

Adaptív tesztek készítésének folyamata

A technológia elterjedése nemcsak mindennapi életünket, hanem az oktatás, ezen belül a mérés-értékelés folyamatát is jelentős mértékben befolyásolja. A hagyományos tesztelést egyre inkább felváltja a számítógép-alapú mérés, mely lehetővé teszi új, innovatív tesztelési eljárások alkalmazását (Molnár, 2010). Adaptív tesztelés során a tanulók nehézségben saját képességszintjükhöz illeszkedő feladatokat kapnak, ezáltal megvalósul a személyre szabott tesztelés (Keng, 2008).

A számítógépre alapozott adaptív tesztelés elméleti alapjainak kidolgozása már a 20. század első éveiben megtörtént, gyakorlati megvalósítása a számítógépek alkalmazásáig váratott magára (Linacre, 2000). Napjainkban, a számítógépek egyre nagyobb mérvű elterjedésével és a valószínűségi tesztelmélet térhódításával megteremtődtek a feltételei az adaptív tesztelés egyre szélesebb körben való felhasználásának. Alkalmazásuk terén az Egyesült Államok vezet, de az utóbbi időben Európa egyre több országában állnak át adaptív tesztekkel történő vizsgáztatásra (Magyar, 2012), illetve ezzel párhuzamosan új generációs értékelési módszerek kidolgozására (Molnár és Latour, 2011; R. Tóth, Molnár, Latour és Csapó, 2011).

Az adaptív teszteknek számos változata létezik, az item-alapú tesztektől a lineáris alteszteket alkalmazó többszakaszos tesztekig, az alapelvet tekintve azonban mindegyik adaptív teszt hasonlóan épül fel. A tanulmány célja áttekintést adni az adaptív tesztek készítésének folyamatáról és bemutatni az adaptív tesztek fő komponenseit, úgymint a kalibrált feladatbank létrehozásának folyamatát, az item-kiválasztási algoritmusokat, a tesztvégzódtetési kritériumokat, valamint a teszt kiközvetítésére használható platformokat.

Az adaptív tesztek készítésének lépései

Az adaptív tesztek működése szigorú algoritmushoz kötött (Linacre, 2000). A tesztelés háttérét kalibrált feladatbank biztosítja, mely a feladatokon túl azok pszichometriai jellemzőit is tartalmazza. A teszt kezdő feladatát ebből a feladatbankból választják ki – ez a legtöbb esetben egy átlagos nehézségű feladatot jelent (Csapó, Molnár és R. Tóth, 2008). A tanuló kezdő itemre adott helyes vagy helytelen választól függően az adaptív algoritmus szerint történik a következő feladat kiválasztása (Molnár, 2013). Helyes válasz esetén minden esetben nehezebb, helytelen válasznál könnyebb feladat következik. Ennek következtében a program algoritmus biztosítja, hogy a soron következő feladat nehézség vonatkozásában egyre közelebb legyen a tanuló képességszintjéhez. A feladatok kiértékelése automatikusan történik. A beépített algoritmus szabályozza azt is, hogy szükséges-e új item kiválasztása, vagy a tesztelés véget ért. A tesztelés végén a tanuló azonnali visszajelzést kap elért eredményéről (Csapó, Molnár és R. Tóth, 2008; Eggen, 2004; Magyar, 2012). Az adaptív tesztelés algoritmusának biztosítását a tesztek

szerkesztési folyamata biztosítja, mely a következő fő komponensekből tevődik össze (Weiss és Kingsbury, 1984; Thompson, 2007; Thomson és Weiss, 2011):

- Megvalósíthatóság, alkalmazhatóság lehetőségeinek felmérése,
- Feladatbank létrehozása,
- Valószínűségi tesztelméleti modell (Item Response Theory, IRT) kiválasztása,
- Itemek előtesztelése, kalibrálása, skálázása,
- Kezdő item(ek) kiválasztása,
- Itemkiválasztási algoritmus meghatározása,
- Végződtetési kritérium,
- A teszt kiközvetítése.

