

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

HÜLBER LÁSZLÓ

**ÁTTÉRÉS A TECHNOLOGIA ALAPÚ MÉRÉS-ÉRTÉKELÉSRE**  
PAPÍR ÉS SZÁMÍTÓGÉP ALAPÚ MATEMATIKA TESZTELÉS  
ÖSSZEHASONLÍTÓ VIZSGÁLATA 1-6. ÉVFOLYAMON

Ph.D-értekezés tézisei

Témavezető:  
Dr. Molnár Gyöngyvér  
egyetemi tanár



Szeged  
2015

## A disszertáció témája és szerkezete

Az ezredforduló óta a nemzetközi és nemzeti mérések iránti igény fokozódó hangsúllyal jelenik meg. A tényeken alapuló rendszerirányításnak és döntéshozatalnak, valamint a különböző intézkedések nyomon követésének kulcsfontosságú elemeit a felmérésekből származó mutatók, mérőszámok biztosítják. A mikroszintű tanórai folyamatoktól kezdve a makroszintű közoktatás egészének irányításáig az érdekelteknek a rendszer hatékony fejlesztéséhez rendszeres, megbízható minőségű, információgazdag visszajelentésekre van szükségük (*Csapó, Molnár, Pap-Szigeti és R. Tóth, 2009*).

Az infokommunikációs technológiák (IKT) az oktatás világában egyre meghatározóbb szerepet töltenek be, így természetes következményként tekinthetünk arra, hogy a mérés-értékelés területén is megkezdődött a technológia integrációja. Az elektronikus tesztelés bevezetésével nem csupán a teszt médiuma változik meg, hanem az infokommunikációs technológiák által olyan lehetőségek nyílnak meg, amelyek meghatározó reformokat indíthatnak el a mérés-értékelés területén (*Lent, 2009*).

A papír alapú tesztelés elérte korlátait, megtörtént a tesztmédium adta potenciál teljeskörű kihasználása (*Scheuermann és Björnsson, 2009*). A technológia számos területen olyan prosperációs lehetőséggel rendelkezik, amelyek egyértelműen kijelölik a mérés-értékelés jövőjét, a továbbfejlődés útját. Két fontos innovációs területet érdemes kiemelni: (1) a multimédiás, interaktív, dinamikus elemek alkalmazhatóságának, az információs és kommunikációs technológiák autentikus használhatóságának következtében mérhetővé válnak a 21. századi képességek (2) az automatizált javítási mechanizmusok, az eredmények azonnali biztosítása, az adaptív tesztalgoritmusok alkalmazása, a motiválóbb tesztkörnyezet és a tesztelésbe bevonhatók körének kibővülésével a kiértékelési folyamat hatékonyabb, testreszabható megvalósítása realizálható (*Csapó, Ainley, Bennett, Latour és Law, 2012*).

A technológia alapú tesztelés fejlődését illetően a négy szakaszt különítenek el a generációs elméletek (*Bunderson, Inouye és Olsen, 1989; Bennett, 1998, 2008; Redecker és Johannessen, 2013*). Az első két generációt a mérés-értékelési folyamat hatékonyságának növeléséhez rendelik a szakértők, viszont a 3. és 4. generációs tesztelés feladatának már a holisztikus, személyre szabott értékelés tanulási folyamatba való integrálását tartják. A különböző generációs elméletek egyetértenek abban, hogy a beágyazott értékelés (*embedded assessment*) korszakába való átmenet a technológia fejlődése mellett egy konceptuális paradigmaváltást igényel az értékelés területén. Ettől a fordulattól várhatók el a mérföldkövet jelentő reformváltozások, amelyek következtében a tanulás és az értékelés gyakorlata szorosan összefonódna (*Redecker és Johannessen, 2013*).

A határozott jövőképek ellenére a megvalósulás hosszú és kihívásokkal teli folyamat, melynek első lépése a papír alapú tesztelésről való áttérés. A mérési intézetek a számítógép alapú mérésre való migrációt nem tartják azonnal elérhető végcélnak, ezért átmeneti állapotokkal közelítenek a megvalósításhoz (*Lent, 2009*). A tradicionális tesztelésről a számítógép alapú tesztelésre való áttérés első lépéseként a papír alapú tesztek digitalizálását javasolják a szakértők (*Pommerich, 2004*). Az eredeti teszt formátumának minél több szempont szerint történő megőrzése mellett összehasonlítható a két tesztkörnyezet. Az ilyen vizsgálati elrendezés biztosítja az olyan komparatiztikus elemzések véghezvitelét, amellyel azonosítani

lehet a tesztmédiium hatását (*media effect*) a teszteredményekre vonatkozóan (*Clariana és Wallace, 2002*).

A disszertáció fő témáját tekintve a papírról számítógép alapú tesztelésre való áttérést állítja fókuszpontjába. A migráció vizsgálatához papír és számítógép alapú matematika tesztelés összehasonlító nagymintás elemzését mutatja be, 1-6. évfolyamos tanulókat vizsgálva.

A jelenlegi trendeket figyelembe véve a technológia alapú mérés-értékelés témában számtalan doktori kutatás fog indulni a közeljövőben. A dolgozat elméleti keretet, kiinduló pontot is kíván adni, általánosságban a technológiai alapú tesztelés témaköréhez.

Ezt a célt valósítja meg az I. fejezet, amely rögzíti a téma terminológiai rendszerét, bemutatja az elektronikus tesztelés típusainak különböző osztályozási módjait. A különböző infokommunikációs technológiákkal végzett mérés-értékelésnek számtalan megvalósulási formája létezik attól függően, hogy a tesztelés folyamatának melyik állapotában, szintjén alkalmazzák, illetve az érintettek közül kik és mire használják az elektronikus megoldásokat (*Molnár, 2010*). Az adott rész egyben segít elhelyezni a dolgozatban alkalmazott online tesztelést a különböző kategóriák viszonylatában. A technika mérésben való implementációjának előnyét számos tanulmány említi, különböző csoportosítási módokkal, eltérő hangsúlyokat alkalmazva, jellemzően a papír alapú médiummal való összehasonlítás mentén. A disszertáció ezen a dialektikán túllépve hierarchikus modellbe rendezi és strukturálja a lehetőségeket, és az azokból származó előnyöket, feltérképezve, hogy milyen kapcsolatban és kölcsönhatásban állnak egymással az egyes részelemek (1.2. fejezet). Az I. fejezet harmadik része bemutatja a technológia alapú tesztelés fejlődéstörténetét, az egymást követő tesztelési generációkat és jellemzőiket, valamint az elektronikus tesztelés jövőjére vonatkozó elképzeléseket, elvárásokat. Ez a rész történeti síkon pozicionálja a disszertáció kutatásához tartozó számítógép alapú tesztelést, megmutatja, hogy milyen előzmények és jövőbeni perspektívák fogalmazhatók meg a dolgozatban bemutatott konkrét mérési típusal kapcsolatban. A fejlődési tendenciák ismeretében nemzetközi és hazai viszonylatban is jellemzésre kerül, hogy az egyes országok és mérési intézetek hol tartanak a papírról számítógép alapú tesztelésre való áttérésben; milyen mértékű az elektronikus tesztelés elterjedtsége.

