

Értékek és értékelés

Értékelés-számítógépalapú tesztelés, online tesztkörnyezet, adaptív tesztelés¹

Az osztály országos felmérésre készül. Tavaly még mindenkinek hoznia kellett két jól fogó tollat, amivel beült az iskolapadba, kézhez kapta a meglehetősen vastag tesztfüzetet, és hozzálátott a feladatok megoldásához. A megoldásokat az arra kijelölt üres helyre, esetleg külön lapra kellett írniuk a diákoknak. Idén az egész osztály az informatika laborba ment kitölteni a tesztet, ahol állandó internetkapcsolat mellett oldották meg a diákok a feladatokat. Két tanulásban akadályozott, diszlexiás gyermek kapott egy-egy fülhallgatót, amin keresztül meghallgathatták a feladatokat, a számítógép olvasta fel nekik. Néhány diák hiányzott az iskolából, ők jövő héten oldják meg a tesztet. A diákoknak nemcsak feleletválasztós (multiple choice) kérdésekre kellett válaszolniuk, hanem esszét is írtak, grafikusan ábrázolandó matematika feladatokat oldottak meg, szimulációkkal dolgoztak, virtuális kísérleteket végeztek és számos eszközt használtak annak érdekében, hogy életszerű problémákat oldjanak meg. A számítógép a szerint adta minden diáknak a következő problémát, hogy az előzőre helyesen válaszolt-e. Ha jó választ adott a diák, akkor nehezebb feladat következett, ha helytelenül, akkor könnyebb. Ennek következtében az értékelésnél részletesebb, pontosabb képet kaptunk arról, hogy mit tud a diák. Mivel a tesztelés alkalmazkodott a tanulók tudásához, kevesebb feladatot kellett megoldaniuk ahhoz, hogy pontosabb eredményt kapjunk. A teszt megoldása után a számítógép rögtön elvégezte az értékelést, szövegelemző program segítségével még az esszék értékelését is. Az eredményt e-mailben elküldte a diáknak, a tanárnak, a felmérést szervezőnek, ahol azonnal adatbázis formájában elérhetőek voltak az eredmények.

Ez nem utópisztikus, 22. századbeli történet, hanem realitás a világ több országában, és hazánkban is hamarosan megvalósuló mérési-értékelési eljárás, módszer, cél. Ahogy és amilyen gyorsan a személyi számítógép és az internet megváltoztatta munkánkat, vásárlási és kommunikációs szokásainkat, úgy tűnik, hamarosan ugyanígy megváltoztatja azt is, ahogy az iskolában értékelünk, ahogy az iskolai, hazai és nemzetközi felméréseket végezzük. A számonkérések és a felmérések papír nélkül történnek majd, és az eredményre sem kell hosszasan várakozni, hanem azonnal rendelkezésre áll.

¹ A fejezet alapját CSAPÓ BENÓ, MOLNAR GYÖNGYVÉR és F TÓTH KRISZTINA: A papír alapú tesztekől a számítógépes adaptív tesztelésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. Iskolakultúra. 2008. 3-4. sz. 3-16. tanulmánya adta

Az utóbbi évek gyors technikai fejlődése, ennek az oktatásban való megjelenése fokozódó igényeket támasztott az értékelés újabb adathéberítési és adatelemzési technikáinak, módszereinek kidolgozása irányában. A számítógép alkalmazása nemcsak leegyszerűsíti a tesztelés folyamatát, hanem olyan hatékony módszereket is lehetővé tesz, amelyeket a hagyományos mérésekkel meg sem lehet közelíteni.

A számítógépes tesztelés szélesebb körű kipróbálására csak az utóbbi években, a számítógépek elterjedése után kerülhetett sor. Tekintettel a számítógépes tesztelés kimeríthetetlen lehetőségeire, kétségtelen, hogy belátható időn belül ki fogja szorítani a hagyományos papír alapú tesztelést. Iskolai kontextusban azonban csak fokozatosan lehet áttérni egy ilyen rendszerre, minden lépésben gondosan ellenőrizve, és kiszűrve a nemkívánatos mellékhatásokat (CSAPÓ, MOLNÁR és F. TÓTH, 2008).

Ebben a fejezetben áttekintjük a számítógépes tesztelés fő formáit, és bemutatjuk az adaptív tesztelés fontosabb lehetőségeit. Sorra vesszük azokat a problémákat is, amelyeket a pedagógiai alkalmazások felvetnek, és felvázoljuk a megoldás érdekében elvégzendő vizsgálatokat. A számítógépes tesztelés rövid történetére tekintettel a hatásvizsgálatok csak a közelmúltban kezdődtek el, és viszonylag kevés általánosítható eredmény áll a rendelkezésre.

