

# Az önértékelés buktatói: a tesztmegoldási erőfeszítés kérdőíves és logadatalalapú elemzése

Csányi Róbert<sup>1,2\*</sup>, Molnár Gyöngyvér<sup>3</sup>

<sup>1</sup>SZTE Neveléstudományi Doktori Iskola, Szeged, Magyarország

<sup>2</sup>Nemzeti Szakképzési és Felnőttképzési Hivatal, Budapest, Magyarország

<sup>3</sup>SZTE Neveléstudományi Intézet, MTA-SZTE Digitális Tanulási Technológiák Kutatócsoport, Szeged, Magyarország

\*Levelező szerző, e-mail: csanyi.robort@edu.u-szeged.hu

Beérkezett: 2023. november 27.; elfogadva: 2024. február 27.

## Összefoglalás

A kutatásban egyetemi hallgatók tesztmegoldási erőfeszítését monitoroztuk és hasonlítottuk össze egyrészt hagyományos önértékelő kérdőíves módszerek alkalmazásával, másrészt a számítógép-alapú tesztfeladatok megoldása során mutatott viselkedést pontosan rögzítő logadatok elemzésével. Az adatfelvételben alkalmazott teszt fiktív, ugyanakkor interaktív és komplex problémákra épített, melyek megoldása során eléztes tudásra nem építhettek a hallgatók. K-közép klaszterelemzést használtunk a tesztmegoldási erőfeszítésben eltérő hallgatók csoportjainak kialakítására. Az eredmények alapján megállapítható, hogy (1) a nem motivált válaszok kiszűrése növeli a teszten nyújtott átlagos teljesítményt, (2) azonosíthatók különböző tesztmegoldási erőfeszítéssel rendelkező hallgatói csoportok és (3) a hallgatók önértékelése, azaz az önértékelő kérdőívek kérdéseire adott válaszok nem teljesen állnak összhangban a hallgatók tényleges tesztmegoldási viselkedésével.

**Kulcsszavak:** tesztmegoldási erőfeszítés, logfájl-elemzés, k-közép klaszteranalízis, tanulói profilok

## The pitfalls of self-assessment: A questionnaire- and log data-based analysis of test-taking effort

Róbert Csányi<sup>1,2</sup>, Gyöngyvér Molnár<sup>3</sup>

<sup>1</sup>Doctoral School of Education, University of Szeged, Szeged, Hungary

<sup>2</sup>National Office of Vocational Education and Training and Adult Learning, Budapest, Hungary

<sup>3</sup>Institute of Education, University of Szeged, MTA-SZTE Digital Learning Technologies Research Group, Szeged, Hungary

## Summary

Students' cognitive test performance is not only determined by their actual knowledge and skills, but can also be strongly influenced by various affective factors, including partial or total lack of test-taking effort. According to Wise and DeMars (2005), unmotivated students scored more than half a standard deviation lower on tests than their motivated peers. Recent studies on test-taking effort have generally used a single method, and only a few studies have examined students' actual test-taking behaviour in multiple ways. The aim of the research is twofold: (1) to comparatively analyse methods used in measuring test-taking effort and (2) to identify and characterise student profiles that can be distinguished based on test-taking behaviour.

The research sample consisted of first-year university students (N=1748) just starting their studies in autumn 2021. Computer-based data collection was carried out using the eDia platform, while students were presented with interactive problems that required no prior knowledge and could be solved in 3–4 minutes. In order to reproduce and understand test-taking and problem-solving behaviour as well as possible, it was necessary to interact with the problem environment to understand, map and solve the problems. Students' test-taking effort was measured in two ways: (1) with a self-report questionnaire on a five-point Likert scale and (2) by analysing log data (time spent on

tasks and number of clicks). We used k-means cluster analysis, which had successfully been used in previous research on test-taking behaviour.

The percentage of unmotivated responses identified by the log data was 2.3%, which is relatively low compared to the typical 5–15%. After excluding unmotivated responses, the percentage of correct responses grew significantly from 61.4% to 63.0%. Three groups of students were identified based on k-means cluster analysis: (1) low effort, (2) high ability with medium effort, and (3) low ability with high effort. Results suggest that self-report questionnaire data are not consistent with students' actual test-taking behaviour.

Identifying test-taking effort and filtering out unmotivated answers/respondents accordingly increases the validity of the tests. Understanding students' test-taking behaviour greatly facilitates the identification of individual differences and provides an opportunity to design personalised learning.

**Keywords:** test-taking effort, logfile analysis, k-means clustering, student profiles

## Témavezetői előszó

A technológia – beleértve a mesterséges intelligencia adta eszközök mérésértékelésben való alkalmazását – számos új lehetőséget, korábban nem, vagy csak korlátozott módszertannal vizsgálható jelenségek alaposabb megértését teszik lehetővé. Csányi Róbert ezt használta ki a Kooperatív Doktori Program keretein belül megvalósított kutatásában, ahol egyrészt a tesztmegoldási erőfeszítés teljesítményre gyakorolt hatását monitorozta általános iskolás, valamint egyetemista elsőéves hallgatók körében, másrészt összevetette az önbevalláson alapuló hagyományos kérdőívalapú módszerek eredményeit a diákok valós viselkedését loggolt, úgynevezett kontextuális adatok elemzésének eredményeivel. Korosztálytól függetlenül megállapításra került, hogy a motiváció, az érdeklődés, a befektetett erőfeszítés jelentős teljesítménybefolyásoló hatással bír, valamint az önbevalláson alapuló válaszok és a kérdőíves kutatások eredményei kevéssé tükrözik, igen csekély mértékben korrelálnak a tesztelt személyek valós viselkedésével. Csányi Róbert kutatási eredményei nemcsak a tesztmegoldási motiváció jelenségének vizsgálata terén jelentősek, hanem általában a kérdőíves vizsgálatok validitása tekintetében is komoly kérdéseket indukál.

*Prof. dr. Molnár Gyöngyvér*  
intézetvezető egyetemi tanár  
Szegedi Tudományegyetem  
Neveléstudományi Intézet

## Vállalati szakértői előszó

A Kooperatív Doktori Program keretében Csányi Róbert a technológiaalapú mérésértékelési módszerek kutatását, oktatásban való felhasználását tűzte ki célul, amelyek a korábbi papír-ceruza alapú tesztekhez képest számos új lehetőséget kínálnak. A technológiaalapú értékelés esetén számos olyan kontextuális adat rögzíthető és elemezhető, amelyek a hagyományos értékelési módszerek esetén elérhetetlenek voltak.

Kutatásának egyik irányvonala a tesztmegoldási erőfeszítés logadatalapú vizsgálata. A motiválatlan válaszok/válaszadók azonosítása és kiszűrése növeli a tesztek validitását, a tanulók tesztmegoldási viselkedésének feltérképezése lehetőséget nyújt egyedi beavatkozások megtételére. A Nemzeti Szakképzési és Felnőttképzési Hivatal

szervezésében minden évben megvalósul a szakképzésben tanuló 9. évfolyamos tanulók bemeneti és kimeneti mérése. A doktorandusz kutatásaira alapozva a kompetenciamérés során tervezzük bevezetni a tesztmegoldási erőfeszítés monitorozását, ami segíti a szakképzésben tanulóknak kompetenciáinak pontosabb felmérését.