A II. fejezet elsőként a disszertáció kutatási problémáját járja körbe. Bemutatja, hogy a tesztmédiium megváltoztatása nem egyszerűen technológiai, logisztika kérdés, hanem számos megoldandó pszichometriai kihívást állít fel az elektronikus tesztelés megbízható alkalmazhatóságához. Célja feltérképezni a médiahatás minden vonatkozását a validitásra, az alkalmazott technológiára, a minta paramétereire, az objektivitásra, a megbízhatóságra vonatkozó kérdések felvetésével. Jellemzi a médiahatás természetét, komponensekre bontja, megmutatja ezek egymáshoz való viszonyát. Rögzíti a médiahatás megismerésének a dolgozatban alkalmazott kutatási stratégiáját, elrendezését. Azonos feltételek mellett, ugyanazon teszt papír és számítógép alapú eredményeit összehasonlítva ismerhetjük meg a közvetítő médium befolyásoló hatásának mértékét. Az ekvivalencia kérdésének tárgyalása következik, mikor igazolódik, milyen esetek állhatnak elő, azok milyen következményekkel bírnak. Felmerül annak a kérdése is, hogy lehet-e egyáltalán ekvivalenciáról beszélni. A problémák, kérdések felvetésére a következő rész (2.2. fejezet) igyekszik választ adni a korábbi médiahatás vizsgálatok eredményeinek bemutatásával, szintetizálásával. Először meghatározásra kerül a disszertáció szempontjából releváns összehasonlító vizsgálatok köre az

elmúlt 35 év tanulmányaiból. Ezután mutatom be, elemzem az eredményeket a technológia jellemzőire, teszt és itemparaméterekre és a minta tulajdonságaira, külön kiemelve, és részletezve a Magyarországra és a matematikára vonatkozó vizsgálatokat. A fejezet harmadik része a médiahatás vizsgálatok kutatási stratégiáit mutatja be, rendszerezi. Részletesen leírja a két fő domináns elrendezési módnak (független és összetartozó minták) a profilját, milyen altípusai vannak, mely eljárások alkalmazhatók, milyen megbízhatósággal bírnak, következtetések levonására alkalmasak. Az utolsó alfejezet (2.4. fejezet) a nemzetközi útmutatások bemutatásával foglalkozik. Rögzítésre kerül, hogy ezen iránymutatások követésével a médiahatás kezelhetősége jelentősen növelhető lenne. A helyzet megoldására *Waters* és *Pommerich* (2007) javasolja az irányelveknél nagyobb szabályozást, standardokat, amelyek biztosítanák, hogy minden elektronikus tesztelésben résztvevő számára érvényesek, megbízhatóak, objektívek legyenek a mérési feltételek.

A III. fejezet a tradicionális disszertáció szerkesztési stílusát követve az empirikus kutatás céljait, hipotéziseit, kutatási stratégiáját mutatja be. Négy átfogó, komplex célt fogalmaz meg, amelyet összesen 22 hipotézisre bontok, általános összehasonlító feltételezésekkel valamint az itemek paramétereire és a mintára vonatkozó tulajdonságok mentén. Részletezem a mérésben résztvevők háttéradatait, amelyeket az országos Központi Statisztikai Hivatal által biztosított adatokkal hasonlítom össze. Ismertetem az adatfelvétel körülményeit: felkérő levelet, regisztrációs eljárást, a mérésre vonatkozó előírásokat, mérési útmutatókat. Az utóbbi folyamat részletes bemutatást indokoltnak tartottam, mivel kevés információ áll rendelkezésre arra vonatkozóan, hogy milyen előkészületi folyamatok tartoznak egy elektronikus tesztelés lebonyolításához. Bemutatom és vizuálisan magyarázom a felhasználásra kerülő mintaillesztési, horgonyzási eljárásokat, a Rasch-modell alkalmazását. A fejezet egyik későbbi médiahatás vizsgálatok számára nagy relevanciával bíró része az a részletes, teljességre törekvő paraméterrendszer, amely a tesztmédiium befolyásoló erejének azonosítására alkalmas. A változórendszer a feladat megoldásának komplex folyamatának egyes lépéseihez rendeli a vizsgálható itemparaméterek körét: információ feldolgozás, feladatmegoldás (*Csikos* és *Csapó*, 2011) és feladatmegoldó tevékenység.

Az utolsó fejezet a hipotézisek mentén leíró jelleggel, fókuszáltan csak a felhasználható eredményeket mutatja be első lépésben. Ezt követi az eredmények részletes értelmezése, reflexió a II. fejezetben bemutatott kutatásokkal, párhuzamok és ellentétek vonása azok mentén. Gazdag illusztrációs anyag igyekszik az adott rész következtetéseit értelmezni. A következő fejezet jelentős gyakorlati relevanciával bír, mivel az elmélet szintetizációja, a vizsgálat során tapasztaltak, és kutatás eredményei alapján fogalmaz meg olyan online matematika tesztelésre vonatkozó javaslatokat, amelyek egyrészt további vizsgálatok tárgyát képezhetik, másrészt hozzájárulhatnak az elektronikus tesztelés jóságmutatóinak javításához.

### **Az empirikus vizsgálat problémaháttere**

A számtalan innovációs lehetőség mellett a mérés-értékelés technológia alapúvá tétele különböző pszichometriai kérdéseket vet fel, problémák megoldását igényli. A teszt kiközvetítésének, a válaszok rögzítésének és javításának gépesítésével felmerül a kérdés, hogy a technológia alkalmazása milyen módosító hatásokat fejt ki a tesztelésre, a tesztek eredményeire vonatkozóan. Ugyanazt mérjük papír és számítógép alapon? Megváltozik-e a

teszt érvényessége? Minden tanuló egyformán reagál-e a tesztmédiium megváltozására? Vannak-e hátrányosan érintett tanulói körök?

A validitási kérdések attól függően változnak, hogy a mért konstruktumnak részét képezi-e a technológiai jártasság, eszközhasználati tudás. Ez a papír alapú tesztek digitalizálásakor azonban nem merülhet fel, mivel a tradicionális tesztmédiumnak nem képezheti részét a technológia használathoz köthető konstruktum. Ennek következtében a disszertáció kutatása során a kérdés, hogy módosul-e, kiegészül-e a teszt érvényessége olyan technológiához köthető tényezőkkel, amelyek nem képezik a tudásmérő tesztek részét.

Az alkalmazott technológiára vonatkozó kérdések eredendője, hogy az iskolákban elérhető számítógépek változatos szoftver és hardver elemekből épülnek fel. Az itt megjelenő kérdés, hogy a különböző konfigurációkon kiközvetített megoldások ekvivalensnek tekinthetők-e, szenved-e hátrányt a kitöltők egy csoportja csak amiatt, mert eltérő a használt infrastruktúra egy vagy több paramétere. Külön-külön a paraméterek vizsgálata nem elegendő, mivel ezek az elemek kölcsönhatásban vannak egymással, és együttesen befolyásolják a teszt kiközvetítését meghatározó változókat (pl.: megjelenítés, sebesség).

