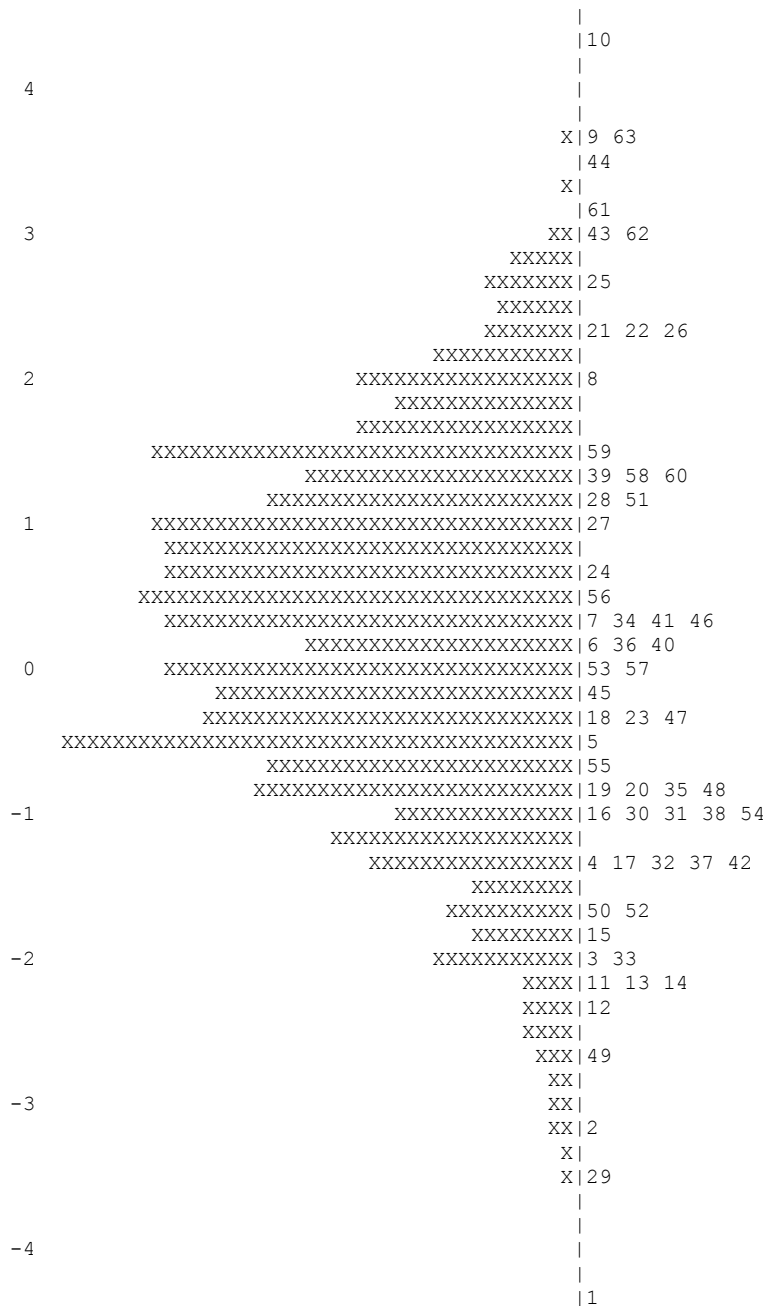
Az adatfelvétel első fázisában minden osztályt véletlenszerűen két részre osztottunk. A diákok egyik része a (n=79) lineáris induktív gondolkodás-tesztet oldotta meg, a másik fele a többszakaszos adaptív tesztet. Két hét elteltével, az adatfelvétel második fázisában, a korábban lineáris tesztet megoldó tanulók az adaptív tesztet, az adaptív tesztet megoldó tanulók lineáris tesztet kaptak. A diákok számára rendelkezésre álló idő mindkét esetben 45 perc volt.

III.1.4 Eredmények és diszkusszió

III.1.4.1 A tesztek reliabilitása

Az adaptív teszt WLE (*Weighted Likelihood Estimate*) személyszeparációs reliabilitásmutatója 0,85, ami magasabb volt, mint a lineáris formátumú teszt megbízhatósági mutatója (0,83). A reliabilitásmutatók alapján megállapítható, hogy a kidolgozott itembank megbízhatóságát tekintve alkalmas 5–8. évfolyamos diákok induktív gondolkodásának, e gondolkodás fejlettségi szintjének meghatározására.

Az adatfelvételek alapján készült személy-item térkép (18. ábra) azt mutatja, hogy az itemek lefedik a minta képességtartományát, tehát a teszt megfelelő volt a korosztály mérésére.



18. ábra: Az adatfelvétel alapján kirajzolható személy-item térkép

A diákok lineáris formátumú teszten mutatott teljesítménye (átlag=500, szórás=100) és az adaptív teszt alapján számolt képességszintje (átlag=489, szórás=100) erősen korrelált egymással ($r=0,82$, $p<0,01$). A páros t próba értéke szerint nem volt szignifikáns különbség az elért eredmények között ($t=-0,23$, $p=0,98$), melyet a *Cohen d* értéke is megerősít ($d=0,10$). A két tesztkörnyezetben egymástól függetlenül meghatározott, azonos diákra vonatkozó képességszintek átlagosan azonosnak tekinthetők, ugyanakkor a korrelációs együttható nagysága eltérésekre is utal.

III.1.4.2 A becsült képességszintek összehasonlítása évfolyamonként és személyenként

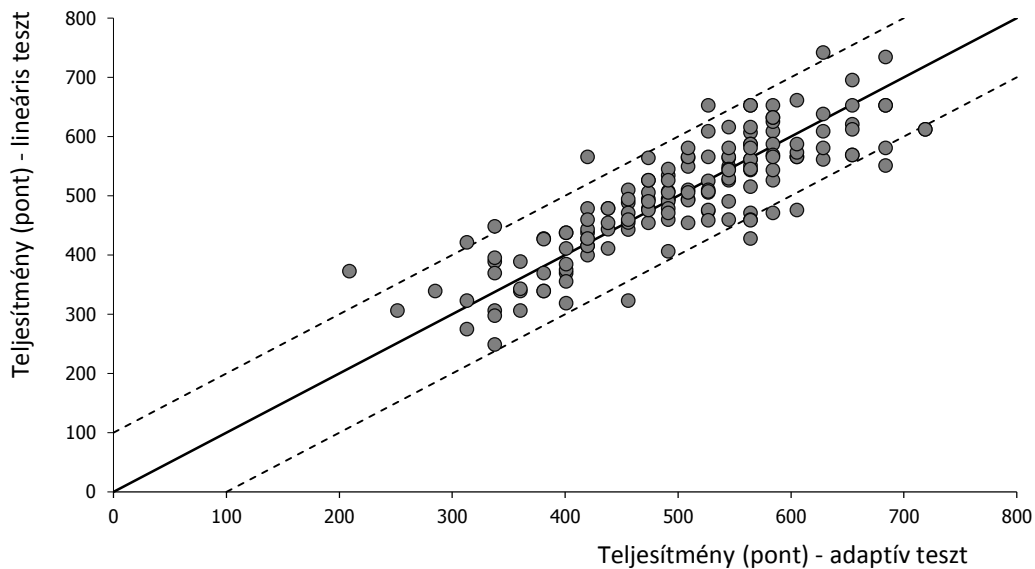
A diákok lineáris, illetve adaptív tesztkörnyezetben mutatott teljesítményének alapstatisztikai mutatóit évfolyamonkénti bontásban a 4. táblázat mutatja. Az 500 pont körül, azaz átlagosan teljesítő diákok között nem volt szignifikáns különbség a kétféle tesztkörnyezetben elért eredmények között. Az átlagosan legmagasabb képességszintű, 8. évfolyamos diákok esetében volt szignifikáns különbség, ők az adaptív teszten szignifikánsan jobban teljesítettek. Ez arra enged következtetni, hogy az adaptív tesztelés elsősorban a magasabb képességtartományú tanulók számára volt előnyösebb.

Évfolyamonkénti bontásban vizsgálva az adaptív teszt az 5. és a 8. évfolyamot különítette el szignifikánsan ($F=7,35$, $P<0,01$; $\{5\}<\{6,7\}<\{8\}$).

4. táblázat. Lineáris, illetve adaptív tesztkörnyezetben mutatott teljesítmények átlaga és szórása évfolyamonkénti bontásban

Évfolyam	N	Lineáris teszt		Adaptív teszt		páros t-próba	p	d
		átlag	szórás	átlag	szórás			
5.	22	439	90	432	95	-0,55	0,59	0,08
6.	44	510	94	500	96	-1,35	0,18	0,11
7.	51	498	106	491	99	-0,76	0,45	0,06
8.	41	524	93	547	86	2,60	0,01	0,26

A 19. ábra a két tesztkörnyezetben nyújtott teljesítmények diákszintű összehasonlítását ábrázolja. Ha a diák képességszintje tesztkörnyezettől függetlenül számszerűen ugyanannak bizonyult, akkor a diákot reprezentáló alakzat a folytonos vonalon helyezkedik el. Amennyiben megállapított képességszintje nem különbözik egymástól szignifikánsan lineáris és adaptív környezetben, az őt reprezentáló jel a szaggatott vonalakon belül található. A szaggatott vonalak által képzett sávon kívül elhelyezkedő diákok számára a lineáris vagy az adaptív tesztkörnyezet bizonyult kedvezőbbnek. Előfordulásuk elenyésző a mintában, azaz különböző mérési hiba alkalmazása mellett, de szignifikanciaszinten belüli, közel azonos képességszint-beclést végeztünk lineáris, illetve adaptív tesztelosztással.



19. ábra: A lineáris és az adaptív teszten nyújtott teljesítmények összehasonlítása diákonkénti bontásban

III.1.4.3 A helyes válaszok aránya a kétféle tesztkörnyezetben

Összehasonlítottuk évfolyamonként a tanulók által elért helyes válaszok arányát a kétféle tesztkörnyezetben (5. táblázat). A tanulók mindegyik évfolyamon több helyes választ adtak az adaptív teszten, mint a lineáris tesztkörnyezetben. Eszerint a tanulók adaptív tesztelésnél a képességszintjüknek megfelelő itemeket kaptak, amikre nagyobb százalékban tudtak helyesen válaszolni, mint amikor vegyesen kaptak könnyebb és nehezebb feladatokat a lineáris teszteléskor.

5. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak aránya évfolyamonkénti szerinti bontásban

Évfolyam	N	Helyes válaszok aránya (%)				d
		Lineáris teszt		Adaptív teszt		
		átlag	szórás	átlag	szórás	
5.	22	23	9,8	46	13,9	1,91
6.	44	33	12,2	53	11,1	1,71
7.	51	30	13,3	52	11,6	1,76
8.	41	35	11,6	59	10,4	2,18

A kétféle tesztkörnyezetben nyújtott teljesítmények képességszint szerinti arányát ábrázolja a 6. táblázat. Mindkét teszt a jó megoldások arányának tipikus mintázatát mutatja,

a képességszint növekedésével párhuzamosan emelkedett a teszteken elért helyes válaszok aránya is.

6. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak aránya képességszint szerinti bontásban

Képességszint	Helyes válaszok aránya (%)						d
	Lineáris teszt			Adaptív teszt			
	N	átlag	szórás	N	átlag	szórás	
250-350	13	10	4,9	15	28	7,2	2,92
351-450	34	19	4,9	29	44	5,3	4,90
451-550	59	32	5,6	60	54	5,1	4,11
551-650	39	42	4,4	43	63	6,5	3,78
651-750	13	52	3,1	11	68	5,8	3,44

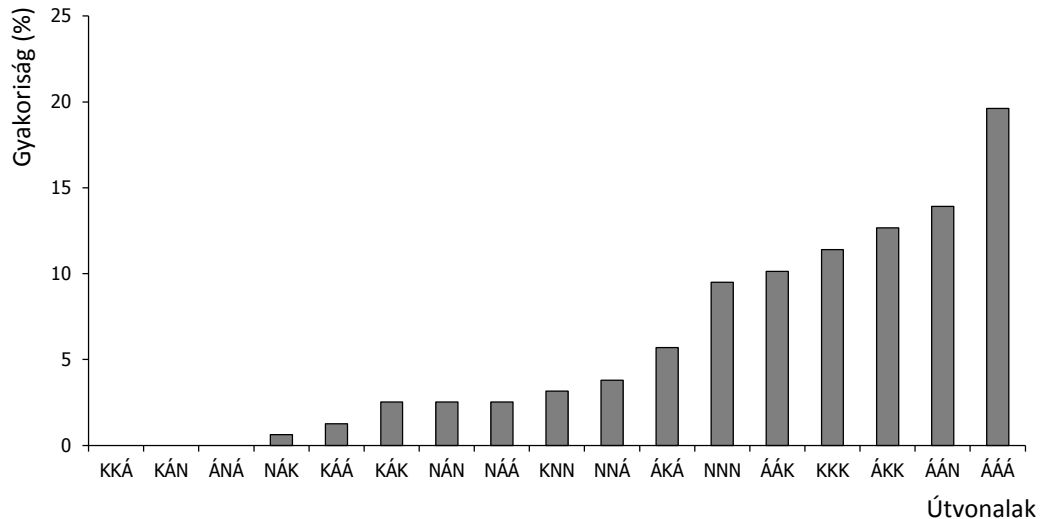
Összességében az adaptív teszt esetében jóval magasabb százalékban válaszoltak a tanulók helyesen (53%), mint a lineáris teszten (31%). Az egész képességszintet nézve minden képességszinten magasabb volt a helyes válaszok aránya, mint a lineáris teszten.

III.1.4.4 A lineáris és az adaptív tesztelés során kiosztott résztesztek nehézségi szintjének változásmintázata

Az adatfelvétel során a többszakaszos adaptív teszt esetében a négy szakaszból összeállítható 17 különböző teszt közül 14-et osztottunk ki (17. ábra). Az esetek ötödében a részteszteken nyújtott teljesítmények alapján kizárólagosan az átlagos nehézségi szintű feladatokból álló tesztet közvetítettük ki. 11%-ban a kizárólagosan könnyű és közel 10%-ban a kizárólagosan nehéz résztesztekből álló, nehézségi szint tekintetében homogén tesztet vettük fel. Mindezek alapján megállapítható, hogy a diákok 40%-a a kezdő részteszten nyújtott teljesítménye alapján egyértelműen besorolható volt a három képességszint egyikébe. Egyetlen egy diák esetében fordult elő, hogy két nehézségi szintet is ugrott a tesztelés folyamán. A kezdő modul után megállapított képességszintje a legmagasabb képességtartományba sorolta őt, ugyanakkor a tesztelés végére átkerült az átlagosnál alacsonyabban teljesítő diákok csoportjába.

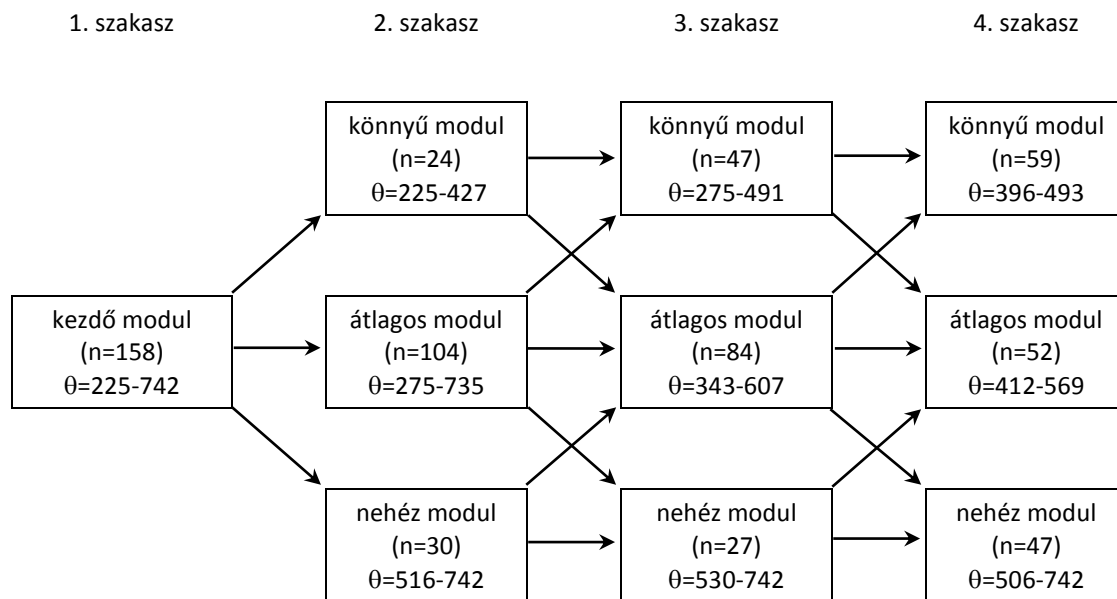
A diákok kétharmadának viselkedését jól jellemzi a hat leggyakoribb útvonal, melyek között a három azonos nehézségi szintű modulból álló tesztek mellett szerepel két olyan útvonal, ahol a két átlagos nehézségi szintű részteszt után a teljesítmények alapján a záró modulon a nehéz, illetve a könnyű részteszt irányába ágazott el a rendszer (20. ábra). A négy

szakaszból álló adaptív tesztelési modell előnye a háromszakaszos modellhez képest az átlagos képességszinthez közeli, ugyanakkor azt vagy nem elérő, vagy kicsit túlteljesítő tanulók pontosabb képességszint-meghatározásában mutatkozott meg.



20. ábra: Az adaptív tesztrendszeren belül a második, harmadik és negyedik szakaszban kiosztott útvonalak gyakorisága (K: könnyű, Á: átlagos, N: nehéz modul)

A szakaszokon belül a tanulók modulonkénti eloszlását képességszint és gyakoriság szerinti bontásban a 21. ábra szemlélteti. A kezdő modulon mutatott teljesítmény alapján a tanulók háromötöd része a teszt második szakaszában közepes nehézségű résztesztet kapott, majd a teszt harmadik szakaszában mutatott teljesítmények alapján a teszt negyedik szakaszában közel azonos módon oszlottak el az átlagos (34%), az átlagnál alacsonyabb (37%) és az átlagnál magasabb (29%) képességszintű diákok. A lineáris formátumú teszttel ellentétben, ahol állandó volt a teszten belüli könnyebb, átlagos és nehezebb feladatok aránya, az adaptív feladat kiosztás során a magasabb képességszintű diákok nagyobb arányban kaptak nehezebb, míg az alacsonyabb képességszintű diákok könnyebb feladatokat. A teszt utolsó szakaszában, a diákokat közel harmadolva azonos átfedéssel, egyértelműen kialakult a három képességszív.



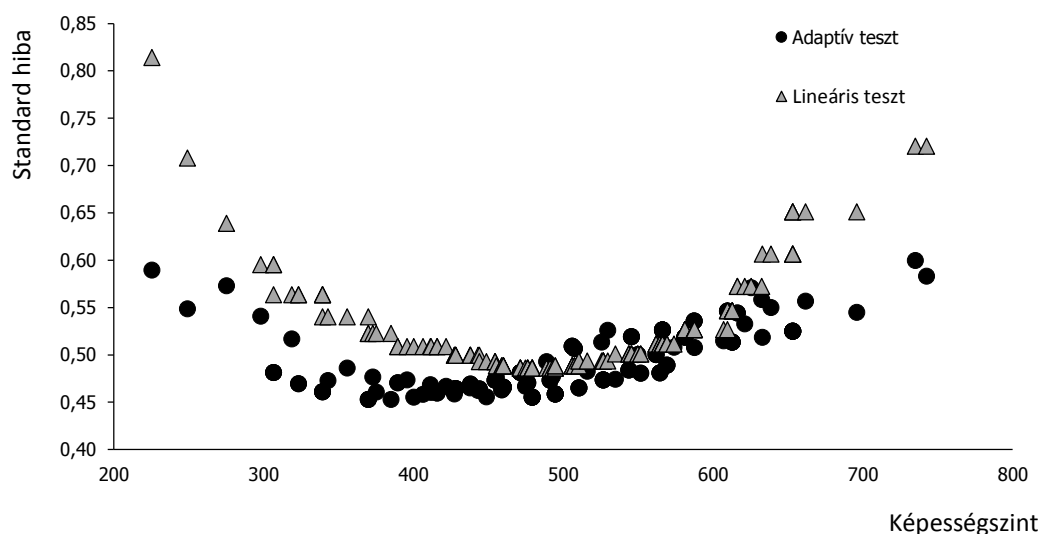
21. ábra: A tanulók gyakorisági és képességszint szerinti eloszlása a szakaszokon és a modulokon belül

III.1.4.5 A lineáris és az adaptív tesztelés során kinyert mérési hibák és kinyert információk nagyságának összehasonlítása

A kutatás felépítése diákszinten lehetővé tette a becslt képességszintek pontosságának összehasonlítását. Diákonkénti bontásban a lineáris, illetve az adaptív tesztkörnyezetben történt képességszint-becslés során elkövetett hiba nagyságát összehasonlítva (Wang és Kolen, 2001; Wang, 2010) megállapítható, hogy a lineáris formátumú teszt alapján történt képességszint-becslés hibáinak nagysága diákszinten átlagosan nagyobb ($t=-7,54$, $p<0,01$; $se_{\text{átlag}}=0,53$), mint ugyanazon diákok adaptív tesztkörnyezetben történt képességszint-becslésének hibája ($se_{\text{átlag}}=0,49$). A hatás-mérték 0,78, vagyis jelentős az eltérés. A teljes minta vonatkozásában pontosabban, kisebb mérési hibával történt adaptív tesztkörnyezetben a diákok képességszintjének becslése. A hiba nagysága nem egyenletesen oszlik el a teljes képességskálán: különböző mintázat várható az alacsonyabb, az átlagos és a magasabb képességtartományban.

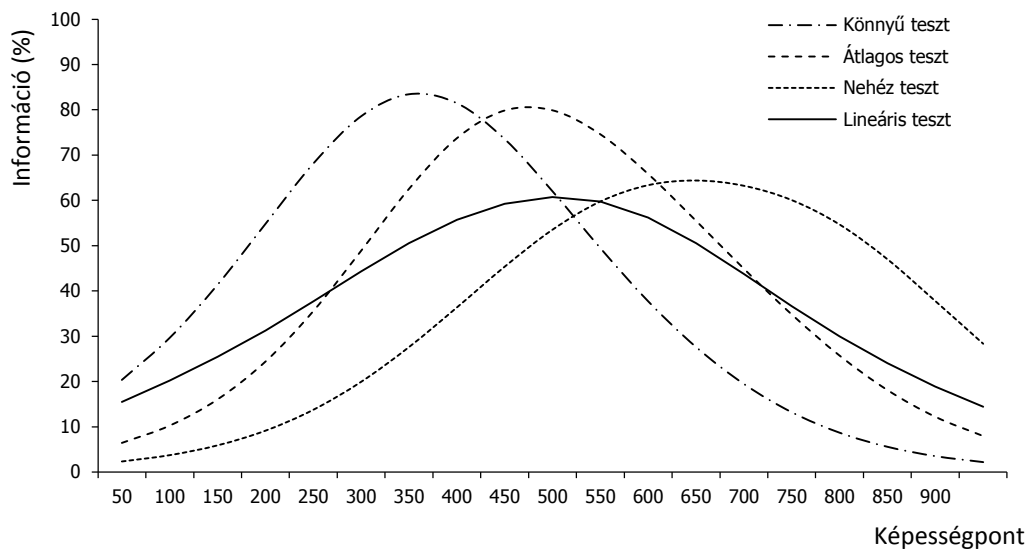
Összehasonlítva a lineáris formátumú és az adaptív teszten elért eredmények sztenderd hibáinak változását (22. ábra), hipotéziseinknek megfelelően, az alacsonyabb és a magasabb képességtartományban nagyobb hibával mér a lineáris formátumú teszt, mint az átlagos képességszintű diákok körében. Ez azt jelenti, hogy az adaptívteszt-algoritmus alkalmazásának előnye ezen képességtartományokban a legjelentősebb, átlagos

képességszintű diákok mérése során közel azonosnak bizonyult a két tesztkörnyezetben becsült képességszintek sztenderd hibáinak alakulása.



22. ábra: Az adaptív és a lineáris formátumú lineáris teszt standard hibáinak alakulása a tanulók képességszintjének függvényében

A pontosság egy másik mutatója a tesztelés során kinyert információ nagysága, amit jelen esetben a (rész)teszteken nyújtott teljesítmények alapján felrajzolt tesztinformációs görbék segítségével jellemzünk. A kinyert információ százalékos nagyságát összehasonlítva, míg a lineáris teszt átlagosan 60%-os információt szolgáltatott, addig az adaptív tesztelés során kinyert átlagos információ nagysága 76% volt. A személyszintű összehasonlítás alapján az eltérés elsősorban az alacsony és a magas képességszintű tanulók esetében volt jelentős, előbbinél közel 34%, utóbbi során közel 24%-kal több volt az adaptív tesztből kinyert információ mennyisége. A 23. ábra görbéi grafikusán szemléltetik, hogy már akár a kizárólagosan könnyű, átlagos, illetve nehéz modul résztesztjeiből összeállított tesztek (a 17 tesztváltozatból csak 3) is több információt szolgáltattak a tesztelés során, mint az egyetlen, sokféle nehézségű feladatot tartalmazó lineáris formátumú teszt. A görbék a 21. ábrán ismerttetett képességszintekkel egybevetve megállapítható, hogy minden képességszinten több információt tudunk kinyerni adaptív tesztek alkalmazásával (a görbék minden esetben a lineáris formátumú teszt által adott információs függvény felett futnak az érintett képességtartományokban). A többletinformáció már abban az esetben is kimutatható, ha a diákokat az első részteszten nyújtott teljesítményük alapján három csoportba soroljuk, majd az adott képességtartományhoz közeli nehézségi szintű feladatokból állítjuk össze a tesztet.



23. ábra: Az adaptív technikával összeállított és a lineáris formátumú, azonos nehézségű részteszteket tartalmazó tesztek információs függvényei

III.1.5 A pilotmérés eredményeinek összefoglalása

A mérés során ugyanazon a mintán felvett 28 ítemes lineáris formátumú és egy összességében 28 ítemes, ám 1-3-3-3 szerkezetű négyszakaszos adaptív teszten elért teljesítményeket, a helyes válaszok arányát, a becsült képességszinteket, azok sztenderd hibáit, a tesztelés során kinyert információ nagyságát, valamint a tesztek jóságmutatóit hasonlítottunk össze.

A két teszt megbízhatóságát a reliabilitásmutatókkal jellemeztük. Mindkét tesztnél a reliabilitásmutatók értékei megfelelőek voltak, tehát mindkét teszt megbízhatóan alkalmazható a korosztály induktív képességének mérésére. Az adaptív teszten mért személyszeparációs reliabilitása magasabb volt, mint a lineáris teszt Cronbach- α mutatója. A személy-item térkép lefedte a vizsgált tanulók képességszintjét, tehát az ítemek nehézsége megfelelt a tanulók képességszintjének.

A tanulók adaptív és lineáris tesztkörnyezetben elért eredményei magasan korreláltak egymással, a személyszintű bontás szerint a legidősebb, 8. évfolyamos tanulók esetében volt szignifikáns eltérés a kétféle teszten elért teljesítmények között. Az alacsonyabb évfolyamos diákok eredményei nem különböztek jelentős mértékben a különböző tesztkörnyezetekben.

Az adaptív tesztelésnél minden képességszinten és évfolyamon magasabb volt a helyes megoldások aránya, vagyis nagyobb sikerélményt okozott a tanulóknak.