Kutatásának megvalósult alkalmazása az Újratervezés Program keretében szervezett távoktatásos képzések során a felnőtt résztvevők tanulási viselkedésének feltérképezése. A résztvevőknek a kurzus elvégzése közben mutatott tanulási viselkedése alapján azonosította az egymástól minőségileg is különböző tanulói profilokat, ami lehetőséget nyújt a felnőtt tanulói viselkedés feltérképezéséhez. A doktorandusz kutatása mind elméleti, mind gyakorlati szempontból jelentős, hozzájárul a tanulási-tesztelési folyamatok megértéséhez, a személyre szabott oktatás fejlesztéséhez.

*Dr. Lőrincz Mónika Mária*  
vállalati szakértő

Nemzeti Szakképzési és Felnőttképzési Hivatal

## Bevezetés

A diákok kognitív teszteken nyújtott teljesítményét nemcsak a tényleges tudásuk és készségeik határozzák meg (*Wolgast-Schmidt-Ranger 2020*), hanem számos affektív tényező is befolyásolhatja. A tesztek tépje jelentős hatással lehet az eredmények érvényességére: ahogy csökken a tét, úgy csökken a tesztmegoldási motiváció szintje (*Wise-Ma-Theaker 2014*). Az egyik legkiemelkedőbb nagymintás nemzetközi vizsgálatban – az intelligencián és a korábbi teszteredményeken túl – a diákok matematikai teszteredményei variációjának 1–29%-át lehetett a tesztmegoldási motivációval magyarázni (*Kriegbaum-Jansen-Spinath 2014*). Wise és DeMars szerint a motiválatlan tanulók több mint fél standard eltéréssel alacsonyabb pontszámot értek el a teszteken, mint motivált társaik (*Wise-DeMars 2005*). Ezt támasztja alá több kutatási eredmény is, amelyek a motiváltabb tesztmegoldók magasabb teljesítményét jelezték (*Finn 2015; Schüttelz-Brauns et al. 2018; Wise-Kong 2005*). Más kutatások szerint azonban nem szükséges a maximális vagy nagyon magas szintű tesztmegoldási erőfeszítés ahhoz, hogy ér-

vényes teszteredményt kapjunk, hanem elegendő egy bizonyos szintű erőfeszítés elérése (Csányi–Molnár 2023; Gignac–Bartulovich–Salleo 2019).

### A tesztek tétje

Egy teszt tétje a tesztmegoldó szemszögéből mutatja meg, hogy a tesztelt személy számára milyen következményekkel jár a teszten nyújtott teljesítmény (Csapó–Molnár–R. Tóth 2008). Az alacsony téttel bíró teszt eredményének nincs jelentős hatása az adott személy előmenetelére (Lindner–Lüdtke–Nagy 2019), míg a magas téttel bíró teszteken nyújtott teljesítmény jelentős következményekkel jár. Magas téttel bíró tesztnek számít a nyelvvizsga vagy az érettségi vizsga, amelyek eredményének hosszú időre, sőt akár egész életre szóló hatása lehet a tesztmegoldóra. Az alacsony téttel bíró tesztek közé sorolhatók a hazai és nemzetközi nagymintás felmérések, így az országos kompetenciamérés, vagy például a legismertebb nemzetközi tanulói teljesítménymérés, a PISA felmérés. Ezen tesztekre a tanulók hivatalosan nem kapnak – nem kaphatnak – osztályzatot, ezért diák szinten nincs tétje ezen felméréseknek. Az elért eredmények alapján a szakemberek az egyes oktatási intézmények, vagy az egész oktatási rendszer teljesítményére vonatkozóan vonnak le következtetéseket (Lindner–Lüdtke–Nagy 2019). Szintén az alacsony téttel bíró tesztek közé tartoznak a tanulási folyamatot közvetlenül támogató, ezért egyre inkább terjedő diagnosztikus tesztek. A számítógép-alapú diagnosztikus tesztek a tanítási-tanulási folyamatba közvetlenül beépítve biztosítják a viszonyítási pontokkal ellátott azonnali visszacsatolást, folyamatosan támogatva a tanulók személyre szabott fejlesztését (Csapó–Molnár 2019).

### Tesztmegoldási erőfeszítés

A tesztmegoldási motiváció értelmezésének egyik általánosan használt megközelítése az elvárás-érték elmélet (Eccles–Wigfield 2002; Wise–DeMars 2005). A modell szerint a tesztmegoldási motivációt a teljesítményre vonatkozó elvárás és a teszt értéke határozza meg. A tesztmegoldók elvárásait (1) a saját képességeikről alkotott elképzelésük és (2) a feladatok nehézsége határozza meg. Az értékek négy összetevőből állnak: (1) a teljesítményérték, azaz a teszt fontossága; (2) a belső érték, amelyet a feladatban való részvétel öröme határoz meg; (3) a hasznossági érték, azaz a feladat jövőbeli céljaihoz való viszonya; és (4) a költség, amelyet a feladat negatív aspektusa határoz meg (pl. a feladattal töltött idő vagy a teszttel kapcsolatos szorongás). A tesztmegoldási motiváció abban az erőfeszítésben nyilvánul meg, amelyet a vizsgázó a feladatok megoldásába fektet. A tesztmegoldási erőfeszítés az az erőforrás-mennyiség, amelyet a tesztmegoldó arra fordít, hogy a lehető legjobb eredményt érje el egy adott teszten (Lundgren–Eklöf 2020).

A tesztmegoldási erőfeszítés mérésére számos módszer létezik, amelyek három fő csoportba sorolhatók: (1) önértékelő kérdőívek, (2) válaszidő-alapú megközelítések és (3) modellalapú elemzések. A kutatásunkban ezek közül az első két típust alkalmaztuk. Az önértékelő kérdőívek a vizsgázók saját megítélésén alapulnak, míg a válaszidő-alapú megközelítések a tesztmegoldók által a feladatok megoldására fordított időt elemzik.

Az önértékelő kérdőívek a legrégebben és a legszélesebb körben használt eszközök a tesztmegoldási erőfeszítés mérésére, jellemzően annak összetevőit mérik Likert-skálán. A legegyszerűbb, legszélesebb körben használt esetben a tanulók a teszt végén, az utolsó feladat elvégzése után kapnak kérdéseket a teszt elvégzésével kapcsolatos erőfeszítéseikről. Ez a megközelítés azt feltételezi, hogy a tanulók tesztmegoldási erőfeszítése statikus, azaz nem változik a teszt során. Ez a megközelítés nem teszi lehetővé a folyamat dinamizmusának nyomon követését (Silm–Pedaste–Täht 2020). A tesztmegoldási erőfeszítés változása vagy akár állandósága is nyomon követhető a teszt elején, a tesztfeladatok között és utána az aktuális erőfeszítés szintjére vonatkozó kérdések megválaszolásával (Penk–Richter 2017). Ez utóbbi kutatási elrendezés lehetőséget biztosít a tesztmegoldási erőfeszítésben bekövetkező változások nyomon követésére. Az önértékelő kérdőívek fontos előnye, hogy viszonylag könnyen használhatók, és akár hagyományos papír-ceruza alapú tesztelésben is megvalósíthatók. Ugyanakkor e megközelítés korlátja a szubjektivitás, mivel nincs tudomásunk arról, hogy a tesztelő válaszai mennyire tükrözik a valós erőfeszítéseiket (Wise–Kong 2005).

