

Hülber László

SZTE Neveléstudományi Doktori Iskola

A papír- és a számítógép alapú tesztelés összehasonlító vizsgálata különböző ite paraméterek menté

A mérés-értékelés IKT eszközökkel való támogatása képes lehet a folyamat teljes és valódi megreformálására. A felelősségteljes áttéréshez azonban pontos információkkal kell rendelkezni arról, hogy a médium megváltoztatás, hogyan befolyásolja az eredményeket. Ez egyrészt indokolt a korábbi papí alapú eredményekkel való összehasonlítás végett, másrésztől nem lehetnek hátrányosan érintett tanulói körök (Molnár, 2010; Lent, 2009). Jelen tanulmány az itemek különböző paramétereinek szempontjából vizsgálja a média befolyásoló erejét a matematika műveltségterületen elsőtől hatodik osztályig.

Az infokommunikációs technológiák szerepe a mérés-értékelés folyamatában

A 21. században az infokommunikációs technológiák a mindennapi életben való érvényesülés általánosan elfogadott eszközeivé váltak. Korunk gyorsuló fejlődési folyamataiban azok a tudományterületek képesek követni a változásokat, amelyekben a technológiai eszközök használata teret nyer. A fejlődéssel lépést tartani akaró oktatási intézmények feladatai közé kell tartozzon a modern eszközök használatához tartozó képességek kifejlesztése (Griffin, Care és McGaw, 2012). Másrésztől a tanítási-tanulási folyamatok hatékonyságát növelhetik, az IKT mint információforrás, kommunikációs csatorna, produktum létrehozására alkalmas platform. Szerepét az oktatásban érintettek egyre komolyabban elismerik, szorgalmazzák használatát és igyekeznek bővíteni a hozzájuk társított feladatköröket.

A mérés-értékelési folyamatok IKT eszközökkel való támogatása is vitathatatlan előnyöket mutat az érintettek többségénél. A mérések hosszú távon alacsonyabb költségűek; az értékelési folyamatok a zárt végű feladatoknál teljes mértékben objektívek, a méréssel egy időben végbemenő számítógépes kiértékelésnek köszönhetően gyorsak és azonnali visszacsatolási lehetőséget biztosítanak a megoldóknak (Csapó és mtsai, 2012). Az értékelési folyamatok leegyszerűsödése hosszú távon megnövelheti a mérések számát, ami precízebb rendszerirányítást tesz lehetővé, mivel elegendő információ áll rendelkezésre a korrekciós eljárások gyakoribb alkalmazásához. Ezek a lehetőségek egyben igényként is megfogalmazódnak a rendszer irányításáért felelősök szempontjából.

Az előnyök mellett az is megfogalmazható, hogy a mérés-értékelés korszerűvé válásához elengedhetetlen a folyamat technológiai alapokra helyezése. Amennyiben a

* A kutatást a TÁMOP-4.2.2/B-10/1-2010-0012 projekt támogatta.

tanulási-tanítási környezetet meghatározza az IKT eszközök használata, értelemszerű következmény, hogy a tudás ellenőrzéséhez is kapcsolódnia kell az eszközhasználatnak (Lent, 2009). A multimédiás itemek használata, a dinamikus problémamegoldó képesség tesztelése, az adaptív mérés stb. megvalósítása papíron mind-mind kivitelezhetetlenek.

A tanulók mai generációjának információ preferenciája egyértelműen a képek, animációk, videók irányába mutat, teljes mértékben elfogadható módon, hiszen az agy működésénél fogva jobban dolgozza fel és raktározza a multimédiás forrásokat (Gyarmathy, 2012). Ezekhez az igényekhez igazodik, másrésztől a fejlődést szolgálja az ilyen típusú információk tesztelésben való felhasználása. A multimédiás elemek alkalmazási lehetősége teremti meg annak az alapját is, hogy a dinamikus problémamegoldó képességet lehessen tesztelni. A megoldók jellemzően egy általuk is megfigyelhető, változtatható szimulált környezet viselkedéséből vonnak le következtetéseket. Egy alapvetően fontosnak tartott képesség mérése válik így lehetővé, amely az életben való boldogulás egyik fokmérője (Greiff, Wüstenberg és Funke, 2012). Az adaptív teszteléssel, amely biztosítja a tesztelt aktuális tudáshoz illeszkedő itemeket, növelhető a tesztelés hatékonysága és a megoldók flow élménye (Csapó, Molnár és R. Tóth, 2008). Ezekkel a lehetőségekkel valósítható meg a tesztelés valódi megreformálása, amely jelentősen túlmutat a papír alapú tesztek számítógépes környezetbe való átültetésén.

Az áttérés biztosítása

A számítógép alapú tesztelés elterjedése a különböző szinteken és eltérő tétek esetén más és más problémaosztályokba ütközik. A nagy tétellel bíró tesztelés esetén hibamentes megbízhatóságot kell biztosítani, amelyhez megfelelő szoftver- és hardver infrastruktúra kiépítése szükséges. A kis tétellel bíró online mérések akkor lehetnek a mindennapi pedagógiai gyakorlat része, ha sikerül a tanárok mérési kultúrájának fejlesztés. (Csapó és mtsai, 2012). A technológiai alapú mérés népszerűségének mikro- és makroszinten is nagy lendületet adhat a PIS A mérések 2015-re tervezett számítógép alapúvá tétele.