A továbbiakban ezen komponensek mentén ismertetem a tesztkészítés folyamatát.

Megvalósíthatóság, alkalmazhatóság lehetőségeinek felmérése

Adaptív tesztelésre való átállás esetén számos gyakorlati és üzleti kérdés merül fel, melyek befolyásolhatják a döntést. Először is nem minden hagyományos teszt konvertálható adaptívvá. (Linacre, 2000; Čisar, 2010). Másodsor fontos megfontolás tárgyát kell képezze, hogy az adaptív tesztre való átállás elvezet-e az elvárt mérészetodikai javuláshoz, vagyis a teszt hosszának és a tesztelési időnek a rövidüléséhez, valamint a precizitás és a tesztbiztonság növekedéséhez (Thompson és Weiss, 2011).

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

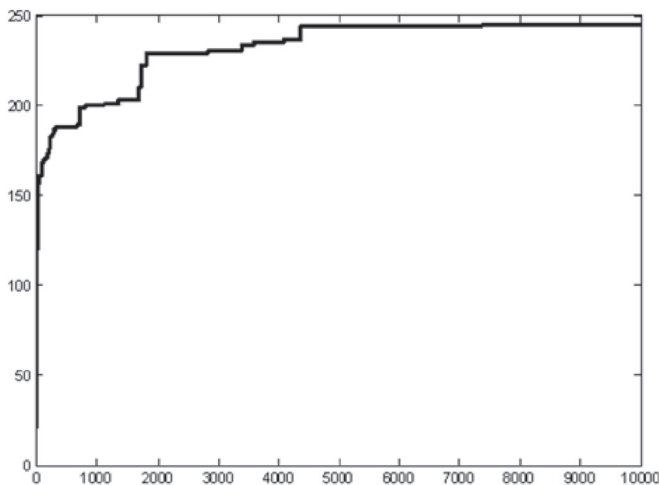
Kalibrált feladatbank létrehozása

Az adaptív tesztek alapvető feltétele a megfelelően kalibrált feladatbank. A feladatbank Milman (1984, 315. o.) definíciója szerint „könnyen hozzáférhető tesztkérdések viszonylag nagy gyűjteménye”. A „viszonylag nagy” azt jelenti, hogy az itemek száma többszöröse a tesztnél előforduló itemek számának, a „könnyen hozzáférhető” pedig azt érti, hogy az itemek indexeltek, paraméterekkel ellátottak, hogy a tesztelés folyamán minél könnyebben elérhető legyenek.

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

Feladatbankok kidolgozásánál fontos azt is tekintetbe venni, hogy a feladatoknak a tanulók képességszintjéhez kell igazodniuk, ezért a jó feladatbankok a képességskála egészét lefedik (Keng, 2008). Segall (2005) szerint az ideális feladatbank a képességskála minden szintjére vonatkozóan elegendő mennyiségű feladatot tartalmaz, és a magas diszkrimináló erővel, valamint alacsony találgatási paraméterrel rendelkező itemek vannak túlsúlyban. Tehát például érdemes kerülni az úgynevezett igaz-hamis állításokat tartalmazó itemeket, ahol a találgatás valószínűsége 50 százalék.

Reckase (2007) az egyparaméteres Rasch-modellt felhasználva próbált módszert kidolgozni az optimális feladatbank méretének meghatározására. Segal-hoz hasonló, ő is hangsúlyozta, hogy a feladatbank mérete szoros összefüggésben van a tesztelésbe bevont kohorsz képesség-eloszlásával.