Amennyiben létezik a tesztelteknek olyan csoportja, amely a technológia bevezetése miatt szenved hátrányokat vagy élvez előnyöket, akkor a tesztelés méltányosságára (*fairness*) vonatkozó kérdések merülnek fel. A pártatlanság és érvényesség szempontjából sem az egyéni felmérést, sem a tanárokat, intézményeket, országokat illetően nem elképzelhető, hogy valaki hátrányba kerüljön a médium megváltoztatása miatt. Mind a szummatív, mind a diagnosztikus méréseket tekintve nem kívánatos állapot, hogy érvénytelen eredmények alapján szülessenek ítéletek. Az oktatási hatóságok, illetve jogi, erkölcsi szempontból sem volna tartható állapot (*Lent, 2009*).

### **Médiahatás vizsgálatok**

A médiahatás megállapítására vonatkozó papír és számítógép alapú összehasonlító vizsgálatok jelentik az első lépést a tesztelés új formájának megismeréséhez. Amennyiben bizonyítható, hogy a két médiumon elért eredmények statisztikailag ekvivalensek vagy ekvivalensé tehetőek, akkor bizonyítható az elektronikus tesztelés alkalmazhatósága (*Schroeders, 2009*). Az *International Test Commission* (Nemzetközi Teszt Szervezet), *American Educational Research Association* (AERA), az *American Psychological Association* (APA) és a *National Council on Measurement in Education* (NCME) a tesztek ekvivalenciájára vonatkozó irányelveiknél rögzítik, hogy egy teszt számítógép és papír alapú változata közötti ekvivalenciát dokumentált bizonyítékokkal kell alátámasztani (*ITC, 2006; AERA, APA és NCME, 1999, 2014*). A fenti dokumentumok és a témában készült tanulmányok alapján, két különböző médiumon felvett azonos tesztek ekvivalenciájához biztosítani kell:

- a két tesztkörnyezetből származó pontok felcserélhetőségét (*AERA, APA és NCME, 1999, 2014*);
- a számítógép és papír alapú teszttel végzett mérések megbízhatóságát, és ezen megbízhatósági mutatók összehasonlíthatóságát (*ITC, 2006*);
- a számítógép és papír alapú tesztelés validitásának egyezőségét (*Hargreaves, Shorrocks-Taylor, Swinnerton, Tait és Threlfall, 2004*);

- itemszintű analízissel igazolt ekvivalenciát a feladatokra vonatkozóan (*Puhan, Boughton és Kim, 2007*);
- a minta releváns háttérváltozóira kiterjedő, egyéni különbségeket azonosító összehasonlításokkal biztosított egyezőséget (*Wolfe és Manalo, 2005*);
- az alkalmazott technológiai paraméterek befolyásoló hatásától való függetlenséget (*CTB/McGraw-Hill, 2003*).

Amennyiben a fenti pontok alapján megállapítható az ekvivalencia, akkor a két teszt párhuzamosnak tekinthető és igazolt, hogy egyik mód sem befolyásolja a konstruktum irreleváns varianciáját. Amennyiben nem bizonyítható az egyezőség, akkor nem azt a konstruktumot méri, vagy eltér a konstruktum-irreleváns variancia, illetve mindkettő egyidejűleg is teljesülhet (*Csapó és mtsai, 2012*). Amennyiben fenn áll, hogy az új tesztmédiám befolyásolja az eredményeket, akkor meg kell határozni (*R. Tóth, 2009*):

- a két tesztkörnyezetben elért eredmények közötti különbség mértékét;
- az eltérés irányát;
- az előnyt élvező vagy hátrányt szenvedő teszteltek körét;
- azokat a változókat, amelyek az eltérésért felelősek.

A médiahatás kutatása egyidős a technológia alapú tesztelés történetével, mégsem fogalmazható meg egy egységes, minden infrastruktúrára, területre, kontextusra, itemformátumra és mintára kiterjedő konklúzió a szakirodalom alapján. Az eredmények ellentmondásainak hátterében az áll, hogy a különböző vizsgálatok más-más konstruktumot, különböző mintán, különféle mintakialakítással és itemformátum segítségével mértek és eltérő hardver/szoftver infrastruktúrán, diverz elemzési technikákkal analizáltak (*Wang és Shin, 2009*). A vizsgálatok eredményei sokszor megkérdőjelezhetőek hiányos dokumentáció miatt, avagy azért, mert bizonyos változóknak nem tulajdonítottak jelentőséget és nem kontrollálták azokat. A 90'-es évek előtti kutatások elavultnak tekinthetők, mert jelentősek az eltérések az akkor és a ma használt számítógépes architektúrák között, vagy például a mostani tesztmegoldók átlagosan nagyobb eszközhasználati tapasztalattal rendelkeznek. A nagyszámú változórendszer hatása és eltérő eredménye miatt az elektronikus tesztelésre történő felelősségteljes átállás érdekében *Pommerich (2004)* azt javasolja, hogy a számítógép alapú tesztelésre való áttérés előtt minden mérési szervezet végezzen összehasonlító tanulmányokat a saját tesztjeivel, technológiai hátterével és mintájával, mert a tanulmányok eredményei nem teljeskörűen általánosíthatóak. A médiahatás vizsgálatok természetéhez továbbá hozzátartozik, hogy állandó jelleggel szükségesek, nem ér véget a kutatások köre, mivel folyamatosan fejlődik a technológia (lásd érintőképernyős eszközök), új tesztelési formák, itemek jelennek meg és változnak a tesztelésben résztvevők jellemzői is.

### **A kutatási célok és hipotézisek**

A disszertációban megjelenő probléma nemcsak elméleti kihívást ad, hanem konkrét gyakorlati relevancia is társítható a célkitűzésekhez. A 20 éves mérési hagyományokkal rendelkező szegedi műhely (MTA-SZTE Képességkutató Csoport, SZTE Oktatásméleti Kutatócsoport, SZTE Neveléstudományi Intézet és SZTE Neveléstudományi Doktori Iskola)

célja, hogy a számtalan tárgykörben végzett kutatásainál az adatfelvételi módszereit a nemzetközi trendeket követve technológia alapúvá tegye. Ahhoz, hogy felelősségteljesen vihesse véghez a migrációt, szükségesek az olyan papír és számítógép alapú tesztelést összehasonlító vizsgálatok, mint ami a jelen disszertáció tárgyát képezi. Ezen konkrét vonatkozáson túl további, magyarországi relevanciát biztosít, hogy más mérési programok például az Országos kompetenciamérések átültetéséhez, jövőbeni számítógép alapú tesztelésekhez is hasznos információkat biztosíthat a dolgozat.

A disszertáció matematika területén, 1-6. évfolyamon hasonlítja össze a papír és számítógép alapon felvett eredményeket. A matematika területére azért esett a választás, mert az alaptantárgyak közé tartozik, meghatározza a tanulók előmenetelét egészen az általános iskolától kezdve a középiskola befejeztéig (*Bennett, Braswell, Oranje, Sandene, Kaplan és Yan, 2008*). Az elsőtől hatodik osztályig tartó intervallum azt az életkori spektrumot fedi le, amikor a diákok számítógépes gyakorlottságában, jártasságában nagy különbségek lehetnek, így jól vizsgálható az életkor és az IKT tapasztalat szerepe. A kutatás céljai általános megfogalmazásúak, több vagy az elsőt tekintve az összes hipotézist átívelik.

A kutatás céljai:

- 1-6. évfolyamos diákok online tesztkörnyezetben való viselkedésének feltérképezése,
- papír és számítógép alapú környezetben nyújtott teljesítmények összehasonlítása,
- azon item paraméterek meghatározása, amelyeknél jellemző a teljesítmények valamilyen irányú eltérése, vagy azonossága,
- olyan részminták azonosítása, amelyek tipikusan azonosan, illetve eltérően viselkednek különböző tesztkörnyezetben.