A logadatokon alapuló módszerek a számítógépes értékelések elterjedésével párhuzamosan jelentek meg. A logadatok olyan számítógép által generált feljegyzések (logok), amelyek a felhasználók tevékenységéhez kapcsolódnak. Ezek a módszerek többnyire a válaszidőn alapulnak, ami azt az időt jelenti, amelyet a tesztmegoldó egy adott feladattal tölt a feladat kiközvetítésétől annak befejezéséig (vagyis attól az időpillanattól, hogy meglátja a képernyőn a feladatot, addig amíg a „tovább” gombra kattint). A hagyományos papír-ceruza teszt csak a tesztmegoldók válaszainak értékelését teszi lehetővé. Számítógépes értékelés esetén rengeteg olyan kontextuális adat (kattintások, feladatokkal töltött idő, előre-hátra ugrálás, szemmozgások stb.) rögzíthető, amelyek a hagyományos papír-ceruza alapú értékelési rendszerekkel elképzelhetetlenek voltak, és elemzésük mélyebb összefüggéseket tárhat fel (Tóth et al. 2017).

A válaszidő-alapú megközelítések mögött az az alapvető feltétele annak elolvasása, majd értelmezése, és csak azt követi a feladat konkrét megoldása. Ebből következően az alacsony tesztmegoldási erőfeszítéssel rendelkező résztvevők kevesebb időt töltenek a feladattal (annak elolvasásával, értelmezésével és megoldásával), ezért gyorsabban reagálnak, mint a magasabb motivációval rendelkezők

(Wise–Kong 2005). A feladatokra fordított időadatok elemzése kiegészíthető más viselkedési adatok elemzésével, például interaktív feladatok esetén a kattintások számával és típusával. A korábbi kutatási eredmények értelmében a kattintások alacsonyabb száma – hasonlóan az időadatokhoz – alacsonyabb motivációs szintre utal (Sahin–Colvin 2020).

A válaszidő-alapú módszereknek számos előnyük van az önértékelő kérdőívekkel szemben: (1) a tesztmegoldási erőfeszítés beavatkozás nélkül mérhető, (2) nem járul pluszmunka a vizsgált személyre, valamint (3) a mérés a vizsgázók valós viselkedésén alapul, nem pedig a saját megítélésükön, így (4) kevésbé lesz torzított az önértékelő kérdőívhez képest. A tesztmegoldási erőfeszítés változása sokkal pontosabban nyomon követhető, mivel a válaszidő adatai minden egyes feladatra vonatkozóan rendelkezésre állnak, nem csak bizonyos időpontokban (Wise–Kong 2005). A válaszidő-alapú módszereknél meg kell határozni egy küszöbidőt. Ha a válaszidő rövidebb, mint a küszöbérték, a válaszról azt feltételezzük, hogy az nem motivált (Wise–Kong 2005). A legegyszerűbb és legrégebben bevált megoldás az állandó küszöbérték alkalmazása, azaz minden egyes elemhez egy adott, előre meghatározott küszöbértéket használnak. Egy kifinomultabb megoldás az itemspecifikus küszöbértékek alkalmazása. Ezeket azon a feltételezésen alapulva határozzák meg, hogy az egyes itemek teljesítéséhez szükséges minimális idő minden egyes item esetében eltérő. Míg egy egyszerű matematikai alapműveletet a tesztmegoldók gyorsan meg tudnak oldani, addig egy összetett problémamegoldó feladat elolvasása, értelmezése és megoldása sokkal több időt vesz igénybe (Goldhammer et al. 2016). Ez azt jelenti, hogy a küszöb nem minden feladat, sőt item esetében azonos, hanem itemként, feladatonként eltérő lehet.

### Kutatási kérdések

A tesztmegoldási erőfeszítés nem elégséges szintje ronthatja az eredmények érvényességét (Rios 2021), mert ha a hibás választ úgy azonosítjuk, hogy a vizsgázó nem tudta megoldani a feladatot, ahelyett hogy motivátlanságként azonosítanánk, az befolyásolja a kapott eredményeket. Általában pozitív korreláció van a tesztmegoldási erőfeszítés és a teszten nyújtott teljesítmény között, de a kapcsolat nem ilyen egyértelmű, ha a tesztelt személyeket klaszterezett csoportokban vizsgáljuk (Lundgren–Eklöf 2020). A diákok tesztmegoldással kapcsolatos profilját számos tényező befolyásolja. Attól függően, hogy mely változók alapján történt az elemzés, a vonatkozó kutatások eltérő számú és tulajdonságú tanulói csoportot azonosítottak.

A jelen kutatás célja a hallgatók tesztmegoldási erőfeszítésének monitorozásán felül az volt, hogy összehasonlítsa az egyszerűsége miatt széles körben alkalmazott kérdőívalapú, önértékelésen nyugvó adatok elemzési

eredményét a diákok pontos viselkedésadatait tartalmazó logadatokon alapuló elemzési módszerek eredményeivel, valamint elemezzük a különböző profilú hallgatók tesztmegoldási viselkedését.

A következő kutatási kérdésekre kerestük a választ:

- 1) Milyen arányban fordulnak elő motivátlatlan válaszok item- és tesztszinten?
- 2) Változik-e a teszten nyújtott teljesítmény, ha a motivátlatlan válaszokat itemszinten töröljük az adatbázisból?
- 3) Milyen tesztmegoldási erőfeszítési hallgatói profilok határozhatók meg az önértékelésre, a logadatokra (a feladatra fordított idő és a kattintások száma) és a teszten nyújtott teljesítményadatokra építve?
- 4) A hallgatói profilok alapján összhangban van-e a hallgatók önértékelése a teszten mutatott valós viselkedésükkel?

## Adatok és módszertan

### Résztevők

Az adatfelvételre 2021 őszén került sor, a résztvevők az egyik vezető hazai egyetem (a bírálathoz elrejtve) 12 karának elsőéves hallgatói voltak, akik akkor kezdték meg tanulmányaikat. A hallgatók részvétele önkéntes volt, a tesztek sikeres teljesítéséért ösztönzőként egy kreditet kaptak. Az adatfelvétel részleteiről az egyetemi tanulmányi rendszeren keresztül kapott tájékoztatást minden nappali tagozatos elsőéves hallgató. A vizsgálatban összesen 1748 hallgató vett részt, akik a célcsoport 46,2%-át képviselték (átlagéletkor = 19,80 év, SD = 1,92), 53,0%-uk nő volt.