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

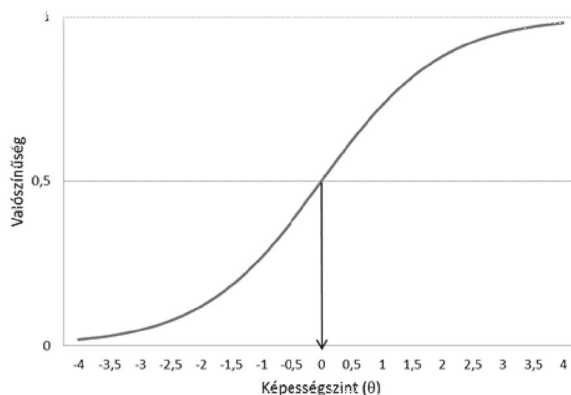
Az 1. ábra azt mutatja, hogy kisebb mintás mérések esetén (100–200 fő) 150–200 item elegendő, nagyobb minta esetén azonban minimum 250-es itemszám szükséges. A feladatok megfelelően kalibrálva kerülnek a feladatbankba. Az item kalibrálása a választott valószínűségi tesztelméleti modell segítségével az adott item paramétereinek becslését jelenti (Eggen, 2007; Weiss, 2011).

IRT modell választása

Az adaptív tesztek a valószínűségi tesztelmélet (Item Response Theory, IRT) felhasználásával készülnek (Thompson és Weiss, 2011). Ennek az az oka, hogy az IRT segítségével a különböző teszteken elért eredmények összehasonlíthatóak lesznek annak ellenére, hogy a tanulók különböző tesztekre oldanak meg (Eggen, 2008; Molnár, 2006). Az IRT így megkönnyíti annak meghatározását, hogy adott képességszintű tanuló milyen valószínűséggel teljesítene adott feladatbankban szereplő feladatok megoldásán, még akkor is, ha csak a feladatbankban szereplő itemek egy bizonyos részét oldja meg (Molnár, 2013).

Míg a klasszikus tesztelméleti mutatók csak az adott tanulócsoport esetén érvényesek, a valószínűségi tesztelmélet lehetővé teszi a mintafüggetlen, illetve tesztfüggetlen összehasonlítást (Csapó, Molnár és R. Tóth, 2008). A valószínűségi tesztelmélet abból indul

ki, hogy a jobb képességűek nagyobb valószínűséggel, az alacsonyabb képességszintűek kisebb valószínűséggel fogják jól megoldani ugyanazt a feladatot, ebből következőleg az a feladat nehezebb, amelyiket kisebb valószínűséggel oldanak meg a diákok (Csapó, Molnár és R. Tóth, 2008). Ez alapján mindegyik itemhez hozzárendelhető egy itemkarakterisztikus görbe, mely a különböző képességszintű tanulók helyes válaszána valószínűségét ábrázolja (2. ábra).