A kutatás hipotézisei általános irányultságú, itemekre és mintára vonatkozó paraméterek mentén fogalmazznak meg feltételezéseket.

Általános irányú hipotézisek:

- Megbízhatóságra vonatkozóan:
  - H1: Az online matematika tesztek belső konzisztenciája megfelelő ( $\text{Cronbach-}\alpha \geq 0,8$ ).
- Papír és számítógép alapú teszteredmények átlagos különbségeire vonatkozóan:
  - H2a: Az átlagos eltérések szintjén nincsenek szignifikáns különbségek a két médiumon elért eredmények között.
  - H2b: A két médium közötti eltérések a minta és az itemek paramétereinek függvényében meghatározottak.

Az itemek tulajdonságaira vonatkozó hipotézisek:

- A feladatok típusával kapcsolatban:
  - H3a: Zárt végű itemtípusoknál a diákok számítógépen teljesítenek jobban.
  - H3b: Nyílt végű szövegalkotó feladatoknál a papír alapú módszer bizonyul könnyebbnek a tanulók számára.

- A feladatok belső tartalmi jellemzőivel és a megoldáshoz szükséges tudáselemekkel kapcsolatban:
  - H4a: A különbségek mértékét befolyásolhatja a feladatok tartalmi dimenziója.
  - H4b: A magasabb szintű gondolkodási műveleteket igénylő feladatok papír alapon jobban teljesítenek.
- A feladatokhoz tartozó információkkal kapcsolatban:
  - H5a: A feladatoknál megjelenő szöveges információ mennyisége (karaktárszám) együtt jár a médiahatás nagyságával.
  - H5b: A grafikus elemeket használó itemek számítógép alapon eredményesebbek.
  - H5c: Táblázatok alkalmazása nem befolyásolja a médiahatást.
  - H5d: Az információk elrendezési módja nincs hatással a két tesztkörnyezet közötti különbségekre.
- A feladat által igényelt válaszok jellemzőivel kapcsolatban:
  - H6a: A feladat által igényelt válasz hossza korrelál a számítógépen való eredményességgel.
  - H6b: A billentyűzet kombinációt igénylő feladatoknál médiahatás tapasztalható

Mintára vonatkozó hipotézisek:

- Az életkor szerepével kapcsolatban:
  - H7a: Az életkor előrehaladtával a papír és számítógép alapú teszteken elért eredmények közötti különbségek csökkennek.
  - H7b: Az egyes évfolyamok között szignifikáns eltérések vannak a különbségek nagyságában.
- A diákok nemével kapcsolatban:
  - H8: A diákok neme befolyásolja a két médiumon elért teljesítmények közti különbségek nagyságát.
- Diákok lakhelyének földrajzi paramétereivel kapcsolatban:
  - H9: A tanulók lakhelyének elhelyezkedése (régió) nem módosítja az eredményességet az egyes médiumok tekintetében
- A tanulók szüleinek iskolai végzettségével kapcsolatban:
  - H10a: Az anya iskolai végzettsége médiahatást befolyásoló háttérváltozó.
  - H10b: Az apa iskolai végzettsége médiahatást befolyásoló háttérváltozó.
- A tanulók tanulmányi előmenetelével kapcsolatban:
  - H11a: A tanulók matematika eredményei (félévi jegy) nem befolyásolják az egyes teszt környezetben elért eredmények közti különbséget.
  - H11b: A tanulók tanulmányi eredményeinek (félévi jegyek) átlaga nem befolyásolja az egyes teszt környezetben elért eredmények közti különbséget.



- A tanulók informatikai képzésével kapcsolatban:
  - o H12: Azok a tanulók, akiknek lehetősége van informatika (számítástechnika) órára, szakkörre járni, eredményesebbek társaiknál az online matematika teszten.

## **A vizsgálat kutatási stratégiája**

### *A mérésben résztvevők és az adatfelvétel*

A kutatás szerkezete szerint a független minták típusán belül azon alkategóriába tartozik, amikor egy korábbi papír alapú mérésnek a tesztjei, illetve eredményei kerülnek felhasználásra a számítógépen felvett adatok összehasonlító vizsgálataihoz. Papír alapon reprezentatív minta állt rendelkezésre, elsőtől hatodik osztályig, minden évfolyamon legalább 5700 fő, összesen 40571 tanuló oldotta meg a 45 percre optimalizált tradicionális tesztek 2010-ben. A számítógép alapú adatfelvétel 2012-ben zajlott, amelyben minden évfolyamon legalább 3000 fő összesen 22715 tanuló vett részt, eDia tesztplatform alkalmazásával szintén egy tanórát igénybe véve. Az iskolák mindkét esetben önkéntesen vállalták az együttműködést. A papír és számítógép alapon dolgozók is kitöltöttek egy kérdőívet, amellyel a mintaillesztéshez, és az elemzésekhez szükséges háttéradatokat vettük fel. Kvalitatív adatgyűjtést végeztünk az online mérések lebonyolítása után a mérésben résztvevő felügyelő tanárokkal, rendszergazdákkal. A nyíltvégű kérdéseket tartalmazó kérdőív a mérés során tapasztalt információk begyűjtését célozta meg.

### *Mérőeszközök*

Papír alapon évfolyamonként átlagosan 30 tesztváltozat került bemérésre. Minden tesztváltozat három klaszterből (részesztből) épült fel. Egy klaszter 3-4 darab 4-5 ítemes matematika feladatot, azaz összesen körülbelül 15 ítemet tartalmazott. A feladatok az SZTE Oktatásméleti Kutatócsoport lektorált feladatbankjából kerültek kiválogatásra.

Számítógép alapon, a papíron bemért ítemek kerültek felhasználásra, de eltérő tesztváltozatok megalkotása volt szükséges, mivel nem minden feladatot lehetett digitalizálni. A feladatok kiválasztásánál, az azokból álló tesztek megszerkesztéséhez olyan szempontrendszer került meghatározásra, amely az összehasonlítás teljességét, az eredmények általánosíthatóságát, a médiahatás természetének mélyebb megértését hivatott szolgálni. Ennek megfelelően a korábbi kutatási eredményeket felhasználva kiegyenlített számban szerepeltettem a médiahatás szempontjából figyelmet érdemlő paramétereknek megfelelő feladatokat. Ezen feltételek közül kiemelhető, hogy a nyílt és a zártvégű feladatok ugyanakkora arányban szerepeltek a gépesített mérés során. A számítógép alapú tesztek megalkotásakor kiemelt irányelvet jelentett, hogy az ítemek minden jellemzőjüket tekintve a papír alapúval azonos módon legyenek reprezentálva.

### *Eljárások*

Miután a két adatfelvétel mintáját nem lehetett azonosnak tekinteni, így a független minta elrendezésnél jellemző mintaillesztési eljárást végeztem az eredmények megbízhatóságának növelése érdekében. Mintaillesztésnél a következő prioritási sorrendet követve évfolyam, nem,

régió, településtípus, anya iskolai végzettsége, apa iskolai végzettsége, tanulmányi átlag (egészre kerekítve) vettem figyelembe a paramétereket.