Azonban az áttérés biztosításához pontosan kell tudni, hogy a közvetítő eszköz megváltoztatás, hogyan befolyásolja az eredményeket. Ez egyrésztől szükséges a korábbi papí alapú eredményekkel való összehasonlítás érdekében, másrésztől nem lehetnek hátrányosan érintett tanulói körök (Molnár, 2010; Lent, 2009). A médium megváltozása miatt feltehető az az alapvető validitási kérdés is, hogy ugyanazt mérjük-e papíron, mint számítógépen. A felelősségteljes migrációhoz megfelelő információval kell rendelkezni a média eredményt befolyásoló erejéről a mintára, a technológiára és a teszt jellemzőire vonatkozóan (CTB/McGraw-Hill, 2003).

Az eddigi kutatások eredményei

A kutatások ellentétes irányú eredményei miatt nehéz szintetizált általános érvényű következtetéseket megfogalmazni. Kim (1999) ötvenegy tanulmány felhasználásával készített átfogó meta-analízisben arra a következtetésre jut, hogy a számítógép- és a papí alapú eredmények közötti egyenértékűség mutatható ki mindaddig, amíg a teszt nem adaptív (Kim, 1999). A közelmúltban két, 1-12. osztályig terjedő matematikai és angol szövegértési meta-analízis (Wang és mtsai, 2007; 2008) fedi le az elmúlt 25 év kutatás eredményeit. Negyvennégy független adatállományt tartalmazó 14 matematikai tanulmány tette lehetővé a különböző médiumon elért pontok összehasonlítását. Hat adatállomány kivonása után, amelyek nagymértékben hozzájárultak a homogenitáshoz, a mért átlagos hatás statisztikailag nem különbözött a nullától. A hatás méretének előre-

jelzéséhez statisztikailag jelentősen hozzájárult, hogy a teszt adaptív vagy sem. A többi vizsgált változónak (a tanulmány tervezete, az életkor, a minta mérete, a teszt típusa, a hálózat alkalmazása) nem volt jelentékeny hatása.

Az angol szövegértés értékelésen mért átlagos hatás sem különbözött statisztikailag a nullától, kivéve azt a hatot a 42 adatállományból, amelyeket azért vettek ki, hogy megszüntessék a hatás méretének heterogenitását. Négy moderátor-változó (a tanulmány tervezete, a minta mérete, a számítógépes algoritmus és a számítógépes gyakorlat) volt hatással a szövegértés tesztmédiák közötti pontkülönbségeire, míg három másik feltételezett változó (az életkor, a teszt típusa és az internet-alapú tesztelés) statisztikailag nem volt jelentős hatással. A kis minta mindkét területen korlátozza az eredmények általánosíthatóságát.

A kizárólag online matematika tesztelésre fókuszáló elemzések közül Bennett és munkatársai egy 2001-ben az Amerikai Egyesült Államokban nyolcadikosok között végzett reprezentatív mérés során szignifikáns különbséget találtak a papí alapú eredmények javára. A tanulók egymás után oldották meg ugyanazokat az itemeket két csoportban, hogy mindkét lehetséges sorrendet vizsgálni lehessen. A vizsgált paraméterek közül az informatikai eszközök használatának sebessége befolyásolta leginkább az eredményeket (Bennett és mtsai, 2008).

Poggio és munkatársai 644 hetedikes tanulót vizsgáltak Kansas államban matematika műveltségterületen. 4 csoport jelentette a mintát: (a) ugyanazt a tesztet, illetve (b) ugyanolyan nehézségű tesztet oldottak meg papír-, majd számítógép -alapon, (c) csak papí alapú tesztet, illetve (d) csak számítógép alapú tesztet oldottak meg. Az eredmények statisztikailag elhanyagolható különbséget mutattak ki, az eltérésben a szülők iskolázottsága jelentette a legerősebb befolyásoló tényezőt (Poggio, 2005).

Azokon a területeken, ahol nagyobb hatást tudott gyakorolni a két médian történő olvasási képesség fejlettségi szintje, a számítógép alapú tesztverzió bizonyult nehezebbnek (Bennett és mtsai, 2008), és hasonlóan alakult az eredmény ott, ahol a szövegalkotás volt a feladat (Horkay és mtsai, 2006). Azonban a leggyakrabban vizsgált háttérváltozók közül a számítógéphasználat gyakorisága, fejlettségi szintje általában nem bizonyult erős teljesítménymódosító tényezőnek (O'Dwyer és mtsai, 2008), de az ellenkezőjét is igazolták már (Martin, 2009). Az eddigi tanulmányok közös pontja, hogy az adaptivitást jelentős módosító tényezőnek tartják. A különböző országokban végzett kutatások eredményeinek eltérését az az elv is eredményezheti, amely levonható következtetésképpen az adatokból, miszerint a hagyományos papíralapú és a számítógép alapú felmérés közötti eltérések minden bizonnyal nem egyeznek meg a különböző országokban. Másképpen megfogalmazva: az eltérések kultúrafüggőek is lehetnek (Björnsson, 2008).

