

**Csányi Róbert<sup>1</sup> – Molnár Gyöngyvér<sup>2</sup>**<sup>1</sup> SZTE BTK Neveléstudományi Doktori Iskola<sup>2</sup> SZTE BTK Neveléstudományi Intézet

MTA-SZTE Digitális Tanulási Technológiák Kutatócsoportot

# A tesztmegoldási motiváció szerepe az alacsony téttel rendelkező tesztek eredményeinek értékelésében

*A mérés-értékelés főbb céljainak átalakulásával, tanulást segítő diagnosztikus funkcióinak előtérbe kerülésével az alacsony téttel bíró tesztek egyre nagyobb szerepet játszanak a tanítás-tanulás folyamatában, valamint az országos és nemzetközi értékelésekben. Miután az alacsony téttel bíró teszteken nyújtott teljesítménynek nincs következménye a tanulók tanulmányi előrehaladására, előfordulhat, hogy az ezeken a teszteken elért eredmények eltérnek a nagy téttel bíró teszteken nyújtott teljesítményektől. Ennek oka lehet, hogy előbbin a tanulók nem feltétlen tesznek meg minden erőfeszítést a feladatok megoldása során, utóbbin a diákok tesztmegoldás közbeni szorongása jelentős teljesítménycsökkenő hatással bírhat. A tanulmány keretein belül az első problémakörrel foglalkozunk részletesebben.*

## Bevezetés

Az iskolai, illetve tágabb értelemben véve az oktatási folyamat során alkalmazott értékelési eljárások hagyományosnak tekinthető célja annak meghatározása, hogy a tanulók mit tudnak, és a megszerzett tudásukat milyen mértékben képesek alkalmazni különböző helyzetekben. Ennek meghatározásához olyan mérőeszközök szükségesek, amelyek megfelelően reprezentálják a mérni kívánt tudásterületeket. A tesztek eredményeit – a mérés típusától függően – (1) az egyes tanulók, (2) a tanulócsoporthoz, (3) az intézmény vagy (4) az egész oktatási rendszer jellemzésére használják. Mind nemzetközi, mind hazai szinten a mérés-értékelés szummatív hegemoniája, megközelítése erőteljes változáson esett át az elmúlt években (Molnár, 2021b). Ezt a folyamatot jelentősen gyorsították a COVID-19-cel kapcsolatos változások, kutatások. Az elszámoltathatóság szempontjából lényeges tesztelés mellett (Koretz, 2018) egyre inkább előtérbe kerül a személyre szabott tanulást és tanítást támogató, tanulói szintű visszacsatolásra fókuszáló diagnosztikus értékelés (Csapó és Molnár, 2019).

Az objektív, megbízható és érvényes mérési eredmény eléréséhez nem elegendők a jó megbízhatósági mutatókkal rendelkező tesztek, valamint azok alkalmazása, nem elegendő a megfelelőség szerint azonos mérési körülmények biztosítása, nem elegendő az objektív, mindenki számára azonos szigorúsággal történő kiértékelés (Molnár, 2013), ugyanis különböző affektív tényezők is erőteljesen befolyásolhatják a tesztelt

személyek teljesítményét (Wise és mtsai, 2014). A motiváció részleges vagy teljes hiánya jelentős hatással van a teszten nyújtott teljesítményre.

Számos kutatás elemezte a teszten nyújtott teljesítmény és a tesztmegoldási erőfeszítés, valamint a motiváció közötti kapcsolatot. Wise és DeMars (2005) elemzése alapján a nem motivált diákok több mint fél szórással gyengébb eredményt értek el a teszteken, mint motivált társaik. Ezt támasztják alá Finn (2015), Schüttpelz-Brauns és munkatársai (2018), illetve Wise és Kong (2005) kutatási eredményei is, amelyek a motiváltabb tesztmegoldók magasabb teljesítményét jelezték. Mindeközben Lee és Chen (2011) kutatási eredményei szerint nem mutatható ki jelentős összefüggés a tesztelt személyek motivációja és képességei között, azaz a teljesítménybeli különbségek oka nem feltétlen a résztvevők között meglévő tudás- és képességszintbeli különbséget jelzik, azok a tesztmegoldási motiváció indikátorai lehetnek (Lee és Chen, 2011). Goldhammer és munkatársai (2016) szerint a tesztfeladatok megoldása közben mutatott magatartás (*test-taking behaviour*) függ mind a tesztelt személy tudásszintjétől, mind a tesztmegoldási elkötelezettségtől. Az egyik legjelentősebb nemzetközi felméréssorozat, az OECD PISA mérés 2015-ös adatfelvételének másodelemzése kapcsán Akyol és munkatársai (2021) megállapították, hogy a nem motivált diákok válaszaik miatti torzítás jelentős, az adatfeldolgozás és adatelemzés során a motivátlanságból eredő változás mindössze felét korrigálják az elemzések. A fenti elemzések hagyományos, szummatív teszteken nyújtott teljesítményekre vonatkoztak, amelyek eredményei bár nem általánosíthatóak, de fontos kiindulópontul szolgálnak az egyre nagyobb teret nyerő kis téttel bíró diagnosztikus teszteredmények értékelése során.

### **A tesztek tékje: alacsony és nagy téttel rendelkező tesztek**

A tanítási-tanulási folyamat akkor a legsikeresebb, amikor a diákok azt tanulják, aminek az elsajátítására felkészültek, aminek az elsajátítására készen állnak, azaz amihez rendelkeznek a szükséges alapokkal (Molnár és Csapó, 2019a). A személyre szabott tanulás szerepének felértékelődésével párhuzamosan egyre fontosabbá vált a tanulók előrehaladásának, hiányosságainak ismerete. Ezt a diákok fejlődésére vonatkozó információt csak gyors, lehetőség szerint azonnali visszajelzésen alapuló, gyakori teszteléssel kapjuk meg. Ennek megvalósítását szolgálják az alacsony téttel rendelkező diagnosztikus tesztek.

Egy teszt tékje definíció szerint azt jelenti, hogy a mérési folyamatban tesztelt személy számára milyen következményei vannak a teszten nyújtott teljesítménynek, azaz mekkora tékje van a teszten nyújtott teljesítménynek a felmért egyén számára (Csapó, Molnár és R. Tóth, 2008). A tesztelt személy számára alacsony téttel rendelkező tesztteredmény definíció szerint nincs jelentős hatással az adott személy előmenetelére (Lindner és mtsai, 2019), míg a magas téttel rendelkező teszteken nyújtott teljesítmény meghatározó következményekkel bír: ide sorolható például a KRESZ-vizsga, vagy a kifejezetten nagy téttel bíró érettségi vizsga, aminek eredménye kihathat a tesztelt személy egész további életére. A felmért személy számára alacsony téttel rendelkező tesztek közé sorolható az Országos kompetenciamérés vagy a PISA felmérés. Előbbi iskolai szinten, utóbbi az oktatási rendszer szintjén rendelkezik komolyabb téttel, miután azok eredményei, az intézményi, valamint nemzetközi értékelések, akár oktatási reformok alapját képezik (Lindner és mtsai, 2019). Megjegyzendő, hogy az alacsony vagy a magas tét a skála két végpontjára utal, hiszen a tét nagyságát tekintve egy folytonos változóról beszélhetünk (Csapó, Molnár és R. Tóth, 2008).

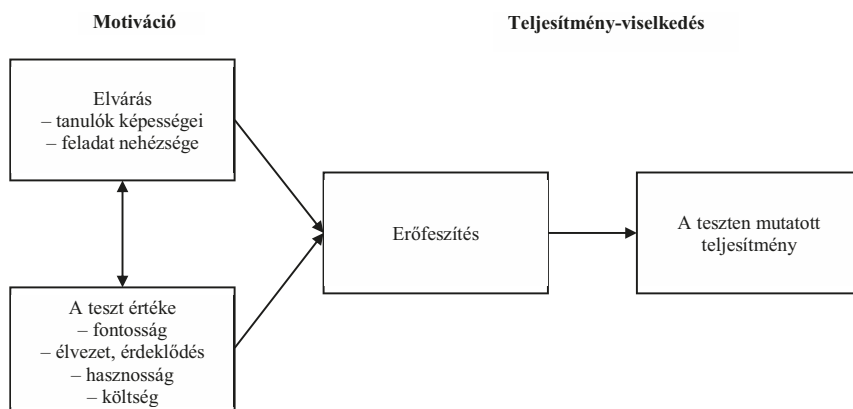
A tesztelt személy számára alacsony téttel rendelkező tesztek három fő kategóriája különböztethető meg (Wise és Kong, 2005): (1) a nemzetközi szervezetek mérés-értékelési programjainak tesztjei, mint például a 15 éves diákok műveltség típusú tudását

mérő, az OECD szakértő csoportjai által kidolgozott PISA tesztek, vagy az OECD felnőttekre (16–64 évesek) fókuszáló felmérésorozata, a PIAAC felmérés tesztjei; (2) a magas téttel bíró tesztek feladatainak kipróbálására szervezett mérések tesztjei; (3) a diákok tanulási folyamatát közvetlenül támogató diagnosztikus értékelést szolgáló tesztek. Ezek eredményei nem befolyásolják a tanulók osztályzatait, előmenetelét, így alacsony a tétjük, ugyanakkor segítenek feltárni a tanulók tudás- és képességszintjét, ezért az oktatási fejlesztések fontos alapját képezik. A teszt tétje alapvetően befolyásolja a tesztelt személyek érdekltségét és motivációját: például a tét növekedésével növekszik annak a valószínűsége is, hogy a tesztmegoldók tiltott segédeszközöket használnak. Ennek következtében a teszt tétjének növekedésével párhuzamosan egyre nagyobb erőfeszítések szükségesek a tesztelés objektivitásának biztosítása érdekében. Ugyanakkor a teszt-eredmény tétjének szerepének csökkenésével arányosan várhatóan csökken a tesztmegoldók tesztmegoldási motivációja, ami hatással lehet teljesítményükre.