A számítógépes és internet alapú tesztelés kulcsfogalmai

A technológia alapú mérés-értékelés

A technológiai alapú mérés (Technology Based Assessment - TBA) magában foglalja az összes olyan mérési-értékelési rendszert, amelyben az adatgyűjtésre valamilyen információs-kommunikációs technológiai eszközt használunk. Annak ellenére, hogy ez az eszköz általában a számítógép, a számítógépes mérés-értékelés halmazát magába foglaló bővebb halmazként mégis megkülönböztetjük ezt a kategóriát. A közvetítő eszköz ebben az esetben nem feltétlen a számítógép, lehet POA, mobiltelefon, szavazórendszer stb. (ezek iskolai alkalmazásáról 1. MOLNÁR, 2007), amelyek egy része alkalmas arra, hogy a nap bármely időszakában bizonyos kérdéseket tegyen fel a mérésben résztvevőnek - attól függetlenül, hogy az illető hol van -, s felkínálja az azonnali válaszadás lehetőségét is. A technológiai alapú mérés szinonim kifejezéseként használatosak az elektronikus tesztelés és e-tesztelés (e-Testing) kifejezések is.

A technológiai alapú mérésen belül a legtöbb lehetőséget természetesen a számítógépes értékelés kínálja, és (ma) ennek alkalmazása a legelterjedtebb.

A számítógépes mérés-értékelés

A számítógépes mérés-értékelés során az alkalmazott teszt a számítógép monitorján jelenik meg (on-screen presentation), a tesztelt személy pedig szintén a számítógép segítségével (billentyűzet, egér stb.) adja meg válaszát. A válaszokat rögzíti a szoftver, és általában az elemzés is számítógéppel történik. A számítógépes tesztelés a helyi hálózaton, illetve interneten keresztül is történhet. Ha semmilyen hálózatot (helyi hálózat, internet) nem vonunk be a tesztelés lefolytatásába, akkor a tesztelést végző programot, feladatlapot minden egyes számítógépre installálni kell. Az esetleges változtatásokat minden egyes számítógépen külön regisztrálni kell, majd az adatokat minden egyes számítógépről be kell gyűjteni.

A számítógépes tesztelés lényegében a számítógép oktatási célú alkalmazásával egy időben jelent meg. A korábban papír alapú feleletválasztós feladatokat minden nehézség nélkül át lehetett ültetni számítógépre, és ahogy a számítógépek fejlődtek, úgy alakultak ki az egyre fejlettebb számítógépes technikák.

A hálózat alapú mérés-értékelés

A hálózatalapú mérés-értékelés a számítógépes tesztelés egy olyan alkalmazása, amikor a teszt, a feladatok, a tesztelést végző program egy adott számítógépes hálózaton belül érhető csak el. Ez a hálózat lehet helyi (LAN) vagy az internet, illetve a kettő kombinációja (JURECKA és HARTIG, 2007). A hálózatban történő mérés egy gyakori alkalmazása, amikor az adott hálózaton belül egyszerre több gépen zajlik a tesztelés, a tesztelést egy külön számítógépről irányítják, ahol az adatok összegyűjtése, elemzése történik. A tesztelés előtt minden egyes adatfelvételben részt vevő gépre installálják a teszteléshez szükséges szoftvert. A kiértékelés szoftvertől függően vagy a helyi számítógépen vagy a központi szerveren történik.

Az internet alapú mérés-értékelés

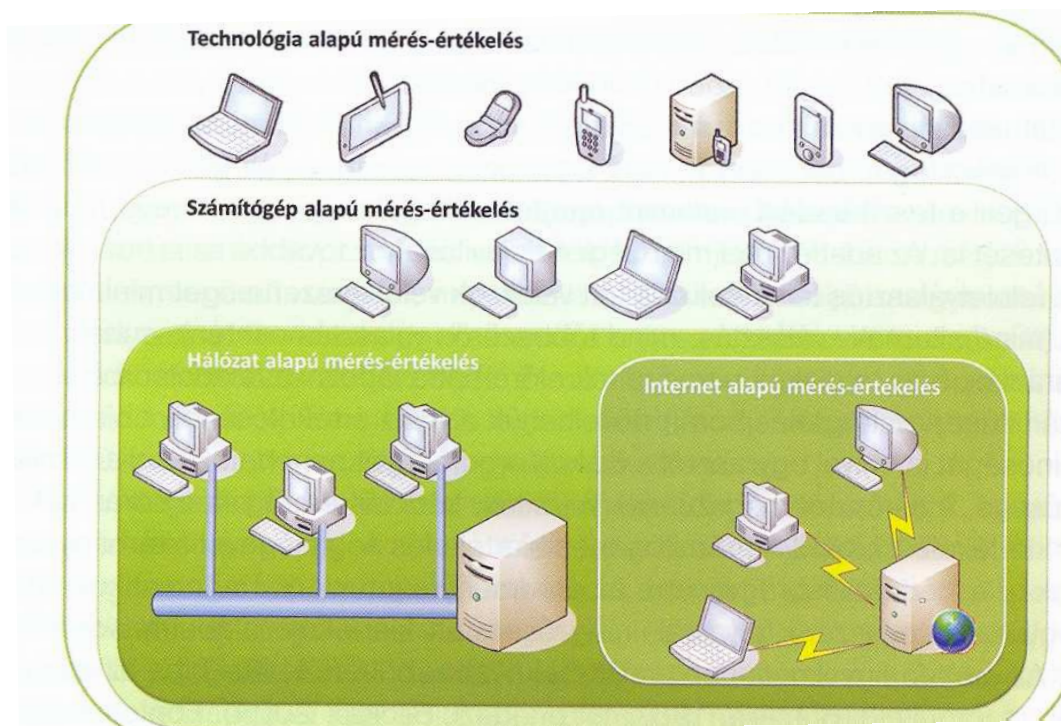
Az internet segítségével folyó teszteléshez az adatfelvételben részt vevő személynek csak internetkapcsolatra és egy internetes böngészőre van szüksége a válaszadáshoz. Nincs szükség arra, hogy a helyi számítógépen fusson a tesztelő program. A vizsgázó az azonosítójával tud belépni a rendszerbe, ott csatlakozik a tesztelő szoftverhez, ami a szerverrel kommunikálva választja ki a diák számára a megoldandó feladatokat. A feladatok és a szoftver is a szerveren található. A válaszok, adatok tárolását és kiértékelését is a központi szerver végzi. Ebből adódóan könnyebb és gyorsabb mind a feladatok, a tesztek, mind az egész feladatbank módosítása, és a szoftvert frissítése is. Fontos előny, hogy ha a szoftver külső gépen fut, nem kell minden iskolának saját szoftverrel rendelkeznie.