A kutatás céljai

Amellett, hogy több összehasonlító tanulmány is rendelkezésre áll, kevés kutatás összpontosít az alkalmazott itemtípus meghatározó szerepére (Johnson és Green, 2006), vagy az ehhez szorosan kapcsolódó teljesítménybeli különbségekre. A fent megemlített tanulmányok legtöbbször igaz, hogy a felhasznált itemek nagy része felelet választós, így kevés információval rendelkezünk a többi altípus viselkedéséről. A vizsgált korosztály tipikusan középiskolában, illetve felsőoktatásban tanulók, azaz olyan populáció, amely feltehetőleg már gyakorlott az informatikai eszközök használatában.

A kutatási céloom, hogy megvizsgáljam, a teszt médiafüggőségét hogyan befolyásolja az alkalmazott itemek egyes paraméterei: tartalmi jellemzői, a műveleti dimenziók, a kontextus, a feladatmegoldás közben működő pszichikus struktúrák, a feladat formai jellemzői. A feladatok külső formai megjelenése mellett, indokolt kérdés, hogy a feladatok

belső jellemzői és a megoldáshoz szükséges tudás elemek, befolyásoló erővel hatnak-e, és ha igen, akkor milyen irányban módosítják az eredményeket. A célok közé tartozik, annak a megállapítása is, hogy léteznek-e jól definiálható azonos és különböző viselkedésű, médiafüggő itemek. Meghatározó tényező-e a kor és a hozzá tartozó számítógépes gyakorlottság abban a korosztályban, amelyben a tanulók elkezdik rendszeresen használni az IKT eszközöket.

Jelen kutatás a matematika műveltségterületre fókuszál, de a későbbiekben tervezett az olvasás-szövegértés és természettudományos területek vizsgálata és ezek összehasonlító elemzése is.

A vizsgált item paraméterek

A tanulmányban vizsgált elméleti paramétereket közé tartoznak a feladatok tartalmi jellemzői, a feladatmegoldás feltételezett kontextusa, a feladatmegoldás közben működtetendő pszichikus struktúrák és a feladat formai jellemzői. Az elemzések három fő dimenziójá, a pszichológiai (gondolkodási), a társadalmi (alkalmazási) és a diszciplináris (szaktárgyi) dimenziók jelentik, az oktatás három fő céljának megfelelően. A pszichológiai dimenzió a gondolkodás fejlesztésére fókuszál, a belső tulajdonságokat állítja középpontba a külső tartalmakkal szemben. A társadalom által elvárt és preferált tudás szempontjára koncentrálni az alkalmazási dimenzió. A szaktudományi dimenzió pedig az adott tudomány képviselői által relevánsnak tartott tudást jeleníti meg (Csíkos és Csapó, 2011; Vidákovich, 2012).

A felmérések tartalmi jellemzői matematika műveltségterületen mindhárom elemzésre használt fő dimenziójában konkrét témakörökhöz kötődnek, ezért értelmezésük azonos (1. táblázat).

1. táblázat. A tartalom paraméter értelmezése három dimenzióban (Vidákovich, 2012)

Szaktudományi dimenzió	Alkalmazási dimenzió	Gondolkodási dimenzió
1. számok, műveletek, algebra 1.1. számok 1.2. műveletek 1.3. algebra 2. relációk, függvények 2.1. sorozatok 2.2. adatpárok, adathármasok 2.3. ábrázolás Descartes-féle koordináta rendszerben		3. geometria 3.1. konstruálások 3.2. transzformációk 3.3. tájékozódás 3.4. mérések 4. kombinatorika, valószínűségszámítás, statisztika 4.1. kombinatorika 4.2. valószínűségszámítás 4.3. statisztika

A pszichikus struktúra elemzésével (2. táblázat) a feladatok megoldásához szükséges értelmi (gondolkodási) műveleteket, műveletsorokat, készségeket vizsgáljuk. A pszichikus struktúrához a szaktudományi és az alkalmazási dimenzióban is Bloom kognitív követelményrendszere alapján a következőket jelenti: ismeret (ráismerés, felidézés), megértés (értelmezés, magyarázat), alkalmazás (átalakítás, kivitelezés), magasabb szintű műveletek (analízis, szintézis, értékelés). A gondolkodási dimenzióban a pszichikus struktúra elemei a rendszerező képesség, a kombinatív képesség, a deduktív gondolkodás és az induktív gondolkodás vizsgált elemei.

2. táblázat. A pszichikus struktúra paraméter értelmezése három dimenzióban (Vidákovich, 2012)

Szaktudományi és alkalmazási dimenzió	Gondolkodási dimenzió	
1. ismeret (ráismerés, felidézés)	1. rendszerező képesség	3. deduktív gondolkodás
2. megértés (értelmezés, magyarázat)	1.1. besorolás, szelektálás	3.1. műveletek
3. alkalmazás (átalakítás, kivitelezés)	1.2. halmazképzés	3.1.1. kapcsolás
4. magasabb szintű műveletek (analízis, szintézis, értékelés)	1.3. felosztás	3.1.2. választás
	1.4. osztályozás	3.1.3. feltételképzés
	1.5. sorképzés	3.2. következtetések
	2. kombinatív képesség	3.2.1. előrelépő következtetés
	2.1. permutálás	3.2.2. visszalépő következtetés
	2.2. variálás	3.2.3. választó következtetés
	2.3. kombinálás	3.2.4. lánckövetkeztetés
	2.4. összes részhalmaz képzése	4. induktív gondolkodás
	2.5. Descartes-szorzat képzése	4.1. kizárás
		4.2. átkódolás
		4.3. analógia képzése
		4.4. sorozat képzése
		4.5. szabály megfogalmazása

A kontextus paraméter (3. táblázat) a feladatmegoldási szituációt jellemzi, ami a feladat által vizualizált szituációt jelenti, mivel a feladatok megoldása nem valós élethelyzetekben, hanem tanulási, iskolai környezetben történik meg.