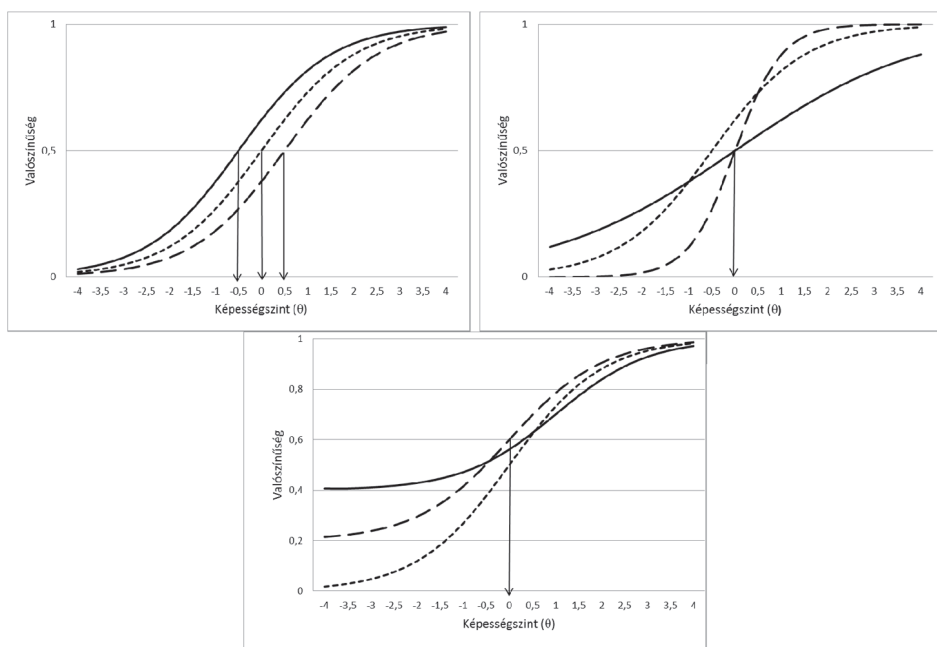
2. ábra. Itemkarakterisztikus görbe

A 2. ábra egy átlagos nehézségű item karakterisztikus görbét ábrázolja. Az ábrázolt item 0 (átlagos) nehézségi szintű (a 0 nehézségi szint a közepes nehézségi szintet jelenti, ettől jobbra a magasabb képességszint, balra az alacsonyabb képességszint található). Erre az itemre a közepes képességszintű tanulók 50 százalékos valószínűséggel lennének képesek helyesen válaszolni. Hasonlóan, a feladatbankban szereplő mindegyik itemhez hozzárendelhető bizonyos nehézségi szint, melyet az alapján definiálnak, hogy milyen képességszint szükséges ahhoz, hogy a helyes megoldás valószínűsége 50 százalék legyen (Molnár, 2013). Így lehetővé válik a tanulók képességszint szerinti és az itemek közös skálán való ábrázolása, mely megkönnyíti a tanulók képességszintjéhez legközelebb álló itemek kiválasztását (Eggen, 2004). A valószínűségi tesztelméleti modelleket leggyakrabban paramétereik száma szerint osztályozzuk (Baker, 2001). Attól függően, hogy az item hány paraméterét veszik számításba, egy-, két-, illetve háromparaméteres logisztikus modellek léteznek.

Az egyparaméteres logisztikus modell (például a Rasch-modell) a személyparaméter mellett egy paramétert tartalmaz, az itemnehézségi mutatót. Az itemeket grafikonon ábrázolva az itemek karakterisztikus görbéi párhuzamosan futnak. Az 3/a ábrán a folytonos vonallal rajzolt görbe a legkönnyebb, a pontozott az átlagos, a szaggatott vonallal ábrázolt a legnehezebb item helyes megoldásának valószínűségi görbét ábrázolja. A példában szereplő itemek nehézségi mutatói: -0,5, 0 és 0,5.

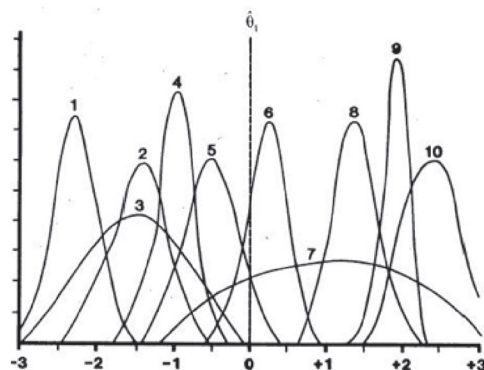
A kétparaméteres modellben az itemnehézségi mutató mellett az itemek diszkriminációs indexe is külön paraméterként szerepel. Az itemek karakterisztikus görbéi metszhetik egymást. A 3/b ábrán a szaggatott vonallal rajzolt item rendelkezik a legnagyobb diszkrimináló erővel, a folytonos vonallal rajzolt különböztet meg a legkevésbé.

A háromparaméteres modell a találgatás valószínűségét is számításba veszi. A 3/c ábra szaggatott és folytonos vonallal rajzolt itemjeinél a görbék aszimptotái 0,2-hez, illetve 0,4-hez tartanak, ugyanakkor a helyes válasz valószínűsége is magasabb, mint 50 százalék (Partchev, 2004; részletesebben lásd: Molnár, 2013).