A számítógépes mérés-értékelés kulcsfogalmainak hierarchikus viszonya

A technológiai alapú mérés-értékelés, a számítógép alapú mérés-értékelés, a hálózat alapú mérés-értékelés és internet alapú mérés-értékelés definíciójuk értelmében egymásra épülnek, hierarchikus kapcsolatban állnak egymással. Az alkalmazott technológia szerint megkülönböztetett szintek egymásra és egymásba épülését a 82. ábra szemlélteti.

A legtágabb fogalom a technológiai alapú mérés-értékelés, ami magába foglalja mind a számítógép alapú, mind a hálózat alapú és az internet alapú mérésértékelést is. A számítógép alapú tesztelés a technológiai alapú mérés-értékelés egy részterületére csupán, de általában minden komputeres (asztali számítógép, notebook, classmate, XO) értékelést magában foglal. Hasonló a hálózat alapú és a számítógép alapú mérés-értékelés egymáshoz való viszonya is, az előbbi csak egy részterülete a számítógép alapú értékelésnek. Az internet alapú mérés-értékelés a legkisebb részterület, amely a hálózatalapú mérés-értékelés halmazába tartozik.

A számítógépes tesztelésre kifejlesztett rendszereket az alkalmazott médiumon kívül egy másik dimenzió mentén is csoportosíthatjuk, a feladatlapok, feladatok, itemek típusa, személyre szabottsága mentén. Ezen változó minden szintje megvalósítható a fent nevezett halmazok, részhalmazok bármelyikében. A továbbiakban e dimenzió mentén különítjük el egymástól az egyes lehetőségeket.



Online környezet

A számítógépes tesztelés legegyszerűbb formája a papír-ceruza tesztek egyszerű, az eredetivel megegyező formában való digitalizálása. Ebben az esetben csak a feladatokat közvetítő eszköz, vagyis a médium változik meg. A feladat a papír helyett a képernyőn jelenik meg, a válaszadás billentyűvel, egérrel, érintőképernyővel, vagy valami egyéb elektronikus eszközzel történik. A tesztelés továbbra is lineáris marad, a feladatok azonos sorrendben jelennek meg minden egyes tesztelt személy előtt. Érintőképernyőt használva papír-ceruza teszteléseivel való egészen közeli hasonlóságot lehet elérni, a vizsgázó - az érintőképernyő technológiájának függvényében - egy digitalizáló vagy egy közönséges toll segítségével jelöli meg választát. Egér- vagy billentyűhasználat esetében már szükség van némi technikai készségre, ha pedig a billentyűzettel hosszabb szövegeket kell bevinni, már számíthat a gépirási készség fejlettsége is. A legtöbb számítógép alapú teszt feleletválasztós feladatokból álló standardizált teszt (JURECKA és HARTIG, 2007).

A számítógépes tesztelés már ezen a szinten is számos előnnyel jár. Annak ellenére, hogy a tesztelt személy számára nem jelent nagy különbséget, a javítás, kódolás, rögzítés munkafázisait ki lehet iktatni, vagy jelentősen le lehet egyszerűsíteni. Objektív feladattechnikát alkalmazva a teszt kiértékelése azonnal megtörténik, az eredmény rögtön rendelkezésre áll. A papír-ceruza tesztelés során emberi munkára van szükség a válaszok javításához, rögzítéséhez, ami magában foglalja az adatvesztés lehetőségét, az adatminőség romlását is.