### Adatfelvétel

A kutatásban alkalmazott összes mérőeszköz (kérdőív és kognitív teszt) kiközvetítése az eDia rendszeren keresztül valósult meg (Csapó–Molnár 2019). Az adatfelvétel az egyetemi tanulmányi és információs központ 150 számítógépet tartalmazó informatikai laboratóriumában zajlott. A tesztek megoldása során a felügyeletet előzetesen felkészített PhD-hallgatók látták el. A Karrierépítés alapozó kurzus (Molnár–Csapó 2019) keretein belül kiközvetítésre került teszt megoldására a beiratkozás hetében, illetve a félév első két hetében volt lehetőség. A teszt elején próbafeladat segítségével a résztvevők megtanulhatták a felhasználói felület használatát. Az eDiába való bejelentkezés után a hallgatóknak 60 percük volt az interaktív problémák megoldására és a kérdőív kitöltésére. Az utolsó kérdés megválaszolása után azonnali visszajelzést kaptak teljesítményükről, egy héttel később pedig részletes, konkrét kurzusajánlásokkal, fejlesztési javaslatokkal egybekötött személyre szabott visszajelzést küldtünk számukra.

## Mérőeszközök

### Komplex problémamegoldás

A vizsgálathoz olyan, széles körben használt és megbízható eszközt kerestünk, amely kizárja a korábbi iskolai tanulás hatását (így figyelmen kívül hagyja a különböző területekkel kapcsolatos, már meglévő attitűdöket), ugyanakkor mindenki számára új, komplex és a problémakörnyezettel történő interakción keresztül hatékonyan feltérképezhető problémamegoldó helyzeteket teremt. A nemzetközi szinten széles körben alkalmazott és validált MicroDYN modellen alapuló problémákra esett a választásunk (Greiff et al. 2013; Greiff et al. 2018; Molnár–Csapó 2018). E problémák egyrészt nehézség tekintetében jól skalázhatóak, másrészt viszonylag rövid idő alatt feltérképezhetők és megoldhatók (Greiff et al. 2013; Funke 2014), miközben a problémakörnyezettel történő interakció által hatékonyan monitorozható a problémamegoldó viselkedés.

A problémamegoldás folyamatában empirikusan is megkülönböztethető a tudás megszerzése és a tudás alkalmazása fázisai (Greiff et al. 2013; Molnár–Greiff 2023). A tudás megszerzése (a problémamegoldás első fázisa) fázisban a problémamegoldók a bemeneti és kimeneti változók közötti összefüggéseket fedezik fel a problémakörnyezettel való szabad interakció révén (pl.: „Anyukádtól két újfajta szörpöt kaptál, amiket összekeverve még finomabb szörpöket tudsz magadnak kikeverni. Találd ki, hogyan befolyásolja a kikevert szörp édességét a zöld és kék színű szörp mennyisége!”). Ebben a fázisban az interakciók számát nem korlátoztuk, de az időkorlát 180 másodperc volt. A megszerzett és értelmezett információk alapján a képernyőn megjelenített, a bemeneti és kimeneti változókat tartalmazó modellben megrajzolták a bemeneti és kimeneti változók közötti kapcsolatokat (Molnár–Csapó 2018). A problémamegoldási folyamat második részében mindezt a tudást a hallgatók a rendszer hatékony működtetésére alkalmazták. 90 másodperces időkorlát mellett a bemeneti változók értékeit úgy kellett változtatniuk, hogy elérjék a problémakörnyezet egy előre adott célállapotát. Ezen idő alatt legfeljebb négy beállítási lehetőséget próbálhattak ki. A példafeladat második fázisában a szörp édességének előre meghatározott értékét legfeljebb négy lépésben úgy állították be, hogy a kétféle szörpöt megfelelő arányban keverték. A teszt tíz, egyre bonyolultabb problémából állt, azaz egyre több bemeneti és kimeneti változóból és egyre több kapcsolatból. A MicroDYN feladatok megbízhatósága a tudás megszerzésének és alkalmazásának mérésére jó volt ( $\alpha = 0,88$ ).

A tanulmány keretein belül bemutatott elemzésekben a problémamegoldási folyamat első fázisában gyűjtött teljesítmény- és logadatokra összpontosítottunk, mivel a lehetséges kattintások számát – ami a tesztmegoldási erőfeszítés jó mutatója lehet – nem maximalizáltuk, és a maximális idő – ami szintén a tesztmegoldási erőfeszítés

fontos mutatója – is jelentősen több volt. Következésképpen mind az idő-, mind a kattintási adatok nagyobb mértékben differenciáltak a hallgatókat, mint a feladatmegoldási folyamat második fázisában gyűjtött logadatok.

### Tesztmegoldással kapcsolatos erőfeszítés kérdőív

Két különböző megközelítést egyesítettünk a tesztmegoldással kapcsolatos erőfeszítések mérésére: az önértékelésen alapuló kérdőíves módszert és a logadatokon alapuló (a feladatra fordított idő és a kattintások száma) megközelítést. Emellett a hallgatók teszten nyújtott teljesítményére vonatkozó adatokat is rögzítettünk. A hallgatókat arra kértük, hogy az általunk tervezett állítás alapján egy ötfokozatú Likert-skálán („Teljes erőbedobással dolgoztam a feladatokon.”) értékeljék a tesztmegoldással kapcsolatos erőfeszítéseiket (self-reported effort, SRE; 1: egyáltalán nem igaz, 5: teljesen igaz). Annak érdekében, hogy ne csak statikus képet kapjunk a hallgatók tesztmegoldási erőfeszítéseiről, hanem a teszt folyamat során az erőfeszítésben bekövetkező változásokat is nyomon kövessük, a hallgatókkal a kognitív teszt során összesen hat alkalommal töltöttük ki az önértékelő kérdőívet. Az elsőt a bemelegítő feladat után, a következő négy alkalommal minden második problémahelyzet után, az utolsót pedig az utolsó feladat megoldása után.

### Eljárások

Kétféle logadatot vontunk be a tesztmegoldási erőfeszítés elemzésébe: (1) a feladatra fordított idő és (2) a kattintások száma. Ezek azt reprezentálták, hogy a hallgatók hogyan viselkedtek a feladatok elvégzése közben. A válaszügy-alapú módszerekben a mért mutató a feladatra fordított idő, vagyis az az idő, amelyet a tesztmegoldó egy feladatra fordít. Ha egy feladatra adott válaszügy kisebb, mint a küszöbérték, akkor azt motiválatlan válasznak minősítjük, ha azonban a küszöbértéknél nagyobb vagy azzal egyenlő, akkor motivált válasznak. Wise és Kong a következő összefüggést vezette be az  $i$  itemhez és a  $j$  vizsgált személyhez kapcsolódó motivált vagy megoldási viselkedés (*solution behavior*,  $SB_{ij}$ ) mérésére:

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

ahol  $T_i$  = az  $i$  itemre vonatkozó küszöbérték (*threshold value*),  $RT_{ij}$  = az  $i$  itemre és  $j$  vizsgázóra vonatkozó válaszügy (*response time*) (Wise–Kong 2005).