Az adaptív teszt 17 tesztváltozata közül hat útvonal fordult elő a legnagyobb hányadban, ezek között is az átlagos nehézségű modulok előfordulási aránya volt a legmagasabb. Mivel a tanulók többsége átlagos képességű, ez megfelel az elvártnak. A könnyű és a nehéz modulok a gyengébb és a magasabb képességű tanulók elkülönítésében vettek részt. A tesztelés során megvalósult a gyengébb és a magasabb képességszintű tanulók elkülönítése, és a tesztelés végére a minta közel azonos arányban történő eloszlásával alakult ki a három képességsáv.

A tesztelés során a teljes minta szintjén több információt nyertünk ki, szignifikánsan pontosabb képességszint-meghatározást végeztünk az adaptívteszt-algoritmus alkalmazásával, mint a hagyományos, lineáris módszerű teszteléskor. A képességszintek becslése során elkövetett hiba nagysága is ezzel párhuzamosan csökkent az adaptívteszt-algoritmus alkalmazása során.

Összefoglalva megállapítható, hogy a mérés a legtöbb hipotézisünket beigazolta. A papíralapú induktív gondolkodást mérő itemek többszakaszos adaptív tesztté konvertálása hatékony és megbízható mérőeszközként funkcionált. A mérés során becsült képességszintek nem különböztek jelentős mértékben a lineáris változatnál elért eredményeknél. A tanulók többsége átlagos képességszintű volt, az adaptív tesztelés során kiosztott útvonalak között is az átlagos modulok előfordulási aránya volt a legjelentősebb. Az ötödik hipotézis részben igazolódott be, mivel képességszinttől függetlenül minden képességszinten magasabb volt a helyes válaszok aránya az adaptív tesztnél. A hetedik hipotézisünk szintén részben igazolódott be, miszerint az adaptív tesztből kinyert információ szignifikánsan magasabb volt, mint a lineáris tesztből kinyert, viszont a különböző képességszinteken elérő volt a mértéke. Az alacsony és a magas képességszinteken jelentősen magasabb mértékű volt, így elsősorban ezeken a képességtartományokon nőtt meg a tesztelés precizitása.

III.2 PROBLÉMAMEGOLDÓ KÉPESSÉG MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL

A pilotkutatás során 5-8 évfolyamos tanulók problémamegoldó képességét vizsgáltuk kétféle tesztelési módszerrel, egyrészt hagyományos, lineáris teszt struktúra, másrészt pedig többszakaszos adaptív teszt alkalmazásával. A vizsgálat során egyrészt a lineáris teszt adaptív formára való átkonvertálhatóságát, másrészt az adaptív változat lineáris tesztváltozattal való

összehasonlításával a mérési hatékonyságát vizsgáltuk az 5-8. évfolyamos korosztály körében (Magyar, 2014c).

III.2.1 Minta

Az adatfelvétel 2013 májusában történt 153 általános iskolai tanuló bevonásával. A minta évfolyamonkénti eloszlását a 7. táblázat mutatja.

7. táblázat. A minta évfolyamok szerinti eloszlása

Évfolyam	N (fő)
5.	33
6.	41
7.	31
8.	48

III.2.2 Mérőeszköz

A problémamegoldó képesség vizsgálatához elkészült számítógépes lineáris és adaptív tesztváltozatokhoz a papír alapon mért paramétereket használtuk. Az itemek nehézségi paraméterei -2,6 és +4,3 logitegységszint között mozogtak. A tesztváltozatok 100 iteméből egy 28 itemű tesztváltozat, és egy szintén 28 item hosszúságú adaptív változat készült kialakításra. A lineáris teszt 11 különböző nehézségű feladatot tartalmazott. Egy-egy feladat egy-öt itemből állt. A feladatok között változatos formában szerepeltek zárt, illetve nyílt végű kérdések. A 24. ábra mutatja a teszt egyik feladatát.

Papír alapon:



Anyuék terveztek egy külföldi utat is. Az eredeti terv szerint összesen

A: 6150 B: 5947 C: 7249 D: 6450
kilométert repültünk volna, de a végén csak az első állomásig, Egyiptomig utaztunk. Ez azt jelentette, hogy a nyaralás során összesen

A: 2213 B: 3548 C: 4426 D: 5204
kilométert tettünk meg repülővel.

Számítógép alapon:

Anyuék külföldi utat terveznek nyárra. Az eredeti terv szerint összesen

A: 6150 B: 5947 C: 7249 D: 6450

kilométert repültünk volna, de a végén csak az első állomásig, Egyiptomig utazunk. Ez azt jelentette, hogy a nyaralás során összesen

A: 2213 B: 3548 C: 4426 D: 5204

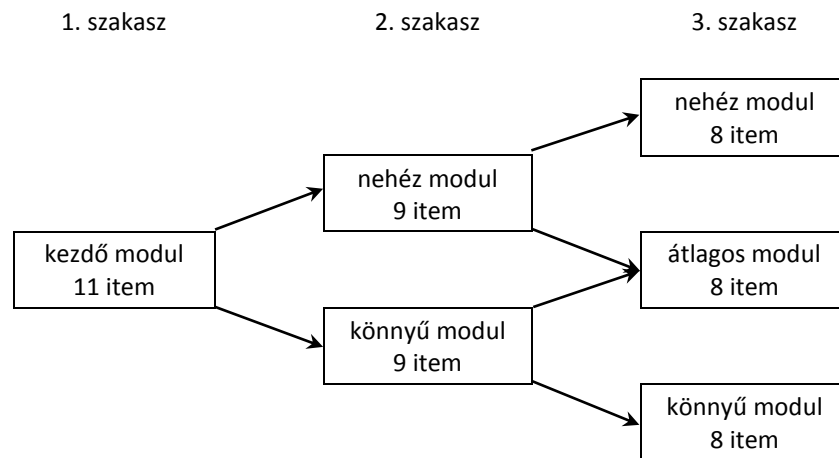
kilométert teszünk meg repülővel.



24. ábra: A problémamegoldó képességet mérő teszt egy mintafeladata

Az adaptív teszthez összesen hat különböző részteszt készült, mely négy különböző útvonalat és így négy tesztváltozatot jelentett. A kezdő modul (1. szakasz) négy feladtból állt, melyek összesen 11 ítemet foglaltak magukba. A második szakaszban két, egyenként kilenc ítemet tartalmazó modul következett. A teszt három különböző nehézségi szintű, egyenként 8-8 ítemet tartalmazó részteszttel fejeződött be. A bevezető teszt 11 íteme mindegyik

tesztváltozatban szerepelt, ezen kívül a második és a harmadik szakaszban is 2-2 item azonos volt. A modulok egymáshoz való viszonyát a 25. ábra mutatja:



25. ábra: Az adaptív teszt szerkezete

Minden részteszt megoldása után rögtön kiértékelésre került az adott modulon elért eredmény, és az algoritmus alapján az adott modulon elért pontszám alapján léphetett a tanuló a következő modulhoz. A tesztelés folyamán dichotóm értékelést alkalmaztunk; az itemekre a helyes válaszért 1, helytelen válaszért 0 pontot adott a rendszer. Az osztópont 50%-os teljesítménynél volt, az ennél alacsonyabb teljesítményt elérőket a könnyebb modul felé léptette tovább a rendszer, az 50% felett teljesítő tanulókat pedig a nehezebb modul felé. Így az első szakasz után fele-fele arányban oszlott meg a minta, a második szakaszt követően pedig 25%-50%-25% lett a tanulók modulok közötti eloszlásának aránya.

III.2.3 Adatfelvétel és eljárások

Az adatfelvétel két részletben, két hetes eltéréssel zajlott. Az első fázisban a tanulók véletlenszerűen kaptak adaptív, illetve lineáris tesztet, a második fázisban fordítva történt a tesztek kiosztása. A tesztek megoldására mindkét esetben 45 perc állt a tanulók rendelkezésére.

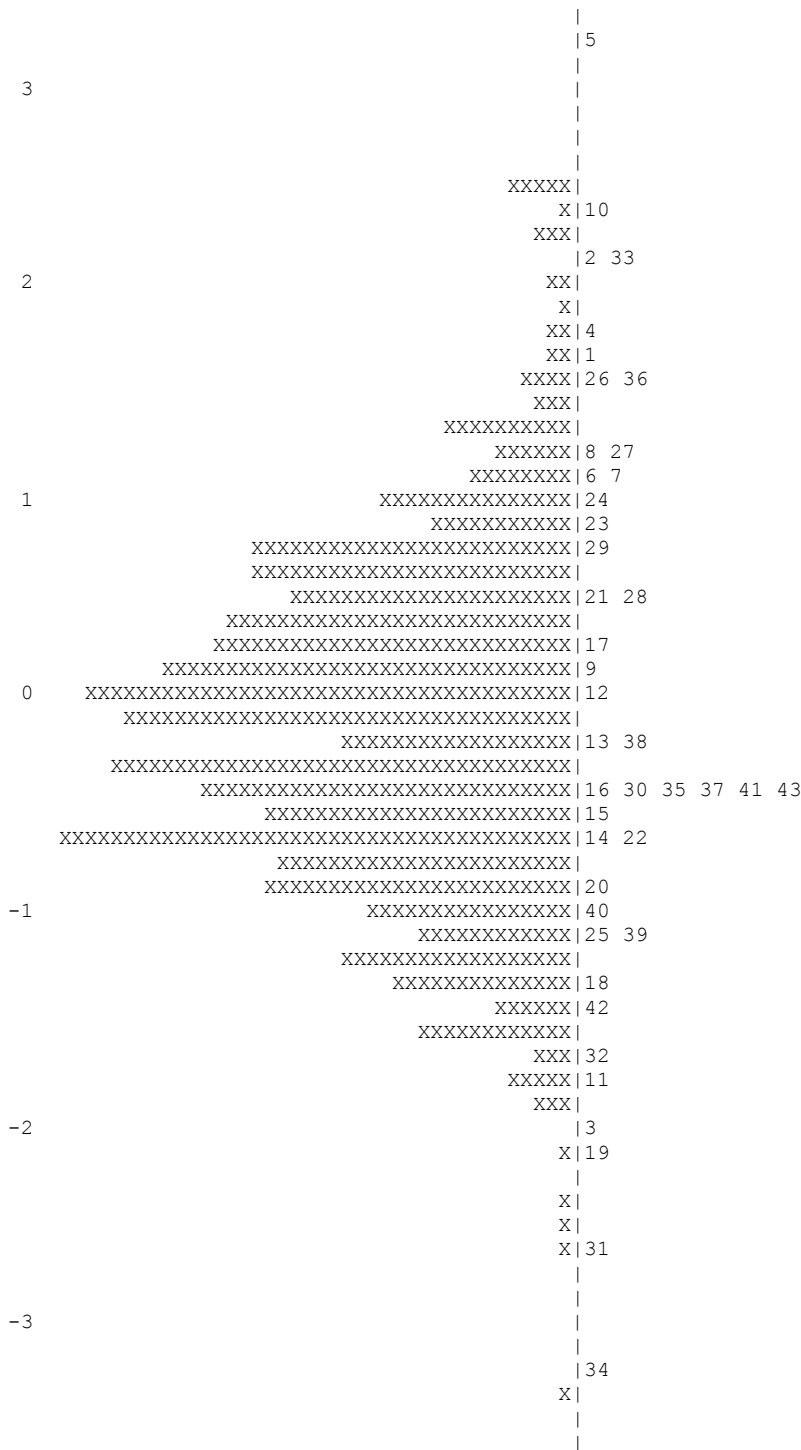
A tanulók képességszintjének megállapítása a *Rasch*-modell segítségével történt, majd a logitegységben kapott értékeket egy 500 pontos átlagú és 100 pontos szórású skálára transzformáltuk.

III.2.4 Eredmények és diszkusszió

III.2.4.1 A tesztek reliabilitása

Az adaptív teszt WLE személy szeparációs reliabilitás-mutatója 0,83 magasabbnak bizonyult, mint a lineáris teszté (0,80). A tanulók két teszten elért eredmények között nem volt szignifikáns különbség ($t=-0,03$, $p=0,98$; $d=0,10$). A teljesítmények erősen korreláltak egymással ($r=0,71$, $p<0,01$), és az elért ami azt mutatja, hogy a kétféle tesztelés nagyon hasonlóan sorolta be a tanulókat, azonban a besorolások közötti különbségek bizonyos eltérésekre utalnak a tesztek működését tekintve.

A személy-item térkép szerint az itemek teljesen lefedik a vizsgált minta képességskáláját, tehát a teszt alkalmas volt a korosztály mérésére (26. ábra).



26. ábra: Az adatfelvétel alapján kirajzolható személy-item térkép

III.2.4.2 A becsült képességszintek évfolyamonkénti és személyszintű összehasonlítása

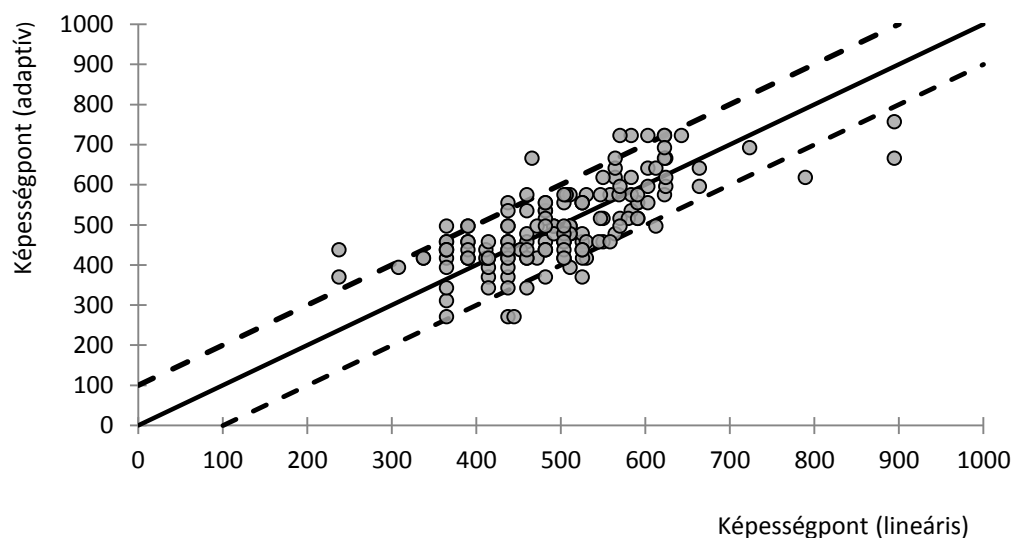
Az 8. táblázat évfolyamonkénti bontásban mutatja a tanulók lineáris, illetve adaptív teszten elért eredményeit. A páros t-próbák eredményei szerint a tanulók kétféle teszten elért teljesítménye között egyik évfolyamon sem volt szignifikáns különbség.

Az adaptív teszten elért eredményeket évfolymonkénti bontásban összehasonlítva megállapítható, hogy szignifikáns különbségek voltak a tanulói teljesítmények között ($F=12,96$, $p<0,01$). A ANOVA elemzés szerint a teszt elsősorban az alacsonyabb és a magasabb évfolyamokon különítette el szignifikánsan a tanulókat ($\{5\}<\{6,7\}<\{8\}$).

8. táblázat. A tanulók lineáris és adaptív teszten elért eredményeinek alapstatisztikai mutatói évfolyamonkénti bontásban

Évfolyam	N	Lineáris teszt		Adaptív teszt		páros t próba	p	d
		átlag	szórás	átlag	szórás			
5.	33	447	76	444	78	0,20	0,83	0,04
6.	41	491	68	492	74	-0,14	0,99	0,01
7.	31	449	70	469	93	-1,56	0,13	0,24
8.	48	576	108	565	104	0,79	0,43	0,10

A kétféle teszten elért eredmények személyszintű összehasonlítását ábrázolja a 27. ábra. A vízszintes tengely mutatja a tanuló lineáris teszten elért eredményét, a függőleges pedig az adaptív teszten nyújtott teljesítményét. Azoknak a tanulóknak, akiknek a teljesítménye nem különbözött szignifikánsan a kétféle tesztkörnyezetben, a szaggatott vonalon belül helyezkedik el a jele. Akiknek az adaptív, vagy a lineáris teszt bizonyult jobbnak, azok a tanulók a szaggatott sávokon kívül helyezkednek el. Számuk a mintában nem jelentős, tehát a kétféle teszteredmény közel azonos képességbecslést végzett.



27. ábra: A kétféle teszten elért eredmények személyszintű bontásban

III.2.4.3 A helyes válaszok aránya adaptív és lineáris tesztkörnyezetben

A tanulók adaptív és lineáris tesztkörnyezetben elért helyes válaszait összegzi a 9. táblázat. A helyes válaszok arányát évfolyamonként vizsgálva megállapítható, hogy az adaptív tesztelés esetén a tanulók a 8. évfolyam kivételével mindegyik évfolyamon több helyes választ produkáltak, mint lineáris teszt esetében. A hatás-méret az 5. és a 7. évfolyamokon mutat közepes mértékű eltérést, a 6. és 8. évfolyamokon nem.

9. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak aránya évfolyamonkénti szerinti bontásban

Évfolyam	N	Helyes válaszok aránya (%)				d
		Lineáris teszt		Adaptív teszt		
		átlag	szórás	átlag	szórás	
5.	33	30	12,6	35	10,2	0,44
6.	41	38	12,0	40	9,7	0,18
7.	31	31	11,6	38	11,8	0,60
8.	48	52	17,4	48	11,6	0,27
összes	153	39	16,6	41	11,9	0,14

A kétféle tesztkörnyezetben elért eredmények képességi szint szerinti arányát ábrázolja a 10. táblázat. Mindkét teszt a jó megoldások arányának tipikus mintázatát mutatja, a képességi szint növekedésével párhuzamosan emelkedett a teszteken elért helyes válaszok aránya is.

10. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak aránya képességi szint szerinti bontásban

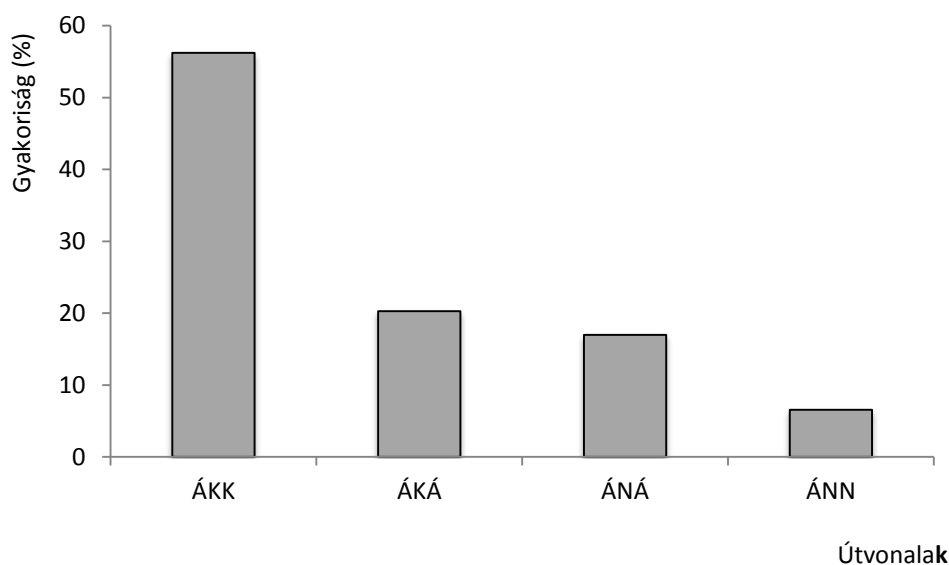
Képességi pont	Helyes válaszok aránya (%)						d
	Lineáris teszt			Adaptív teszt			
	N	átlag	szórás	N	átlag	szórás	
250-350	8	11	3,5	3	13	4,1	0,52
351-450	44	25	3,4	29	27	3,0	0,62
451-550	52	37	4,7	61	39	3,9	0,46
551-650	35	55	4,8	43	49	4,9	0,92
651-750	13	71	3,4	13	57	5,2	0,99
751-850	1	78	0,1	4	73	6,8	0,99

A két tesztet összehasonlítva az adaptív tesztnél az átlag alatti tanulók körében nagyobb volt a helyes válaszok aránya, mint a lineáris teszten, az átlag feletti tanulók

csoportjában viszont fordítva, kisebb arányban fordultak elő a helyes válaszok. Az adaptív teszt az alacsony képességszintű tanulóknak több sikerélményt jelentett, a magas képességszintűek számára viszont nagyobb kihívást jelentett, mint a lineáris teszt.

III.2.4.4 Az adaptív tesztelés során kiosztásra került résztesztek jellemzése

A többszakaszos adaptív teszt esetében a hat résztesztből összesen négyféle különböző adaptív teszt összeállítására volt lehetőség (25. ábra). Az adatfelvétel során ezeknek a teszteknek a kiosztás során alkalmazott gyakoriságát mutatja a 28. ábra. A teszt kiosztás több mint felénél, az esetek 56%-ában az átlagos-könnyű-könnyű útvonal került kiosztásra, a legkevesebben az átlagos-nehéz-nehéz útvonalon haladtak végig (7%). Az átlagos-közepes útvonal az esetek közel háromnegyedében (76%), az átlagos-nehéz pedig közel egynegyed (24%) részben fordult elő. Összességében a harmadik szakasz végére a tanulók 56%-a végzett a könnyű modullal, 37% az átlagossal, és 7% a nehéz modullal. Az adaptivitás előnye elsősorban a kiemelkedő tanulók elkülönítésénél mutatkozott meg.



28. ábra: Az adaptív teszten belül a kiosztásra kerülő útvonalak gyakorisága
(Á: átlagos, K: könnyű, N: nehéz modul)

A különböző évfolyamos tanulók különböző útvonalak közötti megoszlását mutatja a 11. táblázat. Az 5. évfolyamon a tanulók 81%-a az átlagos-könnyű-könnyű útvonalon haladt végig, a 6. és a 7. évfolyamokon a második szakaszban nehéz modult kapó tanulók jellemzően a harmadik szakaszban az átlagos modulon teljesítettek a 3. szakaszban. A 8. évfolyamon már

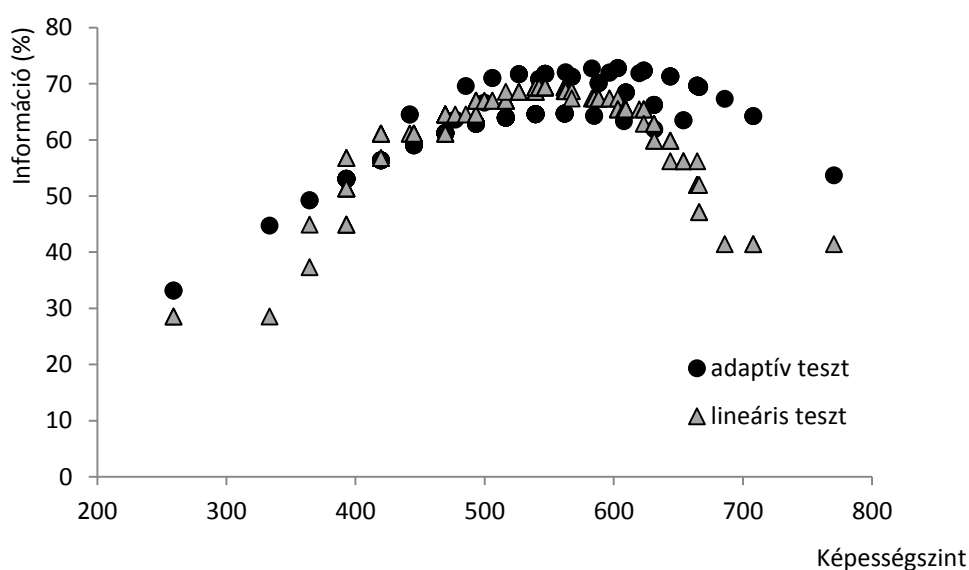
egyenletesebb lett az eloszlás a különböző útvonalak között, a könnyű modult kapó tanulók aránya lecsökkent, és a legjobbak a tesztelés végére jellemzően a nehéz modulokat kapták.

11. táblázat. A tanulók számának megoszlása a különböző útvonalak között évfolyamonkénti bontásban

Tanulók		Útvonalak			
Évfolyam	N	ÁKK	ÁKÁ	ÁNÁ	ÁNN
5	33	27	5	1	0
6	41	21	14	5	1
7	31	21	4	6	0
8	48	17	8	14	9

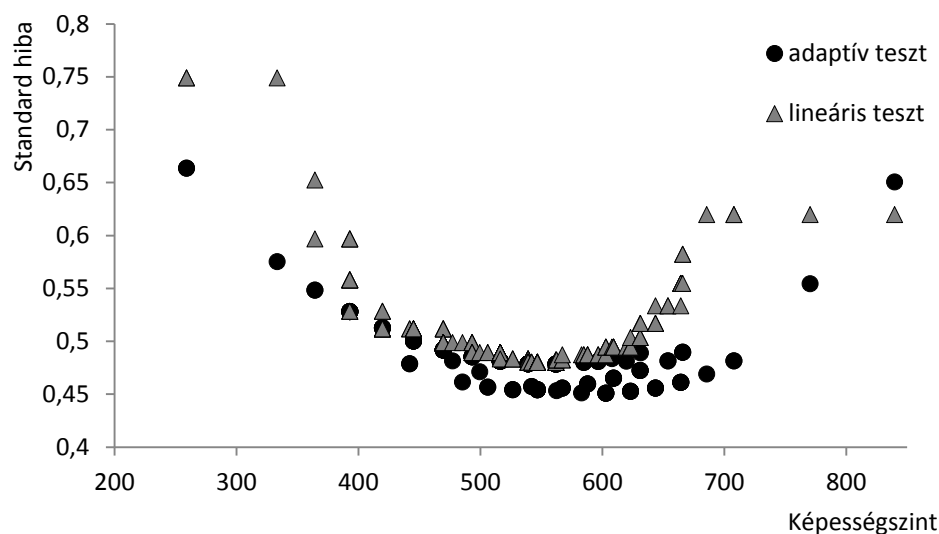
III.2.4.5 A tesztelés során kinyert információk és mérési hibák nagyságának összehasonlítása

A valószínűségi tesztelmélet segítségével mindegyik tanuló esetében kiszámítható, hogy az adott az adott teszt a különböző képességszinteken mekkora információt szolgáltat. A 29. ábra a tanulók képességszintjének függvényében ábrázolja a lineáris, illetve az adaptív teszten kinyert információk mennyiségét. Az adaptív teszt esetében átlagosan 63% volt a kinyert információ mennyisége, a lineáris teszt esetében pedig 61%. (Cohen $d=0,26$). A különbség elsősorban az átlagnál magasabb képességű tanulók esetében jelentős, helyenként 20-25%-kal magasabb az adaptív tesztből kinyert információ mennyisége.



29. ábra: Az adaptív és a lineáris tesztekkel kinyert információk mennyisége

A teszteken mért eredmények standard hibáit összehasonlítva a mérés hibája a teljes minta szintjén nagyobbak bizonyult ($t=6,11$, $p<0,01$; $d=0,48$) a lineáris tesztkörnyezetben ($SE=0,52$), mint az adaptív tesztnél ($SE=0,49$). A tanulók szintjére vetítve (30. ábra) a hiba nagysága összecseng a kinyerhető információ mennyiségével, a magasabb képességtartományban az adaptív teszt jóval alacsonyabb hibaértékeket mutatott, mint a lineáris teszt, azaz az adaptív tesztalgoritmus alkalmazásának előnye e képességtartományban a legjelentősebb.



30. ábra: Az adaptív és a lineáris teszt standard hibáinak alakulása a tanulók képességszintje függvényében

III.2.5 A pilotmérés eredményeinek összefoglalása

A problémamegoldó képesség mérésére fókuszáló adaptív és lineáris tesztkörnyezetben történő kutatás a tesztek mérési pontosságának, a kinyert információk mennyiségének és a becsült hibák nagyságának feltérképezésére és azok összehasonlítására irányult. A kutatás során a bevont tanulók 28 itemből álló lineáris, valamint ugyancsak 28 itemből álló, 1-2-3 szerkezetű háromszakaszos adaptív tesztet oldottak meg két fázisban, véletlenszerű kiosztásban.