### **A tesztmegoldási motiváció: elvárás-érték elmélet**

A tesztmegoldási motiváció értelmezésére jól használható elméleti keretrendszer az elvárás-érték elmélet (*Expectancy-value theory*; Wise és DeMars, 2005). A modell azon a feltételezésen alapul, hogy a motiváció egyrészt az elvárás és az azon alapuló siker valószínűsége, másrészt az érték, ami a teszt hasznosságának függvénye is. Az elvárás-érték elmélet tesztmegoldási motivációra alkalmazott modelljét az 1. ábra mutatja. A tesztmegoldók saját teljesítményükkel kapcsolatos elvárásait egyrészt saját képességeik, pontosabban a saját képességeikről alkotott véleményük, illetve a feladatok nehézsége határozza meg. A teszt hasznossága kapcsán Eccles és Wigfield (2002) négy komponenst különböztetett meg: (1) elérési érték, azaz a teszt fontossága; (2) belső érték, azaz a teszt iránti érdeklődés, annak élvezete; (3) hasznossági érték, a teszt hasznossága; (4) költség (pl. feladatra fordított idő, teszt szorongás).

*Egy teszt tétje definíció szerint azt jelenti, hogy a mérési folyamatban tesztelt személy számára milyen következményei vannak a a teszten nyújtott teljesítménynek, azaz mekkora tétje van a teszten nyújtott teljesítménynek a felmért egyén számára (Csapó, Molnár és R. Tóth, 2008). A tesztelt személy számára alacsony téttel rendelkező teszteredmény definíció szerint nincs jelentős hatással az adott személy előmenetelére (Lindner és mtsai, 2019), míg a magas téttel rendelkező teszten nyújtott teljesítmény meghatározó következményekkel bír: ide sorolható például a KRESZ-vizsga, vagy a kifejezetten nagy téttel bíró érettségi vizsga, aminek eredménye kihathat a tesztelt személy egész további életére. A felmért személy számára alacsony téttel rendelkező tesztek közé sorolható az Országos kompetenciamérés vagy a PISA felmérés. Előbbi iskolai szinten, utóbbi az oktatási rendszer szintjén rendelkezik komolyabb téttel, miután azok eredményei, az intézményi, valamint nemzetközi értékelések, akár oktatási reformok alapját képezik (Lindner és mtsai, 2019).*



1. ábra. Elvárás-érték elmélet a tesztmegoldási motiváció kontextusában (Penk és Richter, 2017 alapján)

Baumert és Demmrich (2001) meghatározása értelmében a tesztmegoldási motiváció (*test-taking motivation*) „hajlandóság a tesztfeladatok kidolgozására”, hajlandóság arra, hogy a tesztmegoldó erőfeszítést és kitartást fektessen a feladatok megoldásába. Wise és DeMars (2005. 2.) a következőképpen definiálta a tesztmegoldási motivációt: a vizsgázók „teljes erőfeszítéssel oldják meg a tesztet azzal a céllal, hogy az pontosan reprezentálja a tudásukat és képességeiket a teszt által lefedett tartalmi területen”. Az elvárás-érték elmélet alapján az erőfeszítés az elvárás és az érték megnyilvánulása, ezért összefügg a teszten nyújtott teljesítménnyel. Összefoglalva, a tesztmegoldási motiváció magában foglalja a vizsgázók erőfeszítéseit, amely a sikerre vonatkozó elvárásaik és a teszt értékének függvénye (Penk és Richter, 2017). A tesztmegoldási motivációval foglalkozó kutatások legnagyobb része nem vizsgálta a tesztmegoldási motiváció összes elemét, azok közül tipikusan csak néhányat (Penk és Richter, 2017), főképp a tesztmegoldó erőfeszítésének feltérképezésére fókuszáltak (ld. alább).

### A tesztmegoldási motivációt befolyásoló tényezők

Az eddigi kutatások (Asseburg és Frey, 2013; Goldhammer és mtsai, 2016; Lindner és mtsai, 2016; Nuutila és mtsai, 2021; Wise és mtsai, 2010) számos olyan tényezőt azonosítottak, amelyek a tesztmegoldási motivációt befolyásolják: ilyenek maguk a feladatok, a tanulmányi terület, a tesztmegoldó személye és a teszthelyzet. Ezen tényezők tulajdonosságainak változtatásával jelentős mértékben befolyásolható a tesztmegoldók tesztmegoldási motivációja és az azzal kapcsolatos erőfeszítések.

#### *A feladat jellemzői*

Korábbi kutatási eredmények alapján (Bergstrom és mtsai, 1994; Lindner és mtsai, 2017) az itemek nehézségének növekedésével csökken a tesztmegoldók motivációs szintje. Véleményünk szerint pontosabb az a megközelítés, miszerint a tesztmegoldók nagyobb erőfeszítést tesznek a képességeikhez illeszkedő, azaz a képességeiknek megfelelő, se nem túl nehéz, se nem túl könnyű feladatok megoldása során (Asseburg és Frey, 2013), mint más esetben (ezen alapul az adaptív tesztelés által biztosított optimális kihívás során tapasztalt flow-élmény; Csapó, Molnár és R. Tóth, 2008). A képeket vagy illusztrációkat

is tartalmazó feladatok esetén motiváltabbak a feladatmegoldók, mint a csak szöveget tartalmazó feladatok megoldása során. Ez a hatás akkor is mérhető, ha a feladat maga megoldható képek, ábrák nélkül is, azok kizárólag a feladat illusztrálását szolgálják (Lindner és mtsai, 2016; Lindner és mtsai, 2019).

### *Itempozíció*

Az itempozíciós hatás, azaz az itemek teszten belüli elhelyezkedése jelentősen befolyásolja a motivált és motiválatlan válaszok arányát. Számos kutatási eredmény szerint a tesztmegoldók tesztmegoldási erőfeszítései a teszt kezdetétől a végéig általában csökkenő tendenciát mutatnak (Attali, 2016; List és mtsai, 2017; Nuutila és mtsai, 2021; Penk és Richter, 2017; Wise és Gao, 2017; Wolgast és mtsai, 2020). Az itempozíciós hatás annyira jelentős, hogy a legjelentősebb nagymintás felmérések során előre meghatározott algoritmus szerint kezelik azt. Például a PISA 2006-os mérésben alkalmazott horgonyzási technika szerint (Molnár, 2013) minden egyes teszt négy résztesztből áll, amelyekben mindhárom mért terület helyet kap. Az előre kialakított 13 résztesztből 13 különböző tesztet alakítottak ki a kutatók úgy, hogy mind a 13 teszt egymással pontosan egy részteszt által legyen összekötöttésben, és mindegyik részteszt mind a négy teszten belüli pozícióban pontosan egyszer szerepeljen. Ennek oka, hogy a résztvevők fáradtsága miatt a tesztmegoldási erőfeszítések tipikusan kisebbek a később megoldandó itemeknél (Goldhammer és mtsai, 2016).

### *Személyi jellemzők*

Nagyobb valószínűséggel fordul elő nem motivált válasz alacsonyabb tanulási képességgel (Goldhammer és mtsai, 2016), alacsonyabb kognitív képességszinttel rendelkező (Goldhammer és mtsai, 2017) vagy alacsonyabb iskolai végzettségű személyek, és/vagy férfiak esetén (Goldhammer és mtsai, 2016; Wise és DeMars, 2010). DeMars és munkatársai (2013) megállapították, hogy a nemek közötti különbség nem egyenletes eloszlású. Több férfi van a motivációs skála alacsony, mint a magasabb végén, azaz a tesztmegoldók között több rendkívül alacsony erőfeszítéssel rendelkező férfit találtak, míg a nők motivációs szintje között nem volt ekkora különbség. A tesztet megoldók életkora szintén befolyásolja a motivált válaszok valószínűségét. Az OECD felnőttek olvasási képességére, matematikai műveltségére és problémamegoldó képességszintjének feltérképezésére fókuszáló felmérésorozatában (PIAAC, 2012) szignifikáns különbség volt a tesztelt személyek tesztmegoldási motivációjában. Tendenciaszerűen az idősebbek kevésbé voltak motiváltak, mint a fiatalok. A problémamegoldó képesség tesztelése során tapasztalták a legerőteljesebb összefüggést: minél magasabb volt a korosztály, annál nagyobb volt a motiválatlan válaszok aránya (Goldhammer és mtsai, 2016). Ugyanezen felmérés keretein belül végzett elemzések rávilágítottak arra is, hogy amennyiben a tesztelt személy anyanyelve és a teszt nyelve különbözött, magasabb volt a motiválatlan válaszok aránya. Szintén szignifikáns különbségek mutatkoztak az egyes országok között a kulturális háttér jellegétől függően (Goldhammer és mtsai, 2016; Goldhammer és mtsai, 2017).

### *Teszthelyzet, a tesztírás időpontja*

A tesztírás időpontjára és az időpont motivációra gyakorolt hatására vonatkozó elemzések eredményei sem erősítik meg egymást minden esetben. Wise és mtsai (2010) nem találtak különbséget a 3–9. osztályos tanulók átlagos tesztmegoldási motivációjában attól függően, hogy a tanév melyik szakaszában (ősz, illetve tavaszi félév) történt a tesztírás, ugyanakkor egy másik kutatásban a motiváció magasabb szintjét állapították

meg tavasszal, mint ősszel (Wise és mtsai, 2014). Nem volt különbség a motivációban attól függően, hogy a hét melyik napján oldották meg a tesztet, ugyanakkor csökkent a motivált válaszok aránya a nap előrehaladtával. Reggel motiváltabbak voltak a diákok, mint délután (Lindner és mtsai, 2019; Wise és mtsai, 2010). Ez utóbbi esetben hasonlóság mutatkozik az itempozíciós hatással, vagyis a később megoldott feladatok, tesztek esetében a motiváció csökkenő szintjét tapasztalhatjuk (List és mtsai, 2017).

### **A tesztmegoldási motiváció növelésének lehetőségei**

Az alacsony szintű tesztmegoldási motiváció kezelésére az egyik lehetőség a tesztek megoldása előtt adódik. Ennek lényege, hogy valamilyen módon megpróbáljuk növelni a tesztmegoldók motivációját, ezáltal várhatóan pontosabb képet kaphatunk valódi képességszintjükéről. A módszereket (Wise és DeMars, 2005) négy csoportra osztotta: (1) a teszt téjének növelése, (2) a teszt jellemzőinek módosítása, (3) visszajelzés adása, illetve (4) külső ösztönzők biztosítása.