3. táblázat. A kontextus paraméter értelmezése három dimenzióban (Vidákovich, 2012)

Szaktudományi dimenzió	Alkalmazási dimenzió	Gondolkodási dimenzió
1. rutinfeladat, nem szöveges	1. realiztikus feladat, nem autentikus	nem releváns
2. rutinfeladat, szöveges	2. realiztikus feladat, autentikus	

A szaktudományi dimenzióban a feladatok rutinfeladatok, ahol a megoldás rutinszerű, nem igényli a hétköznapi életből is ismerhető tények és viszonyok figyelembevételét a tanulóktól. Szöveges változatában az elvégzendő művelet vagy műveletsor szöveges formában jelenik meg, míg a másik változat szöveg nélküli matematikai struktúráként formalizálódik. Az alkalmazási dimenzióban a feladatok realiztikus feladatok, amelyek a tanulók által ismerős szituációval és tartalmakkal bírnak, és az így leírható tudáselemek matematikai modellezése a cél. A realiztikus feladatokon belül az autentikus feladatok sajátossága, hogy egy valóságos problémahelyzetet állít a tanulók elé, akik bevonság érzetük által motiváltabban próbálják a nem létező mechanikus megoldási algoritmus helyett modellalkotással megoldani a feladatot (Csikos és Csapó, 2011). A matematika esetében a gondolkodási dimenzióban a kontextus nem releváns.

A feladattípus paraméter a feladat formai jellemzői (4. táblázat) közé tartozik, ezért független a területtől, dimenziótól, a feladatmegoldó tevékenység típusára utal.

4. táblázat. A feladattípus paraméter értelmezése (Vidákovich, 2012)

Szaktudományi dimenzió	Alkalmazási dimenzió	Gondolkodási dimenzió
1. zárt 1.1. választás 1.1.1. alternatív választás 1.1.2. feleletválasztás 1.1.3. szelektálás 1.2. hozzárendelés 1.2.1. párosítás 1.2.2. halmazba sorolás 1.3. rendezés 1.3.1. relációválasztás 1.3.2. sorba rendezés	2. nyílt 2.1. kiegészítés 2.1.1. kép-kép 2.1.2. szöveg-szöveg 2.1.3. szám-szám 2.1.4. kép-szöveg, szöveg-kép 2.1.5. szöveg-szám, szám-szöveg 2.2. teljes válasz 2.2.1. kép 2.2.2. szöveg 2.2.3. szám 2.2.4. kép és szöveg 2.2.5. szöveg és szám	

A feladatok formai jellemzői közé tartoznak azok megjelenésbeli paraméterei. Így elemzésre került a feladat megoldásához szükséges információ mennyiségének és típusának befolyásoló ereje. Az információ mennyiségben kifejezhető értékei a feladathoz és a konkrét itemhez tartozó karakterek száma, a grafikus elemek mennyisége, táblázatok alkalmazása. Külön kategóriát jelentett azon grafikus elemek használata, amelyek a megoldáshoz szükségesek, illetve egy másikat, ahol csak illusztrációs szerepet töltenek be a rajzok, képek (1.a. és 1.b. ábrák).

Figyeld meg a logikai készlet elemét!

Tegyéél az állítás után (I) betűt, ha igaz. (H) betűt, ha hamis.

Minden lap nagy.

Van közöttük lyukas.

Van olyan, amelyik nem nagy.

Nincs közöttük kicsi, lyukas.

Egyik sem kék.

1.a. ábra. Megoldáshoz szükséges grafikus elem

A tálcán 15-nél kevesebb, de 5-nél több sütemény van.

Karikázd be az igaz állítások betűjelét, húzd át a hamisat!

a) Nincs 15 sütemény a tálcán.

b) 5 sütemény van a tálcán.

c) Ehetek 6 süteményt.

d) Legalább 6 sütemény van a tálcán.

e) Van 15 sütemény a tálcán.

1.b. ábra. Illusztrációs szerepű grafikus elem

Módszerek

Minta

Az összehasonlító elemzéshez papíralapon 2010-ben országosan reprezentatív mintán ($N_{\text{papír}}=20400$) elsőtől hatodik osztályig, három fő műveltségi területen felvett diagnosztikus mérések matematikára vonatkozó adatait használtam fel. Számítógépen 5 megye 17 iskolájának tanulói ($N_{\text{számítógép}}=943$) jelentették a mintát. Ugyanazon évfolyamokon a papíralapú feladatbankból vett itemek kerültek bemérésre.