Mindkét teszt reliabilitása megfelelő volt, azonban az adaptív teszt reliabilitása magasabb volt, mint a lineáris teszté, mely az adaptív tesztrendszer pontosabb képességszint meghatározását jelzi. A tanulók kétféle tesztkörnyezetben elért eredményei erősen korreláltak, a t-próba eredményei szerint nem különböztek egymástól szignifikáns mértékben.

Az évfolyamszintű összehasonlítás szerint sem volt különbség a kétféle tesztkörnyezetben elért eredmények között.

A vizsgálat kitért az adaptív és a lineáris teszteken elért helyes válaszok összehasonlítására. Az évfolyamszintű összehasonlítás szerint a 8. évfolyam kivételével mindegyik évfolyamon magasabb volt a helyes válaszok száma az adaptív tesztkörnyezetben, mint a lineáris teszten. A képességszint szerinti bontást vizsgálva, a képességszint növekedésével emelkedett a helyes válaszok száma, azonban az átlag alatti tanulók az adaptív teszten több helyes választ adtak, mint a lineáris teszten. A magas képességű tanulóknál fordítva alakult, a lineáris teszten tapasztaltnál kisebb arányban fordultak elő helyes válaszok.

Az adaptív teszt hat modulja összesen négy különböző teszt kiosztására adott lehetőséget. Az esetek kevéssel több, mint a felében a tanulók a könnyű modulokon haladtak végig, és a teszt végére is közel a minta fele a könnyű modulon végzett. A résztesztek személyszintű összehasonlítása szerint az adaptivitás előnye elsősorban a magas képességszintű tanulók elkülönítésénél mutatkozott meg.

Az adaptív teszt esetében a minta teljes szintjén több volt a kinyerhető információ mennyisége, és ezzel összhangban kisebb a standard hiba mértéke, mint a lineáris teszten. Az adaptív tesztkörnyezet előnye a magasabb képességtartományban volt számottevő, itt átlagosan 20-25%-kal több volt a kinyert információ mennyisége.

A pilotmérés eredményei alapján több hipotézis beigazolódott, a papíralapú problémamegoldó képességet mérő teszt itemei számítógépes adaptív tesztre való konvertálása hatékonyan és megbízhatóan alkalmazható volt a vizsgált korosztály körében. A kétféle tesztváltozaton elért eredmények nem különböztek jelentős mértékben egymástól. Az adaptív teszten az alacsonyabb képességszinteken magasabb volt a helyes válaszok aránya, a magasabb képességszinteken viszont alacsonyabb, mint a lineáris teszten.

A hatodik hipotézis nem igazolódott be, ugyanis a tanulók többsége a könnyű modulokon haladt végig, és végzett. Ennek az is lehet az oka, hogy az adaptív teszt szerkezete (1-2-3) a második szakaszban nem tartalmazott átlagos nehézségű modult, és a harmadik szakaszban nehezebb volt a közepes modulra való visszalépés lehetősége. Az adaptív teszt elkülönítő ereje így elsősorban a magas képességszintek elkülönítésében mutatkozott meg.

Az utolsó hipotézis részben igazolódott be, mivel az adaptív tesztből kinyerhető információ összességében magasabb volt, mint a lineáris tesztből kinyerhető, ez a különbség azonban elsősorban a magas képességszinteken volt jelentős. A mérési hiba is hasonlóan

alakult, összességében a lineáris teszt esetében volt a becsült hiba mértéke nagyobb, azonban képességszintenkénti bontásban vizsgálva, elsősorban a magas képességtartományban volt jelentős az eltérés.

IV. SZÓOLVASÓ KÉSZSÉG MÉRÉSE ADAPTÍV ÉS LINEÁRIS TESZTEKKEL

A kutatás harmadik szakaszában szóolvasó készséget mérő tesztet konvertáltunk online adaptív és lineáris formátumra, és egy pilot, valamint két nagymintás mérést valósítottunk meg.

IV.2 A SZÓOLVASÓ KÉSZSÉG KRITÉRIUMORIENTÁLT MÉRÉSÉT LEHETŐVÉ TEVŐ TESZTSOROZAT TOVÁBBFEJLESZTÉSÉNEK LEHETŐSÉGEI

A szóolvasó készség kritériumorientált mérését lehetővé tevő rendszer továbbfejlesztése több szempont szerint is megvalósítható. A számítógépek elterjedésével megteremtődtek a feltételek a számítógép alapú tesztek adta lehetőségek minél szélesebb körben való kihasználására (*Molnár, 2010*), mint például a szóolvasó készség mérését megvalósító feladatok kapcsán motiválóbbr környezet (*Thompson és Way, 2007*), azonnali visszajelzés lehetősége, a személy képességszintjéhez illesztett adaptív tesztelési technikával pontosabb vagy ugyanazon pontosság mellett az eredeti rendszer működéséhez képest jóval rövidebb idő alatt történő képességszint meghatározás (*Eggen, 2007; Thompson és Weiss, 2011; Magyar, 2014a; 2014b*). Utóbbi feltétele a feladatok közös nehézségi skálára hozása, ami az eredeti rendszer struktúrájának újragondolását vonta maga után (*Magyar és Molnár, 2014; Magyar, 2014b*).

IV.3 ADAPTÍV TESZTELÉS ALKALMAZÁSA A SZÓOLVASÓ KÉSZSÉG MÉRÉSÉRE

A rendszer átstrukturálásánál alapvető szempont volt, hogy ne változzon a rendszer alapvető struktúrája, vagyis a szóolvasó készség vizsgálata négy szempont szerint történjék (címszóolvasás, toldalékos szóolvasás, 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 képességszintjéről. Ennek következtében amennyiben itemalapú adaptivitást alkalmaztunk volna, nem lehetett volna biztosítani a négy szempont együttes tesztben történő megjelenését, ezért a többszakaszos adaptív rendszerek alkalmazása 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 papíralapú eredményekre, nélkülözhetetlen volt a feladatok ismételt, lineáris, de már számítógép alapú bemérése.

Az adaptív tesztrendszer fejlesztésének lépései a következők voltak: (1) a papíralapú feladatok digitalizálása, feltöltése az eDia rendszerbe, (2) az eredeti lineáris rendszer újrastrukturálása, horgonyrésztesztek alkalmazásával, (3) az online lineáris tesztek felvétele, majd az eredmények alapján a feladatok skálázása, (4) az eredeti papíralapú rendszer kötöttségeihez leginkább igazítható, ugyanakkor a 21. században elvárt mérés-értékelési céloknak leginkább megfelelő adaptív rendszer kiválasztása és a közös nehézségi skálán jellemzett feladatok többszakaszos adaptív tesztrendszerre alakítása, (5) a rendszer viselkedésének kismintán történő tesztelése.

IV.4 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ FELADATBANK KIALAKÍTÁSA ÉS PARAMÉTEREZÉSE

IV.4.1 A papíralapú feladatok digitalizálása


A papíralapú rendszer online formára történő konvertálása során az összes feladatot megtartva igyekeztünk azokat minél kevesebb változtatással digitalizálni, azaz az eDia rendszerbe történő feltöltéssel alapvetően nem változtattuk meg a feladatok típusát. A papír- és számítógép alapú feladatok megjelenítését szemlélteti a négy résztesztből a 31.a-d. ábra.

Amelyik szó jelentése illik a képre, annak betűjelét **karikázd be**, amelyik nem, annak betűjelét **húzd át!**

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!

a) épület	<input type="checkbox"/>
b) kertés	<input type="checkbox"/>
c) ház	<input type="checkbox"/>
d) ól	<input type="checkbox"/>

1.



Előző



épület	<input type="radio"/> Igen	<input type="radio"/> Nem
kertés	<input type="radio"/> Igen	<input type="radio"/> Nem
ház	<input type="radio"/> Igen	<input type="radio"/> Nem
ól	<input type="radio"/> Igen	<input type="radio"/> Nem

Következő

31.a ábra: Címszóolvasás feladat (papíralapú és számítógépes formátumban)

Amelyik szó jelentése illik a képre, annak betűjelét karikázd be, amelyik nem, annak betűjelét húzd át!

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

Előző Következő

31.b ábra: Toldalékos szóolvasás feladat (papíralapú és számítógépes formátumban)

Hasonlítsd össze a kisbetűs és a nagybetűs szavakat!
Amelyik kisbetűs szónak van hasonló jelentésű nagybetűs megfelelője, annak betűjelét **karikázd be**, amelyiknek nincs, annak betűjelét **húzd át!**

Hasonlítsd össze a kisbetűs és nagybetűs szavakat!
Döntsd el, van-e a kisbetűs szavaknak nagybetűs megfelelője!
Ha van, kattints a "Van" felírra, ha nincs, kattints a "Nincs" felírra!

a) varázslat
b) fénylik

7.

FIGYELEM
CSODA

FIGYELEM CSODA

varázslat Van Nincs
fénylik Van Nincs

Előző Következő

31.c ábra: Szinonimaolvasás feladat (papíralapú és számítógépes formátumban)

Keresd meg, hogy a jelentésnek megfelelő szó van-e a nagybetűs szavak között! Ha van, **karikázd be** a jelentés betűjelét, ha nincs, **húzd át!**

Olvasd el a nagybetűs szavakat!
Döntsd el, hogy a kisbetűs jelentéseknek megfelelő szó van-e a nagybetűs szavak között!
Ha van, kattints a szókapcsolatok után lévő "Van" felírra, ha nincs, kattints a "Nincs" felírra!

a) 'hirtelen, erősen megfog'
b) 'a tananyag valamely ismeretkört felölelő része'
c) 'olyan mennyiségű'
d) 'az állam teljes hadereje'
e) 'kellemetlen érzés'

5.

ANNYI, HADSEREG, MEGRAGAD,
TANTÁRGY, UTÓBB

ANNYI, HADSEREG, MEGRAGAD, TANTÁRGY, UTÓBB

hirtelen, erősen megfog Van Nincs
a tananyag valamely ismeretkört felölelő része Van Nincs
olyan mennyiségű Van Nincs
az állam teljes hadereje Van Nincs
kellemetlen érzés Van Nincs

Előző Következő

31.d ábra: Szójelentés olvasás feladat (papíralapú és számítógépes formátumban)

A papíralapú teszten karikázással és áthúzással kellett megadni a diákoknak a választ, azaz, hogy az adott szóval címszóolvasás esetében például jellemezhető-e a mutatott kép vagy sem. Ezt számítógépes formában például az igen vagy nem gombra kattintva tudták megtenni a diákok. Ezzel a válaszadási módszerrel a számítógépes verzió csakis akkor engedte tovább a tanulót, ha bejelölte az adott itemről való döntését, és ezáltal az itemkihagyás lehetősége nem volt megengedett. Miután az eDia rendszer alapaxiómái közé tartozik, hogy minden egyes feladatban a válaszmezőket üresen hagyva 0%-os, a helyes megoldást megadva 100%-os teljesítményt regisztráljon a rendszer, ezért a feladatokat itemenkénti (szavankénti) értékelést alkalmazva nem lehetett jelölőnégyzetes feladatként digitalizálni. (Jelölőnégyzetes feladatként a diákok mindegyik példa esetében az 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 papíralapú formával analóg módon – minden egyes szóról döntést kellett hoznia a diáknak, ezért az igen-nem-es megoldás mellett döntöttünk.

Egy másik jelentős változtatás a teszt egészének megjelenítésében volt. Míg papír alapon a diákok egyszerre több feladatot láttak (részteszttől függően 10-25 darabot) egy tesztoldalon, addig számítógépes formában egyszerre mindig csak egyet. Ennek előnye az volt, hogy a környező itemek nem befolyásolhatták a tanulót a döntésben. A visszalépés lehetősége csak részteszten belül volt engedélyezett.

A teszt pontozása a papíralapú rendszer „szókészlet mutató”-jával azonosan történt, azaz első körben azt vizsgáltuk, hogy a tesztben szereplő szavak hány százalékánál adott helyes választ a tanuló. Minden egyes szóról meghozott döntés nulla vagy egy pontot ért, azaz a címszóolvasás 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.

IV.4.2 A feladatok paraméterezéséhez szükséges lineáris tesztrendszer kialakítása

Az adaptív rendszer felépítésének alapvető feltétele, hogy a rendszerben lévő minden egyes feladat viselkedése, legalább nehézségi, illetve diszkriminációs indexe a többi feladat fényében ismert legyen. A feladatok e mutatóinak meghatározásakor, azaz a feladatok paraméterezése során egyrészt a horgony feladatok hiánya, másrészt az esetleges médiahatás

(Hülber és Molnár, 2013) megléte miatt nem alapozhattunk a korábbi papír alapú adatfelvétel eredményeire, újabb, nagymintás lineáris tesztekkel történő számítógép alapú adatfelvétellel volt szükség, ahol az egyes tesztek között horgony feladatok biztosították a feladatok közös skálára történő konvertálását.

A 850 feladat lineáris 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 horgony itemek rendszerbe történő illesztése (Yousfi és Böhme, 2012); (3) a feladat tesztben elfoglalt pozíciójából (item position effect; Lu, 2010; 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 négyzeten alapuló elrendezése volt a legmegfelelőbb (Frey és Hartig, 2009). A papíralapú rendszer résztesztjeiből klasztereket képeztünk, a klaszterekből bookleteket 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 (teszt első negyedében, második negyedében, harmadik negyedében és a negyedik negyedében) pozícióban szerepelt, még hozzá pontosan egyszer.

A feltételeknek megfelelően 40 bookletet alakítottunk ki, melynek összefűzését, egymáshoz való viszonyát a 32. 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

32. ábra: A BIBD design (balanced incomplete block design; sárga: címszó (c), kék: szójelentés (j), narancs: toldalékos (t), piros: szinonima (s))

IV.4.3 A feladatok paraméterezését szolgáló kutatás mintája

A rendszer komplexitása miatt a feladatok paraméterezéséhez szükséges minta meghatározása során több szempont figyelembevételére is szükség volt. Egy feladat biztos paraméterezése csak akkor valósítható meg, ha arról legalább 300 adat áll rendelkezésünkre, illetve a korábban említett szempontok szerinti elemzés megvalósításának feltétele, hogy szempontonként és feladatonként is legalább 100-100 adatunk legyen (Yousfi és Böhme, 2012). 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 egyes klaszter négy bookletben szerepelt, minden egyes klaszter vonatkozásában minimum 400 diák válasza állt rendelkezésre.

A magyar iskolarendszer jelentős mértékű szelektivitása miatt fennállt az adatok erős klasztereződésének veszélye, ezért a tesztek kiközvetítését egyrészt egy, a magyar iskolarendszert reprezentatívan leképező 40 iskolából álló mintából véletlenszerűen alakítottuk ki, másrészt a kutatásban résztvevő 4-5. évfolyamos osztályokban a diákok véletlenszerűen kapták a 40 booklet egyikét. A 4-5. 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 olvasástanítás. A felső tagozaton viszont számos olyan tantárgy lép be (irodalom, történelem), melyek hosszabb szövegek elolvasását és megértését követelik meg a tanulóktól. Másrészt épp a korábbiakból adódóan – a papíralapú mérések alapján – mind a 4., mind az 5. évfolyamon vannak már diákok, akik képességfejlettségi szintje vetekszik a magasabb képességszintű középiskolásokéval, míg 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.

IV.4.4 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 egyes diáknak egy tanítási óra állt rendelkezésére.

Az adatok elemzése és az itemek paraméterezése a ConQuest program felhasználásával több dimenziós parciális kredit modellel (*Masters, 1982*) történt. Miután az alkalmazott több dimenziós parciális kredit modell a Rasch-modell egy továbbfejlesztett változata, ami egy paraméter szerint, a nehézségi paraméter szerint rendezi az elemeket a képességskálára, így a rendszerben lévő bármely feladat nehézségi indexe összehasonlíthatóvá vált bármely másik rendszerben szereplő feladatével (*Masters, 1982*).

A többdimenziós elemzést azt tette szükségessé, hogy bár a rendszer a szóolvasó készség, mint egy egységes konstruktum mérését valósítja meg, de az eredeti rendszerben kialakított eszközök, résztesztek kifejlesztése különböző mérési szempontok mentén történt. Ennek megfelelően a címszóolvasást, a toldalékos szóolvasást, 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 parciális kredit modellt alkalmaztuk az elemzésekben. Ennek oka, hogy bár a szavak szintjén külön kellett döntést hozni a diákoknak, mégis a feladatok bizonyos szinten összekötötték az egy feladaton belül szereplő szavakat. A teljes teszt EAP/PV reliabilitása 0,91. Az egyes dimenziókon mért reliabilitásmutatókat a *12. táblázat* mutatja.

12. táblázat: A teszt EAP/PV reliabilitása dimenziónkénti bontásban

Részteszt	EAP/PV reliabilitás
Címszóolvasás	0,74
Toldalékos szóolvasás	0,89
Szinonimaolvasás	0,88
Szójelentés olvasás	0,80

Átlagosan a címszóolvasás feladatok voltak a legkönnyebbek, majd ezt követték a toldalékos szóolvasás feladatai, a szinonima és szójelentés feladatok pedig hasonló nehézségűeknek bizonyultak. A feladatok nehézségi indexei a diákok teljes képességskáláját

lefedték, tehát a 850-ból álló feladatbank alkalmasnak bizonyult a különböző képességszintű tanulók szóolvasó készségének mérésére (33. ábra).

	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
2				647.4 754.4
				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	57.4 120.4 188.4 293.2 343.2
1	XXX	XXXX	XXX	213.4 277.2 326.2 337.2 380.2
	XXXX	XXXXX	XXX	27.4 217.4 221.4 245.4 290.2
	XXXX	XXXX	XXXX	22.4 127.4 195.4 203.4 252.2
	XXXXXX	XXXXX	XXXXXX	82.4 84.4 115.4 137.4 157.4
	XXXXXXXX	XXXXXX	XXXXXXXX	81.4 114.4 164.4 211.4 230.4
	XXXXXXXX	XXXXXXXX	XXXXXXXX	89.4 151.4 261.2 264.2 284.2
	XXXXXXXX	XXXXXX	XXXXXXXX	9.4 146.4 201.4 273.2 286.2
	XXXXXXXX	XXXXXXXX	XXXXXXXXXXXXXXXX	105.4 154.4 204.4 206.4 258.2
	XXXXXXXX	XXXXX	XXXXXXXX	117.4 119.3 143.4 196.4 219.4
0	XXXXXXXX	XXXXX	XXXXXXXX	35.4 39.4 87.4 172.4 173.4 184.4
	XXXXXXXX	XXXXX	XXXXXXXX	10.4 30.4 111.4 131.4 132.4
	XXXXXXXX	XXXXX	XXXXXX	6.4 24.4 41.4 62.4 67.4 136.4
	XXXXXX	XXXXX	XXXXXX	108.4 121.4 127.3 150.4 182.4
	XXXXXX	XXXX	XXXXXX	1.4 45.4 54.3 63.4 70.4 78.4
	XXXXXX	XXX	XXXX	27.3 47.4 55.4 65.4 96.4 103.4
	XXXX	XXXX	XXXX	16.4 42.4 91.4 120.3 152.4 157.3
	XX	XXXX	XXXX	23.4 34.4 99.4 102.4 109.4 138.4
	XX	XXX	XXX	22.3 50.4 75.4 92.4 107.4 110.4
-1	XX	XXX	XXXX	4.4 36.4 38.4 40.4 60.4 74.4
	XX	XXX	XX	3.4 53.4 59.4 64.4 66.4 77.4
	X	XX	X	2.4 5.4 27.2 28.4 35.3 56.4 58.4
	X	XXX	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		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
	X			8.4 9.3 18.4 28.3 33.4 34.3 42.2
-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
				2.2 14.2 16.2 46.2 55.1 56.2
-3				1.2 17.1 25.2 34.2 39.2 40.2
				8.1 8.2 9.1 10.1 11.1 11.2 12.1
				1.1 2.1 3.1 4.1 5.1 5.2 6.1 7.1

33. ábra: A feladatok nehézségi szintje a diákok képességszintje fényében – többdimenziós személyitem térkép [címszóolvasás (Címszó), szinonimaolvasás (Szin), szójelentés olvasás (Szó) és toldalékos szóolvasás (Told) dimenziókban]

A többdimenziós parciális kreditmodellel becsült feladatparamétereket 500-as átlagú és 100-as szórású skálára konvertáltuk. A feladatok résztesztenkénti átlagos nehézségi indexét és azok szórását dimenzióként mutatja a 13. táblázat.

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

Dimenziók	Átlag (kéességpont)	Szórás (kéességpont)
Címszóolvasás	397	34
Toldalékos szóolvasás	473	66
Szinonimaolvasás	588	65
Szójelentés olvasás	603	46

IV.5 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ TESZTRENSZER KIALAKÍTÁSA

Az eredeti papíralapú teszt-sorozat tíz különböző tesztváltozatot tartalmazott, minden tesztváltozatban 85 feladat szerepelt. A feladatok dimenziókénti megoszlását mutatja a 34. ábra.



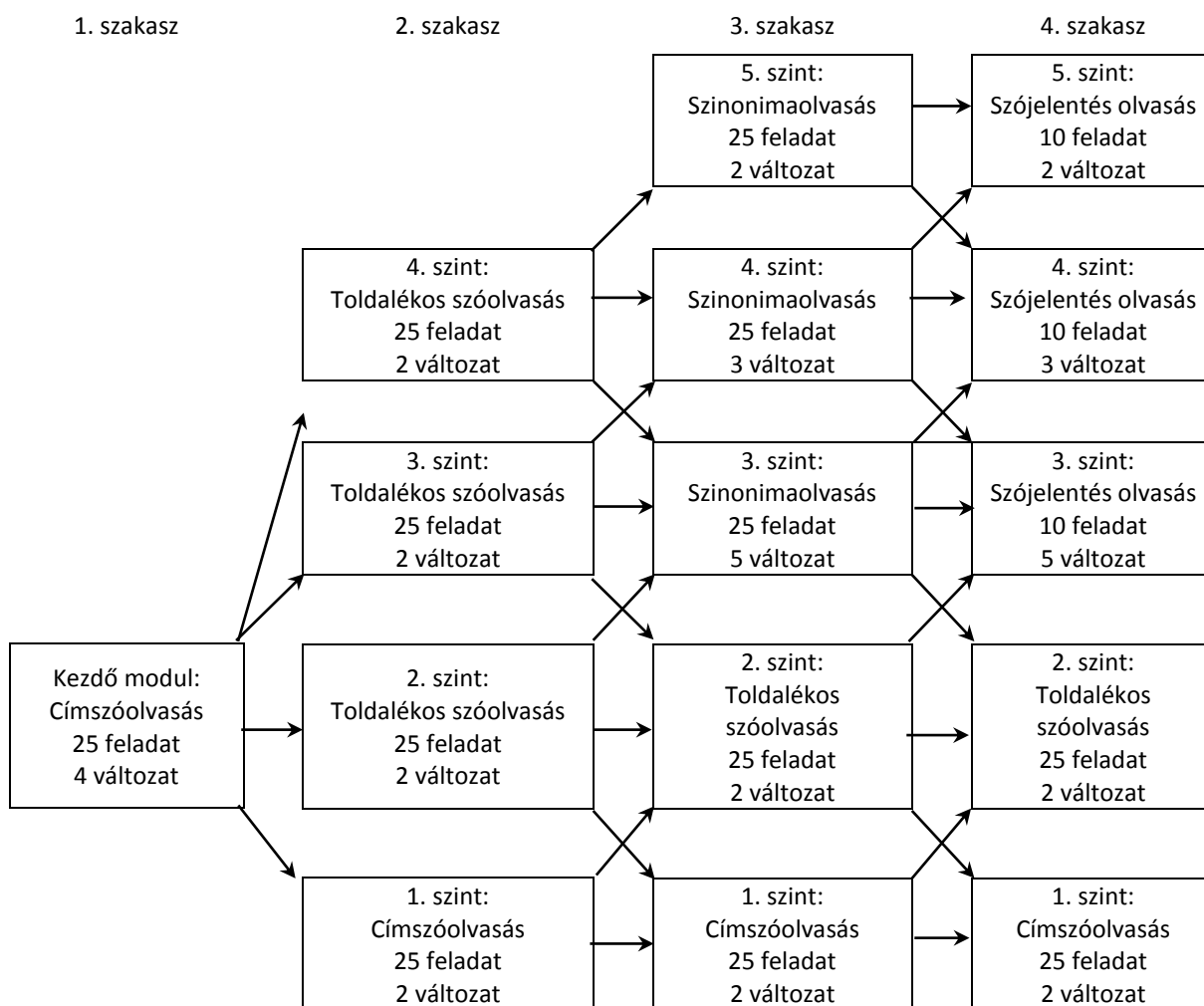
34. ábra: Az eredeti lineáris tesztek felépítése

Az adaptív teszrendszer összeállításánál fő szempontként szerepelt az eredeti teszt szerkezet megtartása, ezért a többféle adaptív elrendezés közül a négyszakaszos, öt különböző szintet megkülönböztető teszrendszer 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 készségfejlettségi szinten különíti el egymástól a diákokat, illetve ezzel párhuzamosan az eltérő nehézségi szintű feladatokat. A teszrendszer kezdő, mindenki által megoldásra kerülő klasztere egy közepes nehézségű címszóolvasás feladatokat tartalmazó klaszter. Ez a klaszter négy azonosan felépített, azonos viselkedésű klaszterből kerül véletlenszerűen kiválasztásra, ezzel

kiküszöbölve a rendszer többszöri használatából adódó gyors elévülést, a tesztre történő emlékezés minimalizálását. E klaszter feladatainak nyújtott teljesítmény függvényében lépnek a tanulók nehezebb vagy könnyebb feladatokat tartalmazó klaszterekre. A címszó, toldalékos szóolvasás és a szinonimaolvasást tartalmazó klaszterek 25-25 feladatot tartalmaznak, a szójelentés-felismerés klaszterek 10 feladatot.

Az elágazási szabály kialakítása során is törekedtünk a *Nagy József* által kidolgozott rendszer megtartására, ezért az elágazási szabályt is ezzel összhangban számítottuk ki. Az első szintről legalább 60%-os teljesítmény elérése szükséges a második szintre való lépéshez, a második szintről 70%-os eredménnyel lehet a harmadik szintre lépni, a harmadik szinten 80%-os teljesítmény az elvárt követelmény, és a negyedik szintről 90% teljesítése után lehet az ötödik szintet elérni.



35. ábra: A szóolvasó készség mérésére alkalmas négyszakaszos adaptív tesztrendszer szerkezete

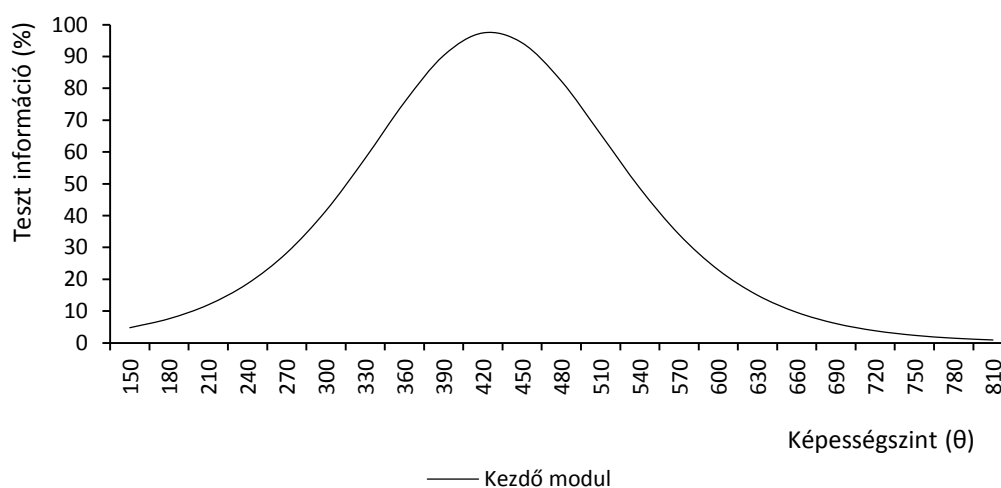
Azzal, hogy a különböző szinteken különböző nehézségű feladatok vannak, a rendszer egyrészt segíti a gyengébb tanulókat, mivel számukra könnyebb a feljebbjutás, másrészt kihívást állít a magasabb szinten lévők elé, mert a magasabb szinteken egyre nehezednek a feladatok. Mivel az eredeti tesztrendszer összes feladatát integráltuk, ezért a tesztrendszer minden egyes szakaszának minden egyes szintjén több azonos felépítésű és nehézségű klaszter szerepel, amelyek közül véletlenszerűen történik annak kiválasztása, hogy melyiket kell a diáknak az adott szakasz adott szintjén megoldani (35. ábra).