Kutatásunkban a küszöbérték meghatározására a nullánál nagyobb helyes válaszarány (*proportion correct greater than zero*,  $P+ > 0\%$ ) módszerét alkalmaztuk, amely több korábbi kutatás során is megfelelően bizonyult (Csányi–Molnár 2023; Goldhammer et al. 2016). Feleletválasztásos tesztek esetén a helyes válasz valószínűsége még véletlenszerű találgatás esetén is nagyobb, mint nul-

la (pl. 0,25 egy olyan teszt esetében, ahol feladatonként négy válaszlehetőség van). A feleletalkotó feladatok esetében, ahol a tesztmegoldónak nem több opció közül kell választania, hanem saját választ kell adnia, a helyes válasz kiválasztásának véletlenszerű esélye nulla. A  $P+ > 0\%$  küszöbérték meghatározásához a válaszokat a válaszidő nagysága szerint növekvő sorrendbe rendezzük. A küszöbérték az a legrövidebb válaszidő, amelynél a helyes válaszok aránya nagyobb, mint nulla (Goldhammer et al. 2016). Például, ha egy adott itemnél a 4, 5, 7 és 8 másodperces válaszok a sorba rendezés után helytelenek, és az első helyes válasz a 9 másodperces válaszidőre érkezik, akkor ez lesz a küszöbérték.

A motiválatlan válaszokat a  $P+ > 0\%$  módszerrel tesztmegoldónként és feladatonként azonosítottuk, majd feladatonként és a teszt egészére vonatkozóan meghatároztuk a motiválatlan válaszok százalékos arányát.

Annak érdekében, hogy meghatározzuk, hogyan változik az átlagos teljesítmény a motiválatlan válaszok kiszűrése után, először az eredeti tesztre vonatkozóan, majd a motiválatlan válaszok eltávolítása után kiszámoltuk az átlagos teljesítményt és a szórást, a két eredményt pedig független mintás  $t$ -próbával hasonlítottuk össze.

K-közép klaszteranalízist alkalmaztunk a hallgatók csoportjainak meghatározásához, hogy azonosítsuk a tesztmegoldási erőfeszítési profiljaikat. A klaszterelemzéshez a teljes, motiválatlan válaszokat is tartalmazó mintát használtuk. A k-közép klaszteranalízist sikeresen alkalmazták a számítógéppel támogatott felméréseken való tesztelési viselkedés korábbi vizsgálataiban (Goldhammer et al. 2017; Lundgren-Eklöf 2020; Qiao-Jiao 2018). A különböző skálák okozta torzítások elkerülése érdekében standardizálást alkalmaztunk. A klaszterelemzés elvégzésekor az egyik legfontosabb kérdés a klaszterek optimális számának meghatározása. Shi és munkatársai (2021) szerint az egyik leggyakrabban használt módszer a könyök módszer (*elbow method*), ami az ábrázolt értékek meredekségének maximális változásánál jelzi az optimális klaszterszámot. A módszer hátránya, hogy ez az érték sok esetben nehezen olvasható le a grafikonról, így nem lehet egyértelműen meghatározni az optimális klaszterszámot. Emiatt az elemzések során a sziluett-módszert (*silhouette method*) alkalmaztunk, ahol a sziluett-érték maximumából azonosítható az optimális klaszterszám (Shi et al. 2021). A sziluett-módszer alapján kutatásunkban a sziluett-érték maximuma három, így az optimális klaszterszám is ez az érték.

## Eredmények

A motiválatlan válaszokat tesztmegoldónként és feladatonként azonosítottuk, a teljes teszten 2,3% volt a motiválatlan válaszok aránya. A teszt előrehaladása során szignifikánsan nőtt a motiválatlan válaszok aránya (Wilks  $\lambda = 0,91$ ,  $F(9,1737) = 36,70$ ,  $p < 0,001$ ,  $\eta^2 = 0,10$ ). A részletes adatokat az 1. táblázat tartalmazza.

Az eredeti teszt esetén a helyes válaszok aránya 61,4% volt ( $SD = 0,49$ ), a motiválatlan válaszok (2,3%) eltávolítása után a helyes válaszok aránya 63,0%-ra nőtt ( $SD = 0,48$ ). A két érték között szignifikáns eltérés van ( $t = 3,06$ ,  $p = 0,002$ ).

A klaszterelemzéshez a változók közül a feladatra fordított időt, a kattintások számát, a teszten elért pontszámot és az önértékelési erőfeszítés mértékét vontuk be az elemzésbe. Az 1. ábrán a változók standardizált értékeinek (Z-pontszámok) klaszterenkénti átlagai láthatók, a 2. táblázat pedig a változók átlagait és szórásait tartalmazza. A varianciaelemzés eredményei alapján megállapítható, hogy a három klaszter között szignifikáns különbség van ( $p < 0,001$ ). A post hoc varianciaelemzések, azaz a klaszterek közötti különbségek vizsgálatára szolgáló elemzések során a varianciák nem homogének, ezért Dunnett-T3 tesztet alkalmaztunk.

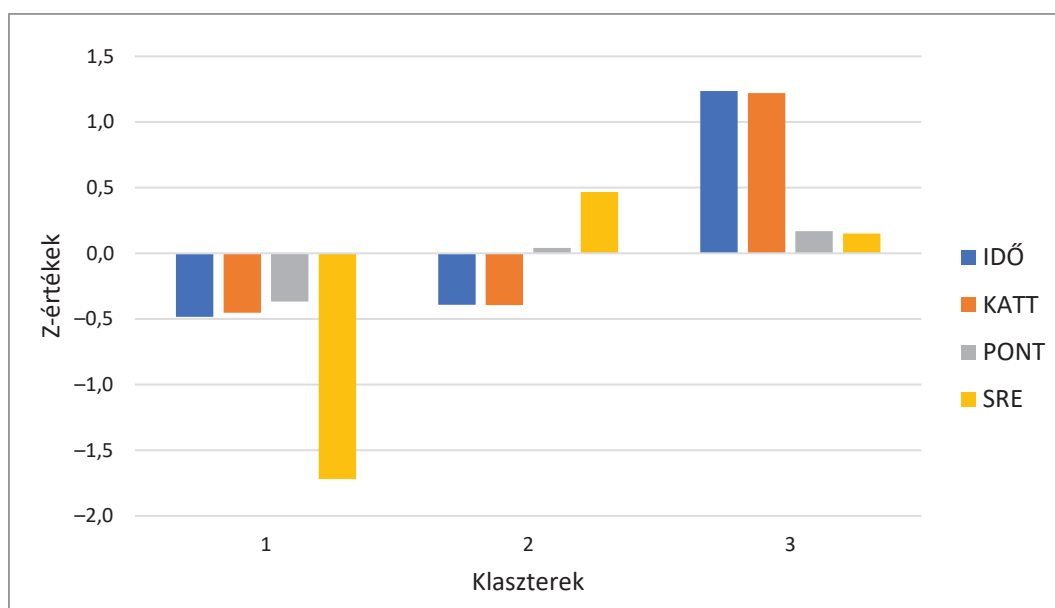
Az első klaszterbe 310 hallgató, a minta 18%-a tartozik. Rájuk jellemző, hogy kevés időt töltöttek a feladatok megoldásával, miközben keveset kattintottak. A 2. klaszterbe tartozó hallgatókhoz képest nem volt szignifikáns különbség a kattintások számában és a feladatokra fordított időben. A három klaszterbe sorolt hallgatók közül átlagosan ők mutatták a legalacsonyabb teljesítményt és jelentősen alacsonyabbra értékelték erőfeszítéseiket, mint a másik két klaszterben lévő társaik.