A tesztelési eljárás

Papíralapon 20 tesztváltozat készült el, összességében 2050 item felhasználásával, ahol minden teszt esetében a reliabilitási mutatók jó értéket mutattak ($\text{Cronbach-}\alpha \geq 0,81$). Számítógép-alapon a fenti 2050 itemből került kiválasztásra 299 item, amely évfolyamonként 3 tesztváltozat formájában került megoldásra, megfelelő reliabilitási értékeket

eredményezve (Cronbach- $\alpha \geq 0,76$). A különböző médiumokon való azonos formai megjelenés fontos alapelvét jelenti az összehasonlító elemzéseknek. Ezért volt szükségszerű a papíralapútól eltérő tesztváltozatok kidolgozása, mert nem minden feladattípust lehetett számítógépen változatlan formában reprezentálni. Mindkét médiumon 45 perc állt a tanulók rendelkezésére, számítógépen a TAO online platform felhasználásával.

A diagnosztikus mérések mellett a korábbi kutatásokban gyakran alkalmazott, Csapó Benő által kidolgozott attitűd-kérdőív (Csapó, 1993–2006) került kitöltésre, így megfelelő információk álltak rendelkezésre a minta háttéradatairól. Ez tette lehetővé a mintaillesztési eljárást is. A jelentős mintaelemszám-eltérés miatt indokolt volt, hogy a számítógépen megoldó 943 főhöz mintaillesztést végezzenek a papíron dolgozó 20400 főből az elemzés megbízhatóságának növelése érdekében. A mintaillesztést az évfolyam, nem, régió, megye, anya és apa iskolai végzettsége és tanulmányi átlag sorrend alapján végezték el. Az évfolyam és a nem befolyásoló ereje, különválasztása társadalomtudományi kutatásokban értelemszerű, a többi tényező – az utolsót kivéve – pedig a PISA jelentések alapján a teljesítményt leginkább meghatározó szociokulturális háttér elemei (OECD, 2004). A fent megnevezett rendszer szerint a tanulók többségét: 919 főt sikerült legalább négy szempont szerint illeszteni.

Mivel a tanulók számítógépen és papíron eltérő tesztváltozatokat oldottak meg, ezért a különböző nehézségű teszten nyújtott nyerspont-alapú teljesítmény összehasonlítása nem alkalmazható. Az eltérő teszteken elért összpontszámok között a kapcsolat nem lineáris, ezért önmagában véve a lineáris transzformáció nem elegendő a közös viszonyítási rendszer kialakításához. Az itemek azonos skálán történő elemzéséhez a valószínűségi tesztmodellek közé sorolható Rasch-modellben használt logisztikus függvény biztosít megfelelő matematikai transzformációt. A különböző tesztváltozatok és évfolyamok egymáshoz való viszonyításának problémáját oldotta fel, hogy közös, összekötő itemek szerepeltek az egyes változatokban (Molnár, 2012).

Az elemzés első lépésben tehát a klasszikus tesztelméleti módszerek helyett a valószínűségiteszt-elméletek ('item response theory') játszottak főszerepet. A ConQuest program használatával az itemek közös skálára kerültek, és a szoftver által meghatározott nehézségi ('estimate') értékek jelentették a további elemzésekhez használt adatok forrását. A klasszikus elemzésekben a korrelációs számítások, két mintás t-próbák mellett több szempontú varianciaanalízist (ANOVA) alkalmaztam.

Annak érdekében, hogy az összehasonlító vizsgálatok teljes értékűek legyenek és a különbözőségekért felelős tényezők magyarázóerejének nagyságára is fény derüljön, a fent felsorolt vizsgálati paraméterek mellett a mintaelméleteket is elvégeztem.

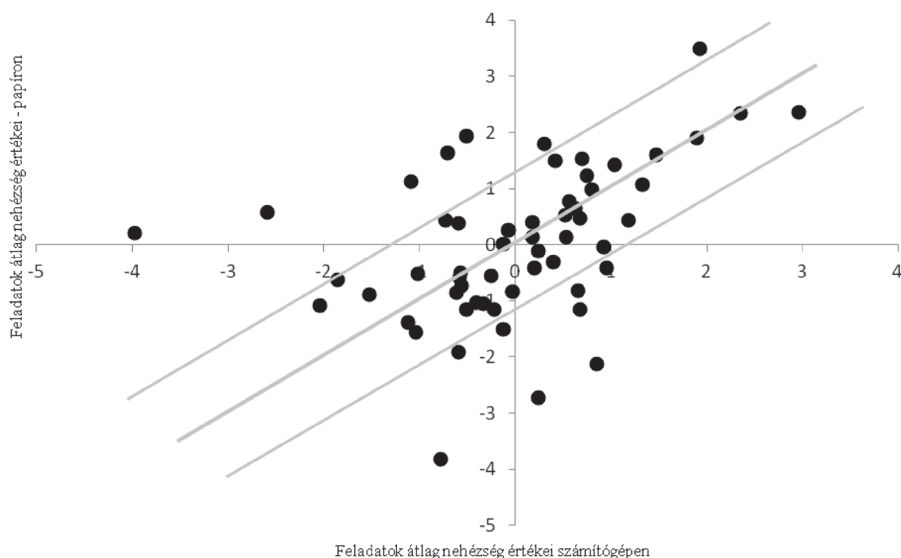
Eredmények

Pontozás

A teszt szerkezetére jellemző kialakítási módban egy feladathoz az ismertető rész után 2–6 darab (tipikusan 4) item tartozott. Mivel ezek az itemek önmagukban nem értelmezhetők teljes feladatként, ezért az összehasonlító vizsgálatokban a feladatszintű elemzés releváns. Ebben az értelmezésben 57 feladat elemzése vált lehetségessé.