Az adatminőség javulásával a mérés egyik minőségi kritériumát, objektivitást növeljük. A számítógép nem fáradt, nem unatkozik, nem frusztrált (BECKER, 2004), nem sűrgeti a tesztbeadást, valamint megtakaríthatjuk a tesztet felvevő tanárok felkészítését is. Az adatfelvétel minőségének javításához továbbá az is hozzájárul, hogy a feleletválasztós feladatokra adott válaszok véletlenszerűségét minimalizálhatjuk (mind alternatív választás, mind többszörös választás esetén), ennek az a magyarázata, hogy a diákok nem tudnak előre-hátra lapozni a feladatsorban.

A számítógépes tesztelés során növelhetjük a teszt értékelésének objektivitását, minőségét is, mivel egyrészt a diákok eredményét nem befolyásolja a javító szigorúsága, másrészt megszűnnek a javítás, kódolás és rögzítés során keletkezett kiértékelési hibák. A számítógépes kiértékelés segítségével akárhányszor futtatjuk le a kiértékelést, ugyanarra az eredményre jutunk. Az automatikus értékelés gyors és egyszerű folyamat, még összetett kiértékelő algoritmusok esetén is. Az emberi figyelmetlenség miatt bekövetkező kiértékelési hiba az esetek 10%-ában fordul elő (BUTCHER, 1987. 17. o., idézi: BECKER, 2004). Fontos megjegyezni, hogy ha automatikusan értékelünk ki, akkor nem csak a feladat javításakor előforduló hibákat zárhatjuk ki, hanem a tradicionális tesztelés alkalmával végzett

adatrögzítéskor bekövetkező elgépelések hibáit is (pl.: 45-öt rögzítenek 54 helyett). Az automatikus kiértékelés egyszerűsíti a dokumentációt és a szervezést, segítségével össze lehet kötni az adatbankokat, és gyors lehívhatóságot (BECKER, 2004) biztosít.

A számítógépes tesztelés segítségével az adatok gyorsan aktualizálhatók, valamint azonnali visszacsatolási lehetőséget nyújt a diákok, tanárok, iskola, régió stb. számára. Az azonnali visszacsatolás pedig hozzájárul az oktatási-tanulási folyamat minőségének javulásához.

A számítógép alapú tesztelés induló költsége természetesen nagyobb, mint egy papír-ceruza tesztelés lebonyolítása, viszont a rendszer kiépítése után a számítógép alapú tesztelés számos megtakarítási lehetőséget kínál. A számítógépes kiértékelés segítségével kiküszöbölhetjük a tesztlapok nyomtatását, fénymásolását, csomagolását, szállítását, válaszlapok készítését stb., ezáltal az eszközököltség is jelentősen csökken. A tesztek javítására nem kell javítókat alkalmazni, a rögzítésre rögzítőket, sőt az alapstatisztikai számítások abban a pillanatban elkészülnek, ahogy a diák befejezte az utolsó ítem megoldását. Számítógépes tesztelése a dokumentációs költségek *2/3-át* meg lehet spórolni (ROSE és mtsái, 1999).

Az elektronikus rendszerre való áttérésnek ezen a fokon már lehetőség adódik a papír alapú és a számítógép alapú tesztelés hatékonyságának, eredményeinek összehasonlítására. A szakirodalomban számos kritikus észrevétellel is találkozunk a számítógépes tesztelések kapcsolatban. Leggyakrabban a számítógépes tapasztalat hiányát és a számítógéptől való idegenkedést említik. Ahogy azonban az IKT eszközök terjednek a hétköznapi életben, ennek a tényeznek a súlya egyre kisebb lesz. Nem szabad azonban megfeledkezni arról, hogy mindaddig, amíg a számítógéphez való hozzáférés tekintetében iskolák, társadalmi csoportok és családok között jelentős különbségek vannak, gondosan meg kell vizsgálni, nem hoz-e az alkalmazott eljárás egyeseket hátrányos helyzetbe. Gondoskodni kell arról, hogy az alkalmazott technika kezelése senkinek ne okozzon nehézséget, és ne vonja el a figyelmét az érdemi feladatmegoldó munkától. Ennek egyik legbiztosabb módja magának a számítógépes tesztelésnek az elterjesztése és gyakori alkalmazása.

A számítógépes tesztelés adta lehetőségeket jobban kihasználjuk, ha gazdagítjuk a tesztelés során alkalmazott feladatok típusát. Alkalmazhatunk multimédiás (hang, mozgókép, animáció, szimuláció, interaktív szimuláció stb.) elemekkel gazdagított feladatokat is, sőt a kiegészítő technológiák alkalmazásával lehetőség nyílik a fogyatékkal élő tanulók tudásának mérésére is. A "látási, hallási és a kézírás készségével kapcsolatos problémák jó része kiküszöbölhető" (KÁRPÁTI, 2002. 8. o.). Számos kutatás fókuszál a háromdimenziós tesztkörnyezet megvalósíthatóságának kérdéskörére. Ebben a háromdimenziós tesztkörnyezetben a

tesztelt személy szabadon mozoghat, cselekedhet, beszélgetéseket folytathat. A legfrissebb kutatási eredmények szerint (pl.: FREY és mtsai, 2007) az ilyen típusú tesztkörnyezethez könnyű alkalmazkodni.