3/a, b, c. ábra. Az egy-, két- és háromparaméteres logisztikus modell itemkarakterisztikus görbéinek egymáshoz való viszonya

Az itemek paramétereit felhasználva számítható ki az iteminformáció-függvény (Weiss, 2011). Egy adott item annál a képességszintnél méri a legpontosabban, ahol a függvény a maximumát éri el, vagyis ahol megegyezik az item nehézségi szintje a személy képességszintjével (Molnár, 2013). A4. ábra egy 10 itemből álló itembank iteminformáció-függvényeit ábrázolja.



4. ábra. Tíz itemből álló feladatbank item-információs függvényei (Weiss, 2011 alapján)

A 4. ábrán jelzett itemek közül a 9-es item szolgáltat a legtöbb információt, de csak a +2 képességtartományban. A 7-es item szélesebb sávon szolgáltat információt, a -1 és +3 közé eső képességtartományban, de a szolgáltatott információ precizitása jóval kisebb, mint a többi item esetében. A tesztelés során, amennyiben átlagos képességű tanuló számára ($\theta=0$) szeretnénk itemet kiválasztani, akkor a 6-os számú item a legmegfelelőbb,

mivel ez szolgáltat a legtöbb információt erről a képességszintről (Weiss, 2011). Az iteminformációs függvénynek az item-kiválasztási algoritmus meghatározásánál lesz kiemelt jelentősége.

Itemek előtesztelése, kalibrálása, skálázása

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

Kezdő item(ek) kiválasztása

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

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

Itemkiválasztási algoritmus

A tesztelés folyamán minden megoldott feladat után újabb feladat kerül kiválasztásra a feladatbankból. Azt, hogy melyik item kerüljön a következő lépésben a tanuló elé, az alkalmazott algoritmus dönti el. A leggyakrabban a legnagyobb valószínűség módszerét ('maximum likelihood estimation', MLE) alkalmazzák, mely szerint mindig a legmagasabb információtartalommal rendelkező item választódik ki (Kingsbury és Zara, 1989) (további módszerek: Swanson és Stocking, 1993; van der Linden és Glas, 2000; Schnipke és Green, 1995; Veerkamp és Berger, 1997; Rijn és mtsai, 2000; Chang és Ansley, 2003; van der Linden, 2005; Barrada és mtsai, 2009).

A tesztelés folyamán fontos szempont az itemek kiválasztási gyakoriságának szabályozása. A következő item kiválasztása az item-információs függvény segítségével történik, és általában a legnagyobb információval rendelkező item választódik ki (Molnár,

2013). Így előfordulhat, hogy némelyik item jóval gyakrabban szerepel, mint a többi. Három tipikus szabályozás létezik, mely ezt megakadályozhatja, az item kiválasztási, a tartalmi és az egymásnak információt szolgáltató itemek feletti szabályozás (Wise, 2000). Az item kiválasztásának szabályozása során az algoritmus azt vezérli, hogy ne hogy bizonyos itemek túl sokszor szerepeljenek, ami veszélyeztetné a tesztbiztonságot. A tartalmi szabályozás a különböző tartalmi elemek azonos mértékű kiválasztását kontrollálja (van der Linden, 2005). Az egymásnak információt szolgáltató itemeken alkalmazott szabályozás kiküszöböli a nagyon hasonló, vagy egymásnak információt szolgáltatatható elemek kiválasztását (Weiss, 2011). A különböző itemkiválasztást szabályozó módszerekről lásd részletesebben Revuelta és Ponsoda (1998), valamint Georgiadou (2007) tanulmányát.

Végződtetési kritérium

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

A tanuló képességszintjét figyelembe vevő módszernél a tesztelés folyamata akkor fejeződik be, amikor az újabb és újabb itemekre adott válaszok alapján a kiközvetített itemek paraméterei már csak minimális mértékben különböznek a tanuló képességszintjétől, vagyis az itemparaméterek közötti eltérés egy előre meghatározott sávon belülre kerül. A standard hibát figyelembe vevő módszer hasonló elv szerint működik, csak itt a mérési hiba mértékét számolják, amely nagyobb a tesztelés elején, és bizonyos számú item után minimálisra csökken (Thompson és Weiss, 2011).