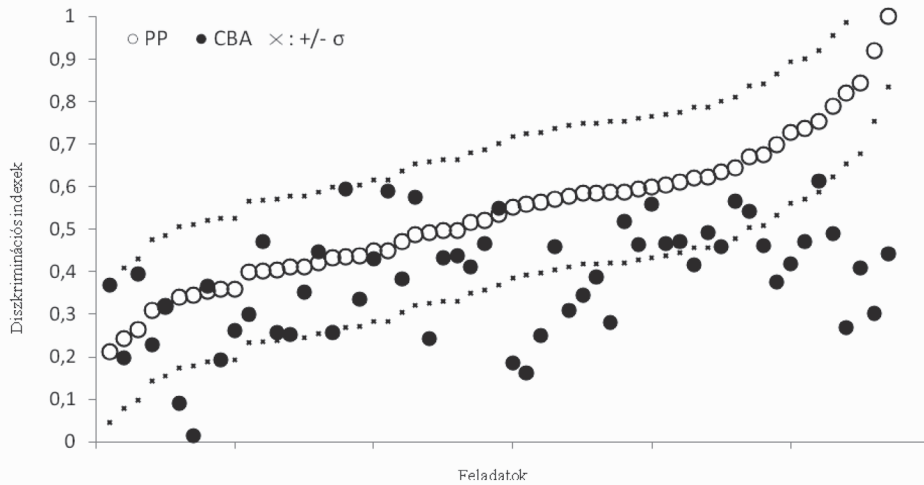
A két médiumon mért eredmények korrelációja

Az egyes feladatokhoz tartozó itemekre kapott nehézségi értékek átlaga alapján kiszámolt számítógép- és papíralapú eredményekhez tartozó korrelációs együttható értéke $r=0,59$ ($p<0,01$). Ez az érték erős kapcsolatot jelent a két médium között, de a 40 százalékos különbség okainak feltárása szükségszerű. Az 1. ábrán az egyes feladatokhoz tartozó két médium átlagos nehézségértékei szerepelnek, a plusz mínusz egy szórás értékek ábrázolásával. Az ezen értékek fölött, illetve alatt elhelyezkedő feladatok felelősek a különbségekért.



1. ábra. Feladatok átlag nehézségeinek korrelációja a két médiumon

A feladatok nehézségértékei mellett jellemző indikátor a feladatokhoz tartozó diszkriminációs index (2. ábra). Az itemdiszkrimináció arra utal, hogy az item mennyire tesz különbséget a magas és az alacsony eredményt elérők között (Crocker, 2006). A diszkriminációs indexek kétmintás t-próbával történő elemzése szignifikáns különbséget eredményez a papíralapú médium javára ($t=-10,03$, $p<0,001$). A 2. ábráról leolvasható, hogy számítógépen a feladatok kisebb mértékben választják szét a jól és a rosszul teljesítő tanulókat.



2. ábra. A feladatok átlag diszkriminációs indexeinek összehasonlítása

A korrelációk tekintetében 15 feladat helyezkedik el az egységnyi szórásértékeken túl, míg diszkriminációs indexek esetében ez az érték 20, a különbségekért felelős feladatok közös halmazába 11 feladat került. Tipikus jellemzőikben, paramétereikben egyértelmű szabályszerűség nem figyelhető meg.

A feladatok belső tartalmi jellemzői és a megoldáshoz szükséges tudáselemek menti összehasonlítás

A varianciaanalízis nem mutat ki szignifikáns különbséget a számítógép- és a papíralapú eredmények között aszerint, hogy a feladat milyen tartalmi jellemzőkkel bír; hogy a feladat szaktudományi, alkalmazási vagy gondolkodási dimenzióhoz köthető; hogy a feladat ismeret, megértés, alkalmazás jellegű vagy magasabb szintű műveleteket igényelt. A feladatok belső jellemzői közül a pszichikus struktúra gondolkodási dimenziójában a tanulók szignifikánsan jobban teljesítettek papíralapon (5. táblázat).

5. táblázat. A feladatok belső paramétereinek befolyásoló hatása

	Varianciaanalízis	
	F	p
Tartalmi jellemzők	0,87	0,46
Műveleti dimenziók	1,97	0,14
Kontextus	2,02	0,11
Pszichikus struktúra szaktudományi, alkalmazási dimenzióban	2,03	0,09
Pszichikus struktúra gondolkodási dimenzióban	17,6	<0,01

Az eltérés következtében t-próbákkal részletesen megvizsgált pszichikus struktúra gondolkodási dimenziójában lévő műveleti elemek közül az induktív gondolkodásnál mutatható ki a szignifikáns különbség (6. táblázat). A deduktív gondolkodást ellenőrző feladat nem állt rendelkezésre a vizsgálatához.