Különböző tanulók kétféle tesztkörnyezetben, eltérő tesztváltozatokat oldottak meg, ezért a teljesítmények összehasonlítását, közös képességskálára való konvertálását, az adatok skálázását a valószínűségi tesztmodellek közé sorolható Rasch-modellben használt logisztikus függvény biztosította. Az alkalmazás feltételét jelentette, hogy az egyes tesztváltozatok össze legyenek kapcsolva, ezért a felhasznált klaszterek három tesztváltozatban szerepeltek, a tesztek elején, közepén és végén. Ezzel az összekapcsolási technikával egyrészt elérhető, hogy minden tesztváltozat kapcsolódjon egymáshoz – minden klaszter egyben horgonyklaszter is – másrészt kiküszöbölhető a feladatok tesztben elfoglalt helyének torzító hatása. A Rasch-modellnek a kétdimenziós változatát használva az egyik dimenziót a számítógép, másik dimenziót a papír alapú tesztkörnyezethez rendeltük, ezáltal függetlenül a két teszt médium eredményeit egymástól, ugyanakkor elérve, hogy azok végeredményben egy skálán ábrázolhatóak legyenek (Hülber és Molnár, 2013). Az ACER ConQuest program (Wu, Adams és Wilson, 1998) használatával meghatározott itemnehézségi értékek (*estimate value*) jelentették az elemzés kiinduló pontját, amelyek klasszikus tesztelméleti módszerekkel kerültek további elemzésre.

Az itemek viselkedésének leírására saját univerzálisan használható szempontrendszert alkottam, melyben a feladat megoldás komplex folyamatának egyes lépéseire rendelttem a vizsgálható itemparaméterek körét. A modell a következő elemekből épül fel (1) a feladatokhoz tartozó információk feldolgozása (információ típusa, mennyisége, elrendezése), (2) a feladatmegoldás során szerepet játszó pszichikus, tartalmi, műveleti, kontextuális jellemzők (Csikos és Csapó, 2011; Vidákovich, 2012), illetve (3) a feladatmegoldó tevékenységhez kötődő itemparaméterek a feladat típusa, a rögzíteni kívánt információ mennyisége, típusa és az eszközhasználati módja.

### **A kutatás eredményei és azok értelmezése**

(H1) Az online tesztek belső konzisztenciája minden évfolyamon, minden tesztváltozat esetében megfelelő volt, a Cronbach- $\alpha$  mutatók 0,8-nál nagyobb értéket vettek fel. A mutatókból képzett átlagok 0,9 körüli értékek, a Rasch-modell által meghatározott személyszeparációs (EAP/PV) megbízhatósági mutatók is megbízhatónak bizonyultak, minden évfolyamra nézve nagyobb értéket vettek fel, mint 0,85.

A mutatók között évfolyamonként nem figyelhetők meg eltérések, mintázatok, a legkisebbek esetén is kiválóak az értékek. Ez alpillérét adja annak a kijelentésnek, hogy az online tesztelés megbízhatóan alkalmazható az iskolába lépéstől kezdődően már 6-7 éves kortól is.

(H2) Az átlagok szintjén nem jelentkezett szignifikáns eltérés az itemek nehézségi értékei között, azonban önmagukban ezek az eredmények nem elegendőek egy végleges ítélet megalkotásához a papír és számítógép alapú tesztelés összehasonlításra vonatkozóan (Wilhelm és Schröeders, 2008). Az itemeknek és a mintának is léteznek olyan csoportjai, amelyeknél szignifikáns eltérések voltak tapasztalhatók.

(H3a) Az első három évfolyam esetén a zártvégű feladatokon belül az alternatív választás típusú feladatoknál jelentkezett szignifikáns eltérés a számítógép alapú médium javára ( $|t_1|=4,16$ ,  $p<0,01$ ;  $|t_2|=2,67$ ,  $p<0,05$ ;  $|t_3|=4,16$ ;  $p<0,01$ ). Az eredmények háttérében az áll, hogy számítógépen átlagosan 15%-kal több válasz született elektronikus tesztkörnyezetben. Ez párhuzamba vonható *Mazzeo és Harvey (1988)*, *Csapó és mtsai (2009)*, *Johnson és Green (2006)* kutatásainak eredményeivel, akik sokkal gyakoribbnak találták a válaszadási próbálkozások arányát számítógépen. Magyarázatot jelenthet, hogy a diákok kevésbé tartják fenyegetett környezetnek a számítógép alapú tesztelést, ezért bátrabban mernek próbálkozni, több válasz születik, ami összességében szignifikánsan több pontszámot eredményez. Ennek a tesztkörnyezettől függő stratégiának a következménye az a papír alapú tesztekénél is megjelenő irányelv, miszerint kívánatos állapot a változatos feladattípusok alkalmazása.

(H4) A feladatok tartalmi elemeit, valamint a megoldáshoz szükséges pszichikus műveleteket tekintve a statisztikai, kombinatorikai tartalmú, magasabb szintű műveletet igénylő feladatok nehezebbnek bizonyultak számítógép alapon, mint papíron. A többi altípusnál nem mutatkozott eltérés az egyes médiumok tekintetében. *Johnson és Green (2006)*, *Hülber (2012)* és *R. Tóth és Hódi (2011)* magyarázatait követve ezek a típusú feladatok igénylik leginkább a jegyzetpapírok használatát, a próbálkozások felvázolását egyfajta kézírással kifejezhető vizuális gondolkodást, ami számítógépen nem megvalósítható. A problémát feltehetőleg az jelenti, hogy a diákoknak nincs elég tapasztalata a számítógépes tesztelésben, nincsenek kialakult megoldási stratégiáik, rutinjuk. Ennek következtében több műveletet igyekeznek fejben elvégezni, nem szeretnék összeegyeztetni a számítógép képernyőjét a jegyzetpapírral, adatokat másolni, megosztani a figyelmüket. A feltételezést erősíti, hogy a felügyelő tanárok szerint nagyon ritka volt a jegyzetpapírok használata.

(H5) A feladatoknál megjelenő szöveges információ mennyisége (karakterszám) ugyanolyan módon befolyásolta az eredményeket számítógép és papír alapon. Az információk elrendezési módja, illetve a táblázatok alkalmazása nem befolyásolta a médiumok közötti különbséget. Az információk típusát tekintve a grafikus elemeket alkalmazó feladatoknál jelentkezett szignifikáns különbség első évfolyamon az elektronikus médium javára ( $|t|=2,01$ ,  $p<0,05$ ). *Richardson, Baird, Ridgway, Ripley, Shorrocks-Taylor és Swan (2002)*, illetve *Hülber (2012)* tanulmányai, ezzel azonos eredményeket állapítottak meg. A háttérben lévő okokra a szerzők számítógépen való szebb megjeleníthetőséget, élethűbb ábrázolást fogalmazznak meg, ami feltehetőleg nagyobb figyelmet, motivációt vált ki (*Gyarmathy, 2012*), több időt szánnak a diákok a feldolgozásukra, így jobban értelmezve oldják meg az ilyen feladatokat. Feltehetőleg papír alapon nem minden esetben nyomtatták színesben a tesztek, míg a számítógépes mérés esetében minden monitor színes volt.