A harmadik megközelítés a tanuló képességszintje helyett a feladatbankot veszi alapul. Ilyen például a minimum-információs kritérium, ahol a tesztelés akkor ér véget, mikor a feladatbankban már nem marad olyan item, mely újabb, az előzőeknél pontosabb információt szolgáltatna a vizsgált személy képességszintjéről (Thompson és Weiss, 2011). A negyedik lehetőség az időkorlát, amikor bizonyos idő elteltével a tesztelés véget ér. További részletek Weiss (2004), Thompson (2007) és Čisar (2010) tanulmányában találhatók.

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

A teszt kiközvetítése

Az adatfelvétel utolsó lépése a tesztek kiközvetítése. Ez történhet különböző, kereskedelmi forgalomban kapható, adaptív tesztek kezelésére alkalmas programok segítségével, de saját kutatócsoportok által kifejlesztett szoftvereken keresztül is végbemehet (lásd például: R. Tóth, Molnár, Latour és Csapó, 2011; Molnár és Latour, 2011; Molnár és Csapó, 2013). Mivel a tesztet használó intézmények különböző infrastrukturális feltételekkel rendelkeznek, az adaptív tesztet közvetítő szoftvernek megfelelően rugalmasnak kell lennie, hogy tudja ezeket a különbségeket kezelni (Way, 2006). A tesztek leggyakrabban online platformokon keresztül működnek, és folyamatos kapcsolatot tartanak a központi szerverrel. A tanulók előre rögzített belépési kóddal kezdenek meg a tesztelést és azonnali visszajelzést kapnak elért eredményükről.

Összefoglalás

Az adaptív tesztelés jelenleg a modern számítógépes technológiát és mérésmetodikai eljárásokat felhasználó eljárások közül az egyik legkifinomultabb és leginnovatívabb értékelési módszer. Az adaptív tesztek szerkesztése azonban meglehetősen összetett feladat. Egyrészt megfelelő szakértelmet kíván, másrészt pedig pénz- és időigényes feladat nagymintás mérésekre is alkalmas adaptív tesztelő rendszer kialakítása. A tesztek készítése, működtetése speciális szoftvereket kíván, melyek fenntartása jelentős anyagi befektetéssel jár. A tesztek működése szigorú algoritmushoz kötött, melyet a teszt készítése folyamán szükséges pontosan kidolgozni. A tanulmány részletesen tárgyalja az adaptív tesztek szerkesztésének főbb lépéseit, és bemutatja a tesztkészítés háttérét képező valószínűségi tesztelmélet néhány vonatkozó elemét is.

Az adaptív tesztek működését megfelelően kalibrált feladatbank biztosítja, melybe az itemek paramétereikkel ellátva kerülnek. A tesztelés során innen választódnak ki a tanulók képességszintjéhez leginkább közel álló itemek. A kezdő item általában egy közepes nehézségű item, majd a megoldástól függően kap a tanuló könnyebb vagy nehezebb itemet a következő lépésben. Az itemek kiválasztását különböző item-kiválasztási algoritmusok szabályozzák, melyek nemcsak az item nehézségét szabhatják meg, hanem különböző szabályozásokat végezhetnek a többi elemen, mint például tartalmi vagy egymásnak információt szolgáltató itemek kiküszöbölése. A legtöbb adaptív teszt pontozására nem alkalmasak a hagyományos pontozási módszerek, ezért leggyakrabban a valószínűségi tesztelméletet használják, és a tanulókat válaszaik függvényében képességszálán helyezik el. A tesztek kiköszvetítése általában online módon történik, melynek az az előnye, hogy a tesztelésbe bevont intézményeknek nem kell speciális szoftvereket telepíteniük, csupán folyamatos internet-elérést kell biztosítaniuk a tesztelési idő alatt.