#### *A teszt téjének növelése*

A tesztmegoldási motiváció növelésének talán legegyszerűbb és legkézenfekvőbb módja, ha növeljük a teszt téjét. Erre az egyik lehetőség, hogy a vizsgázó személyes érdekelttségét növeljük, például úgy, hogy a teszt eredménye beleszámít az osztályzatokba. Ennek a megoldásnak azonban több hátránya is van (Rios, 2021). A teszt személyes téjének növelése egyrészt fokozza a csalás esélyét, így több energiát kell fordítani ennek megelőzésére, másrészt növeli a tesztzorongást. Egy másik lehetőség a teszt téjének növelésére, ha nagyobb szerepet szánunk a vizsgafelügyelőknek. Az ő feladatuk ebben az esetben az, hogy részletes instrukciókat adjanak a tesztmegoldónak például arról, hogy a teszten nyújtott teljesítményük az intézmény és az oktatók megítélése szempontjából nagyon lényeges, így bírva őket motiváltabb feladatmegoldásra (Liu és mtsai, 2015; Baumert és Demmrich, 2001).

#### *A teszt jellemzőinek módosítása*

A tesztmegoldási motiváció növelésének másik, viszonylag egyszerű módja, ha magát a tesztet alakítjuk át oly módon, hogy a résztvevők motiváltabbak legyenek a megoldására. A korábbi kutatások magasabb szintű motivációt találtak azokban az esetekben, ha (1) a feladatok nem voltak túlságosan megterhelőek (DeMars, 2000; Wise, 2006), hanem inkább a résztvevők képességeihez illeszkedőek, annak megfelelő nehézségűek (Asseburg és Frey, 2013; Rios, 2021), ha (2) képeket, illusztrációkat tartalmaztak (Lindner és mtsai, 2016; Lindner és mtsai, 2019), valamint ha (3) érdekesek voltak számukra (Attali és Arieli-Attali, 2015).

A vonatkozó kutatási eredmények alapján összességében érdemes olyan teszteket kidolgozni és alkalmazni, amelyek (1) feladatleírása könnyen érthető, amelyek a (2) feleletalkotó kérdéstípusok mellett feleletválasztós kérdéseket is tartalmaznak (Rios, 2021). (3) Ahol a feladatok nehézsége illeszkedik a tesztelt személyek képességszintjéhez – ezt a számítógépes adaptív teszteszközrendszerével tudjuk biztosítani a tesztelt személyek számára. (4) A feladatokban szereplő képek, illusztrációk segítik, támogatják a feladat megértését, valamint (5) a résztvevők életkorának megfelelő, őket érdeklő, játékos feladatok szintén elősegítik a magasabb szintű tesztmegoldási motivációt. Végül (6) ehhez a megoldáshoz sorolható a tesztszituáció helyes megválasztása is: javasolt a délelőtti írás a délutáni helyett (Lindner és mtsai, 2019; Wise és mtsai, 2010).



### *Visszajelzés, visszajelentés biztosítása*

A résztvevők számára a teljesítményről való visszajelzés szintén motivációnövelő lehet (Baumert és Demmrich, 2001). Minél gyorsabb a visszacsatolás, várhatóan annál nagyobb a motivációs szint emelkedése (Hattie, 2009; Molnár és Csapó, 2019b). Számítógéppel segített oktatás – beleértve a tesztelés folyamatát is – esetén a kontextuális adatok folyamatos monitorozásával (pl. válaszadási idő, kattintások mennyisége) lehetőség van arra, hogy folyamatosan nyomon kövessék a diákok erőfeszítéseit (Molnár, 2021a), sőt figyelmeztessék őket, ha csökkent a motivációs szintjük, és az átgondolt válaszadás helyett elkezdenek gyorsan tippelgetni. Ezen technika alkalmazásával Wise és munkatársai egy kontrollcsoportos kutatás keretein belül magasabb átlagos pontszámot és magasabb motivációs értékeket tapasztaltak (Wise és mtsai, 2006).

### *Külső ösztönzők biztosítása*

A tesztmegoldók számára biztosított külső ösztönző nem csak anyagi ösztönző lehet (pl. pénzjutalom, vagy ajándékutalvány), hanem erkölcsi is (pl. elismerés kiváló teljesítmény esetén). A monetáris ösztönzők motivációra gyakorolt hatása nem egyértelmű. Egyes kutatások szerint a monetáris ösztönzők jelentősen befolyásolták a vizsgázók erőfeszítéseit és általános teljesítményét (Braun és mtsai, 2011), míg más esetben nem értek el erőfeszítésnövelő és azzal párhuzamosan teljesítményjavító hatást (Oneil és mtsai, 2005). A monetáris ösztönzők alkalmazásának hátrányai közé sorolható, hogy egyrészt költségesek, másrészt nem valószínű, hogy ugyanazt a motivációs hatást gyakorolják a különböző anyagi háttérrel rendelkező vizsgázókra (Lau és mtsai, 2009).

A szokásostól eltérő pontozási rendszer szintén képes lehet növelni a tesztmegoldási motivációt. Ide sorolható a hiányzó válaszok „szokásostól” eltérő kezelése, értékelése, vagy a hibás válaszok hagyományostól eltérő értékelése is. A módosítás valódi hatásának tesztelésére az adott kultúrában, iskolai kontextusban tudományos igényességgel kivitelezett kutatások szükségesek. Eltérő pontozási rendszert alkalmazott Verbić és Tomic (2009) kutatásában, amelynek során a diákok 0,25 pontot kaptak a megválaszolatlan itemekért a szokásos 0 pont helyett, függetlenül attól, hogy azok feleletválasztós vagy feleletalkotós itemek voltak. Ezzel a pontozási módszerrel azt érték el, hogy a tesztelt személyek jobban átgondolták a válaszokat, és így csökkentették a motiváltalan válasz esélyét 5-15%-ról 0,3%-ra (Verbić és Tomic, 2009).

Rios (2021) egy metaanalízis során 53 tanulmány adatait elemezve arra a következtetésre jutott, hogy a tesztmegoldási motiváció növelésére alkalmazott módszerek 0,13 szórásnyival növelték mind a tesztmegoldási erőfeszítést, mind a teszten elért eredményeket. A módszerek közül a legjelentősebb hatást az ösztönzők alkalmazása hozta, mintegy 12,5%-kal csökkentette a motiváltalan válaszok arányát.

### **A tesztmegoldási motiváció mérési lehetőségei**

A tesztmegoldási motiváció mérésére többféle módszer létezik: (1) önértékelő kérdőívek alkalmazása, (2) válaszdíó-alapú megközelítések, illetve a kettő ötvözéséből adódó (3) modell-alapú elemzések. Az önértékelő kérdőívek a vizsgázók saját megítélésén alapulnak, a válaszdíó-alapú megközelítés a számítógépes tesztek naplóadatait használja, míg a modell-alapú megközelítés mindkettőt felhasználhatja.

### Önértékelő kérdőívek

Az első és legszélesebb körben alkalmazott megközelítés a tesztmegoldási motiváció és az azzal erőteljes összefüggésben lévő tesztmegoldási erőfeszítés mérésére a Likert-skálát alkalmazó önértékelő kérdőívek használata. A legegyszerűbb esetben a teszt végén, a tesztfeladatok megoldása után teszik fel a tesztmegoldási motivációra vonatkozó kérdéseket a résztvevőknek. Ennek továbbfejlesztett változata, ha a teszt elején, illetve a tesztfeladatok között is válaszolni kell az aktuális motivációs szintre vonatkozó kérdésekre. Utóbbi kutatási elrendezés lehetőséget ad a teszt közbeni motivációs szint változásának nyomon követésére is. Az önértékelő kérdőívek nagy előnye, hogy használhatunk egyszerű, hagyományos, papíralapú tesztlés során is megvalósítható, feldolgozásukhoz nincs szükség kifinomult statisztikai ismeretekre. Fontos követelmény az önértékelő kérdőívekkel szemben, hogy ne legyenek túl hosszúak, mivel az önmagában is csökkentheti a válaszadás motivációját (Schüttpelz-Brauns és mtsai, 2018). A legszélesebb körben használt önértékelő kérdőívek áttekintését az 1. táblázat tartalmazza.

Az önértékelő kérdőívek előnye, hogy azokkal lehetőség van a motiváció minden összetevőjének mérésére. Ugyanakkor, ahogy az 1. táblázatból is olvasható, a legtöbb önértékelő kérdőív nem kérdez rá a motiváció minden elemére. Az önértékelő kérdőívek közül a legegyszerűbb a PISA felmérés során is használt *Effort thermometer*, amely kizárólag a tesztmegoldási erőfeszítést méri három kérdésre adott válasz alapján. A skála másik végén az Online Motivációs Kérdőív helyezkedik el, amely hét alskálát tartalmaz, melyeket 32 kérdéssel mér.