Az azonos szinten lévő modulok átlagos nehézségi szintjében nincs szignifikáns különbség. Az egyes szinteket jellemző átlagos képességszinteket mutatja az 14. táblázat.

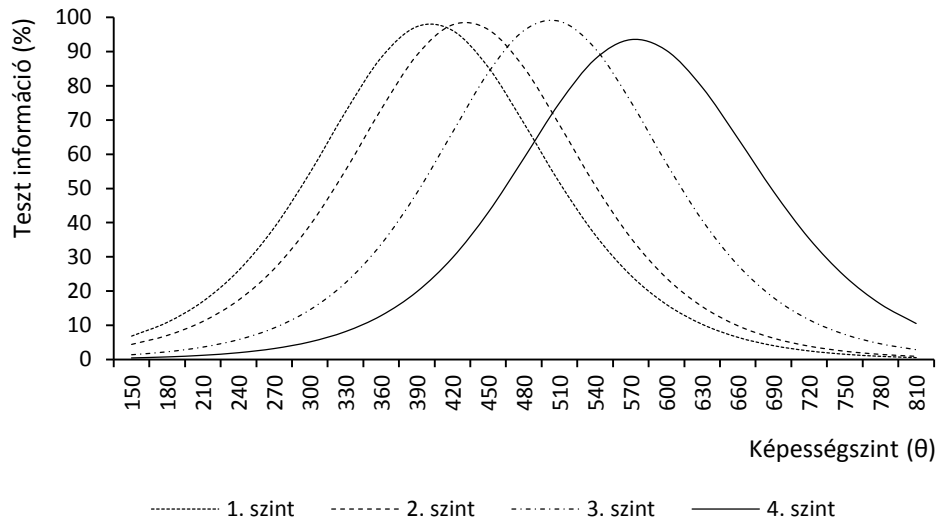
14. táblázat: A különböző szinteken megjelenő feladatok átlagos nehézségi indexei (átlag: 500, szórás: 100)

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

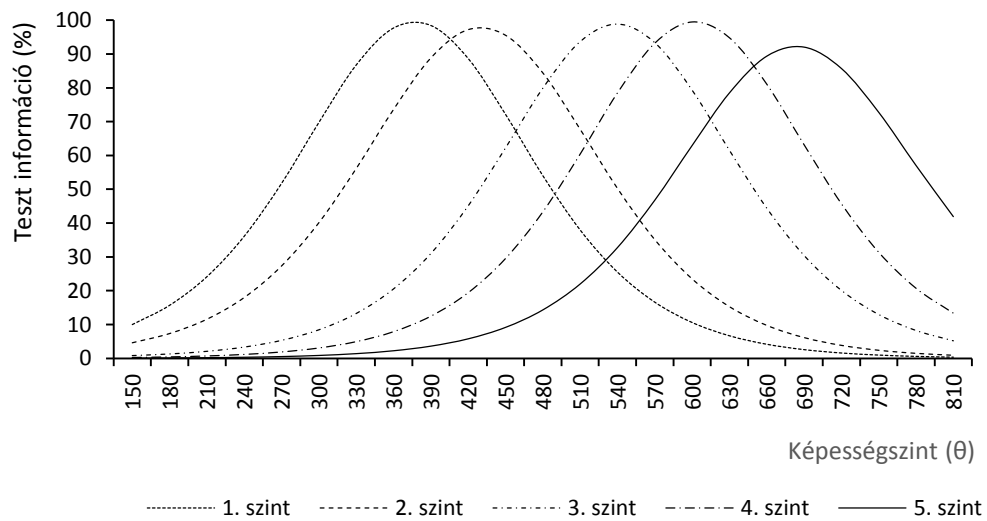
Az egyes szakaszok modul információs függvényeit ábrázolja a 36.a-d ábra. A későbbi szakaszok egyre nagyobb képességszintet fednek le, vagyis egyre több információt szolgáltatnak a képességskála egyre szélesebb tartományáról.



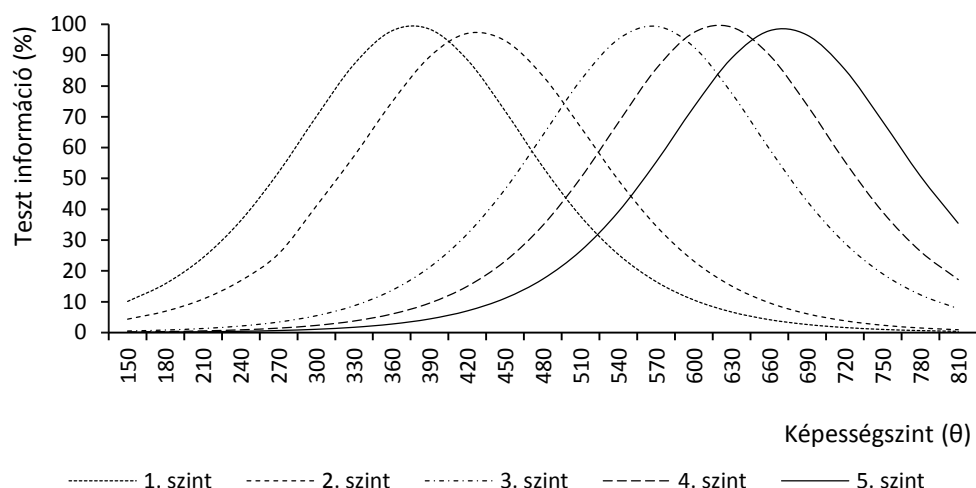
36.a ábra: A kezdő modul információs függvénye



36.b ábra: A második szakasz moduljainak információs függvényei



36.c ábra: A harmadik szakasz moduljainak információs függvényei



36.d ábra: A negyedik szakasz moduljainak információs függvényei

A rendszer alkalmazásával közös képességskálán jellemezhetőek mind a feladatok, mind a tanulók készségszintjük függvényben (Baker, 2001). Minden egyes tanuló képességskálán elfoglalt helye megmutatja, hogy melyek azok a feladatokban szereplő 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).

IV.6 A SZÓOLVASÓ KÉSZSÉG ONLINE ADAPTÍV MÉRÉSÉT MEGVALÓSÍTÓ TESZTRENSZER KIPRÓBÁLÁSA

Az adaptív rendszer megfelelő működését pilotmérés keretében térképeztük fel. A mérés célja az első két hipotézis beigazolása, miszerint (1) a szóolvasási készség papíralapú tesztsorozatának 850 feladatából kifejleszhető többszakaszos adaptív tesztelésre alkalmas teszrendszer, (2) a teszrendszer hatékonyan és megbízhatóan alkalmazható a 4-5. évfolyamos korosztály szóolvasó készség kritériumorientált diagnosztikus mérésére.

IV.6.1 Minta, adatfelvétel

A kismintás adatfelvétel 2014 tavaszán, 154 általános iskolás tanuló részvételével zajlott. A tanulók évfolyam szerinti eloszlását a 15. táblázat tartalmazza.

Az adatfelvétel a tanulók saját iskolájában, a saját internethálózatukon keresztül az eDia rendszer segítségével történt. A feladatok megoldására 45 perc állt a tanulóknak.

rendelkezésére. A rendszerbe a saját mérési azonosítójukkal léptek be a tanulók, és a tesztelés végén azonnali visszajelzést kaptak teljesítményükről.

15. táblázat: A minta évfolyamonkénti eloszlása

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

IV.6.2 Eredmények és diszkusszió

A pilot kutatás során megbízhatónak bizonyult a rendszer, a személyszeparációs reliabilitásmutató mind a teljes rendszer szintjén (0,88), mind az egyes dimenziók szintjén (16. táblázat) megfelelő volt.

16. táblázat: A teszt EAP/PV reliabilitása dimenziókénti bontásban

Dimenzió	EAP/PV reliabilitás
Címszóolvasás	0,75
Toldalékos szóolvasás	0,89
Szinonimaolvasás	0,87
Szójelentés olvasás	0,88

A tanulók személy-ítem térképét mutatja a 37. ábra. Mind a négy dimenzió vonatkozásában a minta készségeloszlása közelíti a normáeloszlást. Az ábra jobb oldalán az ítemek lefedik a minta készségtartományát, tehát a feladatok megfelelőek voltak a vizsgált korosztály készségszintjének.

	Címszó	Szin	Szój	Told
4				206.2
				426.4
3				202.2 234.2 336.2 360.4
				403.3 407.3
		X		209 266 328.3 342.4 346.4 353.3
		X		332.5 333.4 337.3 344.3 383.3
		X	X	354.4 366.4 402.4 415.4 418.5
2		X	X	295.2 357.4 364.4 370.4 374.3
		XX	XX	333.3 350.4 354.3 359.3 360.3
		XX	X	338.2 339.3 345.3 346.3 348.4
		XXX	XXX	55.3 236.2 297.2 303.2 318.2
		XX	XXX	213.2 219.2 221.2 237.2 331.4
	X	XXX	XX	19.3 47.2 50.2 191.2 322.2 323.2
1	XX	XXX	XXXX	X 127.4 137.2 188.2 268.2 275.2
	XXX	XXXX	XXXXX	XXX 57.4 129.3 150.3 176.3 185.2
	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 91.3 107.3 108.3 113.3 165.3
	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 8.2 71.3 75.3 82.2 99.3 111.2
	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 27.3 54.3 64.2 74.2 90.2 97.4
	XXXXXXXXXX	XXXX	XXXX	XXXXXXXXXX 21.3 28.3 41.3 59.3 76.3 89.3
0	XXXXXXXXXX	XXX	XXXX	XXXXXXXXXX 14.3 55.2 56.2 60.3 70.2 80.2
	XXXXXXXXXX	XXX	XXXX	XXXXXXXXXX 4.3 13.2 18.3 34.3 38.3 44.3
	XXXXXX	XXX	XXX	XXXXXXXX 19.2 20.2 30.2 40.2 61.2 69.3
	XXXXXX	XXX	XXX	XXXXXXXX 37.2 39.2 42.2 43.3 46.3 53.2
	XXXX	XX	XXXX	XXX 2.2 5.2 6.2 9.2 12.2 15.2 25.2
	XX	XX	XX	X 11.2 21.2 23.2 24.2 26.2 28.2
-1	X	XX	XX	XX 1.2 29.2 32.2 34.2 35.2 36.2 541
	X	XXXX	XX	3.2 7.2 10.2 22.3 33.2 38.2 51.2
		XXX	XX	X 31.3 49.2 57.2 72.2 87.3 92.2
		XXX	XXX	14.2 27.2 52.2 71.2 85.2 97.2
		X	XX	16.2 18.2 59.2 97.1 99.2 100.2
-2		XX		4.2 17.2 31.2 46.2 48.2 52.1 252
		X	X	43.2 57.1 66.2 75.2 79.2 101.2
		X	X	22.2 27.1 69.2 85.1 89.2 121.2
				44.2 45.2 117.2 118.2 138.2
				87.2 115.2 149.2 157.2 176.2
				146.2 177.2 180.2 223.1 225.1
-3				169.2 173.2 178.2 204.1 216.1
				190.1 220.1 239.1 242 249.1
		X		107.2 185.1 191.1 236.1 264.1
				280.1 303.1 307.1 313.1
				84.2
-4				

37. ábra: A pilot adaptív adatfelvétel alapján kirajzolható többdimenziós személy-ítem térkép [címszóolvasás (Címszó), szinonima (Szin), szójelentés (Szój) és toldalékos szóolvasás (Told)]

A tesztrendszer a kisiskolás diákok szóolvasási készségfejlettségének diagnosztizálására is alkalmasnak bizonyult. A becsült képességszintek képességpont és százalékos teljesítményének átlagát évfolyamonkénti összehasonlításban mutatja a 17. 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

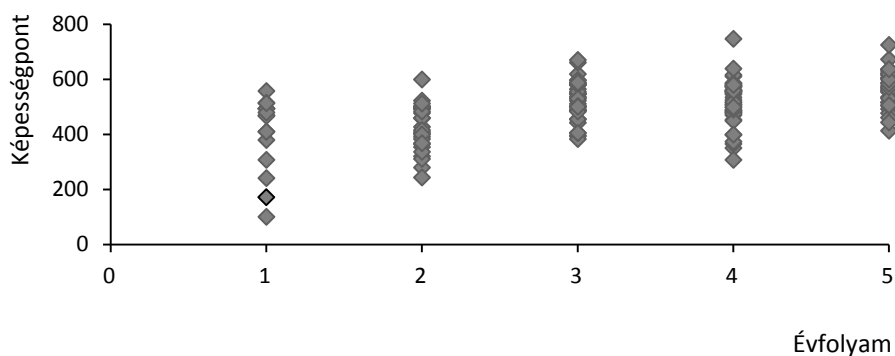
a feladatokat, amelyek megoldási sikeressége a legtöbb információval szolgál a diák készség, illetve képességszintjéről (Molnár, 2013). Ebben az esetben ez az elméleti érték a második szinten (átlagos teljesítmény 430 pont, 13. táblázat) 70% volt, amit igazoltak az empirikus adatok is (első és második 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% körülinek kellett lennie, amit alátámasztanak az empirikus adatok.

17. táblázat. A teszten elért készség szintek átlaga és szórása évfolyamonkénti bontásban

Évfolyam	N	Min (pont)	Max (pont)	Átlag (pont)	Szórás (pont)	Átlag (%)	Szignifikáns különbségeket mutató évfolyamok
1.	16	100	558	407	134	70	
2.	28	280	599	426	84	70	{1,2}<{3, 4}<{5} (F=15,61, p<0,01)
3.	38	384	671	523	72	80	
4.	42	308	748	517	84	78	
5.	30	462	726	565	70	82	

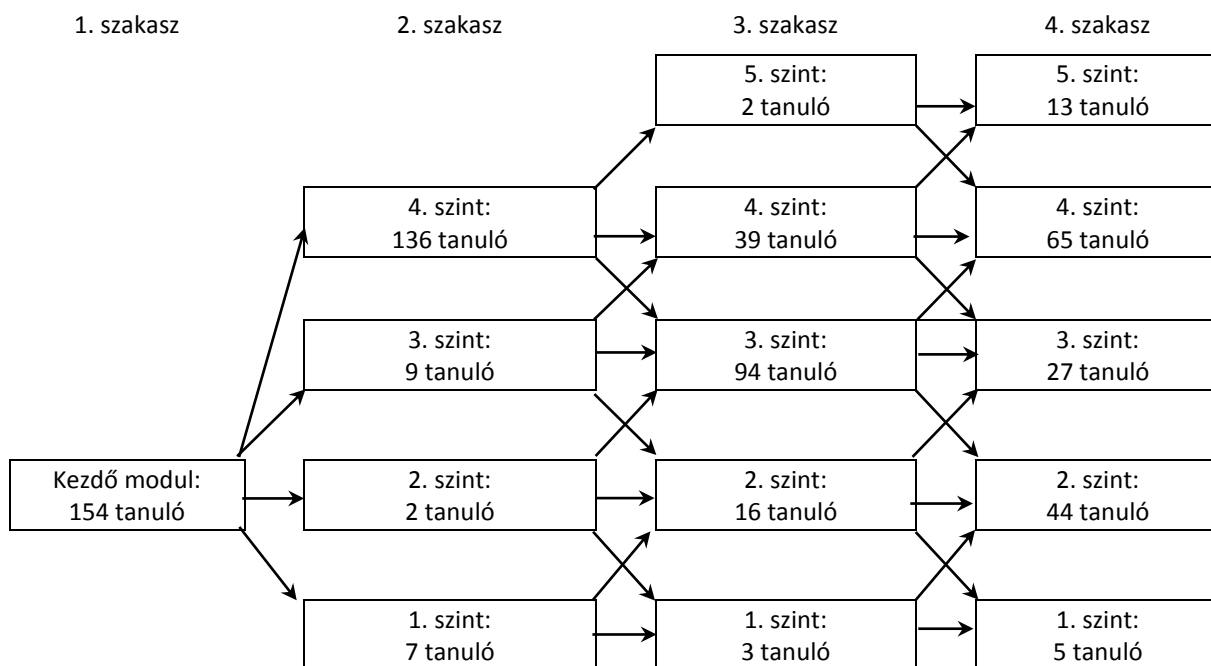
A rendszer mindamellett, hogy alkalmasnak bizonyult az évfolyamok között meglévő átlagos különbségek detektálására, az évfolyamokon belül megjelenő különbségek diagnosztizálását is lehetővé tette. Mind a legalacsonyabb képességszintű első és második évfolyamos diákok, mint a legmagasabb képességszintű 5. évfolyamos diákok készségszint szerinti elkülönítésére alkalmas volt (38. ábra).

A kismintás mérés eredményei alapján a tesztrendszer tág intervallumban, megfelelő információ kinyerése mellett megvalósította a diákok készségszintjéhez illesztett adaptív tesztelést. Az évfolyamokon belül megjelenő készségszintbeli különbségek alapján, illetve az egyes évfolyamok átlagos teljesítményei alapján megállapítható, hogy a feladatok skálázásához kijelölt 4-5. évfolyam és az ott tanuló diákok által lefedett tág készségszint-intervallum megfelelő volt a feladatok skálázásához.



38. ábra: A tanulók készségi szint szerinti eloszlása évfolyamonkénti bontásban

A diákszintű eredmények is alátámasztották a rendszer helyes és tervezett működését. Az adaptív teszt összesen 40 különböző klasztert tartalmazott, melyekből a kismintás adatfelvétel során 38 került kiosztásra. A szakaszokon belül a tanulók azonos tulajdonságokkal jellemezhető klaszterenkénti eloszlását mutatja gyakoriság szerinti bontásban a 39. 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.



39. ábra: A tanulók gyakorisági és képességi szint szerinti eloszlása a szakaszokon és a modulokon belül

IV.6. A PILOTMÉRÉS EREDMÉNYEINEK ÖSSZEFOGLALÁSA

A kismintás kutatás eredményei alapján megállapítható, hogy a rendszer mind évfolyamonkénti, mind diákonkénti bontásban helyesen működött. Az alacsonyabb készségi szintű diákok tipikusan a könnyebb, a magasabb készségiúek a nehezebb klasztereket kapták a tesztelés során, amivel az adatfelvétel során kinyert információ mennyisége javult, miután minden egyes diák a készségi szintjéhez relatív közel álló feladatokat kapott a teszt utolsó moduljában. Az utolsó két modulon 31 tanulónál nem változott a kapott modul szintje a harmadikról a negyedik szakaszba való lépésnél, ami a tanulók egy ötödét jelenti, tehát mindenképpen indokolt volt a négy szakasz alkalmazása.

A mérés beigazolta az első és második hipotézist, mely szerint (1) a szóolvasási készségi papíralapú teszt sorozatának 850 feladatából kifejleszthető többszakaszos adaptív tesztelésre alkalmas tesztrendszer, (2) a tesztrendszer hatékonyan és megbízhatóan alkalmazható az 1-5. évfolyamos korosztály szóolvasó készségi kritériumorientált diagnosztikus mérésére. A következőkben a többi hipotézis beigazolása volt a célunk, és nagymintás vizsgálat keretében vizsgáltuk meg, hogy az adaptív tesztrendszer hatékonyabban alkalmazható-e, mint a lineáris teszt a vizsgált készségi terület mérésére.

V. SZÁMÍTÓGÉP ALAPÚ ADAPTÍV ÉS LINEÁRIS TESZTELÉS ÖSSZEHASONLÍTÓ HATÉKONYSÁGVIZSGÁLATÁRA IRÁNYULÓ NAGYMINTÁS MÉRÉS

A pilotvizsgálatok keretében megvalósult mérések alapján történt a számítógépes adaptív tesztek nagymintás mérésének a megvalósítása. A nagymintás tesztelés célja a 3-7. hipotézisek beigazolása.

V.1 A MÉRÉS MÓDSZEREI

V.1.1 Minta

A nagymintás adatfelvétel 2014 tavaszán 3782 4. és 5. évfolyamos diákok részvételével zajlott. Az elemzésben azon 3220 tanuló eredményét használtuk fel, akik mindkét mérésben részt vettek, mivel így személyszintű összehasonlításra nyílt lehetőség. A minta évfolyam szerinti eloszlását a 18. táblázat tartalmazza.

18. táblázat. A minta évfolyam szerinti eloszlása

Évfolyam	N (fő)
4.	1706
5.	1514

V.1.2 Mérőeszköz

A médiahatás (Wang és Kolen, 2001; Hülber és Molnár, 2013) kiküszöbölése érdekében az adaptív rendszer hatékonyságvizsgálatát nem a papíralapú rendszerrel történő adatfelvétel eredményeivel összevetve végeztük el, hanem az itembank paraméterezéséhez összeállított lineáris tesztrendszer számítógép alapú nagymintás mérésének adataival. Ez alapján a mérőeszközök a negyedik fejezetben ismertetett lineáris tesztváltozatok és az adaptív tesztrendszer voltak. A szóolvasás mérésére kifejlesztett adaptív mérőeszköz a pilotmérés során jól működött, ezért változatlan formában alkalmaztuk a nagymintás mérés során is.

V.1.3 Adatfelvétel és eljárások

Az első adatfelvétel során a tanulók mindegyike a lineáris tesztrendszer egyik változatát oldotta meg. A második adatfelvétel egy hónap múlva történt, amikor minden

tanuló az adaptív tesztrendszer feladatait kapta. Mivel a lineáris tesztben 40 különböző booklet volt, az adaptív tesztben 30 különböző útvonal és 2-5 tesztváltozat volt elérhető, ezért elhanyagolható volt annak esélye, hogy egy tanuló ugyanazt a tesztet kapja.

V.2 AZ EMPIRIKUS VIZSGÁLAT EREDMÉNYEI ÉS DISZKUSSZIÓ

V.2.1 A tesztek reliabilitása

Az eredmények kiterjeszthetőségének, általánosíthatóságának körét első szinten jól jellemzi a teszt reliabilitásmutatójának értéke, aminek meghatározására a WLE személy-szeperációs reliabilitásmutatót használtuk mindkét teszt esetében. Az adaptív teszt reliabilitásmutatója 0,92 ami magasabb, mint a lineáris teszt megbízhatósági mutatója (0,90). A 19. táblázat dimenziók szerinti bontásban mutatja a lineáris és az adaptív teszteken mért reliabilitásokat.

19. táblázat: A tesztek EAP/PV reliabilitása dimenziónkénti bontásban

Dimenzió	WLE személy-szeperációs reliabilitás	
	Lineáris teszt	Adaptív teszt
Címszóolvasás	0,73	0,73
Toldalékos szóolvasás	0,81	0,85
Szinonimaolvasás	0,88	0,91
Szójelentés olvasás	0,87	0,90

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

A tanulók személy-item térképét mutatja a 40.a-b ábra. A minta készségeloszlása mindkét teszten, mind a négy dimenzió vonatkozásában közelíti a normáeloszlást. Mind a lineáris, mind az adaptív teszt esetében a minta készségeloszlása a -2 – +2 készségsvámban helyezkedik el mind a négy dimenzióban. A lineáris teszt esetében mind a négy

készségterületen hasonlóan alakul a minta képességeloszlása, mivel a gyengébb és a magasabb képességtartományba eső tanulók ugyanazokat a feladatokat kapták ugyanolyan arányban. Az adaptív teszt esetében viszont csak a címszó és a toldalékos szó feladatokat kapta meg minden tanuló, a szinonima és a szójelentés feladatokat viszont csak a közepes és a magasabb képességsávba tartozó tanulóknak történtek kiközvetítésre, és őket ezek alapján rangsorolta a rendszer.

A feladatok nehézségi indexeit tekintve mindkét tesztelési módnál a címszóolvasás feladatok kerültek a képességsáv aljára, ezt követték a toldalékos szóolvasás feladatok, majd a szinonima és a szójelentés feladatok bizonyultak a legnehezebbeknek. A feladatok nehézségi indexei lefedték a teljes képességskálát, tehát alkalmasak voltak a tanulók készségmérésére. Az alacsonyabb készségintűek mérésére elsősorban a címszó, és szóolvasás feladatok voltak megfelelőek, a magasabb készségtartományokat a szinonima és szójelentés feladatok tudták – átlagosan magasabb nehézségi szintjük miatt – precízebben mérni. A mérésben a 4. és 5. évfolyamos korosztály vett részt, de mivel a feladatok jóval nagyobb tartományban szóródtak, ezért alacsonyabb és magasabb évfolyamok és készséggel rendelkező tanulók mérésére is alkalmas lehet a rendszer.

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

40.a ábra: A nagymintás lineáris adatfelvétel alapján kirajzolható többdimenziós személy-item térkép [címszóolvasás (Címszó), szinonima (Szin), szójelentés (Szój) és toldalékos szóolvasás (Told)]

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

40.b ábra: A nagymintás adaptív adatfelvétel alapján kirajzolható többdimenziós személy-item térkép [címszóolvasás (Címszó), szinonima (Szin), szójelentés (Szój) és toldalékos szóolvasás (Told)]

Az egyes dimenziók közötti összefüggések erősségét mutatja a 20.a-b táblázat. Mind a lineáris, mind az adaptív tesztek esetében szoros összefüggés volt a különböző dimenziókon elért eredmények között. A legszorosabb összefüggés a szinonima és a szójelentés olvasás

között volt, majd ezt követte a címszóolvasás és a toldalékos szóolvasás dimenzióban elért eredmény, a leggyengébben a címszóolvasás és a szinonimaolvasás függött össze. A dimenziók között magas korrelációk jelzik, hogy releváns volt az alkalmazott adaptív tesztrendszer alkalmazása, a mért dimenziók szorosan összefüggenek, ezért annak ellenére, hogy a tanulók a feladatbank egy bizonyos részhalmazát oldották meg, vagyis készségi szintjüktől függően csak bizonyos dimenziókból kaptak feladatokat, a megoldott itemekből nagy valószínűséggel következtethetünk arra is, hogy a tanuló a többi dimenzióhoz tartozó feladatokon milyen eredményt ért volna el.

20.a táblázat: A lineáris tesz korreláció mátrixa

Dimenziók	Dimenziók			
	Címszó	Szin	Szój	Told
Címszó		0,26	0,20	0,29
Szin	0,49		0,49	0,30
Szój	0,50	0,97		0,21
Told	0,78	0,65	0,58	
Variancia	0,42	0,64	0,40	0,33

Megj: címszóolvasás (Címszó), szinonima (Szin), szójelentés (Szój) és toldalékos szóolvasás (Told)

20.b táblázat: Az adaptív tesz korreláció mátrixa

Dimenziók	Dimenziók			
	Címszó	Szin	Szój	Told
Címszó		0,25	0,26	0,19
Szin	0,51		0,79	0,28
Szój	0,54	0,92		0,30
Told	0,76	0,64	0,68	
Variancia	0,28	0,86	0,85	0,22

Megj: címszóolvasás (Címszó), szinonima (Szin), szójelentés (Szój) és toldalékos szóolvasás (Told)

A diákoknak a lineáris teszten mutatott teljesítménye (átlag=0,04, szórás=0,55) és az adaptív tesz alapján számolt képességi szintje (átlag=0,06 szórás=0,61) magasan korrelált egymással ($r=0,74$, $p<0,01$). A páros t-próba eredménye szerint nem volt szignifikáns különbség a tanulók két teszten elért eredményei között ($t=-1,29$, $p=0,20$, $d=0,03$). Az adaptív rendszerre való kifejlesztés során célunk az eredeti struktúra megtartása volt, ami a mutatók szerint sikeres volt, mivel a tanulók különböző készségi szintekre sorolása hasonlóan történt mindkét tesztrendszer esetében, és az elért eredmények sem különböztek jelentősen a kétféle tesztkörnyezetben.