A tanulmány röviden összefoglalta az adaptív tesztfejlesztés fő lépéseit és általános képet adott az adaptív tesztek felépítéséről. Mivel azonban számos típusú adaptív teszt létezik, a különböző szerkezetű tesztek fejlesztése esetén speciális kérdések merülhetnek fel, melyek eltérő eljárásokat igényelhetnek.

Az adaptív tesztek működését megfelelően kalibrált feladatbank biztosítja, melybe az itemek paramétereikkel ellátva kerülnek. A tesztelés során innen választódnak ki a tanulók képességszintjéhez leginkább közel álló itemek. A kezdő item általában egy közepes nehézségű item, majd a megoldástól függően kap a tanuló könnyebb vagy nehezebb itemet a következő lépésben. Az itemek kiválasztását különböző item-kiválasztási algoritmusok szabályozzák, melyek nemcsak az item nehézségét szabhatják meg, hanem különböző szabályozásokat végezhetnek a többi elemen, mint például tartalmi vagy egymásnak információt szolgáltató itemek kiküszöbölése.

Köszönetnyilvánítás

A tanulmány megírását a TÁMOP 3.1.9-11 kutatási program támogatta.

Irodalomjegyzék

- Baker, F. B. (2001): *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation. University of Maryland, College Park, MD.
- Barrada, J. R. és mtsai (2009): Item selection rules in computerized adaptive testing accuracy and security. *Methodology*, **5**. 1. sz. 7–17.
- Chang, S. W. és Ansley, T. N. (2003): A comparative study of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, **40**. 1. sz. 71–103.
- Čisar, S. M., Radosav, D., Markoski, B., Pinter, R. és Čisar, P. (2010): Computer adaptive testing of student knowledge. *Acta Polytechnica Hungarica*, **7**. 4. sz. 139–152.
- 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.
- Davis, S. (2005): *Exploring a new methodology for setting performance level standards with computerized adaptive tests*. Paper presented as part of the „Computer Adaptive Testing in State Assessment” session at the 35 th Annual National Conference on Large-Scale Assessment. San Antonio, TX.
- Engen, T. J. H. M. és Straemans, G. J. J. M. (2000): Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, **60**. 5. sz. 713–734.
- Engen, T. J. H. M. (2004): *Contributions to the theory and practice of computerized adaptive testing*. Citogroep Arnhem, Netherlands.
- Engen, T. J. H. M. és Verschoor, A. J. (2006): Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement*, **30**. 5. sz. 379–393.
- Engen, T. J. H. M. (2008): Choices in CAT models in the context of educational testing. In: Hartig, J., Klieme, E. és Leutner, D. (szerk.): *Assessments of competences in educational contexts*. Hogrefe & Huber Publishers. 199–217.
- Georgiadou, E. és mtsai (2007) A Review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment*, **5**. 8. sz. 5–38.
- Harris, D. (1989): Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, **8**. sz. 35–41.
- Keng, L. (2008): *A Comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas, Austin.
- Kingsbury, G. G. és Zara, A. R. (1989): Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, **2**. 359–375.
- Lee, J. E. (2011): *Full-metric concurrent calibration for the development of CAT item banks*. A thesis submitted to the faculty of the graduate school of the University of Minnesota.
- Linacre, J. M. (2000): *Computer-adaptive testing: A methodology whose time has come*. MESA Psychometric Laboratory, University of Chicago.
- Magyar Andrea (2012): Számítógépes adaptív tesztlés. *Iskolakultúra*, **22**. 6. sz. 52–60.
- Milman, J. és Arter, J. A. (1984): Issues in item banking. *Journal of Educational Measurement*, **21**. 4. sz. 315–330.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, **16**. 12. sz. 99–113
- Molnár Gyöngyvér (2010): Technológia-alapú mérés-értékelés hazai és nemzetközi implementációi. *Iskolakultúra*, **20**. 7–8. sz. 22–34.
- Molnár Gyöngyvér és Latour, Th. (2011): *Online tesztlés: lehetőségek és kihívások*. IX. Pedagógiai Értékelési Konferencia, Szeged, 2011. április 29–30. 63. o.
- Molnár Gyöngyvér (2013): *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó, Budapest.
- Molnár Gyöngyvér és Csapó Benő (2013): *Az eDia online diagnosztikus mérési rendszer*. XI. Pedagógiai Értékelési Konferencia. Elfogadott előadás.
- Partchev, I. (2004): *A visual guide to item response theory*. <http://www.metheval.unijena.de/irt/VisualIRT.pdf>. Tóth Krisztina, Molnár Gyöngyvér, Latour, Th. és Csapó Benő (2011): Az online tesztlés lehetőségei és a TAO platform alkalmazása. Új Pedagógiai Szemle, **61**. 1–2–3–4–5. sz. 8–22.
- Reckase, M. D. (2007). The design of p-optimal item bank for computerized adaptive tests. In: Weiss, D. J. (szerk.): *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. <http://publicdocs.iacat.org/cat2010/cat07reckase.pdf>
- Revuelta, J. és Ponsoda, V. (1998): A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, **35**. 4. sz. 311–327.
- Rijn, P. W. és mtsai (2000): *A selection procedure for polytomous items in computerized adaptive testing*. Measurement and Research Department Reports (2000-5). Cito, Arnhem.
- Schnipke, D. L. és Green, B. F. (1995): A comparison of item selection routines in linear and adaptive tests.