1. táblázat. Néhány széles körben használt önértékelő kérdőív

Mérőeszköz	Itemek száma	Skála	Alskála	Mért jellemzők	Hivatkozás
Online Motivációs Kérdőív ( <i>Online Motivation Questionnaire</i> , OMQ)	32	4-pontos Likert skála	Hangulat Önhatékonyság Várható siker Feladat vonzereje Észlelt hasznosság Tesztszorongás Erőfeszítés	Tesztmegoldási erőfeszítés, a teszt előtt és után	(Crombach és mtsai, 2003)
Kérdőív a jelenlegi motivációról ( <i>Questionnaire of Current Motivation</i> , QCM)	18	7-pontos Likert skála	A siker valószínűsége A teszt kihívás jellege Érdeklődés Kudarctól való félelem	Tesztmegoldási motiváció	(Freund és mtsai, 2011)
Erőfeszítés hőmérő ( <i>Effort-thermometer</i> )	3	10-pontos Likert skála	Nincs	Tesztmegoldási erőfeszítés összehasonlítva a személyes tapasztalattal, amikor a vizsgázó korábban a legtöbb erőfeszítést alkalmazta	(Butler és Adams, 2007)



Mérőeszköz	Itemek száma	Skála	Alskála	Mért jellemzők	Hivatkozás
Tesztmegoldási erőfeszítés rövid skála ( <i>Test-taking Effort Short Scale</i> , TESS)	3	5-pontos Likert skála	Elérési érték Belső értéke Hasznossági értéke	Tesztmegoldási erőfeszítés rövid formája	(Schüttpelz-Brauns és mtsai, 2018)
Hallgatói vélemény skála ( <i>Student Opinion Scale</i> , SOS)	10	5-pontos Likert skála	Fontosság Erőfeszítés	Motiváció a teszt után	(Thelk és mtsai, 2009)

Az önértékelő kérdőíveknek számos korlátja van. Alkalmazásuk során általában nem tudjuk, hogy mennyire volt őszinte a válasz, ezért könnyen alul- vagy felülbecsülik a tesztelt személy valódi erőfeszítését, ami torzításhoz vezethet. Nem reális az, hogy minden egyes feladat megoldása után kitöltetünk egy motivációs kérdőívet a résztvevőkkel, mert annyi kérdésre senki sem adna őszinte, motivált választ. Ennek következtében a leggyakoribb és legelterjedtebb alkalmazás a tesztelt személyek általános motivációs szintjének mérése, a pontos részletek, a motivációs szint feladatonkénti, itemenkénti változása rejtve marad (Wise és Gao, 2017). Az önértékelő kérdőívek alkalmazásának további jelentős korlátja, hogy azok a személyek, akik nem motiváltak a teszt megoldására, várhatóan nem lesznek motiváltak az önértékelő kérdőív kitöltésére sem. Ez további torzításhoz vezethet (Wise és Kong, 2005). A kérdőív kitöltését befolyásolhatják a társadalmi elvárások, azaz előfordulhat, hogy a tesztelt személyek nem a valóságnak megfelelő válaszokat adnak, hanem olyanokat, amiket elvárnak tőlük (Goldhammer és mtsai, 2016).

#### *Válaszidő-alapú megközelítés*

Az informatikai eszközök jelentős fejlődésével és széles körben való elterjedésével lehetővé vált a technológia oktatásban való alkalmazására. Ezzel párhuzamosan jelentős ütemben fejlődött az oktatási adatbányászat, az oktatással kapcsolatos tevékenységek során gyűjtött logadatok, metaadatok elemzése (Alrababah és Molnár, 2021; Sahin és Yurduguül, 2020). A számítógép-alapú tesztelés hatalmas előnye a hagyományos papíralapú,

*Az önértékelő kérdőíveknek számos korlátja van. Alkalmazásuk során általában nem tudjuk, hogy mennyire volt őszinte a válasz, ezért könnyen alul- vagy felülbecsülik a tesztelt személy valódi erőfeszítését, ami torzításhoz vezethet. Nem reális az, hogy minden egyes feladat megoldása után kitöltetünk egy motivációs kérdőívet a résztvevőkkel, mert annyi kérdésre senki sem adna őszinte, motivált választ. Ennek következtében a leggyakoribb és legelterjedtebb alkalmazás a tesztelt személyek általános motivációs szintjének mérése, a pontos részletek, a motivációs szint feladatonkénti, itemenkénti változása rejtve marad (Wise és Gao, 2017).*

szemtől-szembeni értékelési rendszerekkel szemben, hogy a tesztek megoldása során számos olyan kontextuális adatot tudunk rögzíteni (feladattal eltöltött idő, előre- és hátraugrások, szemmozgás, minden egyes klikkelés stb.), amelyek korábban elképzelhetetlenek voltak, és ezek elemzése mélyebb összefüggésekre világíthat rá (Tóth és mtsai, 2017).

A válaszügy az az időtartam, amit a tesztelt személy az adott feladat kiközvetítésétől annak megoldásának befejezéséig (a tovább gombra kattintásig) az adott feladat megoldásával tölt. Schnipke (1995) kutatási eredménye jelezte először, hogy van összefüggés a válaszügy és a tesztmegoldási erőfeszítés között. A magas tétellel és időkorláttal rendelkező tesztekre adott válaszok vizsgálata során arra a következtetésre jutott, hogy amint a teszt megoldása során telik az idő, és a vizsgázóknak egyre kevesebb ideje marad a teszt befejezésére, a résztvevők nagyon gyors válaszügyeket produkálnak, amit gyors tippelési viselkedésnek nevezett el, szemben a megoldásra törekvő viselkedéssel. Gyors, tippelgető viselkedés az alacsony tétellel rendelkező teszteknel is előfordul, de mivel ebben az esetben nincs időkorlát, ez a fajta viselkedési mintázat a motiválatlanságot jelzi (Wise és Kong, 2005). A válaszügy-alapú módszerek mind azon a feltételezésen alapulnak, hogy az alacsony tesztmegoldási erőfeszítéssel rendelkező résztvevők kevesebb időt töltenek a feladat megoldásával, ezért gyorsabban válaszolnak a kérdésekre, mint a magasabb szinten motiváltak (Wise és Kong, 2005). A gyors találgatás negatívan torzítja a teszt eredményeit, mert alacsonyabb helyes válaszadási arányhoz vezet, mint a megoldásra törekvő viselkedés (Wise és Ma, 2012).

#### A válaszügy-alapú megközelítés indikátorai

A válaszügy alapú megközelítések esetén a legtöbb esetben a mért indikátor az itemhez tartozó válaszügy (Wise és Gao, 2017), amit néhány kutatás kiegészít a kattintások számának elemzésével. Az általános megközelítés szerint meg kell határozni egy idő küszöbértéket, és amennyiben az adott válaszügy ennél a küszöbértéknél rövidebb, akkor azt motiválatlan, míg ha hosszabb, akkor motivált válasznak tekintjük. A küszöbérték pontos meghatározása amiatt fontos, mert ha a küszöbérték hosszabb, mint a valós, akkor a motivált és gyors vizsgázók választát motiválatlanként azonosítjuk (fals negatív), míg ha a küszöbérték rövidebb, mint a valós, a lassú, motiválatlan választ azonosítjuk motiváltként (fals pozitív) (Wise és Ma, 2012). Mindkét hibát nem lehet egyszerre csökkenteni, ha az egyik nő, a másik csökken, és fordítva (Sahin és Colvin, 2020). Wise és Gao (2017) a következő eseteket definiálta nem motivált viselkedésként:

1. gyors találgatás: feleletválasztós vagy más típusú zárt végű itemek esetében a vizsgázó gyorsabban válaszol, mint a küszöbérték,
2. gyors kihagyás: a vizsgázó az adott feladatra nem ad választ, és a küszöbértéknél gyorsabban továbblép a következő feladatra,
3. gyors felületes válasz: nyílt végű itemek esetében a vizsgázó rövidebb ideig foglalkozik a feladattal, mint az item küszöbértéke, és nagyon rövid, elnagyolt, nem releváns választ ad.

Az  $i$  itemhez és  $j$  vizsgázóhoz tartozó motivált, vagy más néven megoldási viselkedés (*solution behavior*,  $SB_{ij}$ ) mérésére Wise és Kong (2005) az alábbi képletet vezette be:

$$SB_{ij} = \begin{cases} 1, & \text{ha } RT_{ij} \geq T_i \\ 0, & \text{ha } RT_{ij} < T_i \end{cases}$$

ahol

$T_i$  = az  $i$  itemhez tartozó küszöbérték, amely különbséget tesz a motivált és a motiválatlan viselkedés között,

$RT_{ij}$  = az  $i$  itemhez és  $j$  vizsgázóhoz tartozó válaszügy.

A tesztmegoldási erőfeszítés aggregált szinten is jellemezhető, egyrészt a személyre, másrészt az adott itemre vonatkozóan (Lindner és mtsai, 2019).

Wise és Kong (2005) vezette be a válaszüidő-erőfeszítés kifejezést (*response time effort*, RTE). Az RTE az adott résztvevőre vonatkoztatva a motivált válaszok összegének és a teszt összes itemének aránya, vagyis a befektetett erőfeszítés teljes szintjét mutatja meg az adott résztvevőre vonatkoztatva.

Az RTE a  $j$  vizsgázóra vonatkoztatva

$$RTE_j = \frac{\sum SB_{ij}}{k}$$

ahol  $k$  = az itemek száma.

Az itemspecifikus megközelítést válaszüidőhűségnek nevezzük (*response time fidelity*; RTF; Wise, 2006). Az RTF az adott itemre vonatkozóan a motivált válaszok összegének az összes vizsgázóhoz viszonyított aránya, azaz egy adott  $i$  item esetén

$$RTF_i = \frac{\sum SB_{ij}}{n}$$

ahol  $n$  = a vizsgázók száma.

Mind az RTE, mind az RTF index 0 és 1 közötti értékeket vehet fel.

Feltételes válaszüidőhűség (*conditional RTF*): a résztvevőket teljesítményük alapján csoportokba soroljuk, majd csoportonkénti bontásban határozzuk meg RTF értéket. Ezáltal elemezhető, hogy van-e összefüggés a résztvevők teljesítményszintje és az RTF értéke között (Lee és Jia, 2014).

A válaszüidő mérésének számos előnye van az önértékelő kérdőívvel szemben (Wise és Kong, 2005). A motiváció beavatkozás nélkül mérhető, így egyrészt nem terheljük többletfeladattal a vizsgázót, másrészt torzítatlan információhoz jutunk. A mérés nem a válaszadó ítéletein, hanem mérésen alapul, így nem lesz elfogult. A motiváció változása sokkal pontosabban nyomon követhető, mert a válaszüidő adatai minden egyes item esetében rendelkezésre állnak, nem csak bizonyos időpillanatokban. Végül, ezen módszereknek magas a belső konzisztenciája (Wise és Ma, 2012).

A válaszüidő mérésén alapuló módszereknek azonban vannak korlátai is. Egyrészt a nagyon jó képességekkel rendelkező tesztmegoldók, akik nagyon gyorsan képesek azonosítani a feladathoz tartozó kulcsszavakat, és pillanatok alatt reagálnak, nem különböztethetők meg azoktól, akik motiválatlanok (Schüttpelz-Brauns és mtsai, 2018), másrészt az alacsony képességszintű vizsgázók, akik nagyon lassan olvasnak, és ezért lassabban adnak motiválatlan választ, nem különböztethetők meg a motivált válaszadóktól (Akyol és mtsai, 2021).