V.2.2 A becsült készség szintek összehasonlítása évfolyamonként és személyenként

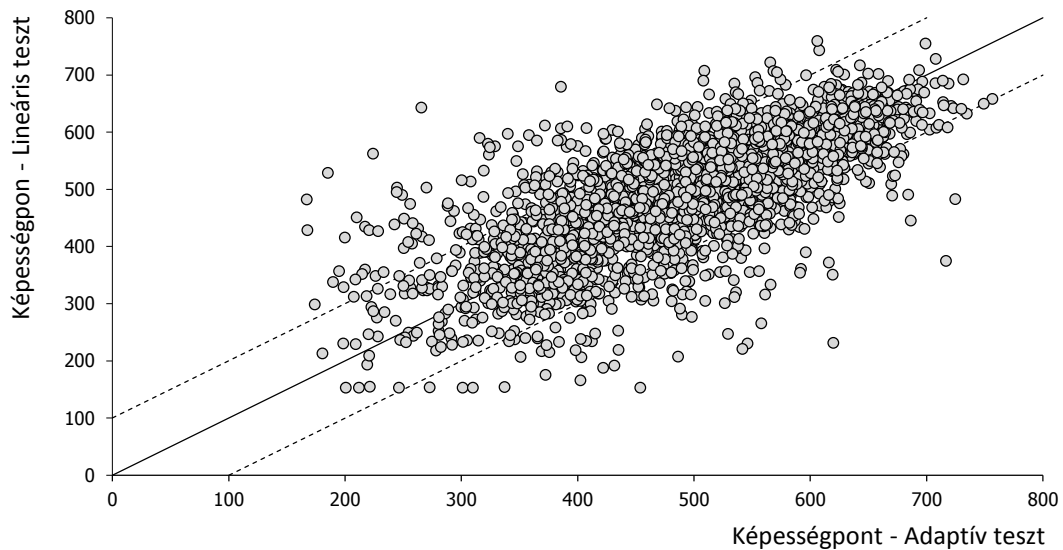
A diákok lineáris, illetve adaptív tesztkörnyezetben mutatott teljesítményének alapstatisztikai mutatóit évfolyamonkénti bontásban a 21. táblázat mutatja. Sem a negyedik évfolyamon, sem az ötödik évfolyamon nem volt szignifikáns különbség a lineáris, illetve adaptív tesztkörnyezetben becsült készség szintek között. Évfolyamonkénti összehasonlítást nézve sem különböztek az eredmények szignifikánsan egyik teszten sem ($t_{\text{lineáris}}=-0,59$, $p=0,55$, $F=0,02$, $p=0,89$, $d=0,16$; $t_{\text{adaptív}}=-1,82$, $p=0,07$, $F=0,67$, $p=0,41$, $d=0,06$), ami azt jelzi, hogy nem fordult elő olyan eset, hogy valamelyik tanuló egyik teszten kiugróan jó eredményt ért volna el, a másikon pedig gyengén teljesített volna. Ez is mindkét teszrendszer alapvetően jó működését mutatja.

21. táblázat. A lineáris, illetve adaptív tesztkörnyezetben mutatott teljesítmények klasszikus tesztelméleti mutatói évfolyamonkénti bontásban

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

A 41. ábra a két tesztkörnyezetben nyújtott teljesítmények diáksintű összehasonlítását ábrázolja. Ha a diák készség szintje tesztkörnyezettől függetlenül számszerűen ugyanannak bizonyult, akkor a diákot reprezentáló alakzat a folytonos vonalon helyezkedik el. Amennyiben megállapított készség szintje nem különbözött egymástól szignifikánsan lineáris és adaptív környezetben, az őt reprezentáló jel a szaggatott vonalakon belül található. A szaggatott vonalak által képzett sávon kívül elhelyezkedő diákok esetében az adaptív tesztkörnyezet szignifikánsan különböző készség szintet állapított meg, mint a lineáris tesztekkel diagnosztizált készség szint. A mintában szereplő tanulók 7%-ánál magasabb, 6%-nál pedig alacsonyabb készség szint került diagnosztizálásra adaptív környezetben, azaz a diákok 13%-nál alapvetően más készség szint került meghatározásra. Jellemzően az adaptív teszten alacsonyabb képességtartományba sorolt diákoknál fordult inkább elő, hogy a lineáris teszt magasabb készség szintet mutatott, valamint az alacsony és átlagos készségű diákok körében fordult ennek fordítottja is elő. Ha az értelmezés során figyelembe vesszük az adaptív tesztkörnyezetben tapasztalt méréselméleti mutatók (mint a

reliabilitás növekedése, vagy a mérési hiba csökkenése) javulását, akkor megállapítható, hogy a diákok e 13%-a számára volt igazán meghatározó az adaptív környezet. A többiek mérés során meghatározott készpésszintje nem különbözött jelentős mértékben egymástól lineáris és adaptív környezetben.



41. ábra: A lineáris és az adaptív teszten nyújtott teljesítmények összehasonlítása diákonkénti bontásban

V.2.3 A helyes válaszok aránya adaptív és lineáris tesztkörnyezetben

Az adaptív tesztrendszer helyes működését jelzi, ha a tanulók teszten nyújtott teljesítménye a képességszinttől függően azonos (a rendszer beállításakor meghatározott, jelen esetben 80% körüli) százalékos szinten mozog. Ebben az esetben ugyanis mindenki a készpésszintjéhez leginkább közel álló feladatokat kapta a tesztben, az alacsony készpésszintűek könnyebb, a magasabb készpésszintűek nehezebb feladatot. A helyes megoldások aránya így minden esetben közel azonos lesz, annak ellenére, hogy az egyes százalékos teljesítmények más-más képességszintet jeleznek. A 22. táblázat mutatja a tanulók adaptív és lineáris tesztkörnyezetben elért helyes válaszainak arányát évfolyamonkénti bontásban. A helyes válaszok aránya mindkét évfolyamon megfelelt az előre beállított 80% körüli értéknek. Az átlagosan elért helyes válaszokat illetően nem volt szignifikáns különbség a két teszt között egyik évfolyamon sem.

22. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak aránya évfolyamonkénti szerinti bontásban

Évfolyam	N	Helyes válaszok aránya (%)				t	p	d
		Lineáris teszt		Adaptív teszt				
		átlag	szórás	átlag	szórás			
4.	1706	81,64	8,88	81,19	6,99	1,68	0,09	0,06
5.	1514	81,69	8,81	81,63	6,99	0,28	0,78	0,01
összes	3220	81,66	8,85	81,39	6,97	1,59	0,11	0,03

A kétféle tesztkörnyezetben elért eredmények készség szint szerinti eloszlását ábrázolja a 23. táblázat. Általánosságban mindkét teszten a készség szint növekedésével párhuzamosan emelkedett a teszteken elért helyes válaszok aránya, bár a növekedés mértéke különbözőnek bizonyult. Adaptív tesztenél az átlag alatti tanulóknál jóval magasabb arányban fordultak elő helyes válaszok, mint a lineáris teszten, átlag feletti tanulók esetében viszont fordítva, kisebb arányban fordultak elő helyes válaszok. Ez arra enged következtetni, hogy az adaptív teszt az alacsony készség szintű tanulóknak több sikerélményt hozott, (kevesebb számukra túl nehéz, nagy valószínűség szerint megoldhatatlan feladatot osztott a rendszer), a magas készségűek számára viszont nagyobb kihívást jelentett (kevesebb könnyebb feladatot kiközvetítve), mint a lineáris teszt.

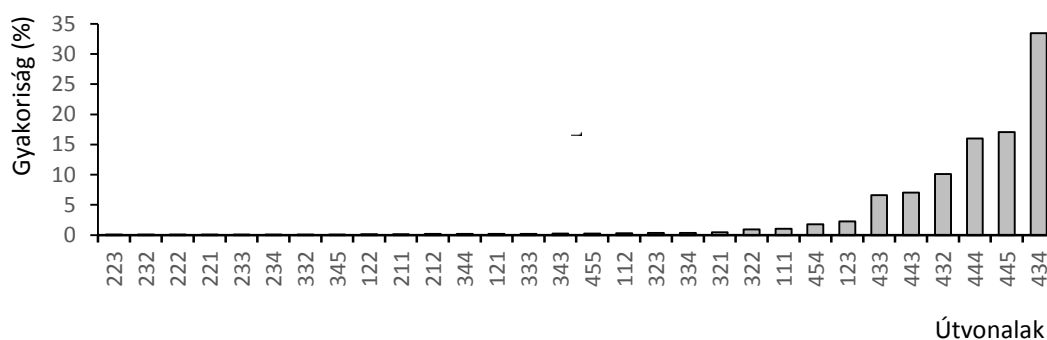
23. táblázat. A tanulók adaptív és lineáris teszten elért helyes válaszainak eloszlása készség szint szerinti bontásban

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

V.2.4 A lineáris és az adaptív tesztelés során kiosztott résztesztek nehézségi szintjének változásmintázata

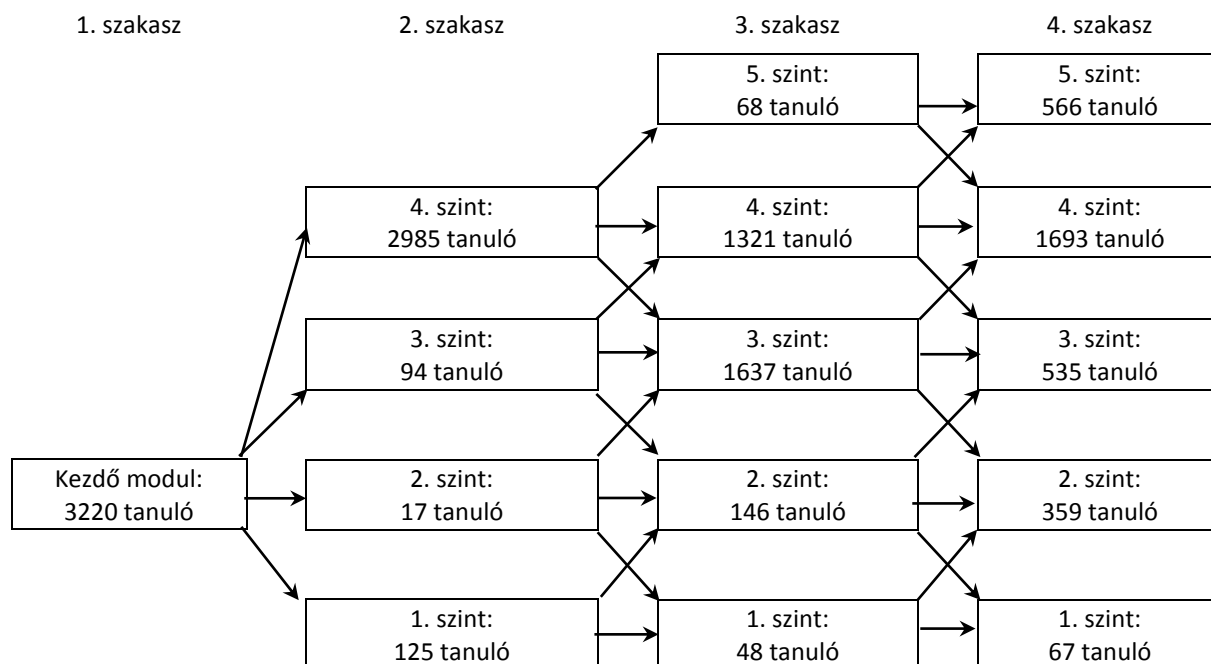
Az adatfelvétel során a többszakaszos adaptív tesztet illetően az öt szakaszból összeállítható 30 különböző teszt mindegyike kiosztásra került. (42. ábra).

A kiosztott részteket illetően a legnagyobb részben az átlagos nehézségű, a 3., illetve 4. szinten szereplő feladatok kerültek kiközvetítésre. A hat leggyakoribb útvonal a 4. és a 3. szinteken haladt át. Az esetek 20%-ában jelentek meg a legnehezebb, 5. szint feladatai, 5%-ban kizárólagosan könnyű feladatokból álló teszteket oldottak meg a tanulók. A kezdő részeszt után jellemzően a 4. szintre kerültek besorolásra a tanulók, majd innen a 3., illetve a 4. szintű modul felé haladtak. A negyedik szakaszban a tanulók közel fele maradt a 4. szinten, harmaduk pedig a 3-as, illetve 5. szinteken végzett. A tanulók 10%-a a könnyebb, 2-es szint felé haladt a 4. szakaszban. Az esetek 75%-ában a tanulók a 3. szakasz után még szintet váltottak, ami indokolja a 4. szakasz szükségességét.



42. ábra: Az adaptív tesztrendszeren belül a második, harmadik és negyedik szakaszban kiosztott útvonalak gyakorisága

A szakaszokon belül a tanulók modulonkénti eloszlását képességszint és gyakoriság szerinti bontásban a 43. ábra szemlélteti. A kezdő modulon mutatott teljesítmény alapján a tanulók legnagyobb része a teszt második szakaszában a 4. szintű feladatokat kapta, majd a teszt 3. és 4. szakaszában egyre egyenletesebben oszlottak el a 2-5. szintek között. A teszt utolsó szakaszában a tanulók közel fele a 4. szinten végzett, a többiek közel azonos arányban oszlottak el a 2., 3. és az 5. szinteken. A legelső szinten viszonylag kevés, 67 tanuló végzett, ők 90%-ban a teszt kezdetétől az 1-es szinten haladtak végig.



43. ábra: A tanulók gyakorisági és képességszint szerinti eloszlása a szakaszokon és a modulokon belül

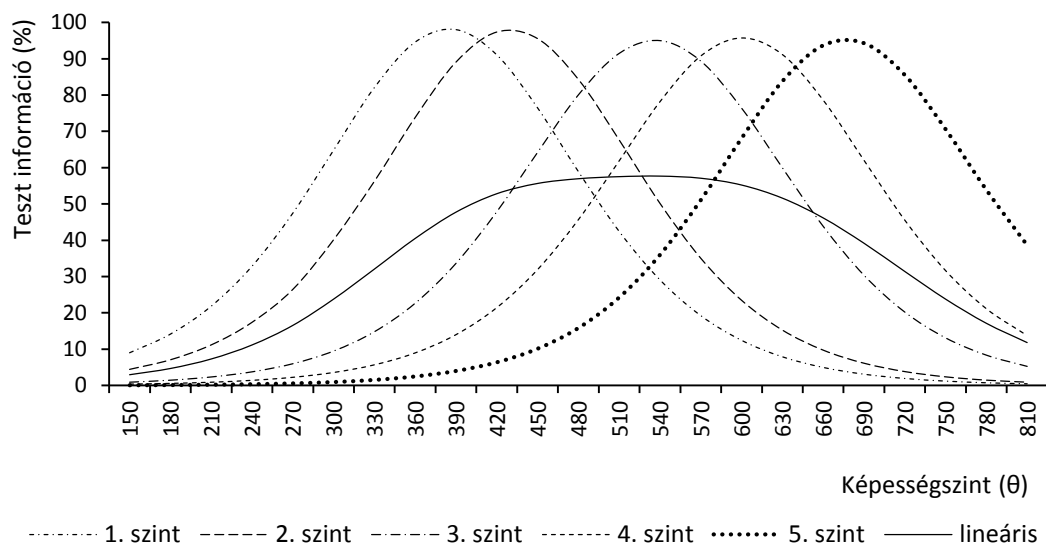
A papíralapú rendszerrel összhangban a lineáris teszt eredményei alapján is besorolhatóak voltak a tanulók az öt szint valamelyikébe. A tanulók a lineáris teszten és az adaptív teszt 4. szakaszában történt szintenkénti besorolását összegzi a 24. táblázat. A két rendszer hasonló arányban végezte el a tanulók besorolását, az adaptív teszt esetében azonban több tanuló jutott el az 5. szintre, és a második szinten is többen végeztek, mint a lineáris teszt esetében, tehát a kiemelkedő és a gyengébb készségszintű tanulók jobb eredményeket produkáltak az adaptív teszt kiosztásnál. Itt mutatkozik meg a pontosabb képességmérés jelentősége, hiszen a tanulók végső besorolását befolyásolhatja, hogy mennyire működik pontosan a rendszer.

24. táblázat: A tanulók lineáris és adaptív teszten való szintenkénti besorolása

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

V.2.5 A lineáris és az adaptív tesztelés során kinyert információ és a mérési hiba nagyságának összehasonlítása

A reliabilitás mellett a teszt mérési pontosságának egy másik mutatója a tesztelés során kinyert információ nagysága, amit jelen esetben a teszteken nyújtott teljesítmények alapján számított tesztinformációk segítségével jellemzünk. A 44. ábra görbéi grafikusán szemléltetik, hogy már akár a kizárólagosan 1. szintű, 2. szintű, 3. szintű, 4., illetve 5. szintű modulok résztesztjeiből összeállított tesztek (a 30 tesztváltozatból csak 5) is több információt szolgáltatott a tesztelés során, mint az egyetlen, sokféle nehézségű feladatot tartalmazó lineáris teszt. A görbéket a 13. táblázatban ismertetett képességszintekkel egybevetve megállapítható, hogy minden képességszinten több információt tudunk kinyerni adaptív tesztek alkalmazásával (a görbék minden esetben a lineáris teszt által adott információs függvény felett futnak az érintett képességtartományokban).



44. ábra: A lineáris és az adaptív technikával összeállított, azonos nehézségű modulokat tartalmazó tesztek információs függvényei

A mérés során minden tanuló megoldotta az adaptív és a lineáris verziót is, így lehetőség volt az eredmények személyszintű összehasonlítására. A tesztek mindkét változatát megíró 3220 tanuló közül 1927 esetben, vagyis közel a tanulók kétharmadánál volt az adaptív tesztből kinyert információ magasabb, mint a lineáris teszt esetében (1293 tanuló). A különbség átlagosan 2,9% volt, nagysága képességszintenként különbözött. A 25. táblázat

mutatja a kinyert információk mennyiségét készségi szintenkénti bontásban. Minden készségi szinten több információ volt kinyerhető az adaptív tesztből, mint a lineáris tesztből. A *Cohen d* értéke is megerősíti ezt: a kinyert információk mértéke az alacsony 400 képességi pont alatti) és a magas (600 képességi pont feletti) képességi sávokban jelentősen magasabb volt az adaptív tesztnél, az átlagos készségi szinteken (400-600 képességi pontok között) közepes mértékű volt.

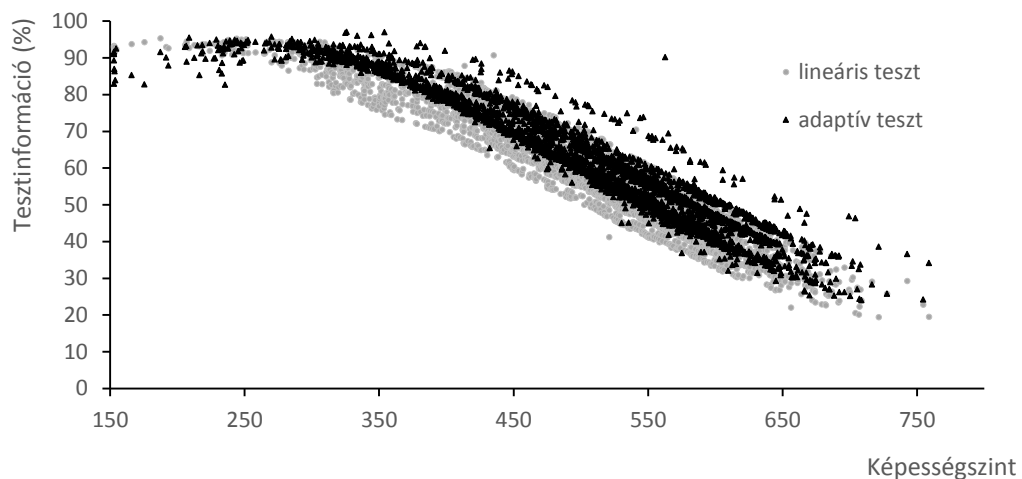
25. táblázat: A tesztekől kinyert információk mennyisége készségi szintenkénti bontásban

Képességi pont	Lineáris teszt			Adaptív teszt			<i>d</i>
	N	információ (%)	szórás (%)	N	információ (%)	szórás (%)	
300 alatt	105	92,0	2,0	116	92,8	2,8	0,33
301-400	450	82,6	5,4	397	87,1	4,1	0,93
401-500	928	67,7	7,3	904	71,3	6,4	0,52
501-600	1235	50,9	7,2	1352	53,8	6,9	0,411
601-700	482	37,0	6,0	435	43,0	6,0	0,99
700 felett	20	25,0	4,0	16	30,0	7,0	0,88

Az adaptív teszt esetében átlagosan 63% volt a kinyert információ mennyisége, a lineáris teszt esetében pedig 59%. A páros *t* próba szerint a különbség szignifikáns ($t=12,61$, $p<0,001$). A 45. ábra a tanulók készségi szintjének függvényében ábrázolja a tanulók lineáris, illetve az adaptív teszten kinyert információk mennyiségét. Mindkét teszt esetében a magasabb készségi szint felé haladva csökkent a kinyert információ nagysága, azonban az adaptív teszten mindegyik készségi szinten magasabb volt.

A kinyert információ csökkenésével a teszt mérési precizitása is csökken. A tesztből kinyerhető információ a tanuló készségi szintjéhez közel álló nehézségi indexű itemek kiközvetítésével növelhető, amely adaptív teszteléssel valósítható meg nagyobb mértékben, hiszen itt érhető el, hogy a tanulók készségi szintjükhöz illeszkedő itemeket kapjanak, ellentétben a lineáris tesztekkel, ahol készségi szinttől függetlenül többféle, különböző nehézségi itemeket kapnak a tanulók. A kutatásban használt tesztek esetében mindkét típusú tesztnél csökkent a kinyerhető információ mennyisége az egyre magasabb készségi szinteken, ami annak tulajdonítható, hogy a tesztelés elején minden tanuló címszóolvasás feladatot kapott, amely itemek elsősorban az alsó készségi tartományban lévő tanulóknál tudtak magasabb mennyiségű információt szolgáltatni (lásd 36. a-d ábra). A tesztelés későbbi szakaszában a lineáris teszt megoldása során minden tanuló egyformán kapott mindegyik

dimenzióból itemeket, így többféle nehézségű feladatot kellett megoldania, melynek csak egy kis része volt a készségszintjéhez illeszkedő, melyek a kinyerhető információt jelentős mértékben csökkentették. Az adaptív tesztelés során a gyengébb készségűek végig címszóolvasás és toldalékos szóolvasás feladatokat oldottak meg, melyek minden iteme nehézségi indexében közel állt a készségszintjükhöz. Ezért volt a gyenge tanulók esetében a kinyert információk mennyisége igen magas. A magasabb készségszintű tanulók viszont a tesztelés első felében címszó- és toldalékos szóolvasás feladatokat kaptak, melyek alacsonyabb nehézségi indexűek voltak, mint a készségszintjük, és csak a tesztelés második fele biztosította számukra a készségszintjükhöz közel álló feladatokat. Ez okozhatta a magasabb készségszinten a kinyerhető információ mértékének a csökkenését, ami még így is magasabb volt, mint a lineáris tesztből kinyert információ.



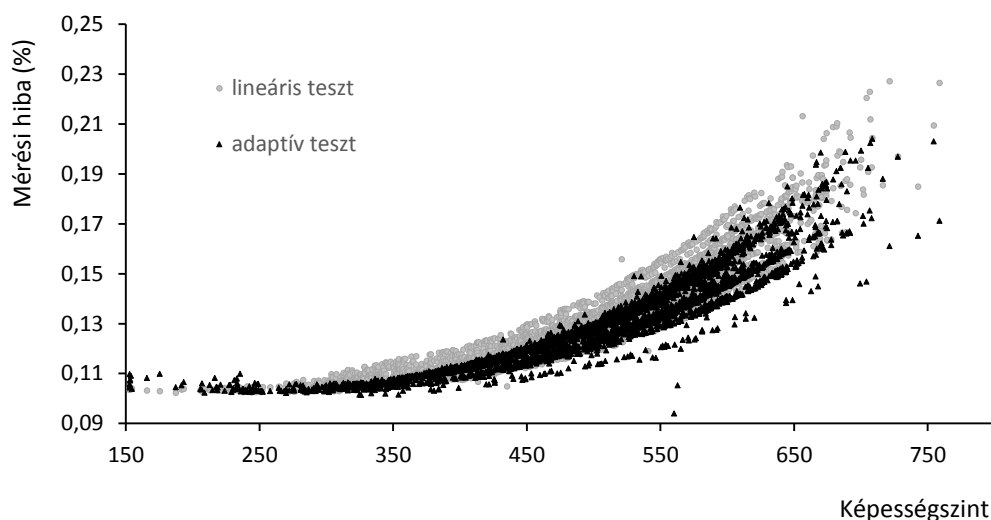
45. ábra: A lineáris és az adaptív teszten kinyert információk a tanulók képességszintjének függvényében

A kinyert tesztinformáció nagyságához hasonlóan a mérési hibák is összehasonlíthatóak diákonkénti bontásban. Vizsgálatunkban a kétféle teszt-környezetben történt készségszint-becslés során elkövetett hiba nagyságát összehasonlítva megállapítható, hogy a lineáris formátumú teszt alapján történt készségszint-becslés hibáinak nagysága diákszinten átlagosan nagyobb ($t=13,52$, $p<0,01$; $SE=0,14$), mint ugyanazon diákok adaptív teszt-környezetben történt készségszint-becslésének hibája ($SE=0,13$). A teljes minta mindegyik készségszinten pontosabban, kisebb mérési hibával történt adaptív teszt-környezetben a diákok képességszintjének becslése (26. táblázat).

26. táblázat: A teszteken mért mérési hibák nagysága képesszintenkénti bontásban

Képességpont	Lineáris teszt			Adaptív teszt		
	N	SE	szórás	N	SE	szórás
300 alatt	105	0,10	0,01	116	0,10	0,01
301-400	450	0,11	0,01	397	0,10	0,01
401-500	928	0,12	0,01	904	0,11	0,01
501-600	1235	0,14	0,01	1352	0,13	0,01
601-700	482	0,16	0,01	435	0,15	0,01
700 felett	20	0,2	0,01	16	0,18	0,01

Összehasonlítva a lineáris és az adaptív teszten elért eredmények sztenderd hibáinak változását (46. ábra), hipotéziseinknek megfelelően, mindegyik képességtartományban nagyobb hibával mért a lineáris teszt, mint az adaptív tesztváltozat.



46. ábra: Az adaptív és a lineáris teszt standard hibáinak alakulása a tanulók képesszintjének függvényében

V.3 A SZÓOLVASÓ KÉSZSÉG VIZSGÁLATÁRA IRÁNYULÓ NAGYMINTÁS MÉRÉS EREDMÉNYEINEK ÖSSZEFOGLALÁSA

A szóolvasó készség mérésére kifejlesztett adaptív tesztrendszer nagymintás tesztelése során történt a lineáris rendszer és az adaptív rendszer hatékonyságának összehasonlítása.