(H3b, H6) Az átlagok szintjén nem mutatkozott különbség a nyíltvégű feladatokon való teljesítés között számítógép és papír alapú médium között. A feladatokra adandó válasz hossza is azonosan korrelált a feladatokhoz tartozó itemnehezséggel mindkét tesztkörnyezetben. Ez alapján feltételezhető volna, hogy a nyíltvégű feladatok alkalmazása ugyanakkora kihívást jelentett mindkét médiumon, azt azonban nem tudjuk, hogy a válaszok begépelése és leírása ugyanannyi időt vett-e igénybe. Feltételezhető a felügyelő tanároktól való visszajelzések alapján, hogy abban az esetben, amikor a tanulók tudják a választ, megküzdnek a helyes válasz begépelésével, így item szinten nem jelentkezik különbség, de az idővesztés miatt a teszt

szintjén hátrányt szenvedhetnek. A harmadik osztálytól megjelenő billentyűkombinációkat igénylő feladatoknál szignifikáns hátrányt szenvedtek a számítógépen dolgozók.

(H7) A különbségek előrejelzése szempontjából az egyik legmeghatározóbb változót a tanulók életkora jelenti. Amíg az első három évfolyamon 0,7 körüliek a korrelációs értékek a két médiumon mutatkozó itemnehézségi értékek között ( $r_1=0,7$ ;  $r_2=0,72$ ;  $r_3=0,69$ ), addig negyedik évfolyamtól egy növekvő együtt járása figyelhető meg az eredményeknek ( $r_4=0,77$ ;  $r_5=0,85$ ). Ez hatodik osztályra eléri 0,92-es értéket, ami erős kapcsolatot feltételez a két médium között. Az első-negyedik, negyedik-ötödik, ötödik-hatodik évfolyam korrelációs együtthatói között a z-próbák szignifikáns különbséget állapítottak meg. Az életkor befolyásoló erejét, avagy a legfiatalabbaknál lévő jelentős különbségeket több tanulmány igazolja (pl.: *Choi és Tinkler, 2002; Ito és Sykes, 2004; Applegate, 1993; Barnes, 2010*). Ezek a tanulmányok egyöntetűen a technológiai jártasság hiányának tudják be, hogy a legfiatalabbak rosszabbul teljesítenek elektronikus alapon. 6-7 éves korban gyakori, hogy a tanulók nem használtak még egyáltalán számítógépet, avagy nem egyértelmű számukra az egér használata, hogy a szövegmezőbe bele kell kattintani, hogy lehessen írni, mi az enter, space szerepe, hogyan lehet eljutni a következő tesztkérdésre, avagy milyen módon kell egyes „speciális” karaktereket bevinni. Gyakorlatlanságukat jól igazolja, hogy a tesztben való mozgás az elsős, másodikosokat tekintve ritka, nem térnek vissza korábbi feladatokra, nem ellenőrzik, módosítják, gondolják végig újra a megoldásaikat, hanem egyszer végig mennek rajta, ami feltehetőleg a papír alapú stratégiáikkal ellentétes és kevésbé eredményes. Évfolyamonként 10-15%-kal nő a tesztben való mozgás gyakorisága, ami valószínűleg egyre jobban közelít a papír alapú megoldási mechanizmusokhoz. A gyakorlatlanságra, a technológiai járatlanságra vonatkozó feltételezéseket megerősítik a tesztek felügyelők visszajelzései is.

(H8) Nemek tekintetében csak első évfolyamon mutatkozott különbség: a lányok számítógépen szignifikánsan rosszabbul teljesítettek, mint papír alapon ( $|t|=2,11$ ,  $p<0,05$ ). Azonos eredményekre jutott *Horne (2007), Johnson és Green (2006), Halldórsson, McKelvie és Björnsson (2009), Csapó és mtsai (2009)*. Magyarázatként a következők fogalmazhatók meg: a fiúk azért teljesítenek jobban, mert jobban motiváltabbak számítógépes környezetben, mert nagyobb a tapasztalatuk és ebből következőleg a jártasságuk is, magabiztosabbak. Feltételezhető tovább, hogy a fiúk fontosabbnak találják a számítógépen való jól teljesítést, mint papíron, a lányokkal ellentétben, akik fordítva viselkednek.

(H9) Matematikailag jelentős eltérések csak az első három évfolyamnál jelentkeztek a tanulók lakhelyének elhelyezkedésére (régió) vonatkozóan, de nem megfigyelhető szisztematikus elrendeződése ezen különbségeknek. Ezek alapján nem tartom a különbségek szempontjából prediktív változónak a földrajzi elhelyezkedést. Papír alapú eredmények alapján tudjuk, hogy a régiókhoz, megyékhez tartozó teljesítménybeli különbségek általában valamilyen más változó hatásához tartoznak, ami megerősíti az előző állítást.

(H10) Az első három évfolyamon jelentkezett különbség a szülők iskolai végzettségének teljesítményt befolyásoló erejével kapcsolatban. A felsőfokú végzettséggel rendelkező szülők gyermekei számítógépen szignifikánsan rosszabbul teljesítettek. *Bennett és mtsai (2008)* valamint *Hódi és R. Tóth (2009)* tanulmányaiban találunk hasonló eredményeket. Indokként megfogalmazható, hogy a jobb teljesítményű tanulók érzékenyebben reagálnak a tesztmédium megváltozására. Nagyobb a megfelelési kényszerük, a jó teljesítés iránti

motiváció, így nagyobb zavart kelthet, hogy nem alkalmazhatják a megszokott, bevált megoldási stratégiáikat, mechanizmusaikat.

(H11) A tanulmányi előmenetel szempontjából a változásokra érzékeny korcsoportot nem lehetett vizsgálni, mivel a tanulók csak második évvégén kapnak először osztályzatot. A következtetések levonását szintén nehezítette, hogy a rosszabb osztályzatok tekintetében kevés mintaelemszám állt rendelkezésre. Harmadiktól-hatodik évfolyamon hármas vagy jobb (kerekített) félévi átlageredménnyel, illetve kettes vagy jobb matematika jeggyel rendelkező tanulók esetében nem jelentkezett szignifikáns eltérés, egy esetet kivéve (4. évfolyam, hármas félévi átlag). A felsorolt hiányosságok miatt a hipotézisek teljeskörű vizsgálatát nem lehetett elvégezni, illetve a kapott eredmények nem utalnak arra, hogy a médiumok közötti különbséget befolyásolná a tanulmányi előmenetel a vizsgálható harmadik osztálytól.

(H12) A tanulóknak alsó tagozaton az adatfelvétel időpontjában nem szerepelt kötelező tantárgyként az informatika. Az első négy évfolyamon nagy mintaelemszám különbségek tapasztalhatók a tekintetben, hogy a diákok tanulnak vagy sem informatikát. A kiegyenlített elemszámmal bíró 3-4. évfolyamon nem volt szignifikáns hatással az online eredményekre az, hogy tanulnak vagy sem informatikát.

### **Javaslatok, további kutatási lehetőségek**

A javaslatok alapját a szakirodalom szintézise, a gyakorlati megvalósítás tapasztalatai, eredményei és a mérésben résztvevő felügyelők, rendszergazdák, mérési koordinátorok visszajelzései jelentik. Ezek a javaslatok a dolgozat gyakorlati relevanciáját növelik, amelyek a számítógép alapú tesztelés megbízhatóbb, érvényesebb és objektívebb megvalósítását szolgálhatják függetlenül a használt tesztelési platformtól, mérési területtől. Az ajánlásaimat egy-egy probléma köré rendezve rögzítettem.