A diákok konkrét válaszáon kívül további adatokat is gyűjthetünk a tesztelés során a tanulókról. Mérhetjük a diákok egyes feladatok megoldásához szükséges idejét, rögzíthetjük reakcióikat, az egész mozgását, a billentyűk lenyomása között eltelt időt, szemmozgásukat, amelyek további adatokat szolgáltatnak a figyelemre, gyorsaságra, olvasási képességre (visszaugrások száma) stb. vonatkozólag. A feladatokból különbözőképpen, tesztelő programtói függően állíthatunk össze tesztek. Eldönthetjük, hogy a diákok vissza tudjanak-e "lapozni" a tesztelés folyamán, hogy egyszerre csak egy feladatot lássanak-e, vagy a papír alapú tesztekhez hasonlóan több feladatot is lássanak egymás alatt. Legyen-e lehetőségük javításra, korlátozzuk-e az egyes feladatok megoldására szánható idő mennyiségét, vagy esetleg csak az egész tesztre vonatkozólag állapítsunk meg korlátot. Lássák-e a diákok, hogy mennyi idejük és kérdésük van még hátra, továbbmehetnek-e akkor is a feladatokon, ha nem adnak választ, vagy mindenképp valamiféle választ kell adniuk a következő elem megkapásához, és még sorolhatnánk az eldöntendő kérdések körét. Egy másik lényeges kérdéskör a teszt összeállításának módjára vonatkozik, ami újabb csoportosítási lehetőséget kínál.

A tesztelés típusai az egyénre szabottság mértéke szerint

A teszt összeállításának legegyszerűbb módja az, ha az előre digitalizált feladatokat statikusan egymás után fűzve adjuk a diákoknak (Linear testing). Ebben az esetben a papír-ceruza teszthez hasonlóan minden egyes diák ugyanazon feladatokat kapja, ugyanabban a sorrendben.

Fejlettebb szint az a lehetőség, ha sok feladattal rendelkezünk, és csak azt határozzuk meg, hogy hány feladatot kapjon az adott diák, a többit a programra bízunk, ami egy random függvénnyel biztosítja, hogy a diákok nagy valószínűséggel különböző tesztekkel oldjanak meg (Linear on-the-Fly testing; AL-NAU, 2007). Ez a tesztelési eljárás nem veszi figyelembe a diák képességszintjét.

Újabb lehetőség, ha részteszteket definiálunk (Testlet testing; AL-NAU, 2007), amelyek minden egyes feladatát megkapja a diák, ha az adott részteszt kiválasztásra kerül a tesztelés folyamán. Az egyes feladatok résztesztbe sorolása történhet például nehézség vagy témakör szerint. A tesztelés ezen típusa sem veszi figyelembe a diák képességszintjét, viszont lehetőséget ad a részteszten belül az egyes feladatok ismételt korrigálására.

Az a negyedik lehetőség, ha különböző feladatcsoportokat definiálunk, - esetleg feladatcsoportokon belül nehézségi index szerinti alakítunk ki csoportokat -, majd a teszt összeállítása során csak azt határozzuk meg, hogy melyik feladat-

csoportból hány és milyen nehézségű feladatot válasszon ki véletlenszerűen a program. Ebben az esetben - megfelelő számú feladat mellett - már nagy valószínűséggel biztosítható, hogy az egymás mellett ülő diákok adott időpontban különböző feladatokat kapnak. A feladatok nehézségének előzetes indexelése mellett még az is biztosítható, hogy mindenki közel azonos nehézségű tesztet oldjon meg. A diákok közel azonos nehézségű, de mégis különböző feladatokat kapnak a tesztelés során akkor is, ha az előre definiált feladatok bizonyos változóit véletlenszerűen generáltatjuk a programmal. Ennek a módszernek az alkalmazása gyakori fizika, kémia és matematikai feladatok esetén. Például a szöveges feladatok szövege és esetleg néhány adata változatlan marad, míg más változókat véletlenszerűen generáltatunk a programmal, s ennek következtében minden egyes diák más-más feladatot old meg. Az eddig említett tesztelési eljárások figyelmen kívül hagyják a diák aktuális képességszintjét. Egy, a korábbiaktól lényegesen eltérő mérés-értékelési eljárás, ha a tesztelés folyamatát egyénre szabjuk, azaz egy adott diák tesztjének nehézségi szintjét befolyásolja a diák képességszintje. Ehhez egy teljes mértékben parametrizált, indexelt és egy azonos nehézségi, illetve képességskálán leírható feladatbank szükséges. Ha a feladatbankból az egyes feladatok kiválasztása a vizsgázó előző válaszainak függvényében történik, adaptív tesztelésről beszélünk.