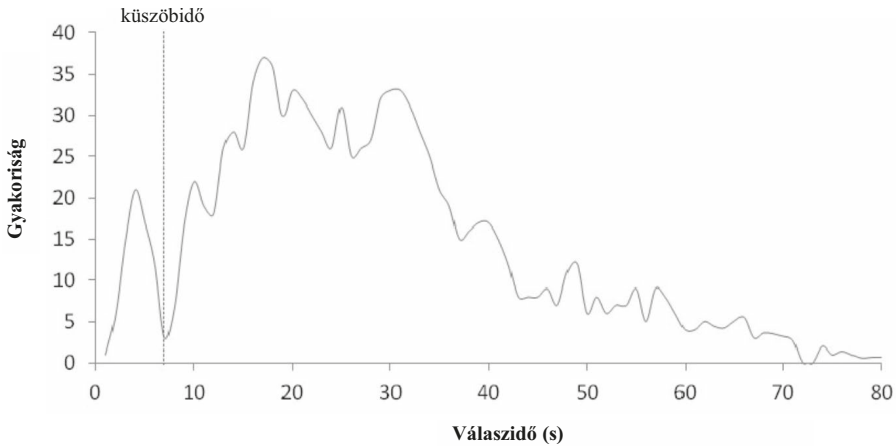
#### A küszöbérték meghatározása

A tesztfeladatokon való sikeres teljesítés első feltétele a feladathoz tartozó instrukciók elolvasása/meghallgatása, illetve az olvasottak/hallottak értelmezése. Amikor a tesztelt személyek olyan rövid időn belül válaszolnak, ami nem teszi lehetővé a teljes szöveg értő elolvasását, illetve hangalapú instrukció esetén a hang teljes meghallgatását, majd a feladat megoldását, motiválatlan magatartásról beszélünk. Ellenkező esetben a detektált viselkedésmintázatot a motivált viselkedésük csoportjába sorolják. A küszöbérték minél pontosabb meghatározása a fals pozitív és fals negatív osztályozás elkerülése érdekében kiemelten fontos (Lindner és mtsai, 2019).

A küszöbérték alkalmazásának leegyszerűsített módja, ha egy előre meghatározott, úgynevezett konstans küszöbértéket állapítunk meg, majd minden item/feladat esetén ezt a küszöbértéket alkalmazzuk. A vonatkozó kutatásokban az első ilyen, általánosan használt küszöbérték a 3 másodperces küszöbszint volt (*Common3*). A három másodperces küszöbérték azonban túl rövidnek bizonyult a hosszabb instrukciót tartalmazó, ezért több olvasást igénylő feladatoknál, ezért alkalmazhatósága erősen korlátozott (Hauser és Kingsbury, 2009). Később bevezették az 5 másodperces küszöbértéket (*Common5*) (Wise és Kong, 2005), amit többek között a PIAAC felmérés során a hiányzó válaszok kódolásához is alkalmaztak, attól függően, hogy az eltelt idő kevesebb volt-e a küszöbértéknél, és csak 0-2 kattintást végzett-e a személy („nem ért el a feladathoz/nem próbálta meg”), vagy meghaladta a küszöbszintet, és/vagy több, mint 0-2 kattintást végzett („nincs válasz”) (OECD, 2019).

A küszöbérték meghatározásának szofisztikáltabb módja az itemspecifikus küszöbértékek alkalmazása. Az itemspecifikus küszöbértékek elmélete azon a feltételezésen alapul, hogy a válaszadáshoz szükséges minimális idő itemenként eltérő. Például amíg a vizsgázók képesek gyorsan elolvasni és értelmezni egy egyszerű számolási feladatot, addig egy hosszú és nehéz problémamegoldó feladat elolvasása és értelmezése sokkal több időt vesz igénybe (Goldhammer és mtsai, 2016). Ez alapján nem egy közös küszöbértéket határoznak meg az összes itemhez, hanem itemenként mást és mást. Az itemspecifikus küszöbértékek meghatározására számos koncepció létezik. A leggyakrabban alkalmazott módszereket az alábbiakban mutatjuk be.

1. Vizuális ellenőrzés (*visual inspection*; VI) módszere: amennyiben ábrázoljuk a válaszütemek eloszlását, megállapítható, hogy amíg a motivált viselkedésnek unimodális eloszlása van, addig a gyors találgatások jelenléte az első néhány másodpercben egy kezdeti frekvenciacsúcs megjelenését eredményezi, így az eloszlás bimodálissá válik (Schnipke és Scrams, 1997). Ezen minta alapján beazonosíthatók a motiválatlan válaszok, mivel a bimodális válaszütem-eloszlásban határozott „rés” van, amint az a 2. ábrán látható (Lee és Chen, 2011). Vannak azonban olyan minták, ahol az itemek jelentős részének unimodális eloszlása van a bimodális helyett, függetlenül a motiválatlan válaszok arányától, így ez a módszer nem minden esetben teszi lehetővé a motiválatlan válaszok azonosítását (Lee és Jia, 2014).
2. A válaszütem-pontossági módszer a feleletválasztós itemekre alkalmazható. Működésének lényege, hogy minden egyes item esetén másodpercenként kiszámítjuk a helyes válaszok arányát (pl. 3 másodpercnél 16%, 4 másodpercnél 21%). Az itemhez tartozó küszöbérték az az időpillanat, amikor a helyes válaszok aránya meghaladja a véletlen találgatásból származó arányt (pl. 0,25-öt, ha négy válaszlehetőség van, 0,2-et, ha öt válaszlehetőség van) (Lee és Jia, 2014). A módszer hátránya az, hogy egy valós tesztszituációban a feladatmegoldók nagyobb valószínűséggel választják a középső lehetőségek egyikét, mint az első vagy az utolsó válaszlehetőséget. Ennek oka az a pszichológiai jelenség, amely megnehezíti az emberek teljesen véletlenszerű viselkedését, így az itemek véletlenszerű kitalálásának pontossága a helyes válasz helyétől is függ (Wise, 2006). Például egy 4 válaszlehetőséggel rendelkező matematikai item megoldása során a véletlenszerű találgatások aránya az „a” válaszopció esetében 0,17, míg a „b” válaszopció esetében 0,24 volt (Wise és Ma, 2012).



2. ábra. Hipotetikus válaszidő-eloszlás egy adott itemre vonatkozóan (Lee és Chen, 2011)

3. Normatív küszöb módszer (*normative threshold*; NT10) (Wise és Ma, 2012). Az NT10 által meghatározott küszöbérték egyenlő a vizsgázók által az itemre fordított átlagos idő 10%-ával, de maximálva van 10 másodpercben. Például, ha egy adott item esetén az átlagos feladattal töltött idő 28 s, akkor az itemhez tartozó küszöbérték 2,8 s, viszont ha az átlagos feladattal töltött idő 150 s, akkor a küszöbérték nem 15 s lesz, hanem 10 s. A fenti szabály alapján lehet meghatározni az NT15 és az NT20 küszöbértéket is.
4. Szabály-alapú küszöbidő: az alkalmazott küszöbidő az item elolvasásához szükséges időtől, azaz az item hosszától függ (karakterben) (Wise és Kong, 2005).

$$T_i = \begin{cases} 3 \text{ s, ha az } i \text{ item rövidebb, mint 200 karakter} \\ 5 \text{ s, ha az item is legalább 200, de legfeljebb 1000 karakter} \\ 10 \text{ s, ha az } i \text{ item legalább 1000 karakter} \end{cases}$$

5. A helyes válaszarány nagyobb, mint 0 (*proportion correct greater than zero*;  $P > 0\%$ ) módszer: feleletalkotó itemek esetén használható, ahol a tesztelt személyeknek meg kell adniuk saját válaszaikat, így a helyes válasz elérésének véletlenszerű esélye nulla. A  $P > 0\%$  küszöb meghatározásához az adott válaszidőhöz tartozó helyes válaszok arányát egy másodperces időközönként kell kiszámítani. A küszöb az a legrövidebb válaszidő, amelynél a helyes válaszok aránya nulla felett van (Goldhammer és mtsai, 2016). A PIAAC mintán alapuló kutatás során Goldhammer és munkatársai (2016) négy módszert hasonlítottak össze (Common3, Common5, VI,  $P > 0\%$ ), és arra a következtetésre jutottak, hogy az alkalmazott validitási vizsgálatok alapján a  $P > 0\%$  módszer volt a legpontosabb.
6. Továbbfejlesztett módszerek (*enhanced methods*): ezek a módszerek a válaszidőn alapulnak, de ki vannak egészítve a válaszviselkedéssel, azaz egyrészt a válasz viselkedésének típusával (pl. kattintások, billentyűleütések, szimuláció futtatása), másrészt a válaszviselkedések összes számával (minden egyes kattintás és billentyűlenyomás összege). A módszerek azon a feltételezésen alapulnak, hogy nem csak a feladattal töltött idő, hanem a válaszreakciók száma is a motiváció mértékéhez köthető, így amennyiben kevesebb válaszreakciók mérhető, az motiválatlan viselkedést jelez. Ezek alapján a válaszviselkedés segíthet pontosabban megkülönböztetni a motivált és a motiválatlan vizsgázókat. A továbbfejlesztett módszerek

bizonyos itemek esetén pontosabb eredményt adnak, mint a kizárólag válaszüdő alapú módszerek, de nem minden item esetén (Sahin és Colvin, 2020).

### Validációs kritériumok

Miután megtörtént a válaszok motiváltak vagy motiválatlanak minősítése, validációs ellenőrzéseket kell végrehajtani, amelyek segítenek eldönteni, hogy az alkalmazott módszer valóban megfelelő-e. A leggyakrabban alkalmazott validációs kritériumokat az alábbiakban foglaltuk össze:

1. A motivált válaszok helyes arányának a véletlen szintnél sokkal nagyobbaknak kell lenniük, míg a motiválatlan válaszok helyes arányának közel kell lennie a véletlen szintjéhez (Lee és Jia, 2014).
2. A motiválatlan válaszok sokkal alacsonyabb arányban helyesek, mint azok, amelyek motiváltként azonosítottak (Wise és Gao, 2017).
3. A küszöbértékek, az egyes itemek olvasási ideje, illetve azok összetettsége között legyen pozitív és szignifikáns kapcsolat. Ennek egyik ellenőrzési módja, hogy összehasonlítjuk a tesztelt személyek által az egyes itemekre fordított idő átlagát vagy annak móduszát a küszöbértékekkel.
4. A teszten nyújtott összteljesítmény alapján különböző átlagteljesítményt mutató csoportokat képzünk, majd elemezzük az így kapott csoportok, illetve a motivált és a nem motivált viselkedés közötti kapcsolatokat (Goldhammer és mtsai, 2016; Lee és Jia, 2014).