- A tanulók technológiai jártasságához kapcsolódó problémák kezelésére:
  - Számítógép használatát bemutató, gyakoroltató tutorialok (oktató programok) elvégzése.
  - Példa-, próbafeladatok elvégzése.
- Médiahatás előrejelzésére vonatkozóan:
  - Kontextuális adatok gyűjtése és elemzése. Például mennyi idő alatt oldanak meg egy egér mozgatásához, kattintásához kapcsolódó feladatot, szövegbevitelt. Tutorial, példafeladatok, és az egész teszt alkalmas lehet ilyen információk begyűjtésére.
- Technológiához, infrastruktúrához kapcsolódó problémák kezelésére:
  - Hardver, szoftver (rendszer) diagnosztikai eszközök használata, amelyek a tesztelés során a háttérben futva felméri, és rögzíti a tesztelésre használt számítógép jellemzőit.
  - Számítógépek azonosítását lehetővé tevő adatbázis felépítése, a diákok azonosítójához a teszteredmények mellett a gépek azonosítója is tárolásra kerülne (ez által biztosítva a technológiai paraméterek kontrollálását).
  - Infrastruktúra hitelesítési, minősítési folyamata külső szakemberek által a nemzetközi irányelvek, útmutatók alapján.

- Rajzolás igénylő feladattípusok számítógépes megvalósítására:
  - Feladatok átalakítása, zártvégűvé tétele egyes variációk felvázolásával (validitás módosító lehet)
  - Digitális ceruza használata.
- A diákok online tesztelésére való áttérésének segítéséhez:
  - Látványos, ergonomikus ütemező eszközök használata, például: feladatrács.
  - Tesztben töltött idő megjelenítése.
- Speciális karakterek beillesztésére, karakterformázások elvégzésére
  - Tutorial alkalmazása.
  - Dedikált gombok, ikonok elhelyezése a tesztelési felületen.
- Teszt felületének ergonómiájára vonatkozóan:
  - Tesztfelület gyermekbarát kialakítása.
  - Eredmények visszajelzése a tanulók számára is értelmezhető, játékos formában.
- A tesztelés megszakadására vonatkozóan:
  - Újratöltés, újraindítás utáni azonnali lehetőség biztosítása a teszt folytathatóságára.
  - Olyan akcióterv kidolgozása, amely minden normáltól eltérő eshetőségre gondolva megadja a követendő cselekvési sort.
- A tesztelést megelőző operatív műveletek elősegítésére:
  - Gazdag illusztrációkkal színesített brosúra biztosítása.
- Tesztszerkesztésre vonatkozóan:
  - Feladatsablonok alkalmazása (sorköz, sorhossz, bekezdések közötti térköz, betűméret, betűtípus, betűszín szövegdobozok, képek elhelyezése, mérete stb.).
  - Ellenőrző mechanizmus, az ettől eltérő kialakítás esetén figyelmeztető előrejelzés biztosítása a médiahatásra vonatkozóan.
- Számítógép alapú tesztelés gyakorlatlanságából eredő negatív hatásaira:
  - Technológia alapú tesztelés népszerűsítése mikroszinten, osztálytermi gyakorlatban is.
  - Együttműködés biztosítása a tanárokkal, továbbképzések tartása, önálló használatra való tesztelési platform biztosítása.

A javaslatok többsége egyben újabb kutatási lehetőségeket is tartalmaz, amivel igazolhatók vagy elvethetők az ajánlott eljárások a tesztelés jóságmutatóira vonatkozó hatása. További kutatási irányt jelöl ki a dolgozat az itemekre vonatkozóan a nyílt végű feladatokkal (pontosabban azok megoldási idejével) kapcsolatosan, a mintára értelmezve a diákok számítógépes tapasztalata, tanulmányi előmenetele mentén, illetve a technológiai paraméterek feldolgozása érdemel külön feldolgozást.

## A disszertáció jelentősége, korlátai

A disszertáció az első országos szintű és nagy elemszámú online tesztelés tapasztalatairól számol be. A kutatás egyedinek és hiánypótlónak mondható, mivel technológia alapon a számítógépes gyakorlottság szempontjából heterogén tudással bíró elemi iskolásokat is magas elemszámmal vizsgálta, a tanulmányi előmenetel szempontjából meghatározó matematika műveltség területén. Az elemzett itemek nem korlátozódtak egy szűk körre (például típus szerint: feleletválasztós itemek), hanem közel lefedték az itemeket jellemző több dimenziós teljes paraméterrendszert. A dolgozat modern eljárásokat használt: teszthorgonyzás, mintaillesztés; a matematika elemzések alapját valószínűségi tesztelméleti számítások jelentik. A disszertáció elméleti téren is kiinduló pontot szeretne nyújtani a számítógép alapú teszteléshez, a terminológia, osztályozási módok, lehetőségek és előnyök, generációs elméletek bemutatásával és szintetizálásával. Hasonlóan a médiahatás vizsgálatoknál a teljességre törekszik, számtalan kutatás, tanulmány többszempontú jellemzésével, elemzésével, kiemelve a matematika területét, a magyarországi tapasztalatokat. Tisztázva azok módszertani kérdéseit, típusait, jellemzőit. Az itemek összehasonlítására olyan összetett modell kerül bemutatásra, amely más kontextussal rendelkező komparatiztikus vizsgálatoknál is kiinduló pontot jelenthet. A dolgozat gyakorlati relevanciáját a területtől, platformtól függetlenül alkalmazható javaslatok tovább növelik.

Az összehasonlító vizsgálat matematika területén történt, így első sorban erre a műveltségterületre általánosíthatóak az eredmények. Más területek eltérő változórendszert eredményezhetnek, nem csak a tartalmi részhez kapcsolódóan. Például olvasás-szövegértés területén a nagyobb információ mennyiség feldolgozása, avagy a hosszabb válaszokat igénylő feladatok megoldása, eltérő karakterisztikájú médiahatást eredményezhet (R. Tóth és Hódi, 2011). Az eredmények általánosíthatóságának a korlátját jelenti, hogy a tesztelésre használt számítógépek paramétereinek megfigyelésére nem nyílt lehetőség a kutatás során; nem azonos minta oldotta meg a tesztek számítógép és papír alapon, nem azonos tesztváltozatokat oldottak meg a tanulók; különböző évfolyamok egymáshoz képest eltérő tulajdonságúak, bizonyos paramétereiket tekintve különböző mintaelemszámmal oszlanak el; az információ egyenlőtlen eloszlása miatt nem lehetett tanulói képességszintekkel számolni.

## Irodalom

- American Educational Research Association (AERA), American Psychological Association (APA) és National Council on Measurement in Education (NCME) (1999, 2014): *Standards for educational and psychological testing*. Washington, Amerikai Egyesült Államok.
- Applegate, B. (1993): Construction of geometric analogy problems by young children in a computer-based test. *Journal of Educational Computing Research*, 9. 1. sz. 61–77.
- Barnes, S. K. (2010): Using computer-based testing with young children. Előadás. NERA Conference Proceedings 2010. Paper 22.
- Bennett, R. E. (1998): *Reinventing assessment: Speculations on the future of large-scale educational testing*. Policy Information Center, Educational Testing Service, Princeton, NJ.