- Journal of Educational Measurement*, **32**. 3. sz. 227–242.
- Segall, D. O. (2004). Computerized adaptive testing. In: Kempf-Lenard, K. (szerk.): *The Encyclopedia of Social Measurement*. Academic Press, San Diego, CA. <http://iacat.org/sites/default/files/biblio/se04-01.pdf>
- Spray, J. A. és Reckase, M. D. (1994): *The selection of test items for decision making with a computer adaptive test*. Paper presented at the annual meeting of the National Council on measurement in Education, New Orleans.
- Swanson, L. és Stocking, M. L. (1993): A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, **17**. sz. 151–166.
- Thompson, N. A. és Prometric, T. (2007): A practitioner's guide for variable-length computerized classification testing. *Practical Assessment Research and Evaluation*, **12**. 1. sz. <http://pareonline.net/getvn.asp?v=12&n=1>
- Thompson, N. A., és Weiss, D. A. (2011): A framework for the development of computerized adaptive tests. *Practical Assessment Research and Evaluation*, **16**. 1. sz.
- van der Linden, W. J. és Glas, C. A. W. (2000, szerk.): *Computerized adaptive testing: Theory and practice*. Kluwer, Boston.
- van der Linden, W. J. (2005): A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, **42**. 3. sz. 283–302.
- Veerkamp, V. J. J. és Berger, M. P. F. (1997): Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, **22**. 2. sz. 203–226.
- Wainer, H. (2000): *Computerized adaptive testing: A primer*. 2nd Edition. Erlbaum, Hillsdale, NJ.
- Way, W. D., Davis, L. L. és Fitzpatrick, S. (2006): *Practical questions in introducing computerized adaptive testing for K-12 assessments*. Pearson.
- Weiss, D. J. és Kingsbury, G. G. (1984): Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, **21**. 4. sz. 361–375.
- Weiss, D. J. (2004): Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, **2**. sz. 70–84.
- Weiss, D. J. (2011): Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, **2**. 1. sz. 1–27.
- Weiss, D. J. (2011): Item banking, test development, and test delivery. In: Kurt, F. G. (szerk.): *The APA handbook on testing and assessment*. American Psychological Association, Washington DC. in press. http://www.assess.com/docs/Weiss_Handbook_Chapter.pdf
- Wise, S. L. és Kingsbury, G. G. (2000): Practical issues in developing and maintaining a computerized adaptive testing program. *Psicológica*, **21**. sz. 135–155.