A számítógépes adaptív tesztelés

A számítógépes tesztelés igazán nagy lehetősége az adaptivitás, vagyis az, ha lehetőség van arra, hogy a vizsgázók attól függően kaphassanak újabb feladatokat, hogy hogyan oldották meg az előzőt. A számítógépes adaptív tesztelés (Computerized Adaptive Testing - CAT) a teljesítmények sokkal finomabb felbontását, mérését teszi lehetővé. Elméletileg tíz feladat megoldásával 210, azaz 1024 lehetőség közül választhatjuk ki, hogy pontosan milyen a vizsgázó képessége egy adott területen. Ez persze elméleti lehetőség, mert a gyakorlatban ehhez az kellene, hogy legyen 1024 olyan feladat, amelyik nehézsége egyenletesen fedi le a felméréndők képesség-tartományát. Ilyen feladatbankot azonban szinte lehetetlen elkészíteni, mivel a feladatok pontos nehézségét csak empirikus úton lehet meghatározni, és nem lehet "rendelésre" gyártani előre meghatározott nehézségű feladatokat. Mindenesetre ez a becslés jelzi az adaptív tesztelés elméleti lehetőségeit, de egyben a megvalósítás korlátait is.

A hagyományos papír-ceruza tesztelés során, illetve a tesztek digitalizált formában történő felvétele esetében is minden egyes személy ugyanazokat a feladatokat kapja, ugyanabban a sorrendben. Ezzel szemben az adaptív tesztelés során minden egyes személy a számára leginkább diagnosztikus erővel bíró feladatokat

kapja, azaz elhanyagolható annak a valószínűsége, hogy minden egyes személy ugyanazon feladatokat ugyanabban a sorrendben oldja meg. Ezáltal új lehetőségek nyílnak meg a mérés-értékelés területén.

A vizsgáztatás, mérés-értékelés e formáját analógiába állíthatjuk a szóbeli vizsgáztatással, ahol a vizsgáztató a kérdéseit gyakran a vizsgázó képességeihez igazítja. Ha a vizsgázó egy közepes nehézségű kérdésre helyes választ ad, akkor a vizsgáztató következő kérdése általában nehezebb, helytelen válasz esetén könnyebb kérdés következik. A vizsga végén az értékelés annak függvényében történik, hogy milyen nehézségű kérdésekre tudott még helyesen válaszolni a vizsgázó. Ha csak nehéz kérdéseket fogalmazna meg a vizsgáztató, akkor az alacsonyabb képességű vizsgázók értékelése nehézkessé válna, míg csak könnyű kérdések esetén nem lehet a jobb képességű vizsgázókat differenciálni.

Az adaptív tesztelés során a fentiekhez hasonló módon történik a feladatok kiválasztása, de a szóbeli vizsgával ellentétben néhány tényező tekintetében pontosabb, egzaktabb módon (FREY, 2007). A tesztelés során kiválasztásra kerülő kérdéseket a korábban kiválasztásra került feladatokra adott válaszok határozzák meg. Ez az eljárás azt a célt szolgálja, hogy minden egyes személy elé csak olyan feladatok kerüljenek, amelyek a lehető legnagyobb információval, diagnosztikus erővel bírnak az adott személy vizsgált képességszintje tekintetében, azaz amelyek lehetőleg a legközelebb vannak valós képességszintjéhez. A kiválasztás a legtöbb esetben a feladatok nehézsége alapján történik. A magasabb képességszintű egyének nehezebb, az alacsonyabb képességszintűek átlagosan könnyebb feladatokat kapnak a tesztelés során. Ezzel az eljárással elkerülhető, hogy az alacsonyabb képességszintűeket túl nehéz feladatokkal frusztráljuk, illetve a magasabb képességszintűek tesztelésre szánt idejét a könnyebb feladatok megoldásával töltsük ki. A feladatok kiválasztása előzetesen meghatározott algoritmus alapján történik. Ez az algoritmus egy olyan szabályrendszer, ami meghatározza az első és a rákövetkező feladatok kiválasztását, továbbá specifikálja a tesztelés befejezésének kritériumait is.

A számítógépes adaptív tesztelés összességében kevesebb feladat használatával és rövidebb idő alatt pontosabb képességszint-meghatározást tesz lehetővé. A technológia kihasználásával növelhetjük a tesztelés során felhasznált feladatok típusát például azzal, ha multimédiás elemeket is alkalmazunk. A teszt adaptivitásánál fogva nő a tesztbiztonság, mivel a jól és rosszul megoldott feladatok, illetve az előre meghatározott algoritmus függvényében személyre szabott tesztet tölt ki mindenki, azaz megszűnik a sűgás, lesés és előre kondicionált feladatok problémája, viszont megmarad a standardizált mérés. Ebből adódóan gyakran ismételhető, nem szükséges minden egyes mérés során új tesztet kidolgozni, mert a rendszer az előre kifejlesztett adatbankból válogatja össze a diák képesség-

szintjének leg pontosabb meghatározásához szükséges tesztet. Ezért a rendszer alkalmas arra, hogy a tanulókat megfelelő gyakorisággal felmérje, ezáltal állandó visszajelzést biztosítson az aktuális fejlettségük állapotáról.