- Bennett, R. E. (2008). *Technology for large-scale assessment*. ETS Report No. RM-08-10. Educational Testing Service, Princeton, NJ.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. és Yan, F. (2008): Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, **6**. 9. sz. 4–38.
- Bunderson, V. C., Inouye, D. K. és Olsen, J. B. (1989): The four generations of computerized educational measurement. In: Linn, R. L. (szerk.): *Educational Measurement*. Macmillan, New York. 367–407.
- Choi, S. W. és Tinkler, T. (2002): Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting. Előadás. In: Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Clariana, R. és Wallace, P. (2005): Test mode familiarity and performance-gender and race comparisons of test scores among computer-literate students in advanced information systems courses. *Journal of Information Systems Education*, **16**. 2. sz. 177–182.
- CTB/McGraw-Hill (2003): *The computer-based or online administration of paper-pencil tests*. CTB/McGraw-Hill, Amerikai Egyesült Államok.
- Csapó, B., Ainley, J., Bennett, R., Latour, T. és Law, N. (2012): Technological issues of computer-based assessment of 21st century skills. In: McGaw, B. és Griffin, P. (szerk.): *Assessment and teaching of 21st century skills*. Springer, New York. 143–230.
- Csapó Benő, Molnár Gyöngyvér, Pap-Szigeti Róbert és R. Tóth Krisztina (2009): A mérés-értékelés új tendenciái: a papír és számítógép alapú tesztelés összehasonlító vizsgálatai általános iskolás, illetve főiskolás diákok körében. In: Perjés István és Kozma Tamás (szerk.): *Új kutatások a neveléstudományokban. Hatékony tudomány, pedagógiai kultúra, sikeres iskola*. Magyar Tudományos Akadémia, Budapest. 2009. 99–108.
- Csíkos Csaba és Csapó Benő (2011): A diagnosztikus matematika mérések részletes tartalmi kereteinek kidolgozása: elméleti alapok és gyakorlati kérdések. In: Csapó Benő és Szendrei Mária (szerk.): *Tartalmi keretek a matematika diagnosztikus értékeléséhez*. Nemzeti Tankönyvkiadó, Budapest. 59–99.
- Gyarmathy Éva (2012): Ki van kulturális lemaradásban? In: Tóth-Mózer Szilvia, Lévai Dóra és Szekszárdi Júlia (szerk.): *Digitális nemzedék konferencia 2012 Tanulmánykötet*, ELTE Eötvös Kiadó, Budapest. 9–12.
- Halldórsson, A., McKelvie, P. és Björnsson, J. (2009): Are Icelandic boys really better on computerized tests than conventional ones: Interaction between gender test modality and test performance. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 178–193.
- Hargreaves, M., Shorrocks-Taylor, D., Swinnerton, B., Tait, K. és Threlfall, J. (2004): Computer or paper? That is the question: Does the medium in which assessment questions are presented affect children's performance in mathematics? *Educational Review*, **46**, 1. sz. 29–42.
- Horne, J. (2007): Gender differences in computerised and conventional educational tests. *Journal of Computer Assisted Learning*, **23**. 47–55.



- Hülber László (2012): A papír és a számítógép alapú tesztelés összehasonlító vizsgálata különböző item paraméterek mentén. *Iskolakultúra*, **12.** 12. sz. 13–26.
- 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.
- International Test Commission (ITC) (2005): *International guidelines on computer-based and internet delivered testing*. International Test Commission. Bruxelles.
- Ito, K. és Sykes, R. C. (2004): Comparability of scores from norm-reference paper-and-pencil and web-based linear tests for grades 4–12. Előadás. Annual meeting of the American Educational Research Association, 2004. április 12–16., San Diego, Amerikai Egyesült Államok.
- Johnson, M. és Green, S. (2006): On-line mathematics assessment: The impact of mode on performance and question answering strategies. *Journal of Technology, Learning and Assessment*, **4.** 5. sz. 4–33.
- Lent, v. G. (2009): Risks and benefits of CBT versus PBT in high-stakes testing. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 83–91.
- Mazzeo, J. és Harvey, A. L. (1988): *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature*. College Entrance Examination Board, New York.
- Molnár Gyöngyvér (2010): Papír- és számítógép-alapú tesztelés összehasonlító vizsgálata problémamegoldó környezetben. In: Perjés István és Kozma Tamás: *Új Kutatások a Neveléstudományokban*. Aula Kiadó, Corvinus Egyetem, Budapest. 135–144.
- Pommerich, M. (2004): Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment*, **2.** 6. sz. 1–45.
- Puhan, P., Boughton, K. és Kim, S. (2007): Examining differences in examinee performance in paper and pencil and computerized testing. *Journal of Technology, Learning and Assessment*, **6.** 3. sz. 1–21.
- Redecker, C. és Johannessen, Ø. (2013): Changing assessment – Towards a new assessment paradigm using ICT. *European Journal of Education*, **1.** 48. sz. 79–96.
- Richardson, M. T., Baird, J.-A., Ridgway, J., Ripley, M., Shorrocks-Taylor, D. és Swan, M. (2002): Challenging minds? Students' perceptions of computer-based World Class Tests of problem solving. *Computers in Human Behavior*, **18.** 6. sz. 633–649.
- R. Tóth Krisztina (2009): Papír-ceruza és számítógépes tesztek eredményeinek összehasonlító vizsgálata. In: Vajda Zoltán (szerk.): *Bölcsészmuhely 2009*. JATEPress, Szeged. 125–136.
- R. Tóth Krisztina és Hódi Ágnes (2011): Számítógépes és papír-ceruza teszteredmények összehasonlító vizsgálata az olvasás-szövegértés területén. *Magyar Pedagógia*, **111.** 4. sz. 313–332.

- Scheuermann, F. és Björnsson, J. (2009): *The transition to computerbased assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg.
- Schroeders, U. (2009): Testing for equivalence of test data across media. In: Scheuermann, F. és Björnsson, J. (szerk.): *The transition to computerbased assessment: New approaches to skills assessment and implications for large-scale testing*. Office for Official Publications of the European Communities, Luxemburg. 164–170.
- Vidákovich Tibor (2012): A feladatok paraméterezése. Kézirat.
- Wang, S. és Shin, C. D. (2009): Comparability of computerized adaptive and paper-pencil tests. *Test, Measurement & Research Service. Bulletin*, **13**. sz. 1–7.
- Waters, S. D. és Pommerich, M. (2007): Context effects in internet testing: A literature review. Előadás. 22nd Annual Conference of the Society for Industrial and Organizational Psychology, 2007. április 7., New York City, Amerikai Egyesült Államok.
- Wilhelm, O. és Schroeders, U. (2008): Computerized ability measurement: Some substantive dos and don'ts. In: Scheuermann, F. és Pereira, A.G. (Szerk.): *Towards a research agenda in computer-based assessment. Challenges and needs for European Educational Measurement*. European Commission Joint Research Centre, Ispra. 76-84.
- Wolfe, E. W. és Manalo, J. R. (2005): *An investigation of the impact of composition medium on the quality of TOEFL writing scores*. Educational Testing Service, Amerikai Egyesült Államok.
- Wu, M. L., Adams, R. J. és Wilson, M. R. (1998): *ACER ConQuest: Generalised item response modelling software*. ACER press, Melbourne.