A második klaszterbe 1000 hallgató (57%) tartozik. Az ebbe a klaszterbe tartozó hallgatók szintén kevés időt töltöttek a feladatok megoldásával, miközben keveset kattintottak, azaz az 1. klaszter hallgatóihoz hasonlóan alacsony erőfeszítést fordítottak a feladatok elvégzésére. Jó eredményeket értek el, a 3. klaszter hallgatóihoz képest nem volt szignifikáns különbség átlagos problémamegoldó teljesítményükben. Ők értékelték a legmagasabbra a teszten nyújtott erőfeszítéseiket.

A harmadik klaszterbe viselkedésmintázata alapján 438 hallgató került (25%). Ők töltötték a legtöbb időt a problémák megoldásával, és ők kezdeményezték a legtöbb interakciót a problémakörnyezettel, azaz ők kattintottak a legtöbbször. Problémamegoldó teljesítményük azonos volt a második klaszterben lévő hallgatókéhoz,

1. táblázat | A motiválatlan válaszok százalékos aránya feladatonként és a teljes tesztre vonatkozóan

Feladat sorszáma	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Átlag
Motiválatlan válaszok aránya	0,1%	0,1%	0,3%	0,4%	1,5%	1,4%	4,6%	3,2%	3,5%	7,8%	2,3%



**1. ábra** | Hallgatói profilok a feladatra fordított idő, a kattintások száma, az elért pontszám és az önértékelési erőfeszítés alapján.  
IDŐ: a feladatra fordított teljes idő; KATT: a kattintások összes száma; PONT: a teszten elért pontszám; SRE: önértékelési erőfeszítés  
Forrás: saját szerkesztés

**2. táblázat** | A hallgatók logadataiból, a pontszámokból és az önértékelési erőfeszítésekből generált tesztmegoldási erőfeszítési profilok jellemzői

Változók	1. klaszter (N = 310)		2. klaszter (N = 1000)		3. klaszter (N = 438)		F*	Szignifikánsan különböző klaszterek**
	M	SD	M	SD	M	SD		
1. IDŐ	-0,48	0,68	-0,39	0,57	1,24	0,94	914,87	{1, 2} < {3}
2. KATT	-0,45	0,73	-0,39	0,54	1,22	0,98	866,01	{1, 2} < {3}
3. PONT	-0,37	1,13	0,04	0,94	0,17	0,97	28,93	{1} < {2, 3}
4. SRE	-1,72	0,78	0,47	0,42	0,15	0,73	1665,15	{1} < {3} < {2}

*Megjegyzés:* \* Minden F-érték szignifikáns ( $p < 0,001$ ). \*\* A szignifikánsan különböző klaszterek esetén a „<” a különbség irányát jelöli ( $p < 0,05$ ). A klaszterek közötti összehasonlító oszlop a Dunnett-T3 teszt alapján szignifikánsan különböző klasztereket mutatja. IDŐ: a feladatra fordított teljes idő; KATT: a kattintások összes száma; PONT: a teszten elért pontszám; SRE: önértékelési erőfeszítés  
Forrás: saját szerkesztés

de teljesítményük magasabb volt, mint az első klaszterben lévők. A második klaszterbe sorolt hallgatóknál átlagosan alacsonyabb szintűre értékelték erőfeszítéseiket.

## Összefoglalás és konklúzió

Kutatásunk célja az volt, hogy (1) különböző módszerek alkalmazásával és az eredmények összehasonlításával feltérképezzük első évfolyamos egyetemi hallgatók tesztmegoldással kapcsolatos erőfeszítéseit, (2) definiáljunk olyan objektíven mérhető változókat, amelyek a tesztmegoldási erőfeszítés hatékony indikátorai, (3) elemezzük a különböző – ezen indikátorok mentén felépített – profilú hallgatók problémamegoldó sikerességét, valamint (4) összehasonlítsuk a hallgatók önértékelését a teszten mutatott valós viselkedésükkel.

A legtöbb kutatás egyetlen elv szerint vizsgálja a tesztmegoldási erőfeszítést – az általunk áttekintett tanulmányok közül csak kevés olyan kutatást azonosítottunk, amely több, egyszerre alkalmazott módszert tartalmazott. Ezt támasztja alá a Silm, Pedaste és Täht (2020) által készített metaanalízis is, amelyben az elemzett tanulmányoknak mindössze a tizede alkalmazott többféle megközelítést. A kutatásunkban önértékelő kérdőívet használtunk a tesztmegoldási erőfeszítés mérésére, valamint logadatokon alapuló módszereket alkalmaztunk.

A  $P+ > 0\%$  módszer alapján a teljes tesztre vonatkozóan 2,3% volt a motiválatlan válaszok aránya, ami viszonylag alacsony érték a tipikusnak tekinthető 5–15% közötti arányhoz képest (Lee-Chen 2011). Ennek vélhetően az volt az oka, hogy a teszt alapvetően ugyan alacsony tétellel rendelkezett, hiszen a teszten nyújtott teljesítménynek

nem volt következménye a hallgatókra vonatkozóan, azonban a hallgatók nagyrészt kíváncsiak voltak a saját teljesítményükre, illetve azokra a területekre, amelyeken fejlődniük kell a tanulmányaik sikeres folytatásához. Ennek következtében a teszt alacsony tétje ellenére többnyire motiváltak voltak a feladatok megoldása során.

A feladatok megoldása során a tesztmegoldási erőfeszítés csökkent, azaz a motiválatlan válaszok aránya nőtt, ami összhangban van a korábbi kutatásokkal és a kognitív terheléssel magyarázható (Goldhammer et al. 2016; Penk–Richter 2017; Wolgast–Schmidt–Ranger 2020). Az önértékelő kérdőíves vizsgálatok nagy része esetén egy alkalommal, a feladatok megoldása után történik a kérdőív kitöltése, ami így nem ad valós képet a vizsgázók tényleges tesztmegoldási erőfeszítéséről. Pontosabb eredményt kaphatunk, ha a kérdőív több alkalommal is kiköszvetítésre kerül a tesztmegoldás során. Túl sok alkalommal azonban nem érdemes kitölteni a kérdőívet, mert az önmagában is csökkentheti a tesztmegoldási erőfeszítést. A legkifinomultabb lehetőséget a logadatok alkalmazása nyújtja, mivel egyrészt minden egyes feladat esetén mérhető a tesztmegoldási erőfeszítés, másrészt nem a résztvevők megítélésén, hanem a tesztmegoldási viselkedésükön alapul.