### Motivációs szűrés

A nem motivált válaszok kezelésére széles körben alkalmazott módszer a motivációs szűrés, amelynek során vagy a nem motivált válaszokat, vagy a nem motivált tesztmegoldók összes adatát törlik, így csak a motivált adatok maradnak a mintában, csak ezeket veszik figyelembe (Liu és mtsai, 2015). Több kutatás megállapította, hogy a teszt érvényessége és átlagos pontszáma is javul a motivációs szűrés alkalmazásával (Wise és Kong, 2005; Wise és DeMars, 2010).

Válaszszintű szűrés esetén a tesztmegoldók nem motivált válaszait törlik a mintából. Rios és munkatársai (2017) szerint ez a módszer robusztus azon motiválatlan válaszok esetén, amelyek összefüggnek a tesztelt személy képességszintjével. A módszer alkalmazásának azonban van korlátja is. Az a feltevés, hogy minden motiváltként megjelölt válasz felhasználható a feladatmegoldók képességének pontos becslésére, nehezen bizonyítható, hiszen csak válaszüdők állnak rendelkezésre, egyéb adatok, amelyek ezt jobban alá tudnák támasztani, nem (pl. szemmozgás követése, EEG).

Vizsgálószintű szűrés esetén a nem motiváltak minősített feladatmegoldók összes adatát törlik. Az adott személy akkor minősíthető motiválatlanak, ha a motiválatlan válaszainak száma meghalad egy előre meghatározott határértéket (általában 10, 15 vagy 20%-ot). Ez a kritérium a tesztelés során elfogadható motiválatlan viselkedés összarányát határozza meg, azonban nem minden esetben elég pontos. A tesztmegoldók hajlamosak olyan típusú részben motivált, részben nem motivált viselkedést tanúsítani, amelyet ez a kritérium nem képes detektálni. Például abban az esetben, ha 15%-nál húzzuk meg a motiválatlan válaszok határát, és a tesztelt személy motivált válaszokat ad egy 60 itemes teszt első 52 iteménél, majd az utolsó nyolc item esetén motiválatlan válaszokat ad, akkor a motiválatlan viselkedést ez a kritérium nem jelzi. Emiatt érzékenyebb mutatókat érdemes bevezetni, például az itemek gördülő részhalmazainak (*rolling subsets*) vizsgálatát. Egy 60 itemes teszt 10 itemes gördülő részhalmaza az 1-től 10-ig, majd 2-től 11-ig terjedő részhalmaz, egészen az 51-től 60-ig terjedőig. Általában egy  $n$  elemű teszt  $m$  elemű



részhalmazainak száma  $(n - m) + 1$  (Wise és mtsai, 2014). Ez alapján további kritériumok határozhatók meg gördülő részhalmazokkal, pl. motiválatlannak minősítjük a tesztmegoldót, ha legalább 3 motiválatlan válasz van a 10 itemes gördülő részhalmazok 20%-ában (Hauser és Kingsbury, 2009).

A személyszintű szűrés alkalmazása azon a feltételezésen alapul, hogy a tesztet megoldók erőfeszítései nincsenek összefüggésben a képességekkel, azaz ugyanannyira motiváltak a teszt megoldása során a legjobb és a legrosszabb képességű személyek. Ha ez a feltételezés nem igaz, és az alacsonyabb képességszintű személyek kevésbé motiváltak, akkor többnyire az ő adataikat törlik, ami torzításhoz vezet (túlzott átlagos pontszámnövekedés) (Wise és DeMars, 2010).

Rios és munkatársai (2017) a valószínűségi tesztelmélet eszközrendszerével modellezték a válaszsintű, illetve a személyszintű szűrés hatását. Abban az esetben, ha a képesség nincs összefüggésben a motiválatlan válaszok arányával, a személyszintű szűrés pontos eredményeket ad, azonban ha van összefüggés a kettő között, akkor a valós mértéknél jobban megemeli a teszteredmények átlagát. A modellezett körülmények között a válaszsintű szűrés alkalmazásával pontosabb, validabb eredményeket kaptak.

Lee és Jia (2014) egy öt lépésből álló folyamatot javasolt a válaszdő-alapú módszerek alkalmazása esetén. A folyamat során (1) feltáró elemzést kell végezni a válaszdő tulajdonságainak vizsgálatához, (2) küszöbértékek beállítása, a megfelelő módszer kiválasztása, (3) legalább három érvényességi ellenőrzés elvégzése a küszöbértékek igazolásához, (4) a tesztelési magatartás meghatározása: személyspecifikus viselkedés (RTE), elemspecifikus viselkedés (RTF), és itemenként csoportosított (feltételes RTF) és (5) motivációs szűrés, azaz a motiválatlan válaszok/válaszadók törlése az adatbázisból.

#### *Modell-alapú megközelítés*

A modell-alapú megközelítés azon a feltételezésen alapul, hogy a motivált tesztmegoldók válaszainak mintázata összefüggést mutat az itemek nehézségével. A tesztelt személyek válaszmintázatát egy elméleti modellel hasonlítják össze, és abban az esetben, ha rossz az illeszkedés, az a normálistól eltérő viselkedésre utal, például a motiválatlan válaszadásra. A megközelítés alapja a valószínűségi tesztelmélet (*item response theory*) eszközrendszerére. A logisztikus modellek alapfeltételezése, hogy az alacsony képességű személyek alacsonyabb valószínűség mellett oldják meg ugyanazt a feladatot helyesen, mint a

*A modell-alapú megközelítés azon a feltételezésen alapul, hogy a motivált tesztmegoldók válaszainak mintázata összefüggést mutat az itemek nehézségével. A tesztelt személyek válaszmintázatát egy elméleti modellel hasonlítják össze, és abban az esetben, ha rossz az illeszkedés, az a normálistól eltérő viselkedésre utal, például a motiválatlan válaszadásra. A megközelítés alapja a valószínűségi tesztelmélet (item response theory) eszközrendszerére. A logisztikus modellek alapfeltételezése, hogy az alacsony képességű személyek alacsonyabb valószínűség mellett oldják meg ugyanazt a feladatot helyesen, mint a magasabb képességszintű személyek. Minél magasabb a tesztelt személy képességszintje, annál nagyobb valószínűséggel ad helyes választ a feltett kérdésre.*

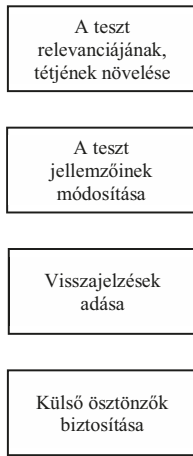
magasabb képességszintű személyek. Minél magasabb a tesztelt személy képességszintje, annál nagyobb valószínűséggel ad helyes választ a feltett kérdésre. A valószínűségi tesztelmélet modelljei közül a legegyszerűbb az egyparaméteres logisztikus modell (1PL; más néven Rasch-modell), amelyben a személyparaméteren kívül mindössze egyetlenegy paraméter: az itemnehézségi mutató szerepel. Ennek következtében a helyes válasz valószínűsége kizárólag a személy képességszintje és az item nehézségi indexének függvényeként adódik (Molnár, 2013). A kétparaméteres logisztikus modell (2PL) esetén ezekhez hozzáadódik az item diszkriminációs paramétere, amely azt mutatja meg, hogy az adott item mennyire erősen tudja elkülöníteni a tesztelt személyeket egymástól, azaz milyen meredek az itemkarakterisztikus görbe (Molnár, 2013). A három paraméteres logisztikus modell (3PL) kiegészül még egy találgatási paraméterrel is, miután például egy négy válaszlehetőséget tartalmazó feleletválasztós item esetén a helyes válasz valószínűsége véletlenszerű tippelés esetén is 0,25 (Molnár, 2003).

A modell-alapú megközelítés legfontosabb előnye, hogy nem a feladatmegoldók önértékelésén, hanem a teszten mutatott teljesítményük megfigyelésén alapul, emiatt kisebb lehet a torzítás mértéke. A megközelítésnek azonban vannak korlátai is. Egyrészt a normálistól eltérő mintázatot nemcsak a motiválatlan válaszadás okozhatja, hanem egyéb tényezők, pl. csalás vagy szerencsés tippelgetés is. A másik fontos korlátja, hogy az önértékelő kérdőívekhez hasonlóan nem képes itemenként jellemezni a motiváció szintjét, hanem csak globális képet ad, így kevésbé gyakori módszer a tesztmegoldási motiváció értékeléséhez (Wise és Smith, 2016).