Az azonos feladatbankon alapuló eredmények a közös nehézségi, illetve képességskálán definiált feladatok miatt viszonyíthatóak egymáshoz, azaz a tanuló korábbi fejlettségi szintjével összevethető az aktuális eredménye, még akkor is, ha összességében minden egyes alkalommal más feladatot oldott meg a diák. Ezzel kiküszöbölődik a longitudinális fejlődésvizsgálatok egyik alapproblémája, mely szerint ugyanazt a tulajdonságot többször egymásután ugyanazzal a teszttel kell felmérni, azonban igma tesztfeladatok egyre ismerősebbek lesznek, ami torzíthatja az eredményeket.

Összevethető a többi diák azonos mérésben megoldott teszteredményével, illetve az adatbank felépítése és az adott képességterület skálázása során meghatározott tudományosan kidolgozott standardokkal. Ennek következtében a papír-alapú keresztmetszeti vizsgálatok megvalósítására könnyen megvalósítható a standardizált longitudinális vizsgálat.

Az GAT azáltal teszi személyre szólóvá a mérést, hogy minden tanuló többségében a saját képességszintjének megfelelő feladatokat old meg. Ezáltal a mérés egésze sokkal szélesebb képességsávot tud átfogni, mint a papír alapú fix tesztek, mégis minden egyes esetben érzékenyebb, azaz az előre meghatározott fix, statikus tesztekénél kisebb különbségeket ki tud mutatni. A képességszinthez közel eső feladatok minden diák számára optimális kihívást jelentenek, így a munka nem válik unalmassá, és nem okoz túlzott szorongást sem. A tesztelési folyamat az optimális tapasztalatok (a flow-élmény, I. CSÍKSZENTMIHÁLYI, 1997) sávjában marad. Mindez előnyösen hat az érdeklődésre és a motivációra, aminek a tesztek gyakori alkalmazásánál meghatározó jelentősége van. A felsorolt előnyös tulajdonságok nagyon vonzóvá teszik a GAT alkalmazását, azonban egy jól működő GAT rendszer kidolgozása rendkívül bonyolult feladat. Még abban az esetben is, ha a mérendő tulajdonság egyszerűen leírható, a feladatok empirikus nehézségét csak megfelelő mintán való kipróbálással lehet meghatározni. Az elkészült feladatok jelentős részéről már az első kipróbálás során kiderül, hogy valamilyen szempontból hibásak, nem differenciálnak, nem illeszkednek a modellbe stb. A szűrőn átjutó feladatoknak pedig nem megfelelően szóródik a nehézsége a felmériendő spektrumon. A fejlesztés újabb fordulóiban további feladatok készülnek, már szándékoltan könnyebbek vagy nehezebbek a még "üres" képességtartományok lefedésére. Egy feladat elkészítése során a nehézségével "beletalálni" egy adott képességtartományba szinte lehetetlen, ezért általában több tucatnyi feladatot el kell készíteni, ki kell próbálni, mire közülük legalább egy megfelel az elvárásoknak. Nehezíti az elvégzendő fejlesztő munkát,

ha mindezt iskolai kontextusban kell elvégezni, hiszen így bizonyos tudást csak a tanév megfelelő szakaszában lehet felmérni, így korrekciós fejlesztő ciklusokra esetleg csak egy újabb év múlva kerülhet sor.

A számítógép alapú mérés-értékelés és a PISA-vizsgálatok

Az egyik legjelentősebb nemzetközi felmérésorozatot, a PISA is fokozatosan áttér a statikus papír alapú tesztelésről a számítógép alapú tesztelésre. Az első kör (2000, 2003 és 2006) adatfelvételeinek médiuma nagyrészt a papír volt. Az áttérés első lépcsőfokaként a 2003-as felmérés mérőeszközei között szerepeket egy számítógép-használati szokásokra és különböző nehézségű számítógépes műveletek ismeretére vonatkozó kérdőív. A PISA 2006-os vizsgálatban már opcionálisan szerepelt a természettudományi tudás számítógépes felmérése (Computer Based Assessment of Science - CBAS). A 13 ország részvételével lezajlott pilot felmérés eredményeiből kiderült, hogy a kétféle médiummal (papír alapú és technológia alap ú) elért eredmények között komoly különbségek voltak. A PISA második körének (2009,2012,2015) első vizsgálatában, a 2009-es felmérésben az elektronikus szövegek olvasása (Electronic Reading Assessment, ERA²) már a szövegértés terület önálló részkálaja lesz (OECD, 2007). A következő felmérési ciklusokban a számítógép alap ú mérés mind nagyobb szerepet kap, és belátható időn belül, tervek szerint a harmadik mérési körre (2018, 2021, 2024) teljesen megszűnik a papír alapú felmérés és a PISA vizsgálatok alapja is az adaptív tesztelés lesz.

A PISA szakértői ettől azt várják, hogy csökken a szervezési költség és a diákok tesztelés során igénybe vett ideje is. Hosszútávon számos további előnye is lesz a számítógép-alapú tesztelés bevezetésének: lehetőség nyílik a gondolkodás olyan aspektusainak mérésére, amit papír alapú tesztelései nem lehet megvalósítani.

Perspektívák és problémák

Mint minden új, a hagyományostól eltérő módszer bevezetésekor, a számítógépes tesztelés esetében sem csupán a lehetőségekre, hanem a problémák és veszélyek elemzésére is figyelmet kell fordítani.