A motiválatlan válaszok kiszűrése után szignifikánsan, 61,4%-ról 63,0%-ra nőtt a helyes válaszok aránya. Ez összhangban van számos korábbi kutatással, amelyek szerint a motiválatlan válaszok kiszűrése növeli a tanulók teszten nyújtott átlagos teljesítményét (Wise–Kong 2005; Wise–DeMars 2010).

A hallgatókat tesztmegoldási erőfeszítésüket jellemző objektív és szubjektív változók mentén profiloztuk, így három, minőségében is különböző csoportot különböztettünk meg. Az első klaszterbe tartozó hallgatók kevés időt töltöttek a feladatok megoldásával, valamint ezalatt kevés interakciót végeztek. Mivel a feladattal töltött idő és a kattintások száma megfelel a feladatokba fektetett erőfeszítésnek, ezért ők alacsony erőfeszítést fejtettek ki a feladatok megoldása során. Ők érték el a legalacsonyabb teljesítményt, és az erőfeszítéseiket is lényegesen alacsonyabbra értékelték, mint a másik két klaszterben lévő hallgatók. Ebben a klaszterben az önértékelés adatai összhangban vannak a logadatokkal. Viselkedési adataik elemzése alapján minőségében más klaszterbe kerültek azon hallgatók, akik ugyanolyan alacsony szintű erőfeszítéssel oldották meg a feladatokat, mint az első klaszter hallgatói, azonban az erőfeszítéseiket a legmagasabbra értékelték (2. klaszter). A harmadik klaszter résztvevői fordították a legtöbb időt a feladatok megoldására, miközben a legtöbb interakciót hajtották végre, azaz a legtöbbet kattintottak. Eredményeik nem térnek el szignifikánsan a második klasztert alkotó hallgatókétól, viszont alacsonyabbra értékelték erőfeszítéseiket, mint a második klaszterbe tartozó társaik.

A második klaszterbe tartozó hallgatók lényegesen rövidebb idő alatt és kevesebb kattintással értek el hasonló

eredményt, mint a harmadik klaszterben részt vevők. Ez arra utal, hogy a második klaszterbe tartozó résztvevők magasabb képességszintűek, mint a harmadik klaszterbe tartozók. A 2. klaszterben a magasabb képességszintű hallgatók szignifikánsan kevesebb idő alatt kevesebbet kattintottak, miközben erőfeszítéseiket magasabbra értékelték, mint a 3. klaszterben lévő alacsonyabb képességű társaik, akik szignifikánsan többet kattintottak több idő alatt. Ez arra utal, hogy a résztvevők válaszaik nem tükrözik teljes mértékben a valós tesztmegoldási viselkedésüket, ami az önértékelő kérdőívek korlátait jelzi. Az önértékelés és a valós tesztmegoldási viselkedés közötti különbség részben a társadalmi elvárásokra vezethető vissza, amelyek arra készíthetik a hallgatókat, hogy azt válaszolják, amit elvárnak tőlük, nem pedig a valódi gondolataikat, érzéseiket. Az is elképzelhető, hogy a harmadik klaszterbe tartozó, gyengébb képességű résztvevők, akiknek gyengébb képességeik miatt általában több erőfeszítésre van szükségük a feladatok elvégzéséhez, alábecsülték erőfeszítéseiket a teszt során. Egy másik lehetséges magyarázat a diákok egy részének nem megfelelő önismerete. Az eredmények azt mutatják, hogy az önértékelő kérdőívre adott válaszok nem teljesen egyeznek a válaszadók tényleges tesztmegoldási magatartásával, amit alátámaszt *Silm, Pedaste és Täht* (2020) metaanalízise is.

A klaszterelemzés egyik előnye, hogy pontosabb betekintést nyújt a részletekbe, mint ha csak a tesztmegoldási erőfeszítés és a teljesítmény közötti korrelációt vizsgálnánk. A klaszterelemzés alapján kizárólag az 1. klaszterbe tartozó hallgatók önértékelése volt összhangban a teszten mutatott viselkedésükkel. A 2. klaszterbe tartozó hallgatók közepes erőfeszítéssel, míg a 3. klaszterbe tartozók nagy erőfeszítéssel értek el jó eredményt. Ez azt jelzi, hogy a jó eredmény eléréséhez nem szükséges maximális erőfeszítést befektetni a feladatok megoldásába, hanem elegendő egy bizonyos szintű erőfeszítést elérni. Ez összhangban van *Gignac, Bartulovich és Salleo* (2019), valamint *Stenlund, Lyrén és Eklöf* (2018) eredményeivel is, akik szerint a legjobb teljesítményt nyújtók magas szintű kockázatvállalással és viszonylag alacsony motivációval rendelkeznek. *Goldhammer és munkatársai* (2017) azt mutatták ki, hogy a magasabb képességszintű tanulóknak kevesebb erőfeszítésre volt szükségük a problémák sikeres megoldásához, ami szintén összhangban van az eredményeinkkel.

A hagyományos papír-ceruza tesztek esetén a tanulók értékelésére kizárólag a feladatokon elért eredményük áll az oktatók rendelkezésére. A számítógéppel támogatott oktatás, vizsgáztatás során olyan kontextuális adatok gyűjthetők a tanulási-vizsgázási folyamatokról, amelyek által pontosabb képet kaphatunk a tanulók viselkedésével kapcsolatosan. A tanulók klaszterezésével olyan hasonló tulajdonságokkal rendelkező tanulói csoportokat tudunk kialakítani, amelyek alapján személyre szabottabbá válhat az oktatás.



## Limitációk

A kutatásunknak számos limitációja van. Az egyik, hogy a teszt kizárólag interaktív problémamegoldó feladatokat tartalmazott. Emiatt az elemzések egy része nem alkalmazható más típusú tesztek, pl. egyszerű választásos feladatok esetén, ahol a válasz egyetlen kattintással megjelölhető. Másik fontos limitáció, hogy a kutatásunk során a tudásmegszerzési fázist elemeztük, míg a legtöbb teszt esetén a tudásalkalmazási fázis vizsgálata valósul meg. További limitációt jelent a kutatás mintaválasztása, mivel kizárólag önkéntes alapon jelentkező elsőéves egyetemisták vettek részt a felmérésen.