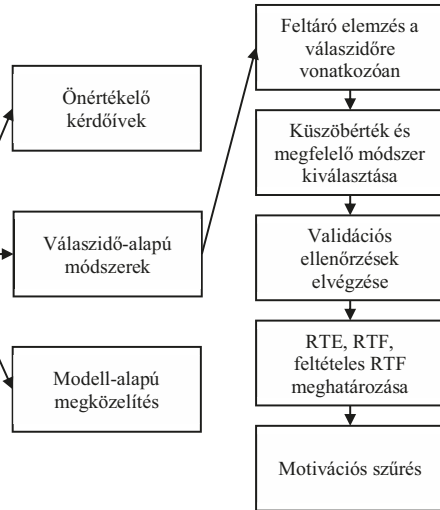
### **A tesztmegoldási motiválatlanság kezelési lehetőségei**

Az alacsony tétellel rendelkező tesztek esetén a tesztmegoldók részéről jelentkező nem teljes mértékű motiváltság jelentős hatással lehet a teszten elért eredményekre. Az ezen hatások csökkentésére szolgáló módszerek alapvetően kétfelől közelítik meg a problémát. Az első megközelítés szerint valamilyen módon növelni kell a tesztmegoldók motivációját, megszüntetve ezzel magát a motiválatlanságot és az ebből adódó torzítást. A feladatmegoldók motivációja növelhető a teszt relevanciájának növelésével, a teszt jellemzőinek módosításával, az eredményekről való visszajelzések adásával, valamint külső ösztönzők biztosításával. A második megközelítés szerint a teszt megoldása után meg kell határozni, hogy a válaszok közül melyek voltak a motiválatlanok, majd a motiválatlan válaszokat/válaszadókat el kell távolítani az elemzés alapját képező adatbázisból/ mintából, így csak a motivált válaszok/válaszadók eredményei képezik az értékelés alapját. A válaszok motivált/motiválatlan voltának itemszintű meghatározása a válaszdíő-alapú módszerek esetén lehetséges, az önértékelő kérdőívek és a modell-alapú megközelítés nem adnak erre lehetőséget. A válaszdíő-alapú módszerek esetén a feltáró elemzés után meg kell határozni az egyes itemekhez a küszöbértéket, validációs ellenőrzéseket kell végezni, meg kell határozni az RTE, RTF és feltételes RTF értékét, majd el kell távolítani a mintából a motiválatlan válaszokat/válaszadókat. A kétféle megközelítés lehetőségei, valamint a válaszdíő-alapú módszerek esetén használt folyamat a 3. ábrán látható.

### Tesztelés előtti módszerek



### Tesztelés utáni folyamat



3. ábra. A tesztmegoldási motiváció kezelésének lehetőségei (saját szerkesztés)

## Összegzés

Az alacsony téttel rendelkező tesztek eredményeinek nincsenek komoly következményei a tanulók tanulmányi eredményeire, későbbi életére, így az ezeken a teszteken elért eredmények nem minden esetben tükrözik teljes mértékben a tanulók képességeit, mivel előfordul, hogy a tanulók nem tesznek meg minden erőfeszítést a tesztek megoldása során. Az alacsony szintű motiváció tesztek torzító hatását kétféle módon lehet kezelni.

Az első lehetőség már a teszt előkészítése, összeállítása során adódik, ezek célja, hogy a résztvevők motivációs szintjét növeljék. A teszt jellemzőinek megfelelő megválasztása, a résztvevők számára részletes, pontos instrukciók adása, az elért eredményekről visszajelzések nyújtása rövid időn belül, valamint külső ösztönzők biztosítása mind-mind képesek növelni a motivációs szintet, ezáltal a valósághoz jobban közelítő eredményeket biztosítva a tesztelt személyek tudásáról.

A másik lehetőség a számítógépen kiközvetített tesztek megoldása után a logadatok, elsősorban a válaszidő elemzése. Ezek a módszerek azon a koncepción alapulnak, hogy a sikeres feladatmegoldáshoz el kell olvasni/meg kell hallgatni a feladatot, értelmezni kell azt, majd meg kell oldani. Minden egyes feladathoz meghatározható egy adott küszöb-idejű, amely alatt motiválatlannak, afölött pedig motiváltak minősítjük a válaszokat.

A feldolgozott tanulmányok alapján a következő következtetések vonhatók le: (1) az alacsony téttel bíró tesztek alkalmazása esetén jelentős lehet azon személyek aránya, akik nem kellőképpen motiváltak a tesztek megoldására, (2) a motiváltabb személyek teljesítménye jellemzően magasabb, mint nem motivált társaiké, (3) a nem motivált válaszok/válaszadók kiszűrése növeli a teszteredmények validitását, (4) a tesztmegoldók motivációjának növelése szintén növeli a validitást, és (5) a legjobb eredményt a különböző módszerek kombinálásával kaphatjuk (Liu és mtsai, 2015; Rios, 2021).

## Támogatás

Az Innovációs és Technológiai Minisztérium Kooperatív Doktori Program Doktori Hallgatói Ösztöndíj Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

A tanulmány megírását az OTKA K135727 kutatási projekt támogatta.

## Irodalom

- Akyol, P., Krishna, K. & Wang, J. (2021). Taking PISA Seriously: How accurate are low-stakes exams? *Journal of Labor Research*. DOI: [10.1007/s12122-021-09317-8](https://doi.org/10.1007/s12122-021-09317-8)
- Alrababah, S. & Molnár, G. (2021). Analyzing contextual data in educational context: Educational data mining and logfile analysis. *Journal of Critical Reviews*, 8(January), 261–273.
- Asseburg, R. & Frey, A. (2013). Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychological Test and Assessment Modeling*, 55, 92–104.
- Attali, Y. (2016). Effort in low-stakes assessments: What does it take to perform as well as in a high-stakes setting? *Educational and Psychological Measurement*, 76(6), 1045–1058. DOI: [10.1177/00131644166634789](https://doi.org/10.1177/00131644166634789)
- Attali, Y. & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers and Education*, 83, 57–63. DOI: [10.1016/j.compedu.2014.12.012](https://doi.org/10.1016/j.compedu.2014.12.012)
- Baumert, J. & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, 16, 441–462. DOI: [10.1007/bf03173192](https://doi.org/10.1007/bf03173192)
- Bergstrom, B., Gershon, R. & Lunz, M. E. (1994). Computerized adaptive testing exploring examinee response time using hierarchical linear modeling. *The Annual Meeting of the National Council on Measurement in Education*.
- Braun, H., Kirsch, I. & Yamamoto, K. (2011). An experimental study of the effects of monetary incentives on performance on the 12<sup>th</sup>-Grade NAEP reading assessment. *Teachers College Record*, 113, 2309–2344.
- Butler, J. & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement*, 8(3), 279–304.
- Crombach, M. J., Boekaerts, M. & Voeten, M. J. M. (2003). Online measurement of appraisals of students faced with curricular tasks. *Educational and Psychological Measurement*, 63(1), 96–111. DOI: [10.1177/0013164402239319](https://doi.org/10.1177/0013164402239319)
- Csapó, B. & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, 10. DOI: [10.3389/fpsyg.2019.01522](https://doi.org/10.3389/fpsyg.2019.01522)
- Csapó Benő, Molnár Gyöngyvér & R. Tóth Krisztina (2008). A papír alapú tesztekől a számítógépes adaptív tesztesítésig: a pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, 18(3–4), 3–16.
- DeMars, C. (2000). Test stakes and item format interactions. *Applied Measurement in Education*, 13(1), 55–77. DOI: [10.1207/s15324818ame1301\\_3](https://doi.org/10.1207/s15324818ame1301_3)
- DeMars, C. E., Bashkov, B. M. & Socha, A. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Research and Practice in Assessment*, 8, 69–82.
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. DOI: [10.1146/annurev.psych.53.100901.135153](https://doi.org/10.1146/annurev.psych.53.100901.135153)
- Finn, B. (2015). Measuring Motivation in Low-Stakes Assessments. *ETS Research Report Series*, (2), 1–17. DOI: [10.1002/ets2.12067](https://doi.org/10.1002/ets2.12067)
- Freund, P. A., Kuhn, J. T. & Holling, H. (2011). Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, 51(5), 629–634. DOI: [10.1016/j.paid.2011.05.033](https://doi.org/10.1016/j.paid.2011.05.033)
- Goldhammer, F., Martens, T., Christoph, G. & Lüdtke, O. (2016). Test-taking engagement in PIAAC. *OECD Education Working Papers*, 133, 1–67. DOI: [10.1787/5jlzfl6fhxs2-en](https://doi.org/10.1787/5jlzfl6fhxs2-en)
- Goldhammer, F., Martens, T. & Lüdtke, O. (2017). Conditioning factors of test-taking engagement in PIAAC: an exploratory IRT modelling approach considering person and item characteristics. *Large-Scale Assessments in Education*, 5(1). DOI: [10.1186/s40536-017-0051-9](https://doi.org/10.1186/s40536-017-0051-9)
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hauser, C. & Kingsbury, G. G. (2009). Individual score validity in a Modest-Stakes adaptive educational testing setting. *The Annual Meeting of the National Council on Measurement in Education*. <https://www.nwea.org/resources/individual-score-validity-modest-stakes-adaptive-educational-testing-setting/>