A nagymintás mérés céljai a következők voltak:

- az adaptív és lineáris tesztek mérési pontosságának összehasonlítása;
- a becsült képesszintek évfolyam és személyszintű összehasonlítása;

a kétféle tesztkörnyezetben elért helyes válaszok arányának összehasonlítása;
az adaptív tesztelés során kiosztott itemek, illetve résztesztek nehézségi szintjének,
ennek változásmintázatainak jellemzése;

a lineáris és az adaptív tesztelés során kinyert információ és a mérési hiba nagyságának
összehasonlítása képességszint szerinti bontásban.

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

A tanulók jelentős része átlagos szóolvasó képességgel rendelkezik, ennek megfelelően a
közepes nehézségű résztesztek szerepeltek a legnagyobb gyakorisággal az adaptív
teszt kiosztásnál, azonban a harmadik, illetve még a negyedik szakaszban is sok tanuló
esetében módosult a szint, ami indokolja az öt különböző nehézségi szintű modul
alkalmazásának szükségességét.

A mérés során összehasonlítottuk a tesztinformációkat, illetve a mérési hibák
nagyságát, és mindkét esetben az adaptív teszttel mértünk magasabb tesztinformációt és
kisebb mérési hibát a teljes képességskálán.

ÖSSZEGZÉSEK, KÖVETKEZTETÉSEK

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

Az adaptív teszteknek számos típusa létezik, az egyik legpreferáltabb típus a többszakaszos adaptív tesztstruktúra, mely során több szakaszban itemek helyett itemcsoportokat, azaz modulokat osztanak ki, melyek tulajdonképpen különböző nehézségi szintű rövid fix tesztek, melyek egy-egy bizonyos képességtartomány mérését célozzák. A tesztípus egyesíti magában a hagyományos lineáris és az adaptív tesztek tulajdonságait, mivel egyrészt a kérdéseket a tanuló képességszintjéhez igazítja, másrészt lehetőséget ad az itemek modulon belüli sorrendjének előzetes meghatározására. A modulok előre tervezhetőek és szerkeszthetőek, így nagyobb kontrollt biztosítanak a teszt adminisztráció számára, így kiküszöbölhetővé válik, hogy az itemek egymásnak információt szolgáltatassanak. További fontos előnyük, hogy a modulokon belül a tanulóknak lehetőségük van a visszalépésre és javításra, mivel adaptivitás csak a modulok között valósul meg, így ez nem veszélyezteti a teszt algoritmusát és segíti a tanulókat a minél magasabb pontszám elérésében. Az itemalapú adaptív tesztekhez képest jóval kevesebb adminisztrációt és számítógépes számítás igényelnek.

A papíralapú tesztekéről az adaptív tesztelésre való átállás során fontos megvizsgálandó kérdés, hogy az átállás biztosítja-e az elvárt szintű mérésmethodikai javulást, és a hatékonyabb képességmérést. A vonatkozó nemzetközi vizsgálatok szerint az adaptív tesztek reliabilitása magasabb a lineáris tesztekénél, a kinyerhető információk mennyisége szintén több, a mérési hiba viszont alacsonyabb, és ezáltal a lineáris tesztekénél jóval pontosabb mérést tesznek lehetővé. A legtöbb vizsgálat eredménye szerint a többszakaszos teszt precizitása valamivel elmaradt az item, illetve a tesztlet alapú tesztekétől, viszont a tesztfejlesztés folyamán alkalmazható nagyobb adminisztratív kontroll előnyössé teszi ezt a tesztípust, ezért papíralapú tesztelésről való átállás esetén a legpreferáltabb típusú adaptív tesztkonstrukció. A nemzetközi kutatásokon elért eredmények elsősorban szimulált adatbázisokra alapozottak, néhány pilotmérés történt empirikus adatok felhasználásával, egyetemista hallgatók bevonásával.

Kutatásunk célja a papíralapú tesztelésről adaptív tesztelésre való átállás feltételeinek megvizsgálása volt empirikus vizsgálat keretében az általános iskolai korosztály körében. A kutatás során több képességterület, valamint különböző szerkezetű adaptív változat esetében az adaptív tesztek hatékonyságát, mérési precizitását hasonlítottuk össze a hagyományos, lineáris tesztkörnyezetben folyó teszteléssel szemben.

Az adatfelvétel több fázisban zajlott. Különböző képességterületeket érintő három pilotmérést és két nagymintás mérést végeztünk el. A mérések az 1-8. évfolyamos tanulók körében zajlottak három különböző képességterületen. A pilotmérések során a tanulók induktív gondolkodásának, problémamegoldó képességének és szóolvasó készségének fejlettségi szintjének mérése valósult meg, mely méréseket két nagymintás adatfelvétel követte, melyek során szóolvasó képesség mérését valósítottuk meg. A felhasznált tesztek minden esetben előzőleg papír alapon alkalmazott tesztek online formára történő konvertálásával kerültek kidolgozásra az eDia rendszerben. A tesztek felvétele osztálytermi környezetben történt, egy tanítási óra, azaz 45 perc alatt. A személyszintű összehasonlítás biztosítása érdekében minden tanuló kétféle tesztváltozatot oldott meg, egy hagyományos lineárisat és egy adaptívat. Mivel az adaptív tesztelés mérésmethodikai hátterét a valószínűségi tesztelmélet biztosítja, az elemzések során az egyparaméteres Rasch-modellt alkalmaztuk.

Az elemzések során a vonatkozó nemzetközi kutatásoknak megfelelően összehasonlítottuk a tesztek jószágmutatóit, a becsült képességszinteket adaptív és lineáris tesztkörnyezetben, megvizsgáltuk a kétféle tesztelés során kiosztott résztesztek nehézségi

szintjének változásmintázatát és a helyes válaszok arányát, valamint összevetettük a kinyert információkat és a mérési hibák nagyságát mindkét tesztelési módra vonatkozóan.

HIPOTÉZISEK BEIGAZOLÓDÁSA

A vizsgálat során felvetett hipotézisek és beigazolódásuk:

1) A korábban papír alapon alkalmazott tesztekből kifejleszthető többszakaszos adaptív tesztelésre alkalmas tesztrendszer.

A kutatás során három, korábban papír alapon alkalmazott tesztrendszer online formára való konvertálása történt. A kifejlesztett tesztek mindegyik esetben megfelelő jóságmutatókkal rendelkeztek, a tanulók kétféle tesztkörnyezetben elért eredményei szoros korrelációt mutattak, és a t-próba eredményei szerint csak egy-egy évfolyam tekintetében volt az elért eredmények között szignifikáns különbség. A mérések során beigazolódott első hipotézisünk, miszerint a korábban papír alapon alkalmazott rendszerek átkonvertálhatóak számítógépes adaptív verzióra.

2) Az adaptív tesztrendszerek hatékonyan és megbízhatóan alkalmazhatóak az 1-8 évfolyamos korosztály diagnosztikus mérésére.

A tesztek hatékonyságának és megbízhatóságának egyik mutatója a reliabilitásmutató. Mivel adaptív teszt kiosztásnál többféle tesztváltozat kerül kiközvetítésre, és a tanulók az itemeknek csak bizonyos részhalmazát oldják meg, ezért a Cronbach- α reliabilitásmutató nem volt használható. Ehelyett az adaptív teszt esetén ennek kiterjesztését, a személyszeparációs reliabilitásmutatót használtuk. Az adaptív teszt WLE (*Weighted Likelihood Estimate*) személyszeparációs reliabilitásmutatói mindegyik tesztnél megfelelőnek bizonyultak, így beigazolódott a második hipotézis, mely szerint a kidolgozott adaptív tesztek megbízhatóságuk szempontjából alkalmasak az 1–8. évfolyamos diákok készségének és képességeinek diagnosztikus mérésére.

3) Az adaptív rendszer a képességek pontosabb mérését teszi lehetővé.

A tesztek mérési pontosságát jól jellemzi a mérési hiba nagysága, valamint a kinyert információ mértéke (*Wang és Kolen, 2001; Wang, 2010; Molnár, 2013*). A tesztek közül

kinyerhető információ nagyságát a tesztet megoldó tanulók átlagos képességszintje és az itemek nehézségi szintje közötti különbségek segítségével jellemtük. A kinyert információ nagyságát akkor tekintettük maximálisnak, ha a feladatok nehézségi szintje és az azokat megoldó diákok képességszintje azonos volt. Minél távolabb volt egymástól ez a két érték, annál kevesebb volt a tesztelés során kinyert információ mértéke. Mindhárom mérésnél az adaptív tesztből kinyerhető információk mértéke szignifikánsan nagyobb volt, mint a különböző nehézségű itemekből álló lineáris teszténél, tehát az adaptív változatok pontosabb, precízebb mérést valósítottak meg, mint a lineáris tesztváltozatok. Hasonlóan, a mérési hibák szignifikánsan alacsonyabbak voltak az adaptív teszt kiosztásoknál, mely szintén igazolja az adaptív tesztek pontosabb működését. A mérési eredmények tehát beigazolták a harmadik hipotézisünket.

4) A becsült képességszintek nem különböznek jelentős mértékben online adaptív és lineáris tesztelésnél.

Mindhárom esetben megvizsgáltuk az adaptív és a lineáris teszt kiosztásnál becsült képességszintek közötti eltérések nagyságát. A kétféle teszt kiosztásnál az elért eredmények magas korrelációs együtthatókat mutattak, amely jelzi, hogy a kétféle teszt kiosztás hasonlóan sorolta be a tanulókat. A t próbák eredményei szerint csak egy-egy évfolyamnál volt különbség a teszteken elért eredmények között, ami arra enged következtetni, hogy általában nem volt jelentős különbség a kétféle képességbecslés között, viszont bizonyos korosztály, illetve képességszintű tanulók esetében előnyösebb lehet egyik, vagy másik teszt változat, mely indokolja az előzetes vizsgálatok szükségességét.

5) Adaptív tesztelésnél az alacsonyabb képességtartományban magasabb a helyes válaszok aránya, mint lineáris teszt környezetben, átlag feletti tanulóknál viszont fordítva, kisebb arányban fordulnak elő helyes válaszok.

Mindkét teszt kiosztásnál a képességszint növekedésével párhuzamosan nőtt a helyes válaszok száma is. Az adaptív teszt kiosztásnál azonban a növekedés mértéke eltérő volt a lineáris tesztelés során tapasztaltaktól. Az alacsony képességtartományban jelentős mértékben megnőtt a helyes válaszok aránya, aminek az a magyarázata, hogy az alacsonyabb képességszintű diákok könnyebb feladatokat kaptak, melyekkel könnyebben megbírták, így nagyobb sikerélményre tehettek szert. A magas képességtartományban az induktív tesztet

kivéve fordítva történt, a magasabb képességszintű tanulók nehezebb feladatokat kaptak, így kevesebb helyes válasz született, ezzel együtt nagyobb kihívást jelentett számukra a teszt megírása. Az induktív tesztelésnél mindegyik képességtartományban nőtt a helyes válaszok aránya. Az eredmények részlegesen igazolták az ötödik hipotézisünket.

6) A tanulók többsége átlagos képességszintű, ezért adaptív teszt kiosztás esetében az átlagos nehézségű itemek/résztesztek szerepelnek leggyakrabban.

Az adaptív tesztelés során kiosztásra került útvonalakat megvizsgáltuk, minden esetben a közepes nehézségű modulok szerepeltek leggyakrabban. Mivel a tanulók többsége átlagos képességszintű, ezért ez megfelelt az elvárásainknak. A tesztbiztonság növelése érdekében ezért célszerű a közepes nehézségű modulok számát növelni. A hatodik hipotézisünk beigazolódott.

7) A tesztekből kinyert információ minden képességszinten szignifikánsan magasabb, a mérési hiba viszont szignifikánsan alacsonyabb az adaptív tesztrendszer esetében, mint lineáris tesztekénél.

Mindegyik kutatás esetében elemeztük a tesztekből kinyert információ mennyiségét. Az adaptív tesztrendszer alkalmazásával kinyert információ adatfelvételtől függetlenül magasabb volt, mint a lineáris tesztekből kinyert információ nagysága. Az eltérések mértéke azonban változó volt a különböző képességtartományokban. Az induktív gondolkodás fejlettségi szintjét célzó mérésnél az alacsony és a magas képességtartományokban volt a kinyert információ mértéke jelentősen több, mint a lineáris teszt esetében. A problémamegoldó képességet mérő tesztek esetében a rendszer elsősorban a magas képességű egyéneket különítette el, tehát a kinyert információ mennyisége a magas képességsávban volt jelentősen több. A szóolvasás mérése alapján megállapítható, hogy az adaptív és a lineáris teszteken mért információk eltérése egyenletesen oszlott el a teljes készségskálán, minden készségszinten több volt az adaptív tesztből kinyert információ mennyisége, mint a lineáris teszt esetében. A mérési hiba mértéke is ennek megfelelően alakult mindegyik mérés esetében. A hetedik hipotézis részben igazolódott be, a minta nagyságától függően a különböző képességszinteken más-más mértékű lehet a kinyerhető információ mennyisége és a mérési hiba nagysága. A kutatás eredményei szerint a kisebb minták esetében elsősorban az alacsony és a magas képességtartományban jelentősen

magasabb az adaptív tesztből kinyerhető információ mértéke, nagyobb mintán ez kiegyenlítődik, és a képességtartomány minden szintjén egyenletesen több információ nyerhető ki az adaptív tesztből, mint a lineáris változatból. A mérési hiba mértéke is ennek megfelelően alakul.

EREDMÉNYEK ÖSSZEFOGLALÁSA, ÉRTELMEZÉSE

Az eredmények a szakirodalmi kutatásokkal összhangban kimutatták, hogy az előzőleg papír alapon használt tesztek adaptív formára konvertálhatók, és hatékonyan alkalmazhatók az általános iskolai korosztály képességszintjének a meghatározására. Az adaptív tesztelési technikánál az alacsonyabb képességgel rendelkező tanulók több helyes választ adtak, így több sikerélményt jelentett számukra a tesztelés, a magasabb képességszintű tanulók számára pedig kihívást jelentett az adaptív tesztek megoldása. Az adaptív tesztek a képességek pontosabb mérését teszik lehetővé, kedvezőbb jószágmutatókkal rendelkeznek, alkalmazásukkal a teljes minta szintjén több információt nyertünk ki, szignifikánsan pontosabb képességszint-meghatározást végeztünk, mint hagyományos, lineáris teszteléssel. A minta nagyságától függően azonban a különböző képességszinteken más-más mértékű lehet a kinyerhető információ mennyisége és a mérési hiba nagysága. A kisebb minták esetében elsősorban az alacsony és a magas képességtartományban volt jelentősen magasabb az adaptív tesztből kinyerhető információ mértéke, nagyobb mintán ez kiegyenlítődött, és a képességtartomány minden szintjén egyenletesen több információ volt kinyerhető az adaptív tesztből, mint a lineáris változatból. A képességszintek becslése során elkövetett hiba nagysága is ezzel összecsengett, és a becsült hiba nagysága a minta méretének függvényében változott.

A kutatás gyakorlati jelentősége jelentősége, hogy a kifejlesztett tesztek osztálytermi környezetben, tanórák keretében felhasználhatóak, azonnali visszajelzést biztosítanak a tanulók és a pedagógus számára. A tanulók képességbecslése pontosabbá válik, ami elsősorban a kritériumorientált mérések esetében jelentős mértékben befolyásolhatja a tanulók elért eredményeit. A tanulók különböző feladatokat kapnak, ami által növekszik a tesztbiztonság. Mivel a feladatok előre paraméterezett itembankból kerülnek kiközvetítésre, a tanulók eredményei közös képességskálán jellemzhetőek, így anélkül, hogy minden tanuló

mindegyik feladatot megoldaná, nagy valószínűséggel megmondható, hogy a többi feladaton hogyan teljesítene.

Összességében a kutatás eredményei szerint az adaptív tesztek több képességterület vonatkozásában megbízhatóan alkalmazhatóak az általános iskolai korosztály körében, pontosabb képességmeghatározást tesznek lehetővé, motiválják az alacsonyabb képességű tanulókat és kihívást jelentenek a magasabb képességűek számára.

A KUTATÁS EREDETISÉGE ÉS KORLÁTAI

A kutatás egyedisége, hogy az adaptív tesztelés hatékonyságát vizsgáló legtöbb kutatással szemben nem szimulált adatbázison, hanem valódi mért adatok segítségével hasonlította össze a lineáris és az adaptív tesztkörnyezetben becsült képességszintek alakulását, továbbá az azonos minta alkalmazása lehetővé tette a diákszintű összehasonlítást is. A mérésekben a 6-14 éves korosztály vett részt, mely szintén egyedinek mondható az adaptív tesztekkel foglalkozó kutatások között. Az eredmények alátámasztották a szimulációs kísérletekben is tapasztaltakat, miszerint jelentős mértékű mérési precizitás érhető el adaptívteszt-algoritmus alkalmazásával a hagyományos lineáris tesztekhez képest.

A kutatási eredményeink általánosíthatóságának korláta, hogy az adatfelvételek során három teszt esetében vizsgáltuk az adaptív tesztelésre való átállás lehetőségét, az alkalmazott itembankok mérete, az itemek típusa különböző volt. Ezek a jellemzők befolyásolhatják a képességszint becslés során kinyert és kinyerhető információ mennyiségét, ezért a kinyert információ mértékének pontosabb meghatározásához további kutatások szükségesek a különböző méretű és tartalmi lefedésű itembankok felhasználásával.

TOVÁBBI KUTATÁSI IRÁNYVONALAK

- 1) A kutatás során az adaptív tesztelésre való átállás lehetőségeit induktív gondolkodás, problémamegoldó képesség és szóolvasó készséget mérő teszteken vizsgáltuk. Szükséges lenne további képességterületekre vonatkozóan is megvizsgálni az átkonvertálás lehetőségeit.

- 2) Az alkalmazott itemek a papír alapú formátummal megegyező formátumban kerültek online formára való konvertálásra. Innovatív itemtípusok, illetve multimédiás alkalmazások esetén is indokolt lehet hasonló vizsgálat lefolytatása.
- 3) A kutatásban többszakaszos adaptív tesztekre való konvertálás történt. Habár többféle szerkezetű többszakaszos tesztet vizsgáltunk (1-3-3-3, 1-2-3, 1-4-5-5), a továbbiakban célszerű lenne egy-egy képességet mérő tesztnél többféle szerkezetű többszakaszos teszt működését megvizsgálni, és hatékonyságukat összehasonlítani.
- 4) A többszakaszos tesztek mellett az itemalapú adaptivitás lehetőségeinek feltárására is célszerű lenne vizsgálatot lefolytatni.
- 5) Az adaptív tesztelés visszahatása az osztálytermi környezetre, a tanulók számára készített visszajelzések alapján a tanárok feladata a további fejlesztések tervezése során.

KÖSZÖNETNYILVÁNÍTÁS

Köszönöm *Molnár Gyöngyvér*, témavezetőm szakmai segítségét, hasznos iránymutatását, építő együttműködését a disszertáció megírásában, valamint a közös publikációk lehetőségét.

Köszönöm *Csapó Benő*, a Neveléstudományi Doktori Iskola vezetőjének a lehetőséget, hogy részt vehettem a Doktori Iskola és az Oktatáselméleti Kutatócsoport munkájában.

Köszönöm *Nagy József* Professzor Úr támogatását és segítségét a mérőeszköz adaptálásában és továbbfejlesztésében.

Köszönöm *Józsa Krisztián* és *Vígh Tibor* házi védésem opponenseinek az építő javaslataikat, segítő kritikáikat.

Köszönöm az Oktatáselméleti Kutatócsoport munkatársainak a mérések megszervezésében és lebonyolításában végzett támogatását.

Köszönöm a mérésekben közreműködő tanárok és tanulók segítségét.

Köszönöm a Neveléstudományi Doktori Iskola minden oktatójának a színvonalas kurzusokon való részvétel lehetőségét, mellyel segítették számomra a kutatásom megvalósítását.

IRODALOM

- Adema, J. J. (1990): The Construction of customized two-stage tests. *Journal of Educational Measurement*, **27**. 3. sz. 241–253.
- Al-A'ali, M. (2007): Implementation of an improved adaptive testing theory. *Educational Technology & Society*, **10**. 4. sz. 80–94.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999): *Standards for educational and psychological testing*. American Educational Research Association, Washington, DC.
- American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment (1986): *Guidelines for computer-based tests and interpretations*. Author, Washington, DC.
- Amstrong, R. D. (2002): *Routing rules for Multiple-Form Structures*. (Computerized Testing Report 02-08). Law School Admission Council, Newtown, PA.
- Amstrong, R. D., Jones, D. H., Koppel, N. B. és Pashley, P. J. (2004): Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement*, **28**. 3. sz. 147–164.
- Baker, F. B. (2001): *The Basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation. University of Maryland, College Park, Maryland, MD.
- Balázsi Ildikó és Ostorics László (2011): *PISA 2009 Digitális szövegértés Olvasás a világhálón*. Oktatási Hivatal, Budapest.
- Balázsi Ildikó, Ostorics László, Szalay Balázs, Szepesi Ildikó és Vadász Csaba (2013): *PISA 2012 Összefoglaló jelentés*. Oktatási Hivatal, Budapest.
- Barrada, J. R. Olea, J., Ponsoda, V., és Abad, F. J. (2009): Item selection rules in computerized adaptive testing accuracy and security. *Methodology*, **5**. 1. sz. 7–17.
- Beller, M. (2013): Technologies in large-scale assessments: New directions, challenges, and opportunities. In: von Davier, M., Gonzalez, E., Kirsch, I., és Yamamoto, K. (szerk.) *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Springer, Netherlands. 25–45.
- Bennett, R. E., Persky, H., Weiss, A. R. és Jenkins, F. (2007): *Problem solving in technology-rich environments: A report from the NAEP technology-based assessment project*. NAEP, Washington, DC.

- Blomert, L. és Csépe Valéria (2012): Az olvasástanulás és –mérés pszichológiai alapjai. In: Csapó Benő és Csépe Valéria (szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 17–86.
- Burghof, K. L. (2001): Assembling an item-bank for computerised linear and adaptive testing in Geography. *International Education Journal*, **2**. 4. sz. 74–83.
- Brossman, B. G. és Guille, R. A. (2014): A Comparison of multi-stage and linear test designs for medium-size licensure and certification examinations. *Journal of Computerized Adaptive Testing*, **2**. 2. sz. 18–36.
- Chang, S. W. és Ansley, T. N. (2003): A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, **40**. 1. sz. 71–103.
- Čisar, S. M., Radosav, D., Markoski, B., Pinter R. és Čisar, P. R. (2010): Computer adaptive testing of student knowledge. *Acta Polytechnica Hungarica*, **7**. 4. sz. 139–152.
- Clauser B. és Linacre J. M. (1999): Relating Cronbach and Rasch Reliabilities. *Rasch Measurement Transactions*, **13**. 2. sz. 696–697.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Crotts, K. M., Zenisky, A. L., Sireci, S. G. és Li, X. (2013): Estimating measurement precision in reduced-length multi-stage adaptive testing. *Journal of Computerized Adaptive Testing*, **1**. 4. sz. 67–87.
- Csapó, B., Ainley, J., Bennett, R. E., Latour, T. és Law, N. (2012): Technological issues for computer-based assessment. In: Griffin, P., McGaw, B. és Care, E. (szerk.): *Assessment and teaching of 21st century skills*. Springer, New York. 143–230.
- Csapó Benő (1994): Az induktív gondolkodás fejlődése. *Magyar Pedagógia*, **94**. 1–2. sz. 53–80.
- Csapó Benő (2001): Az induktív gondolkodás fejlődésének elemzése országos reprezentatív felmérés alapján. *Magyar Pedagógia*, **11**. 3. sz. 373–391.
- Csapó Benő (2002): Az új tudás képződésének eszközei: az induktív gondolkodás. In: Csapó Benő (szerk.): *Az iskolai tudás*. (2. kiadás). Osiris Kiadó, Budapest. 261–29.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papíralapú tesztekől a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, **18**. 3–4. sz. 3–16.

- Csapó Benő és Csépe Valéria (2012, szerk.): *Tartalmi keretek az olvasás diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest.
- 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.
- Csapó Benő és Molnár Gyöngyvér (2012): Gondolkodási készségek és képességek. In: Csapó Benő (szerk.): *Mérlegen a magyar iskola*. Nemzeti Tankönyvkiadó, Budapest. 407–440.
- Davis, L. L., és Dodd, B. G. (2003): Item Exposure Constraints for Testlets in the Verbal Reasoning Section of the MCAT. *Applied Psychological Measurement*, **27**. 5. sz. 335–356.
- Davis, S. L. (2005): *Exploring a new methodology for setting performance level standards with computerized adaptive tests*. 35th Annual National Conference on Large-Scale Assessment, San Antonio, Texas.
- Diao, Q. és Reckase, M. (2009): Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing. In: Weiss, D. J. (szerk.): *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. <http://publicdocs.iacat.org/cat2010/cat09diao.pdf>. Letöltés dátuma: 2015.01.12.
- 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. (2004): *Contributions to the theory and practice of computerized adaptive testing*. Citogroep Arnhem, Netherlands.
- 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 dátuma: 2013.12.12.
- Eggen, T. J. H. M. és Straetmans, G. J. J. M. (2000): Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, **60**. 5. sz. 713–734.

- Eggen, T. J. H. M. és Verschoor, A. J. (2006): Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement*, **30**. 5. sz. 379–393.
- 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. 103–133.
- Georgiadou, E., Triantafillou, E. és Economides, A. (2007): A Review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment*, **5**. 8. sz. 5–38.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. és Reckase, M. D. (1984): Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, **21**. 4. sz. 347–360.
- Greiff, S., Wüstenberg, S. és Funke, J. (2012): Complex Problem Solving. More than reasoning? *Intelligence*, **40**. 1. sz. 1–14.
- Guille, R. A., Becker, K. A., Zhu, R. X., Zhang, Y., Song, H. és Sun, L. (2011): *Comparison of asymmetric early termination MST with linear testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Gupta, S. S., Vijn, V. és Vats, K. (2010): *Design and implementation of adaptive assessment system using item response theory model*. ASCNT, CDAC, Noida, India. 115–124.
- Hambleton, R. K. és Xing, D. (2006): Optimal and nonoptimal computer-based test designs for making pass–fail decisions. *Applied Measurement in Education*, **19**. 3. sz. 221–239.
- Han, K. T. (2012): Fixing the c parameter in the three-parameter logistic model. *Practical Assessment, Research and Evaluation*, **17**. 1. sz. 2–24.

- Harris D. (1989): Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, **8**. 1. sz. 35–41.
- 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.
- Hendrickson, A. (2007): An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, **26**. 2. sz. 44–52.
- Hoe, L. S., Kiong, L. N., Sam, H. K. és Usop, H. B. (2009): *Improving educational assessment: A computer-adaptive multiple choice assessment using NRET as the scoring method*. US-China Education Review, **6**. 5. sz. 51–60.
- 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, Department of Educational Psychology, Minneapolis, MN.
- Jodoin, M., Zenisky, A. és Hambleton, R. K. (2006): Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, **19**. 3. sz. 203–220.
- 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.
- Keng, L., Tusng-Han, H., Tzu-An, C. és Dodd, B. (2000): *A comparison of item and testlet selection procedures in computerized adaptive testing*. The University of Texas, Austin, TX.
- Keng, L. (2008): *A comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas, Austin, TX.
- Kim, H., és Plake, B. S. (1993): *Monte Carlo simulation comparison of two-stage testing and computerized adaptive testing*. Presented at the Annual Meeting of the National Council on Measurement in Education, Atlanta.