6. táblázat. Pszichikus struktúra gondolkodási dimenziójában lévő műveletek befolyásoló hatása

	N	Kétmintás t-próba	
		t	p
Rendszerező képesség	31	-0,52	0,61
Kombinatív képesség	20	-0,2	0,84
Deduktív gondolkodás	0	-	-
Induktív gondolkodás	28	3,62	<0,01

A feladatok formai paramétereire szerinti összehasonlítás

A két szempontos varianciaanalízis nem mutat ki szignifikáns különbséget a nyílt és a zárt végű eredmények közötti különbségben ($F=1,57$; $p=0,21$), de általánosításra kevésbé alkalmasak az eredmények az eltérő mintaelemszám miatt ($N_{Zárt}=183$; $N_{Nyitott}=94$). Önmagukban megvizsgálva a fő feladattípusokon történő teljesítést sem a zárt ($t=-0,797$ $p=0,426$), sem a nyílt ($t=0,886$ $p=0,377$) végű feladatoknál nincs jelentős eltérés a két médium között.

A feldolgozandó információ mennyiség szempontjából vizsgált első tényező a feladatokhoz és az adott itemekhez tartozó karaktermennyiségek összege. A két médiumon megvizsgálva az item-nehézségek és a karakterszámok közötti korrelációs együtthatókat elmondható, hogy gyenge a kapcsolat, és a két médium között számottevő különbség nem áll fenn ($r_{Számítógép}=0,32$, $p<0,01$; $r_{Papír}=0,25$, $p<0,01$).

Grafikus elemek alkalmazása nélkül nincsen szignifikáns különbség a számítógép- és a papíralapú eredmények között ($t=-1,039$; $p=0,3$), azonban a megoldáshoz segítséget adó avagy szükséges grafikus elemek alkalmazásánál (88 item) szignifikáns különbség áll fenn a médiumok között a ($t=2,22$, $p<0,05$) a számítógép javára. A táblázat alkalmazása nem befolyásolta az eredményeket.

A mintára irányuló elemzések

A mintára irányuló elemzésekben az évfolyamonkénti teljesítmény összehasonlításánál a helyes és helytelen megoldások arányát vettem figyelembe (7. táblázat). A minta eloszlása az egyes évfolyamokon nem egyenletes. Szignifikáns különbségek a papíralapú tesztelés javára második, ötödik és hatodik évfolyamokon adóttak.

7. táblázat. Az évfolyam befolyásoló hatása

Évfolyam	N	Kétmintás t-próba	
		t	p
1	49	0,3	0,76
2	84	-2,57	<0,05
3	156	0,57	0,57
4	196	0,95	0,35
5	238	-6,1	<0,01
6	214	-7,1	<0,01

A több szempontos varianciaanalízis szerint a nemnek nincsen befolyásoló hatása az eredményekre. Külön-külön megvizsgálva a fiúk és a lányok teljesítményét az egyes médiumokon mindkét esetben szignifikáns különbség adódott a papíralapú tesztelés javára (8. táblázat).

8. táblázat. A nem befolyásoló hatása

Médium	nem	N	Kétmintás t-próba	
			t	p
Számítógép	fiúk	487	-1,71	0,88
	lányok	456		
Papír-ceruza	fiúk	511	-0,23	0,81
	lányok	432		

Az eredmények értelmezése, következtetések

Mivel a feladatok belső jellemzői (tartalom, műveleti dimenziók, kontextus) tapasztalhatóan nem hatnak befolyásoló erővel, ezért a fókuszpont a külső formai jegyekre összpontosul. A feladatokhoz tartozó jellemző szövegmennyiség feldolgozása nem okoz problémákat, nem jelent hátrányt a számítógépről történő olvasás, a feladatok értelmezése. Az információ képi típusánál pedig éppen ellenkezőleg: számítógépen eredményesebb a feldolgozás. A háttérben lévő okra magyarázatot adhat, hogy számítógépen a tanulók nagyobb jelentőséget tulajdonítottak a kép megfigyelésének, jobban felkeltette az érdeklődésüket.

A hipotézis részét képezte, hogy a két médiumon eltérőek lesznek az eredmények a szövegalkotó és a zárt végű feladatok között. A feltételezés okát az adta, hogy külső szempontból vizsgálva itt jelentkezik a legnagyobb különbség a két médium között. A válaszadásra használt eszköz teljes mértékben más formát ölt és a megszokott ceruza (toll) helyett a számítógép billentyűzete és egeré más használati módot, tudást igényel. A feltevéshez hozzátartozott, hogy a kor előrehaladtával csökkennek ezek a különbségek, köszönhetően annak, hogy nő a számítógépes eszközök használatában való jártasság. A hipotézis nem igazolódott, hasonlóan az ismertetett kutatások eredményeihez. A jól definiálható azonos és különböző viselkedésű, médiafüggő itemek közös jellemzőinek meghatározásához további vizsgálatok szükségesek.

Az eredmények megfelelő kiindulási alapot adtak egy nagymintás online matematika mérés teszt szerkesztéséhez. A teszt összeállításánál meghatározó szerepet játszott,

Ezen kutatás eredményei szerint a nehezebb feladatok számítógépen történő megoldása jelent nehézséget a tanulóknak. Ezt igazolja, hogy a felsőbb évfolyamokon és a magasabb szintű gondolkodási műveleteket igénylő feladatoknál jelentkeztek a szignifikáns eltérések. Elképzelhető, hogy a tanulóknak a papírra történő jegyzetelés, azaz a próbálkozási lehetőség hiánya okozza a zavart. Ez kiküszöbölhető, ha a számítógépes mérés közben a tanulóknak biztosítanak papírt az egyéni számításokhoz.

hogy minden évfolyamon a végek szerint egyenlő arányban szerepeljenek a nyílt és zárt feladatok. A feladatbank lehetőségeinek és a számítógépen való azonos reprezentációs igények figyelembevételével az altípusok szerepeltetésében is a kiegyenlítettiségre törekedtem. A megoldáshoz szükséges grafikus elemeket tartalmazó feladatok számát is magasan tartottam, hogy igazolni tudjam ezen kutatás eredményeit. A minta összeállításánál határozott cél volt, hogy a kisebb korosztályba tartozók is nagyobb számban képviseltesék magukat, hogy jobban lehessen vizsgálni a kor befolyásoló erejét.