- Koretz, D. (2018). Moving beyond the Failure of Test-Based Accountability. *American Educator*, 41(4), 22–26. <http://ezproxy.lib.uconn.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1164385&site=ehost-live>
- Lau, A., Swerdzewski, P., Jones, A., Anderson, R. & Markle, R. (2009). Proctors Matter: Strategies for Increasing Examinee Effort on General Education Program Assessments. *The Journal of General Education*, 58(3), 196–217. DOI: 10.1353/jge.0.0045
- Lee, Y.-H. & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*, 53(3), 359–379. [http://www.psychologie-aktuell.com/fileadmin/download/ptam/3-2011\\_20110927/06\\_Lee.pdf](http://www.psychologie-aktuell.com/fileadmin/download/ptam/3-2011_20110927/06_Lee.pdf)
- Lee, Y.-H. & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-Scale Assessments in Education*, 2(1), 8. DOI: 10.1186/s40536-014-0008-1
- Lindner, C., Nagy, G., Ramos Arhuis, W. A. & Retelsdorf, J. (2017). A new perspective on the interplay between self-control and cognitive performance: Modeling progressive depletion patterns. *PLOS ONE*, 12(6), e0180149. DOI: 10.1371/journal.pone.0180149
- Lindner, M. A., Ihme, J., Saß, S. & Köller, O. (2016). How Representational Pictures Enhance Students' Performance and Test-Taking Pleasure in Low-Stakes Assessment. *European Journal of Psychological Assessment*, 34(6), 376–385. DOI: 10.1027/1015-5759/a000351
- Lindner, M. A., Lüdtkke, O. & Nagy, G. (2019). The Onset of Rapid-Guessing Behavior Over the Course of Testing Time: A Matter of Motivation and Cognitive Resources. *Frontiers in Psychology*, 10, 1–15. DOI: 10.3389/fpsyg.2019.01533
- List, M. K., Robitzsch, A., Lüdtkke, O., Köller, O. & Nagy, G. (2017). Performance decline in low-stakes educational assessments: different mixture modeling approaches. *Large-Scale Assessments in Education*, 5(1). DOI: 10.1186/s40536-017-0049-3
- Liu, O. L., Rios, J. & Borden, V. (2015). The Effects of Motivational Instruction on College Students' Performance on Low-Stakes Assessment. *Educational Assessment*, 20(2), 79–94. DOI: 10.1080/10627197.2015.1028618
- Molnár Gyöngyvér (2003). Az ismeretek alkalmazásának vizsgálata modern tesztelméleti (IRT) eszközökkel. *Magyar Pedagógia*, 103(4), 423–446.
- Molnár Gyöngyvér (2013). *A Rasch modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában*. Gondolat Kiadó.
- Molnár Gyöngyvér (2021a). Az IKT szerepe a felsőoktatás megújításában. *Magyar Tudomány*, 182(11).
- Molnár, G. (2021b). Challenges and developments in technology-based assessment: possibilities in science education. *Europhysics News*, 52(2), 16–19. DOI: 10.1051/epn/2021202
- Molnár Gyöngyvér & Csapó Benő (2019a). A felsőoktatási tanulmányi alkalmasság értékelésére kidolgozott rendszer a Szegedi Tudományegyetemen: elméleti keretek és mérési eredmények. *Educatio*, 28(4), 705–717. DOI: 10.1556/2063.28.2019.4.4
- Molnár, G. & Csapó, B. (2019b). How to Make Learning Visible through Technology: The eDia-Online Diagnostic Assessment System. In Lane, H., Zvacek, S. & Uhomoihi, J. (szerk.), *CSEU 2019. Proceedings of the 11th International Conference on Computer Supported Education. Volume 2*. Scitepress. 122–131. DOI: 10.5220/0007754101220131
- Nuutila, K., Tapola, A., Tuominen, H., Molnár, G. & Niemivirta, M. (2021). Mutual relationships between the levels and changes in interest, self-efficacy, and perceived difficulty during a task. *Preprint*. DOI: 10.31234/osf.io/pd528
- Oneil, H., Abedi, J., Miyoshi, J. & Mastergeorge, A. (2005). Monetary Incentives for Low-Stakes Tests. *Educational Assessment*, 10(3), 185–208. DOI: 10.1207/s15326977ea1003\_3
- Penk, C. & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation and Accountability*, 29(1), 55–79. DOI: 10.1007/s11092-016-9248-7
- Rios, J. (2021). Improving Test-Taking Effort in Low-Stakes Group-Based Educational Testing: A Meta-Analysis of Interventions. *Applied Measurement in Education*, (March), 1–22. DOI: 10.1080/08957347.2021.1890741
- Rios, J. A., Guo, H., Mao, L. & Liu, O. L. (2017). Evaluating the Impact of Careless Responding on Aggregated-Scores: To Filter Unmotivated Examinees or Not? *International Journal of Testing*, 17(1), 74–104. DOI: 10.1080/15305058.2016.1231193
- Sahin, F. & Colvin, K. F. (2020). Enhancing response time thresholds with response behaviors for detecting disengaged examinees. *Large-Scale Assessments in Education*, 8(1), 5. DOI: 10.1186/s40536-020-00082-1
- Sahin, M. & Yurduguül, H. (2020). Educational Data Mining and Learning Analytics : Past , Present and Future. *Bartın University Journal of Faculty of Education*, 9(February), 121–131.
- Schnipke, D. L. (1995). Assessing speededness in computer-based tests using item response times. *Annual Meeting of the National Council on Measurement in Education*.
- Schnipke, D. L. & Scrams, D. J. (1997). Modeling Item Response Times With a Two-State Mixture Model: A New Method of Measuring Speededness. *Journal of Educational Measurement*, 34(3), 213–232. DOI: 10.1111/j.1745-3984.1997.tb00516.x



- Schüttpelz-Brauns, K., Kadmon, M., Kiessling, C., Karay, Y., Gestmann, M. & Kämmer, J. E. (2018). Identifying low test-taking effort during low-stakes tests with the new Test-taking Effort Short Scale (TESS) – development and psychometrics. *BMC Medical Education*, 18(1), 101. DOI: [10.1186/s12909-018-1196-0](https://doi.org/10.1186/s12909-018-1196-0)
- Technical Report of the Survey of Adult Skills (PIAAC) (3rd Edition)*. (2019). [https://www.oecd.org/skills/piaac/publications/PIAAC\\_Technical\\_Report\\_2019.pdf](https://www.oecd.org/skills/piaac/publications/PIAAC_Technical_Report_2019.pdf)
- Thelk, A., Sundre, D., Horst, S. & Finney, S. (2009). Motivation Matters: Using the Student Opinion Scale to Make Valid Inferences About Student Performance. *Journal of General Education*, 58(3), 129–151. DOI: [10.1353/jge.0.0047](https://doi.org/10.1353/jge.0.0047)
- Tóth, K., Rölke, H., Goldhammer, F. & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills. In Csapó, B. & Funke, J. (szerk.), *The nature of problem solving. Using research to inspire 21st century learning*. OECD. 193–209. DOI: [10.1787/9789264273955-14-en](https://doi.org/10.1787/9789264273955-14-en)
- Verbić, S. & Tomić, B. (2009). *Test item response time and the response likelihood*. [https://www.researchgate.net/publication/23951908\\_Test\\_item\\_response\\_time\\_and\\_the\\_response\\_likelihood](https://www.researchgate.net/publication/23951908_Test_item_response_time_and_the_response_likelihood)
- Wise, S. L. & DeMars, C. (2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment*, 10, 1–17. DOI: [10.1207/s15326977ea1001\\_1](https://doi.org/10.1207/s15326977ea1001_1)
- Wise, S. L. (2006). An Investigation of the Differential Effort Received by Items on a Low-Stakes Computer-Based Test. *Applied Measurement in Education*, 19(2), 95–114. DOI: [10.1207/s15324818ame1902\\_2](https://doi.org/10.1207/s15324818ame1902_2)
- Wise, S. L., Bholá, D. S. & Yang, S.-T. (2006). Taking the Time to Improve the Validity of Low-Stakes Tests: The Effort-Monitoring CBT. *Educational Measurement: Issues and Practice*, 25(2), 21–30. DOI: [10.1111/j.1745-3992.2006.00054.x](https://doi.org/10.1111/j.1745-3992.2006.00054.x)
- Wise, S. L. & DeMars, C. E. (2010). Examinee Noneffort and the Validity of Program Assessment Results. *Educational Assessment*, 15(1), 27–41. DOI: [10.1080/10627191003673216](https://doi.org/10.1080/10627191003673216)
- Wise, S. L. & Gao, L. (2017). A General Approach to Measuring Test-Taking Effort on Computer-Based Tests. *Applied Measurement in Education*, 30(4), 1–12. DOI: [10.1080/08957347.2017.1353992](https://doi.org/10.1080/08957347.2017.1353992)
- Wise, S. L. & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18(2), 163–183. DOI: [10.1207/s15324818ame1802\\_2](https://doi.org/10.1207/s15324818ame1802_2)
- Wise, S. L. & Ma, L. (2012). Setting response time thresholds for a CAT item pool: the normative threshold method. *Paper Presented at the 2012 Annual Meeting of the National Council on Measurement in Education*, (March), 1–24.
- Wise, S. L., Ma, L., Kingsbury, G. G., & Hauser, C. (2010). An investigation of the relationship between time of testing and test-taking effort. *National Council on Measurement in Education*, (March), 1–18.
- Wise, S. L., Ma, L. & Theaker, R. A. (2014). Identifying non-effortful student behavior on adaptive tests: Implications for test fraud detection. *Test Fraud: Statistical Detection and Methodology*, (January), 175–185.
- Wise, S. L. & Smith, L. F. (2016). The validity of assessment when students don't give good effort. In Brown, G. T. L. & Harris, L. R. (szerk.), *Handbook of Human and Social Conditions in Assessment*. Routledge. 204–220.
- Wolgast, A., Schmidt, N. & Ranger, J. (2020). Test-Taking Motivation in Education Students: Task Battery Order Affected Within-Test-Taker Effort and Importance. *Frontiers in Psychology*, 11, 559683. DOI: [10.3389/fpsyg.2020.559683](https://doi.org/10.3389/fpsyg.2020.559683)

### Absztrakt

A mérés-értékelés főbb céljainak átalakulásával, tanulást segítő diagnosztikus funkcióinak előtérbe kerülésével az alacsony tétellel bíró tesztek egyre nagyobb szerepet játszanak a tanítás-tanulás folyamatában, valamint az országos és nemzetközi értékelésekben. Mivel az alacsony tétellel bíró teszteken nyújtott teljesítménynek nincs következménye a tanulók tanulmányi előrehaladására, előfordulhat, hogy az ezeken a teszteken elért eredmények eltérnek a nagy tétellel bíró teszteken nyújtott teljesítményektől. Ennek oka lehet, hogy előbbin a tanulók nem feltétlen tesznek meg minden erőfeszítést a feladatok megoldása során, utóbbin a diákok tesztmegoldás közbeni szorongása jelentős teljesítménycsökkenő hatással bírhat. A tanulmány keretein belül az első problémakörrel foglalkozunk részletesebben. A megfelelő motivációval rendelkező diákok aránya számos tényezőtől függ. A szakirodalomban elérhető kutatási eredmények szerint a motiváltabb személyek teljesítménye jellemzően magasabb, mint nem motivált társaiké. Az eredmények validitásának növelésére számos módszer alkalmazható az alacsony tétellel bíró tesztek esetén, amelyek két nagy csoportba sorolhatók. Az első csoportba sorolt módszerek a tesztek megoldása előtt alkalmazandók, céljuk, hogy a tesztet megoldók motivációját növeljék, és így valódi tudásukra vonatkozóan pontosabb, megbízhatóbb értékelést valósítsanak meg. A másik csoportba sorolt módszerek mérik a feladatmegoldók teszt során nyújtott motivációját, majd a teszt megoldása után kiszűrjük a nem motivált válaszokat, vagy épp a nem motivált személyeket ezzel növelve a teszt validitását. A kutatási eredmények alapján megállapítható, hogy mind a nem motivált válaszok/válaszadók kiszűrése, mind a vizsgázók motivációjának növelése növeli a teszteredmények validitását. A legpontosabb eredmény pedig a különböző módszerek kombinálásával érhető el.