- Kingsbury, G. G. és Hauser, C. (2004): *Computerized adaptive testing and the No Child Left Behind*. Presented at the Annual Meeting of the American Educational Research Association. San Diego, CA.
- Kingsbury, G. G. és Zara, A. R. (1989): Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, **2**. 4. sz. 359–375.
- Kingsbury, G. G. és Wise, S. L. (2011): Creating a K-12 adaptive test: Examining the stability of item parameter estimates and measurement scales. *Journal of Applied Testing Technology*, **12**. 5. sz. 1–27.
- Klauer, K. J. (1990): A process theory of inductive reasoning tested by the teaching of domain-specific thinking strategies. *European Journal of Psychology of Education*, **5**. 2. sz. 191–206.
- Kolen, M. J. (1999-2000): Threats to score comparability with applications to performance assessments and computerized adaptive tests. *Educational Assessment*, **6**. 2. sz. 73–96.
- Kolen, M. J. és Brennan, R. L. (2010): *Test equating, scaling, and linking*. Springer, New York.
- Lee, J. E. (2011): *Full-metric concurrent calibration for the development of CAT item banks*. University of Minnesota Digital Conservancy, Minnesota, MN.
- Linacre, J. M. (1997): Kr-20/Cronbach alpha or Rasch reliability: Which tells the truth? *Rasch Measurement Transactions*, **11**. 3. sz. 580–581.
- Linacre, J. M. (2000): *Computer-adaptive testing: A methodology whose time has come*. MESA Psychometric Laboratory, University of Chicago, Chicago, IL.
- Lord, F. M. (1971a): The self-scoring flexilevel test. *Journal of Educational Measurement*, **8**. 3. sz. 147–151.
- Lord, F. M. (1971b): A theoretical study of the measurement of effectiveness of flexilevel tests. *Educational and Psychological Measurement*, **31**. 4. sz. 805–813.
- Lu, R. (2010): *Impacts of local item dependence of testlet items with the multistage tests for pass-fail decisions*. Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, Maryland, MD.
- Luecht, R. M. és Sireci, S. G. (2011): *A Review of models for computer-based testing. Research Report 2011-12*. College Board, New York, NY.
- Magyar Andrea (2012): Számítógépes adaptív tesztelés. *Iskolakultúra*, **22**. 6. sz. 52–60.

- Magyar Andrea (2013): Többszakaszos adaptív tesztek felépítése, működése. *Oktatás-Informatika*, 1-2. sz. <http://www.oktatas-informatika.hu/2013/11/magyar-andrea-tobbszakaszos-adaptiv-tesztek-felepitesi-mukodese>. Letöltés dátuma: 2014.02.11.
- Magyar Andrea (2014a): Adaptív tesztek készítésének folyamata. *Iskolakultúra*, **24.** 4. sz. 26–33.
- Magyar Andrea (2014b): Szóolvasási készséget mérő adaptív tesztelésre alkalmas feladatbank fejlesztése. VI. Oktatás-informatika Konferencia tanulmánykötete. 404–412. http://www.eltereader.hu/media/2014/03/VI_OKTINF_Tanulmanykotet_READER.pdf Letöltés dátuma: 2014.10.12.
- Magyar Andrea (2014c): A problémamegoldó gondolkodás vizsgálata adaptív tesztek alkalmazásával. A VIII. Kiss Árpád Emlékkonferencia előadásainak szerkesztett változata. Tartalmi összefoglalók. Debreceni Egyetem Neveléstudományi Intézete, Debrecen. 211–221.
- Magyar Andrea és Molnár Gyöngyvér (2013): Adaptív és lineáris formátumú tesztek alkalmazásának összehasonlító hatékonyságvizsgálata. *Magyar Pedagógia*, **113.** 3. sz. 181–193.
- Magyar Andrea és Molnár Gyöngyvér (2014): A szóolvasási készség adaptív mérését lehetővé tevő online tesztrendszer kidolgozása. *Magyar Pedagógia*, **114.** 4. sz. 259–279.
- Masters, G. N. (1982): A Rasch model for partial credit scoring. *Psychometrika*, **47.** 2. sz. 149–174.
- Meijer, R. R. és Nering, M. R. (1999): Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, **23.** 3. sz. 187–194.
- Milman, J. és Arter, J. A. (1984): Issues in item banking. *Journal of Educational Measurement*, **21.** 4. sz. 315–330.
- Molnár Gyöngyvér (2003): Az ismeretek alkalmazásának vizsgálata modern tesztelméleti (IRT) eszközökkel. *Magyar Pedagógia*, **103.** 4. sz. 423–446.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, **106.** 12. sz. 99–113.
- Molnár Gyöngyvér (2006a): Az induktív gondolkodás fejlesztése kisiskolás korban. *Magyar Pedagógia*, **106.** 1. sz. 63–80.

- Molnár Gyöngyvér (2006b): Az ismeretek alkalmazhatóságának korlátai: komplex problémamegoldó gondolkodás fejlettsége 7. és 11. évfolyamon. *Magyar Pedagógia*, **106.** 4. sz. 329–344.
- Molnár Gyöngyvér (2006c): *Tudástranszfer és komplex problémamegoldás*. Műszaki Kiadó, Budapest.
- Molnár Gyöngyvér (2008): Kisiskolások induktív gondolkodásának játékos fejlesztése. *Új Pedagógiai Szemle*, **108.** 5. sz. 51–64.
- Molnár Gyöngyvér (2009): Kisiskolás diákok számára kidolgozott induktív gondolkodás fejlesztő program hosszabb távú hatása. In: Perjés István és Kozma Tamás (szerk.): *Új kutatások a neveléstudományokban. Hatékony tudomány, pedagógiai kultúra, sikeres iskola*. Magyar Tudományos Akadémia, Budapest. 118–129.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, **20.** 7–8. sz. 22–34.
- Molnár Gyöngyvér (2012): A problémamegoldó gondolkodás fejlődése: az intelligencia és szocioökonómiai háttér befolyásoló hatása 3–11. évfolyamon. *Magyar Pedagógia*, **112.** 1. sz. 41–58.
- Molnár Gyöngyvér és Csapó Benő (2011): Az 1–11 évfolyamot átfogó induktív gondolkodás kompetenciaskála készítése a valószínűségi tesztelmélet alkalmazásával. *Magyar Pedagógia*, **111.** 2. sz. 127–140.
- Molnár Gyöngyvér és Thibaud Latour (2011): *Online tesztelés: lehetőségek és kihívások*. IX. Pedagógiai Értékelési Konferencia. Szeged, 2011. április 29–30. 63.
- Molnár Gyöngyvér (2013): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér és Csapó Benő (2013): *A eDia online diagnosztikus mérési rendszer*. XI. Pedagógiai Értékelési Konferencia. Szeged, 2012. április 11–13. 82.
- Molnár Gyöngyvér és Magyar Andrea (2015): A számítógép alapú tesztelés elfogadottsága pedagógusok és diákok körében. *Magyar Pedagógia*, **115.** 1. sz. 49–66.
- 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.

- 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 dátuma: 2013.06.04.
- 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.
- Olea, J., Revuelta, J., Ximénez, M.C. és Abad, F. J. (2000): Psychometric and psychological effects of review on computerized fixed and adaptive tests. *Psicológica*, **21**. 1. sz. 157–173.
- Paek, P. (2005): *Recent trends in comparability studies*. PEM Research Report 05-05. Pearson Educational Measurement, Upper Saddle River, NJ.
- Partchev, I. (2004): A visual guide to item response theory. <http://www.metheval.unijena.de/irt/VisualIRT.pdf>. Letöltés dátuma: 2013.02.13.
- Pásztor-Kovács Anita, Magyar Andrea, Hülber László, Pásztor Attila és Tongori Ágota (2013): Áttérés online tesztelésre – a mérés-értékelés új dimenziói. *Iskolakultúra*, **23**. 11. sz. 86–100.
- Patsula, L. N. (1999): *A comparison of computerized adaptive testing and multistage testing*. Electronic Doctoral Dissertation. <http://scholarworks.umass.edu/dissertations/AAI9950199>. Letöltés dátuma: 2013.02.11.
- Puhan, G. és Gierl, M. J. (2003): *Evaluating the comparability of English- and French-speaking examinees on a science achievement test administered using two-stage testing*. Presented at the Annual Meeting of the National Council on Measurement in Education (NCME) at the Symposium entitled, "Test Adaptations and Translations: Developments and Evaluation Advances" Chicago, Illinois, USA. 2003. április 22–24.

- Pyper, A. és Lilley, M. (2010): *A comparison between the flexilevel and conventional approaches to objective testing*. CAA Konferencia, University of Hertfordshire, Hertfordshire.
- Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen.
- Reckase, M. D. (2007): The design of p-optimal item bank for computerized adaptive tests. In: Weiss, D. J. (szerk.): *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*.
<https://www.psych.umn.edu/psylabs/catcentral/pdf%20files/cat07reckase.pdf>.
 Letöltés dátuma: 2012.02.12.
- Reckase, M. D. (2009): Computerized adaptive testing using MIRT. In: Reckase, M. D. (szerk.): *Multidimensional item response theory*. Springer, Dordrecht. 311–339.
- Revuelta, J. és Ponsoda, V. (1998): A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, **35**. 4. sz. 311–327.
- Rijn, P. W., Eggen, T. J. H. M., Hemker, B. T. és Sanders, P. F. (2000): *A selection procedure for polytomous items in computerized adaptive testing*. Measurement and Research Department Reports (2000-5), Cito, Arnhem.
- Rotou, O., Patsula, L., Manfred, S. és Rizavi, S. (2003): *Comparison of multi-stage tests with computerized adaptive and paper and pencil tests*. Presented at the annual meeting of the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), Chicago, IL.
- R. Tóth Krisztina, Molnár Gyöngyvér, Thibaud Latour és Csapó Benő (2011): Az online tesztelés lehetőségei és a TAO platform alkalmazása. *Új Pedagógiai Szemle*, **61**. 1–5. sz. 8–22.
- Scheuermann, F. és Björnsson, J. (2009, szerk.): *The Transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg.
- Scheuermann, F. és Pereira, G. A. (2008, szerk.): *Towards a research agenda on computer-based assessment*. Office for Official Publications of the European Communities, Luxemburg.
- Schnipke, D. L. és Green, B. F. (1995): A comparison of item selection routines in linear and adaptive tests. *Journal of Educational Measurement*, **32**. 3. sz. 227–242.

- Segall, D. O. (2004). Computerized adaptive testing. In: Kimberly, K. L. (szerk.): *Encyclopedia of social measurement*. Elsevier Academic Press, San Diego, CA. <http://www.psych.umn.edu/psylabs/catcentral/pdf%20files/se04-01.pdf>. Letöltés dátuma: 2012.02.01.
- Sereci, S. G. (2003): *Computerized adaptive testing: An introduction*. In: Wall, J. L. ésWalz G. R. (szerk.): *Measuring up: Assessment issues for teachers, counsellors and administrators*. CAPS Press, Greensboro. 685–694.
- Stocking, M. L. (1994): *Three practical issues for modern adaptive testing item pools* (ETS Research Report RR-94-5). Educational Testing Service, Princeton, NJ.
- Stocking, M. L. (1996): An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics*, **21**. 4. sz. 365–389.
- Swanson, L. és Stocking, M. L. (1993): A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, **17**. 2. sz. 151–166.
- Sympson, J. B. és Hetter, R. D. (1985): *Controlling item exposure rates in computerized adaptive testing*. Proceedings of the 27th Annual Meeting of the Military Testing Association. Navy Personnel Research and Development Center, San Diego, CA. 973-977.
- 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.
- Thompson, T. és Way, D. (2007): *Investigating CAT designs to achieve comparability with a paper test*. Presented at the Applications and Issues Paper Session. Minneapolis, MN.
- Tian, J., Miao, D. és Zhu Xia, G. J. (2007): An Introduction to the computerized adaptive testing. *US-China Education Review*, **4**. 1. sz. 72–81.
- van der Linden, W. J. (2001): Computerized test construction. In: Smelser, N. J. és Baltes, P. B. (szerk.): *International Encyclopedia of Social and Behavioral Sciences*. Elsevier Science Ltd., Oxford. 2477–2480.
- van der Linden, W. J. (2005): A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, **42**. 3. sz. 283–302.

- van der Linden, W. J., Ariel, A. és Veldkamp, B. P. (2006): Assembling a computerized adaptive testing item pool as a set of linear tests. *Journal of Educational and Behavioral Statistics*, **31**. 1. sz. 81–99.
- van der Linden, W. J. (2008): Some new developments in adaptive testing technology. *Zeitschrift für Psychologie*, **216**. 1. sz. 3–11.
- van der Linden, W. J. és Glas, C. A. W. (2000): *Computerized adaptive testing: Theory and practice*. Kluwer, Boston.
- van der Linden, W. J. és Glas, C. A. W. (2010): *Elements of Adaptive Testing*. Springer, New York.
- Veerkamp, V. J. J. és Berger, M. P. F. (1997): Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, **22**. 2. sz. 203–226.
- Vidákovich Tibor (1993): *Diagnosztikus tesztbankok*. Pedagógiai diagnosztika 2. Alapműveltségi Vizsgaközpont, Szeged. 7–23.
- Vispoel, W. P., Hendrickson, A. B. és Bleiler, T. (2000): Limiting answer review and change on computerized adaptive vocabulary tests. Psychometric and attitudinal results. *Journal of Educational Measurement*, **37**. 1. sz. 21–38.
- Wainer, H. (2000a): *Computerized adaptive testing: A primer* (2nd Edition). Erlbaum, Hillsdale, NJ.
- Wainer, H. (2000b): CATs: Whither and whence. *Psicológica*, **21**. 1. sz. 121–133.
- Wainer, H. és Kiely, G. (1987): Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, **24**. 3. sz. 185–201.
- Wan, L., Keng, L., McCarty, K. és Davis, L. (2009): Methods of comparability studies for computerized and paper-based tests. *Test, measurements and research services bulletin*, **12**. 10. sz. 1–4.
- Wang, H. (2010): Comparability of computerized adaptive and paper-pencil tests. *Test, measurements and research services bulletin*, **13**. 1. sz. 1–7.
- Wang, T. és Kolen, M. J. (2001): Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement*, **38**. 1. sz. 19–49.
- Wang, S., Jiao, H., Young, M. J., Brooks, T. és Olson, J. (2008): Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments. *Educational and Psychological Measurement*, **68**. 1. sz. 5–24.

- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): *Practical questions in introducing computerized adaptive testing for K-12 assessments*. Pearson, San Antonio.
- Weiss, D. J. (1973): *The stratified adaptive computerized ability test* (Research Report 73–3). Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, Minneapolis, MN.
- Weiss, D. J. (1974): *Strategies of ability measurement*. (Research Report 74). Psychometric Research Program, Department of Psychology, University of Minnesota, Minneapolis, MN.
- Weiss, D. J. (2004): Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, **37**. 2. sz. 70–84.
- Weiss, D. J. (2011): Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, **2**. 1. sz. 1–27.
- Weiss, D. J. (2013): Item banking, test development, and test delivery. In: Geisinger, K. F. (szerk.): *The APA handbook on testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association, Washington DC. 185–200.
- Weiss, D. J. és Kingsbury, G. G. (1984): Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, **21**. 4. sz. 361–375.
- Wise, S. L. és Kingsbury, G. G. (2000): Practical issues in developing and maintaining a computerized adaptive testing program. *Psicológica*, **21**. 1. sz. 135–155.
- Yan, D., von Davier, A. A. és Lewis, C. (2014): *Computerized multistage testing: Theory and applications*. CRC Press, New York.
- Yi, Q. Wang, T. és Ban, J. C. (2001): Effects of scale transformation and test-termination rule on the precision of ability estimation in computerized adaptive testing. *Journal of Educational Measurement*, **38**. 3. sz. 267–292.
- 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 Modeling*, **54**. 4. sz. 366–396.
- Yu, C. H., Jannasch-Pennell, A. és DiGangi, S. (2008): A non-technical approach for illustrating item response theory. *Journal of Applied Testing Technology*, **9**. 2. sz. 1–32.

- Zenisky, A. L. és Hambleton, R. K. (2004): *Effects of selected multi-stage test design alternatives on credentialing examination outcomes*. Paper presented at the annual meeting of NCME, San Diego, CA.
- 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.
- Zheng, Y. (2012): Multistage adaptive testing for a large-scale classification test: Design, heuristic assembly, and comparison with other testing modes, ACT research report series.

ÁBRAJEGYZÉK

1. ÁBRA: ADAPTÍV TESZTELÉS ALGORITMUSA (LINARCE, 2000 ALAPJÁN)	13
2. ÁBRA: ITEMKARAKTERISZTIKUS GÖRBE (MOLNÁR, 2013 ALAPJÁN).....	15
3. ÁBRA: AZ EGYPARAMÉTERES LOGISZTIKUS MODELL ITEMKARAKTERISZTIKUS GÖRBÉJE	17
4. ÁBRA: A KÉTPARAMÉTERES LOGISZTIKUS MODELL ITEMKARAKTERISZTIKUS GÖRBÉJE	18
5. ÁBRA: A HÁROMPARAMÉTERES LOGISZTIKUS MODELL ITEMKARAKTERISZTIKUS GÖRBÉJE.....	19
6. ÁBRA: TESZT KARAKTERISZTIKUS GÖRBE.....	20
7. ÁBRA: EGYPARAMÉTERES ITEMKARAKTERISZTIKUS ÉS INFORMÁCIÓS GÖRBÉJE (MOLNÁR, 2013 ALAPJÁN)	21
8. ÁBRA: HAT KÜLÖNBÖZŐ ITEM ITEM-INFORMÁCIÓS FÜGGVÉNYE	22
9. ÁBRA: A FELADATBANKHOZ SZÜKSÉGES ITEMSZÁM A TANULÓK SZÁMÁNAK FÜGGVÉNYÉBEN (RECKASE, 2007 ALAPJÁN)	25
10. ÁBRA: AZ ÁRNYÉKTESZT MEGKÖZELÍTÉS GRAFIKUS ÁBRÁZOLÁSA (A SÖTÉT TERÜLET MUTATJA A KIKÖZVETÍTETT ITEMET, A VILÁGOSABB TERÜLET AZ EGY-EGY ITEM MEGOLDÁSÁT KÖVETŐEN ÚJRASTRUKTURÁLT LINEÁRIS TESZTET; VAN DER LINDEN, ARIEL ÉS VELDKAMP ALAPJÁN, 2006)	30
11. ÁBRA: 1-3-3-3 SZERKEZETŰ NÉGYSZAKASZOS TESZT (MOLNÁR, 2013 ALAPJÁN).....	36
12. ÁBRA: OSZTÓPONTOK MEGHATÁROZÁSA AZ NC MÓDSZER SZERINT: 10 ITEMBŐL ÁLLÓ MODULNÁL 4 ÉS 6 PONTNÁL HELYEZKEDNEK EL AZ OSZTÓPONTOK	38
13. ÁBRA: MINTAFELADATOK AZ INDUKTÍV GONDOLKODÁST MÉRŐ TESZTBŐL (CSAPÓ, 2002 ALAPJÁN)	53
14. ÁBRA: MINTAFELADATOK AZ INDUKTÍV GONDOLKODÁST MÉRŐ TESZTBŐL (MOLNÁR, 2006 ALAPJÁN).....	54
15. ÁBRA: A KOMPLEX PROBLÉMAMEGOLDÓ FELADATLAP EGY MINTAFELADATA	56
16.A ÁBRA: ZÁRT TÍPUSÚ FELADAT MEGJELENÍTÉSE PAPÍR ALAPON ÉS SZÁMÍTÓGÉP ALAPON	63
16.B ÁBRA: NYÍLT VÉGŰ FELADAT MEGJELENÍTÉSE PAPÍR ALAPON ÉS SZÁMÍTÓGÉP ALAPON.....	63
17. ÁBRA: A TÖBBSZAKASZOS ADAPTÍV TESZT SZERKEZETE	64
18. ÁBRA: AZ ADATFELVÉTEL ALAPJÁN KIRAJZOLHATÓ SZEMÉLY-ITEM TÉRKÉP	66
19. ÁBRA: A LINEÁRIS ÉS AZ ADAPTÍV TESZTEN NYÚJTOTT TELJESÍTMÉNYEK ÖSSZEHOSSZONLÍTÁSA DIÁKONKÉNTI BONTÁSBAN	68
20. ÁBRA: AZ ADAPTÍV TESZTRENSZEREN BELÜL A MÁSODIK, HARMADIK ÉS NEGYEDIK SZAKASZBAN KIOSZTOTT ÚTVONALAK GYAKORISÁGA (K: KÖNNYŰ, Á: ÁTLAGOS, N: NEHÉZ MODUL)	70
21. ÁBRA: A TANULÓK GYAKORISÁGI ÉS KÉPESSÉGSZINT SZERINTI ELOSZLÁSA A SZAKASZOKON ÉS A MODULOKON BELÜL	71
22. ÁBRA: AZ ADAPTÍV ÉS A LINEÁRIS FORMÁTUMÚ LINEÁRIS TESZT STANDARD HIBÁINAK ALAKULÁSA A TANULÓK KÉPESSÉGSZINTJÉNEK FÜGGVÉNYÉBEN.....	72
23. ÁBRA: AZ ADAPTÍV TECHNIKÁVAL ÖSSZEÁLLÍTOTT ÉS A LINEÁRIS FORMÁTUMÚ, AZONOS NEHÉZSÉGŰ RÉSZTESZTEKET TARTALMAZÓ TESZTEK INFORMÁCIÓS FÜGGVÉNYEI	73
24. ÁBRA: A PROBLÉMAMEGOLDÓ KÉPESSÉGET MÉRŐ TESZT EGY MINTAFELADATA.....	76
25. ÁBRA: AZ ADAPTÍV TESZT SZERKEZETE	77
26. ÁBRA: AZ ADATFELVÉTEL ALAPJÁN KIRAJZOLHATÓ SZEMÉLY-ITEM TÉRKÉP	79
27. ÁBRA: A KÉTFÉLE TESZTEN ELÉRT EREDMÉNYEK SZEMÉLYSZINTŰ BONTÁSBAN	80
28. ÁBRA: AZ ADAPTÍV TESZTEN BELÜL A KIOSZTÁSRA KERÜLŐ ÚTVONALAK GYAKORISÁGA (Á: ÁTLAGOS, K: KÖNNYŰ, N: NEHÉZ MODUL).....	82
29. ÁBRA: AZ ADAPTÍV ÉS A LINEÁRIS TESZTEKBŐL KINYERT INFORMÁCIÓK MENNYISÉGE	83
30. ÁBRA: AZ ADAPTÍV ÉS A LINEÁRIS TESZT STANDARD HIBÁINAK ALAKULÁSA A TANULÓK KÉPESSÉGSZINTJE FÜGGVÉNYÉBEN.....	84

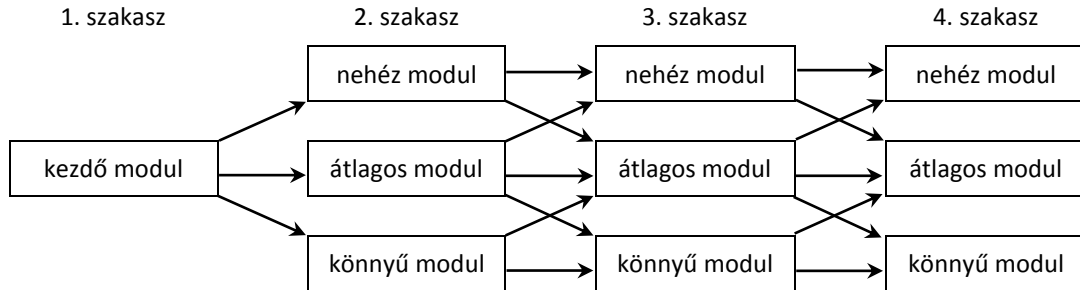
31.A ÁBRA: CÍMSZÓOLVASÁS FELADAT (PAPÍRALAPÚ ÉS SZÁMÍTÓGÉPES FORMÁTUMBAN).....	88
31.B ÁBRA: TOLDALÉKOS SZÓOLVASÁS FELADAT (PAPÍRALAPÚ ÉS SZÁMÍTÓGÉPES FORMÁTUMBAN)	89
31.C ÁBRA: SZINONIMAOLVASÁS FELADAT (PAPÍRALAPÚ ÉS SZÁMÍTÓGÉPES FORMÁTUMBAN).....	89
31.D ÁBRA: SZÓJELENTÉS OLVASÁS FELADAT (PAPÍRALAPÚ ÉS SZÁMÍTÓGÉPES FORMÁTUMBAN)	89
32. ÁBRA: A BIBD DESIGN (BALANCED INCOMPLETE BLOCK DESIGN; SÁRGA: CÍMSZÓ (C), KÉK: SZÓJELENTÉS (J), NARANCS: TOLDALÉKOS (T), PIROS: SZINONIMA (S)).....	92
33. ÁBRA: A FELADATOK NEHÉZSÉGI SZINTJE A DIÁKOK KÉPESSÉGSZINTJE FÉNYÉBEN – TÖBBDIMENZIÓS SZEMÉLY-ITEM TÉRKÉP [CÍMSZÓOLVASÁS (CÍMSZÓ), SZINONIMAOLVASÁS (SZIN), SZÓJELENTÉS OLVASÁS (SZÓJ) ÉS TOLDALÉKOS SZÓOLVASÁS (TOLD) DIMENZIÓKBAN]	94
34. ÁBRA: AZ EREDETI LINEÁRIS TESZTEK FELÉPÍTÉSE.....	95
35. ÁBRA: A SZÓOLVASÓ KÉSZSÉG MÉRÉSÉRE ALKALMAS NÉGYSZAKASZOS ADAPTÍV TESZTRENSZER SZERKEZETE	96
36.A ÁBRA: A KEZDŐ MODUL INFORMÁCIÓS FÜGGVÉNYE	97
36.B ÁBRA: A MÁSODIK SZAKASZ MODULJAINAK INFORMÁCIÓS FÜGGVÉNYEI.....	98
36.C ÁBRA: A HARMADIK SZAKASZ MODULJAINAK INFORMÁCIÓS FÜGGVÉNYEI.....	98
36.D ÁBRA: A NEGYEDIK SZAKASZ MODULJAINAK INFORMÁCIÓS FÜGGVÉNYEI.....	99
37. ÁBRA: A PILOT ADAPTÍV ADATFELVÉTEL ALAPJÁN KIRAJZOLHATÓ TÖBBDIMENZIÓS SZEMÉLY-ITEM TÉRKÉP [CÍMSZÓOLVASÁS (CÍMSZÓ), SZINONIMA (SZIN), SZÓJELENTÉS (SZÓJ) ÉS TOLDALÉKOS SZÓOLVASÁS (TOLD)].....	101
38. ÁBRA: A TANULÓK KÉSZSÉGSZINT SZERINTI ELOSZLÁSA ÉVFOLYAMONKÉNTI BONTÁSBAN	103
39. ÁBRA: A TANULÓK GYAKORISÁGI ÉS KÉPESSÉGSZINT SZERINTI ELOSZLÁSA A SZAKASZOKON ÉS A MODULOKON BELÜL.....	103
40.A ÁBRA: A NAGYMINTÁS LINEÁRIS ADATFELVÉTEL ALAPJÁN KIRAJZOLHATÓ TÖBBDIMENZIÓS SZEMÉLY-ITEM TÉRKÉP [CÍMSZÓOLVASÁS (CÍMSZÓ), SZINONIMA (SZIN), SZÓJELENTÉS (SZÓJ) ÉS TOLDALÉKOS SZÓOLVASÁS (TOLD)].....	108
40.B ÁBRA: A NAGYMINTÁS ADAPTÍV ADATFELVÉTEL ALAPJÁN KIRAJZOLHATÓ TÖBBDIMENZIÓS SZEMÉLY-ITEM TÉRKÉP [CÍMSZÓOLVASÁS (CÍMSZÓ), SZINONIMA (SZIN), SZÓJELENTÉS (SZÓJ) ÉS TOLDALÉKOS SZÓOLVASÁS (TOLD)].....	109
41. ÁBRA: A LINEÁRIS ÉS AZ ADAPTÍV TESZTEN NYÚJTOTT TELJESÍTMÉNYEK ÖSSZEHASONLÍTÁSA DIÁKONKÉNTI BONTÁSBAN	112
42. ÁBRA: AZ ADAPTÍV TESZTRENSZEREN BELÜL A MÁSODIK, HARMADIK ÉS NEGYEDIK SZAKASZBAN KIOSZTOTT ÚTVONALAK GYAKORISÁGA	114
43. ÁBRA: A TANULÓK GYAKORISÁGI ÉS KÉPESSÉGSZINT SZERINTI ELOSZLÁSA A SZAKASZOKON ÉS A MODULOKON BELÜL.....	115
44. ÁBRA: A LINEÁRIS ÉS AZ ADAPTÍV TECHNIKÁVAL ÖSSZEÁLLÍTOTT, AZONOS NEHÉZSÉGŰ MODULOKAT TARTALMAZÓ TESZTEK INFORMÁCIÓS FÜGGVÉNYEI.....	116
45. ÁBRA: A LINEÁRIS ÉS AZ ADAPTÍV TESZTEN KINYERT INFORMÁCIÓK A TANULÓK KÉPESSÉGSZINTJÉNEK FÜGGVÉNYÉBEN.....	118
46. ÁBRA: AZ ADAPTÍV ÉS A LINEÁRIS TESZT STANDARD HIBÁINAK ALAKULÁSA A TANULÓK KÉPESSÉGSZINTJÉNEK FÜGGVÉNYÉBEN	119