Irodalomjegyzék

- 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.
- Björnsson, J. K. (2008): Changing Icelandic national testing from traditional paper and pencil to computer-based assessment: Some background, challenges and problems to overcome. In: Scheuermann, F. és Guimaraes Pereira, A. (szerk.): *Towards a research agenda in computer-based assessment: Challenges and needs for European Educational Measurement*. 6–9.
- Crocker, L. (2006): Introduction to measurement theory. In: Green, J. L., Camilli, G. és Elmore, P. B. (szerk.): *Handbook of complementary methods in education research*. Erlbaum, Mahwah, NJ. 371–384.
- Csapó Benő, Molnár Gyöngyvér és R. Tóth Krisztina (2008): A papír alapú tesztekől a számítógépes adaptív tesztlésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 18. 3–4. sz. 3–16.
- 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.
- Csikos Csaba és Csapó Benő (2011): A diagnosztikus matematika felmérések 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. 141–168.
- CTB/McGraw-Hill (2003): The Computer-Based or Online Administration of paper-Pencil Tests. *Journal of Technology, Learning, and Assessment*, 2. 1. sz.
- Greiff, S., Wüstenberg, S. és Funke, J. (2012): Dynamic Problem Solving: A new measurement perspective. *Applied Psychological Measurement*, 36. 3. sz. 189–213.
- Gyarmathy Éva (2012): *Ki van kulturális lemaradásban?* Előadás: Digitális Nemzedék Konferencia Budapest, 2012. 02. 11.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B. és Yan, F. (2006): Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, 5. 2. sz.
- Hülber László (2011): Technológia alapú mérés-értékelés. *Új Katedra*, 9. sz. 11–15.
- Johnson, M. és Green, S. (2006): On-line Mathematics Assessment: The Impact of Mode on Performance and Question Answering Strategies. *The Journal of Technology, Learning, and Assessment*, 4. 5. sz.
- Kim, J. P. (1999): *Meta-analysis of equivalence of computerized and P&P tests on ability measures*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chigaco, IL.
- 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.
- Martin, R. (2009): Utilising the Potential of Computer-delivered Surveys in Assessing Scientific Literacy. 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. 172–166.
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, 20. 7–8. sz. 22–34.
- Molnár Gyöngyvér (megjelenés alatt): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest. Közlésre elfogadva.
- OECD (2004): *Learning for Tomorrow's World – First Results from PISA 2003*. OECD, Párizs.
- O'Dwyer, L., Russell, M., Bebell, D. és Tucker-Seeley, K. R. (2008): Examining the Relationship between Students' Mathematics Test Scores and Computer Use at Home and at School. *The Journal of Technology, Learning, and Assessment*, 6. 5. sz.

Griffin, P., Care, E. és McGaw, B. (2012): The Changing Role of Education and Schools. In: McGaw, B. és Griffin, P. (szerk): *Assessment and teaching of 21st century skills*. Springer, New York. 1–17.

Poggio, J., Glasnapp, D. R., Yang, X. és Poggio, A. J. (2005): A comparative evaluation of score results from computerized and paper and pencil mathematics testing in a large scale state assessment program. *Journal of Technology, Learning, and Assessment*, 3. 6. sz.

Vidákovich Tibor (2012): A feladatok paraméterezése. Kézirat.

Wang, S., Jiao, H., Young, M. J., Brooks, T. és Olson, J. (2007): A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67. 219–238.

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: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68. 5–24.

Tudósok a megismerés színterein

A romantikus tudományok és a 18–19. századi tudós-sztereotípiák

Szerkesztette GURKA DEZSŐ

Ezzel a kötettel befejező részéhez érkezett négyrészes sorozatunk, amely a 18–19. század fordulójának egyes filozófiai, tudományos, irodalmi és képzőművészeti jelenségeit kívánta bemutatni. Az itt szereplő tanulmányok többsége – éppúgy, mint korábban *A romantika terei*, a *Göttingen dimenziói* és *A középkor vetületei* című kötetekben megjelent írók is – a kora romantika és a német idealizmus kialakulása kapcsán megsokasodó diszciplináris kölcsönösségek és tematikus átfedések feltérképezését tűzte ki céljául. A tanulmánygyűjteményben szereplő filozófiai szempontú megközelítések főként a konstellációkutatás módszereit alkalmazzák, a tudománytörténeti vizsgálódások konceptuális keretét pedig a romantikus tudományok fogalma képezi, illetve az a Nicholas Jardine által kidolgozott fogalmi apparátus, amely elsőként tette lehetővé a pozitívizmus reprezentánsai által zavaros időszaknak minősített kora 19. század történeti-hermeneutikai aspektusú értelmezését.

ISBN 978 963 693 411 8
300 oldal, kartonált, A/5
2500 Ft