A számítógépes tesztelés megvalósításának egyik alapfeltétele a megfelelő hardver és szoftveres környezet megteremtése az iskolákban és a tesztelés központjában. Az iskolákban a csoportos teszteléshez legalább egy, erre a célra használható számítógépekkel berendezett tanteremre van szükség. Ha ezeket a tantermeket a számítógépes tesztelés céljaira kellene létrehozni, az vállalhatatlan beruházást jelentene, és a fejlesztés költségei csak sok év után térülnének meg.

² Elérhető: https://mypisa.acer.edu.au/index.php?option=com_content&task=view&id=66&Itemid=4 51

Egészen más a helyzet, ha ezek a tantermek már ott vannak az iskolában, és többek között erre a célra is használhatók, hiszen így azonnal jelentkezik a költséghatékonyság előnye. A központi hardver és szoftver felállítása, a feladatbank kifejlesztése a papír-ceruza tesztek elkészítésénél költségesebb, de a karbantartása és alkalmazása már kevésbé költséges.

Az adaptív teszteléshez elegendő iskolánként egy tanteremmel számolni, ahol a párhuzamos osztályok egymás után oldhatják meg a feladatokat. Az adaptív feladatkiosztás biztosítja, hogy a tanulók sokféle feladattal találkoznak, ezért nem kell azzal a problémával számolni, hogy a párhuzamos osztályokban tanuló diákok elmondják egymásnak a feladatokat. Az online teszteléshez elegendő egy böngészőprogram, aminek segítségével elérhető a központi szerveren futó tesztelő program és feladatbank. A szabályosan felszerelt gépekre tehát lényegében semmit nem kell a tesztelés érdekében telepíteni. Ebből a szempontból Magyarországon hamarosan meglesznek az online tesztelés iskolai feltételei, így ma már fel lehet vetni az online tesztelés elterjesztésének kérdését.

Nehezebb kérdés a társadalmi feltételek megteremtése. Időbe telik, amíg minden érintett (diákok, tanárok, szülők, döntéshozók) megismeri és elfogadja a tesztelés új lehetőségeit. A személyre szabott számítógépes, online tesztelés Amerikában már jelenős múlttal rendelkezik, Európában azonban még csak most kezdődtek meg a szélesebb körű iskolai alkalmazással kapcsolatos kísérletek. Rendkívül fontos, hogy mielőtt bármilyen komoly tétellel bíró számítógépes tesztelés elkezdődik, lehetőség legyen a rendszer megismerésére, és az alkalmazás feltételeiről szakmai konszenzus alakuljon ki.

A számítógép alapú tesztelések kapcsolatosan az egyik legtöbbet vitatott kérdés a diákok és a tesztelést vezető személy informatikai jártasságának (ICT literacy, ICT familiarity) teszteredményeket befolyásoló hatása, amelyek kulturális, etnikai és a nemek közötti teljesítménykülönbségek, az emberek között lévő digitális szakadék (digital gap) hatásának felerősödéséhez vezethetnek. Ez a problémakör további validitási kérdéseket is felvet, mivel ezen a módon az informatikai jártasság, vagy a számítógéptől való félelem szintje megjelenik a teszteredményekben is, holott az nem képezte a vizsgálat tárgyát. Az ezen a területen végzett kutatások sem szolgálnak egységes eredménnyel. A kutatási eredmények alapján egyrészt van összefüggés a teszt eredménye és a személy informatikai jártassága között (pl.: TSENG, TIPLADY és WRIGHT, 1998), másrészt ez a befolyásoló hatás nem szignifikáns (pl.: POWERS és O'NEILL, 1993). Általánosabban is megfogalmazhatjuk a kérdést, vajon a tesztelés médiája az informatikai jártasság szintjétől függetlenül bír-e befolyásoló erővel.

Feltehetjük a kérdést, vajon ugyanazt a tudást méri a papír alapú és a számítógépalapú teszt, illetve meddig méri ugyanazt a tudást. Összehasonlíthatóak-e a különböző médiumon felvett teszteredmények? Ezek a kérdések már számos

kutatást indukáltak és a mai napig is foglalkoztatják a kutatókat. Az egyes konkrét vizsgálatok ugyanis nem adnak még általánosítható választ a problémára. Feltehető, hogy minél inkább megfeleltethető

egymásnak flexibilitásban, feladattípusok, alkalmazott elemek tekintetében a papíron, illetve számítógép segítségével kitöltött tesztek, annál kisebb a médiahatás. Ezt a feltevést azonban konkrét elemzésekkel kell igazolni, és meg kell határozni, milyen mértékűek az említett hatások. Minél inkább kihasználjuk a számítógép adta lehetőségeket, a számítógép előtt írt és a hagyományos tesztek különböző feladattípusain elért eredmények annál inkább eltérnek egymástól. Ezért az online és papír alapú tesztek eredményeinek összehasonlításakor olyan metrikákat indexeket kell meghatároznunk, amelyek lehetővé teszik a tesztpontszámok átváltását.