TÁBLÁZATOK JEGYZÉKE

1. TÁBLÁZAT: ADAPTÍV TESZTEKEN VÉGZETT ÖSSZEHASONLÍTÓ VIZSGÁLATOK	40
2. TÁBLÁZAT: A MINTÁK ÉVFOLYAMOK SZERINTI MEGOSZLÁSA.....	60
3. TÁBLÁZAT. A MINTA ÉVFOLYAM SZERINTI ELOSZLÁSA.....	62
4. TÁBLÁZAT. LINEÁRIS, ILLETVE ADAPTÍV TESZTKÖRNYEZETBEN MUTATOTT TELJESÍTMÉNYEK ÁTLAGA ÉS SZÓRÁSA ÉVFOLYAMONKÉNTI BONTÁSBAN.....	67
5. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ARÁNYA ÉVFOLYAMONKÉNTI SZERINTI BONTÁSBAN	68
6. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ARÁNYA KÉPESSÉGSZINT SZERINTI BONTÁSBAN .	69
7. TÁBLÁZAT. A MINTA ÉVFOLYAMOK SZERINTI ELOSZLÁSA	75
8. TÁBLÁZAT. A TANULÓK LINEÁRIS ÉS ADAPTÍV TESZTEN ELÉRT EREDMÉNYEINEK ALAPSTATISZTIKAI MUTATÓI ÉVFOLYAMONKÉNTI BONTÁSBAN.....	80
9. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ARÁNYA ÉVFOLYAMONKÉNTI SZERINTI BONTÁSBAN	81
10. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ARÁNYA KÉPESSÉGSZINT SZERINTI BONTÁSBAN	81
11. TÁBLÁZAT. A TANULÓK SZÁMÁNAK MEGOSZLÁSA A KÜLÖNBÖZŐ ÚTVONALAK KÖZÖTT ÉVFOLYAMONKÉNTI BONTÁSBAN	83
12. TÁBLÁZAT: A TESZT EAP/PV RELIABILITÁSA DIMENZIÓNKÉNTI BONTÁSBAN	93
13. TÁBLÁZAT. AZ ITEMEK ÁTLAGOS NEHÉZSÉGI INDEXEI A KÜLÖNBÖZŐ DIMENZIÓKBAN	95
14. TÁBLÁZAT: A KÜLÖNBÖZŐ SZINTEKEN MEGJELENŐ FELADATOK ÁTLAGOS NEHÉZSÉGI INDEXEI (ÁTLAG: 500, SZÓRÁS: 100).....	97
15. TÁBLÁZAT: A MINTA ÉVFOLYAMONKÉNTI ELOSZLÁSA	100
16. TÁBLÁZAT: A TESZT EAP/PV RELIABILITÁSA DIMENZIÓNKÉNTI BONTÁSBAN	100
17. TÁBLÁZAT. A TESZTEN ELÉRT KÉSZSÉGSZINTEK ÁTLAGA ÉS SZÓRÁSA ÉVFOLYAMONKÉNTI BONTÁSBAN	102
18. TÁBLÁZAT. A MINTA ÉVFOLYAM SZERINTI ELOSZLÁSA.....	105
19. TÁBLÁZAT: A TESZTEK EAP/PV RELIABILITÁSA DIMENZIÓNKÉNTI BONTÁSBAN	106
20.A TÁBLÁZAT: A LINEÁRIS TESZ KORRELÁCIÓ MÁTRIXA.....	110
20.B TÁBLÁZAT: AZ ADAPTÍV TESZT KORRELÁCIÓ MÁTRIXA.....	110
21. TÁBLÁZAT. A LINEÁRIS, ILLETVE ADAPTÍV TESZTKÖRNYEZETBEN MUTATOTT TELJESÍTMÉNYEK KLASSZIKUS TESZTELMÉLETI MUTATÓI ÉVFOLYAMONKÉNTI BONTÁSBAN.....	111
22. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ARÁNYA ÉVFOLYAMONKÉNTI SZERINTI BONTÁSBAN.....	113
23. TÁBLÁZAT. A TANULÓK ADAPTÍV ÉS LINEÁRIS TESZTEN ELÉRT HELYES VÁLASZAINAK ELOSZLÁSA KÉSZSÉGSZINT SZERINTI BONTÁSBAN	113
24. TÁBLÁZAT: A TANULÓK LINEÁRIS ÉS ADAPTÍV TESZTEN VALÓ SZINTENKÉNTI BESOROLÁSA	115
25. TÁBLÁZAT: A TESZTEKBŐL KINYERT INFORMÁCIÓK MENNYISÉGE KÉSZSÉGSZINTENKÉNTI BONTÁSBAN	117
26. TÁBLÁZAT: A TESZTEKEN MÉRT MÉRÉSI HIBÁK NAGYSÁGA KÉSZSÉGSZINTENKÉNTI BONTÁSBAN	119

MELLÉKLETEK

1. SZÁMÚ MELLÉKLET: AZ INDUKTÍV GONDOLKODÁST MÉRŐ TESZT FELADATAI



1. szakasz, kezdő modul:

A következő feladatokban azt kell megtalálni, melyik szó illik legjobban a kérdőjel helyére. Kattints a megfelelő szóra! Csak egy szót jelölj meg!

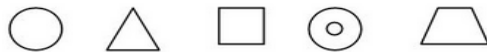
Baleset : Figyelmetlenség = Fagyás : ?

- Lehülés
- Olvadás
- Havazás
- Megszilárdulás
- Dér



2. szakasz, könnyű modul:

Kattints rá arra, amelyik a kakukktojás!



2. szakasz, közepes modul:

A következő feladatokban azt kell megtalálni, melyik szó illik legjobban a kérdőjel helyére. Kattints a megfelelő szóra! Csak egy szót jelölj meg!

Jókedv : Vidámság = Egyedüllét : ?

- Nevetés
- Távolság
- Bánat
- Magány
- Együttlét



2. szakasz, nehéz modul:

Folytasd a következő betűsorokat! Írd mindegyik sor végén a vonalra azt a két betűt, amelyek a sor folytatásaként legjobban illik oda! A sorokban csak az abc következő betűi fordulnak elő:

a b c d e f g h i j k l m n o p q r s t u x y v w z.

b d c e d f e



3. szakasz, könnyű modul:

A következő feladatokban azt kell megtalálni, melyik szó illik legjobban a kérdőjel helyére. Kattints a megfelelő szóra! Csak egy szót jelölj meg!

Falu : Település = Ceruza : ?

Szerszám

Radír

Papír

Ácsceruza

Írószköz



3. szakasz, közepes modul:

A következő feladatokban írd a feladat végén levő vonalra azt a számot, amelyik legjobban illik oda!

5->75 :: 3->45 :: 6->



3. szakasz, nehéz modul:

Folytasd a következő betűsorokat! Írd mindegyik sor végén a vonalra azt a két betűt, amelyik a sor folytatásaként legjobban illik oda! A sorokban csak az abc következő betűi fordulnak elő:

a b c d e f g h i j k l m n o p q r s t u x y v w z.

b e g j l o q



4. szakasz, könnyű modul:

Folytasd a sorozatot! Kattints rá arra, amelyik a kérdőjel helyére illik!



4. szakasz, közepes modul:

A következő feladatokban írd a feladat végén levő vonalra azt a számot, amelyik legjobban illik oda!

7->63 :: 4->36 :: 9->



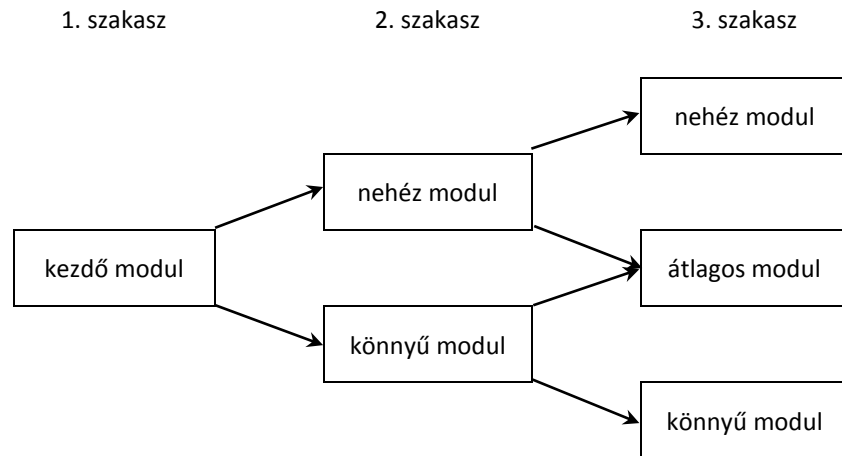
4. szakasz, nehéz modul:

Folytasd a következő számsorokat! Írd mindegyik sor végén a vonalra azt a két számot, amelyek a számsor folytatásaként legjobban illik oda!

1 10 26 51 87 136



2. SZÁMÚ MELLÉKLET: A PROBLÉMAMEGOLDÓ GONDOLKODÁST MÉRŐ TESZT FELADATAI



1. szakasz, kezdő modul:

Boeing 737-500 adatai

Törzs hossza: 29, 8 m
 Szárny fesztávolság: 28, 9 m
 Magasság: 11, 1 m
 Törzsszélesség: kb. 8m
 Utastér hossza: 22,3 m
 Maximális felszállótömeg: 57 t
 Utazómagasság: 11000 m
 Utazósebesség: 860-930 km/h
 Maximális hatótávolság: 4700 km
 Üzemanyag: 20 100 liter
 Hajtómű típusa: CFM 56
 Hajtómű teljesítménye: 2 x 9070 kp
 Hajózási személyzet: 2 fő
 Légiutas-kísérők: 4 fő

[← Vissza](#)

Vajon az átlagsebességünk meghaladta a 700 km/h-t?
 Az átlagsebesség kb.: km/h volt,
 annak ellenére, hogy amikor elértük az utazómagasságot,
 ennél jóval gyorsabban mentünk.
 Az adatok alapján mennyit fogyaszthatott átlagosan a
 gép? (Tegyük fel, hogy a maximális utazótávolságnál
 kiürül a tank.) Körülbelül:

430 liter
 330 liter
 530 liter
 230 liter

Körülbelül mennyi ideig tarthat, amíg feltankolják a
 gépet, ha percenként 220 liter kerozint tudnak beletöl-
 teni?

2 óra
 1,5 óra
 1 óra
 fél óra

Legalább mekkora hangárban fér el a gép?

862 m2
 630 m2
 320 m2
 331 m2

Hány ember férhetett bele, ha a tájékoztatóban lévő
 adatok szerint az utastérben egy emberre kb. 1,6 m2
 terület jut?

111
 123
 131
 149

A végén a fáradságtól álomba merültem...

[Mehet →](#)

2. szakasz, könnyű modul:

Holnap átjönnek a barátaim, Anna, Peti, Juli és Béci. Mindegyikőjüknek szeretnék adni egy kis ajándékot, mindenkinek mást, de csak 1000 Forint zsebpénzem maradt. Kattints arra, amiket választhattam. Végül sikerült venni mindenkinek egy kis apróságot.

A screenshot of a shopping cart interface. It displays eight different snack items with their respective prices in Hungarian Forint (Ft). The items are arranged in two rows. The first row contains: 'JóReggelt!' (150 Ft), 'BOCI' (250 Ft), and 'PILÓTA KAKAOS' (370 Ft). The second row contains: 'PISZTÁCIA' (350 Ft), 'Orsi Pórák' (380 Ft), 'Cujóri Edes' (380 Ft), and 'tibi' (240 Ft). At the bottom left, there is a 'Vissza' button, and at the bottom right, there is a 'Mehet' button.

Item	Price (Ft)
JóReggelt!	150
BOCI	250
PILÓTA KAKAOS	370
PISZTÁCIA	350
Orsi Pórák	380
Cujóri Edes	380
tibi	240

2. szakasz, nehéz modul:

Már kezdtem elfáradni a sok számolásban, szerencsére már csak egyféle kenyeret és tejfölt lehetett kapni, de a rizsnél, a chipseknél és a kávénál megint nagy volt a választék. Apu még megkért, hogy gyorsan számoljam ki, melyik rizst, chips-et és kávé-t éri meg jobban megvenni? Kattints arra, amelyik a relatív legolcsóbb!

A screenshot of a shopping cart interface. It displays ten different food items with their respective prices in Hungarian Forint (Ft). The items are arranged in three rows. The first row contains: 'Rizs' (1kg: 189 Ft), 'SOS' (500g: 105 Ft), 'Kávé' (1kg: 1990 Ft), and 'Dolce Gusto' (250g: 265 Ft). The second row contains: 'Dolce Gusto' (2kg: 395 Ft), 'Rizs' (500g: 85 Ft), 'Cimo Sajtos' (250g: 500 Ft), 'Lays' (250g: 520 Ft), 'Kávé' (500g: 489 Ft), and 'McCafé' (1/2kg: 550 Ft). The third row contains: 'Sun Chips' (250g: 480 Ft) and 'Fritos' (250+50g: 550 Ft). At the bottom left, there is a 'Vissza' button, and at the bottom right, there is a 'Mehet' button.

Item	Price (Ft)
Rizs	1kg: 189
SOS	500g: 105
Kávé	1kg: 1990
Dolce Gusto	250g: 265
Dolce Gusto	2kg: 395
Rizs	500g: 85
Cimo Sajtos	250g: 500
Lays	250g: 520
Kávé	500g: 489
McCafé	1/2kg: 550
Sun Chips	250g: 480
Fritos	250+50g: 550

3. szakasz, könnyű modul:

Hazafelé még beugrottunk egy cukrászdába fagyit venni. Mielőtt megjöttek a vendégek, megettem a fagyit két kilencedét, a blokkot meg eldugtam anyu elől. A két kilenced jobban hangzott, mint a megmaradt gombócok száma:

db

Dobó Cukrászda
 6798 Szeged
 Kikerics u. 999.
 Tel./fax: 657-8794

Téli nyitva tartás: Nyári nyitva tartás:
 Nov1-től ápr. 30-ig Máj. 01-től okt. 31-ig
 09⁰⁰ – 22⁰⁰ 09⁰⁰ – 23⁰⁰

Dobó Cukrászda



DODO CUKRÁSZDA
 Szeged, Kikerics u. 999.

Vanília gombóc	90	12 db	1080
Puncs gombóc	90	12 db	1080
Csokoládé gombóc	90	12 db	1080

Összesen: 3240

[Vissza](#)

[Mehet](#)

3. szakasz, közepes modul:

Apród pizzák				
Név	megnevezés	28 cm	32 cm	40 cm
▶ Sajtós	pizzaszósz, sajt, paradicsomleves	650	850	950
▶ Sonkás	pizzaszósz, sonka, sajt	650	850	950
▶ Gombás	pizzaszósz, gomba, sajt	650	850	950
▶ Kolbásos	pizzaszósz, kolbász, sajt	650	850	950
▶ Szalámi	pizzaszósz, szalámi, sajt	650	850	950
▶ Caribi	pizzaszósz, sonka, bókusi gyümölcsök, sajt	650	850	950
▶ Erdő kapitánya	pizzaszósz, sonka, gomba, sajt	650	850	950
▶ So-ku	pizzaszósz, sonka, kukorica, sajt	650	850	950

Főnemes pizzák				
Név	megnevezés	28 cm	32 cm	40 cm
▶ Hawaii	pizzaszósz, sonka, ananász, kukorica, sajt	750	950	1090
▶ Son-go-ku	pizzaszósz, sonka, gomba, kukorica, sajt	750	950	1090
▶ Spencer	tejtú, 10% tejzsír, bacon, mozzarella	750	950	1090
▶ Mexikói	chiliszósz, sonka, bab, kukorica, pfeferoni paprika, sajt	750	950	1090

Másnap délelőtt átjött négy haverom, 11-kor már nagyon éhesek voltunk, rendeltünk egy-egy pizzát. Anna és Juli közösen kértek egy sonkás pizzát, a fiúk pedig egy-egy kicsi gombásat, én egy közepes mexikóit. Ez vajon mennyibe kerülhetett?

3460 Ft
 3950 Ft
 3180 Ft
 2460 Ft

Miután megérkezett a pizzánk, olyan jó illata volt, hogy anyuék is úgy döntöttek, ők is pizzát esznek ebédre. Apa beült a kocsiba és hozott magának egy nagyobb Erdő kapitánya pizzát, anyának pedig egy kicsit kisebb ananászos pizzát. Mennyi pénzt kellett apunak legalább vinni a pizzériába?

1700 Ft
 2100 Ft
 2500 Ft
 2900 Ft

[Vissza](#)

Régen kellett már ennyi problémát megoldanom!

[Mehet](#)

3. szakasz, nehéz modul:

VALUTAÁRFOLYAM

Devizanem	Egység	Vétel	Eladás	Közép
 CAD	1.00	179.38	196.28	187.83
 CHF	1.00	160.21	170.81	165.51
 EGP	1.00	36.16	40.30	38.23
 EUR	1.00	251.44	265.92	258.68
 GBP	1.00	364.63	388.73	376.68
 HRK	1.00	31.78	38.85	35.31
 JPY	100.00	175.75	192.31	184.03
 NOK	1.00	30.81	33.71	32.26
 USD	1.00	210.04	223.92	216.98

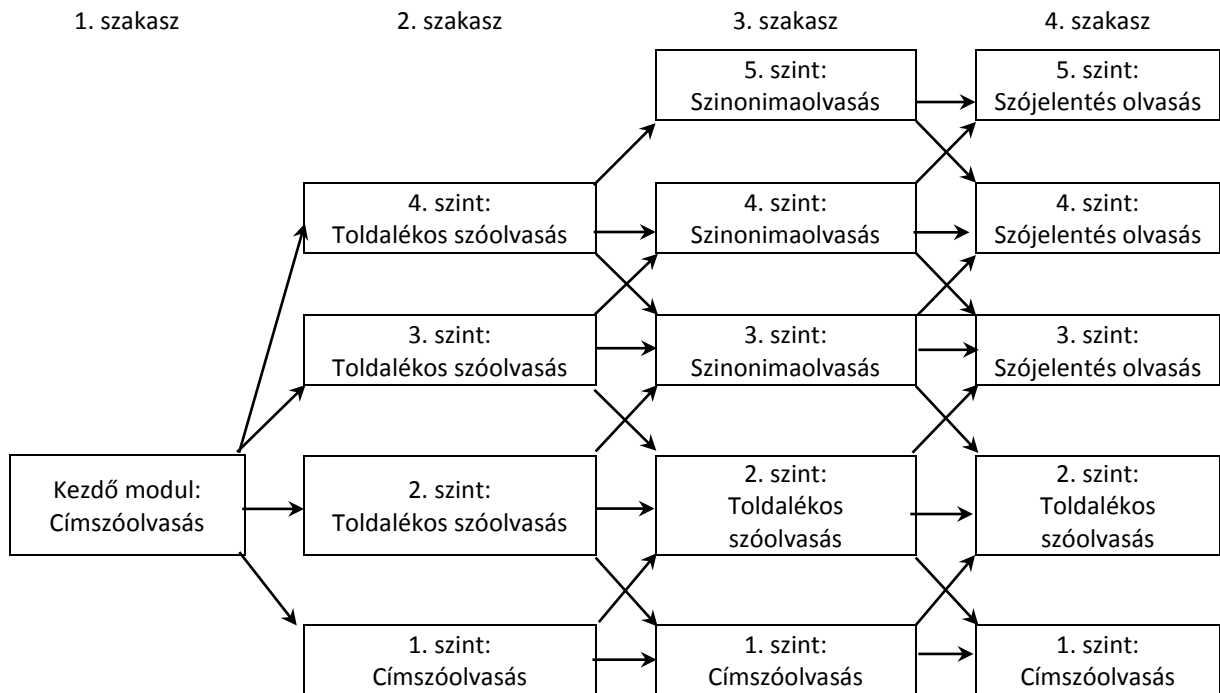
Az út csodálatos volt, viszont a 9000 egyiptomi fontból nem sok maradt meg. Apával elmentünk, hogy visszaváltsuk a hazahozott 890 egyiptomi fontot. A fontért a pénztáros leszámolt forintot.

Igazságtalanság, ha ugyanezt a pénzt ismét meg szeretném venni a pénztárban forintba kerülne.

[← Vissza](#)

[Mehet →](#)

3. SZÁMÚ MELLÉKLET: A SZÓOLVASÓ KÉSZSÉGET MÉRŐ TESZT FELADATAI



1. szakasz, kezdő modul, (címszóolvasás):

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!



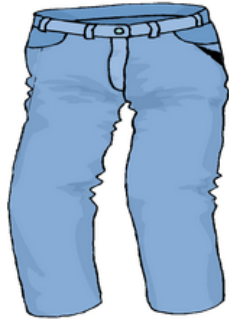
gyönyörű	<input type="radio"/> Igen	<input type="radio"/> Nem
izom	<input type="radio"/> Igen	<input type="radio"/> Nem
kerül	<input type="radio"/> Igen	<input type="radio"/> Nem
nyom	<input type="radio"/> Igen	<input type="radio"/> Nem

◀ Előző

Következő ▶

2. szakasz, 1. szint, (címszóolvasás):

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!



- öregasszony Igen Nem
- nadrág Igen Nem
- messzi Igen Nem
- sok Igen Nem

[← Előző](#)

[Következő →](#)

2. szakasz, 2. szint, (toldalékos szóolvasás):

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!



- alatta Igen Nem
- megdöbben Igen Nem
- lakóval Igen Nem
- érintett Igen Nem

[← Előző](#)

[Következő →](#)

2. szakasz, 3. szint, (toldalékos szóolvasás):

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!



ezután

Igen

Nem

átvilágítás

Igen

Nem

piszkosan

Igen

Nem

cigarettazik

Igen

Nem

[← Előző](#)

[Következő →](#)

2. szakasz, 4. szint, (toldakékos szóolvasás):

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!



bizonyítékképp

Igen

Nem

egészítette

Igen

Nem

megszabadul

Igen

Nem

nélkül

Igen

Nem

[← Előző](#)

[Következő →](#)

3. szakasz, 1. szint, (címszóolvasás):

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!



kopog

Igen

Nem

hátter

Igen

Nem

sátor

Igen

Nem

mély

Igen

Nem

[← Előző](#)

[Következő →](#)

3. szakasz, 2. szint, (toldalékos szóolvasás):

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!



területen

Igen

Nem

nevelte

Igen

Nem

elolvasáskor

Igen

Nem

aratás

Igen

Nem

[← Előző](#)

[Következő →](#)

3. szakasz, 3. szint, (szinonimaolvasás):

Hasonlítsd össze a kisbetűs és nagybetűs szavakat!

Döntsd el, van-e a kisbetűs szavaknak nagybetűs megfelelője!

Ha van, kattints a "Van" felírra, ha nincs, kattints a "Nincs" felírra!

FÜRGE

KÖRÜLBELÜL

kb.

Van

Nincs

gyors

Van

Nincs

[◀ Előző](#)

[Következő ▶](#)

3. szakasz, 4. szint, (szinonimaolvasás):

Hasonlítsd össze a kisbetűs és nagybetűs szavakat!

Döntsd el, van-e a kisbetűs szavaknak nagybetűs megfelelője!

Ha van, kattints a "Van" felírra, ha nincs, kattints a "Nincs" felírra!

SZÁN

LÉTREHOZ

alapít

Van

Nincs

sajnál

Van

Nincs

[◀ Előző](#)

[Következő ▶](#)

3. szakasz, 5. szint, (szinonimaolvasás):

Hasonlítsd össze a kisbetűs és nagybetűs szavakat!

Döntsd el, van-e a kisbetűs szavaknak nagybetűs megfelelője!

Ha van, kattints a "Van" felírra, ha nincs, kattints a "Nincs" felírra!

TŰNIK

DOKUMENTUM

múlik

Van

Nincs

lexikon

Van

Nincs

[◀ Előző](#)

[Következő ▶](#)

4. szakasz, 1. szint, (címszóolvasás):

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!



olvas

Igen

Nem

újság

Igen

Nem

füvel

Igen

Nem

fű

Igen

Nem

[◀ Előző](#)

[Következő ▶](#)

4. szakasz, 2. szint, (toldalékos szölvlasás):

Dönts 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!



látogatás Igen Nem

sürgősen Igen Nem

rosszkedvű Igen Nem

alábbi Igen Nem

[Előző](#)

[Következő](#)

4. szakasz, 3. szint, (szójelentés olvasás):

Olvasd el a nagybetűs szavakat!

Dönts el, hogy a kisbetűs jelentéseknek megfelelő szó van-e a nagybetűs szavak között!
Ha van, kattints a szókapcsolatok után lévő "Van" felíratra, ha nincs, kattints a "Nincs"
felíratra!

ÁLLANDÓ, BIZONYÍT, FÁJ, LEGFELJEBB, ONNAN

arról a helyről Van Nincs

fájdalom érződik Van Nincs

tartósan változatlanul maradó Van Nincs

nem több, mint... Van Nincs

valaminek az igaz voltát igazolja Van Nincs

[Előző](#)

[Következő](#)

4. szakasz, 4. szint, (szójelentés olvasás):

Olvasd el a nagybetűs szavakat!

Döntsd el, hogy a kisbetűs jelentéseknek megfelelő szó van-e a nagybetűs szavak között!
Ha van, kattints a szókapcsolatok után lévő "Van" felíratra, ha nincs, kattints a "Nincs" felíratra!

KIFEJEZ, LÉNYEGES, MILLIÁRD, PROFESSZOR, TÁMOGAT

valakinek a szülőföldjén, hazájában

Van

Nincs

egyetemi tanár, orvos

Van

Nincs

orvosi kórtüneteket megerősítő

Van

Nincs

gondolatot szavakba foglal

Van

Nincs

alapvetően fontos

Van

Nincs

[◀ Előző](#)

[Következő ▶](#)

4. szakasz, 5. szint, (szójelentés olvasás):

Olvasd el a nagybetűs szavakat!

Döntsd el, hogy a kisbetűs jelentéseknek megfelelő szó van-e a nagybetűs szavak között!
Ha van, kattints a szókapcsolatok után lévő "Van" felíratra, ha nincs, kattints a "Nincs" felíratra!

ÁTÉL, EGYÉB, KOTTA, RENDSZERVÁLTÁS, VÁROS

most már, íme

Van

Nincs

külön meg nem nevezett

Van

Nincs

200 méternél nem magasabb felszíni forma

Van

Nincs

legjobb esetben

Van

Nincs

belsőleg azonosul vele

Van

Nincs

[◀ Előző](#)

[Következő ▶](#)