## Köszönetnyilvánítás

*A tanulmány az Innovációs és Technológiai Minisztérium Kooperatív Doktori Program Doktori Hallgatói Ösztöndíj Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült. A kutatást az OTKA K135727, illetve az MTA Közoktatás-fejlesztési Kutatási Pályázata támogatta (KOZOKT2021-16).*



## Irodalomjegyzék

- Csapó B., & Molnár G. (2019) Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, Vol. 10. <https://doi.org/10.3389/fpsyg.2019.01522>
- Csányi R., & Molnár G. (2023) How do test-takers rate their effort? A comparative analysis of self-report and log file data. *Learning and Individual Differences*, Vol. 106. 102340. <https://doi.org/10.1016/j.lindif.2023.102340>
- Csapó B., Molnár G., & R. Tóth, K. (2008) A papíralapú tesztetől a számítógépes adaptív tesztelésig. A pedagógiai mérés-értékelés technikájának fejlődési tendenciái. *Iskolakultúra*, Vol. 18. No. 3–4. pp. 3–16.
- Eccles, J. S., & Wigfield, A. (2002) Motivational beliefs, values, and goals. *Annual Review of Psychology*, Vol. 53. No. 1. pp. 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Finn, B. (2015) Measuring motivation in low-stakes assessments. *ETS Research Report Series*, Vol. 2015. No. 2. pp. 1–17. <https://doi.org/10.1002/ets2.12067>
- Funke, J. (2014) Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology*, Vol. 5. pp. 1–3. <https://doi.org/10.3389/fpsyg.2014.00739>
- Gignac, G. E., Bartulovich, A., & Salleo, E. (2019) Maximum effort may not be required for valid intelligence test score interpretations. *Intelligence*, Vol. 75. pp. 73–84. <https://doi.org/10.1016/j.intell.2019.04.007>
- Goldhammer, F., Martens, T., Christoph, G., & Lüdtke, O. (2016) Test-taking engagement in PIAAC. *OECD Education Working Papers*, No. 133. <https://doi.org/10.1787/5jlzfl6fthxs2-en>
- Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., & Tóth, K. (2017) Relating product data to process data from computer-based competency assessment. In: *Competence Assessment in Education: Research, Models and Instruments*, pp. 407–425. <https://doi.org/10.1007/978-3-319-50030-0>
- Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018) Students' exploration strategies in computer-simulated complex problem environments: A latent class approach. *Computers & Education*, Vol. 126. pp. 248–263. <https://doi.org/10.1016/j.compedu.2018.07.013>
- Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013) Complex problem solving in educational contexts—Something beyond *g*: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, Vol. 105. No. 2. pp. 364–379. <https://doi.org/10.1037/a0031856>
- Kriegbaum, K., Jansen, M., & Spinath, B. (2014) Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences*, Vol. 43. pp. 140–148. <https://doi.org/10.1016/j.lindif.2015.08.026>
- Lee, Y.-H., & Chen, H. (2011) A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*, Vol. 53. No. 3. pp. 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)
- Lindner, M. A., Lüdtke, 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*, Vol. 10. pp. 1–15. <https://doi.org/10.3389/fpsyg.2019.01533>
- Lundgren, E., & Eklöf, H. (2020) Within-item response processes as indicators of test-taking effort and motivation. *Educational Research and Evaluation*, Vol. 26. No. 5–6. pp. 275–301. <https://doi.org/10.1080/13803611.2021.1963940>
- Molnár, G., & Csapó, B. (2018) The efficacy and development of students' problem-solving strategies during compulsory schooling: Logfile analyses. *Frontiers in Psychology*, Vol. 9. pp. 1–17. <https://doi.org/10.3389/fpsyg.2018.00302>
- Molnár Gy. & Csapó B. (2019) A felsőoktatási tanulmányi alkalmasság értékelésére kidolgozott rendszer a Szegedi Tudományegyetemen: elméleti keretek és mérési eredmények. *Educatio*, Vol. 28. No. 4. pp. 705–717.
- Molnár, G. & Greiff, S. (2023) Understanding transitions in complex problem-solving: Why we succeed and where we fail. *Thinking Skills and Creativity*, Vol. 50. 101408. <https://doi.org/10.1016/j.tsc.2023.101408>
- 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*, Vol. 29. No. 1. pp. 55–79. <https://doi.org/10.1007/s11092-016-9248-7>
- Qiao, X., & Jiao, H. (2018) Data mining techniques in analyzing process data : A didactic. *Frontiers in Psychology*, Vol. 9. 2231. <https://doi.org/10.3389/fpsyg.2018.02231>
- Rios, J. A. (2021) Improving test-taking effort in low-stakes group-based educational testing: A Meta-analysis of interventions. *Applied Measurement in Education*, Vol. 34. No. 2. pp. 85–106. <https://doi.org/10.1080/08957347.2021.1890741>
- Sahin, F., & Colvin, K. F. (2020) Enhancing response time thresholds with response behaviors for detecting disengaged examinees. *Large-Scale Assessments in Education*, Vol. 8. No. 5. <https://doi.org/10.1186/s40536-020-00082-1>

- 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*, Vol. 18. 101. <https://doi.org/10.1186/s12909-018-1196-0>
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021) A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, Article number 31. <https://doi.org/10.1186/s13638-021-01910-w>
- Silm, G., Pedaste, M., & Täht, K. (2020) The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review*, Vol. 31. 100335. <https://doi.org/10.1016/j.edurev.2020.100335>
- Stenlund, T., Lyrén, P. E., & Eklöf, H. (2018) The successful test taker: exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education*, Vol. 33. No. 2. pp. 403–417. <https://doi.org/10.1007/s10212-017-0332-2>
- Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017) Educational process mining: New possibilities for understanding students' problem-solving skills. In: Csapó, B. & J. Funke (eds.) The nature of problem solving. Using research to inspire 21st century learning. Paris: OECD. pp. 193–209. <https://doi.org/10.1201/9781003160618-1>
- Wise, S. L., & DeMars, C. E. (2005) Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, Vol. 10. No. 1. pp. 1–17. [https://doi.org/10.1207/s15326977ea1001\\_1](https://doi.org/10.1207/s15326977ea1001_1)
- Wise, S. L., & DeMars, C. E. (2010) Examinee Noneffort and the Validity of Program Assessment Results. *Educational Assessment*, Vol. 15. No. 1. pp. 27–41. <https://doi.org/10.1080/10627191003673216>
- Wise, S. L., & Kong, X. J. (2005) Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, Vol. 18. No. 2. pp. 163–183. [https://doi.org/10.1207/s15324818ame1802\\_2](https://doi.org/10.1207/s15324818ame1802_2)
- Wise, S. L., Ma, L., & Theaker, R. A. (2014) Identifying non-effortful student behavior on adaptive tests: Implications for test fraud detection. In: Kingston, N., & Clark, A. (eds.) *Test Fraud: Statistical Detection and Methodology*, pp. 175–185.
- 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*, Vol. 11. 559683. <https://doi.org/10.3389/fpsyg.2020.559683>

A cikk a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>) feltételei szerint publikált Open Access közlemény, melynek szellemében a cikk bármilyen médiumban szabadon felhasználható, megosztható és újraközölhető, feltéve, hogy az eredeti szerző és a közlés helye, illetve a CC License linkje és az esetlegesen végrehajtott módosítások feltüntetésre kerülnek. (SID\_1)