

STEPHEN KLEIN – ROGER BENJAMIN – RICHARD SHAVELSON –
ROGER BOLUS

FELSŐOKTATÁSI TANULMÁNYI ÉRTÉKELÉS (CLA): TÉNYEK ÉS HIEDELMEK

Az oktatásügyi kormányzat (*Secretary of Education*) által felkért, A *Jövő Felsőoktatása Bizottság* (lásd *U.S. Department of Education*, 2006), mely a Spellings-bizottság néven vált ismertté, a következőképpen jellemezte a CLA-t: „az egyik legátfogóbb országos szintű kezdeményezés annak mérésére, hogy a hallgatók valójában mennyit tanulnak a különböző felsőoktatási intézményekben”, illetve, hogy a CLA „elősegíti a bizonyítékon alapuló értékelés kultúrájának terjedését a felsőoktatásban” (22. old.). A bizottság ezen túl azt javasolta, hogy „a felsőoktatási intézmények mérjék hallgatóik tanulását olyan eszközökből származó minőségértékelési adatok alkalmazásával, mint például a *Collegiate Learning Assessment* (CLA), ami a hallgatók tanulmányi fejlődését méri felsőoktatási intézményekben...” (23. old.). Az Amerikai Főiskolák és Egyetemek Szövetsége, egyetértésben az Amerikai Állami Főiskolák és Egyetemek Szövetségével (AACU) (2006, 4. old.) arra jutott, hogy:

A hozzáadott-értéket mérő közvetlen értékelés legjobb példája a Felsőoktatási Tanulmányi Értékelés (*Collegiate Learning Assessment*, CLA), mely a RAND Hozzáadott-Érték Értékelési Kezdeményezésből (*Value Added Assessment Initiative*, VAAI) nőtt ki, és 2004 tavasza óta áll a főiskolák és egyetemek rendelkezésére. A teszt túlmutat a feleletválasztós formátumon, és valós teljesítményt igénylő feladatokat támaszt, amelyek összetett anyagok elemzését és írásbeli válaszokat (például emlékeztető vagy ajánlás készítését) követelnek meg a hallgatóktól.

Amikor a hallgatói tanulmányok egy viszonylag újszerű értékelése ilyen dicséretet és figyelmet kap, jogos, hogy minden szempontból alaposan megvizsgálják. Így többek közt megkérdézik, hogy mit mér, mik a szakmai értékei és haszna. Bizottságok vizsgálják az eszközöket, az erősségeikről és gyengeségeikről – néha a versenytársak – formálnak megalapozott vagy téves véleményt. *Banta és Pike* (2007) például azt vitatta, hogy a CLA által alkalmazott, elvén alapuló tanulásértékelés helyes-e, *Kuth* (2006) pedig kételyeit fejezte ki a CLA megbízhatóságával és hasznosságával kapcsolatban.

Következésképpen úgy tűnik, itt az ideje, hogy – ha röviden is – bemutassuk a CLA program megközelítését az értékeléshez és a pontozás ismertetéséhez, az alkalmazott mutatókat, a CLA módszer és más kiterjedt mérési programok közti hasonlóságokat és eltéréseket, és végül a CLA mérések néhány fontos statisztikai jellemzőjét, köztük azok megbízhatóságát.

A CLA szemlélete az értékelésben

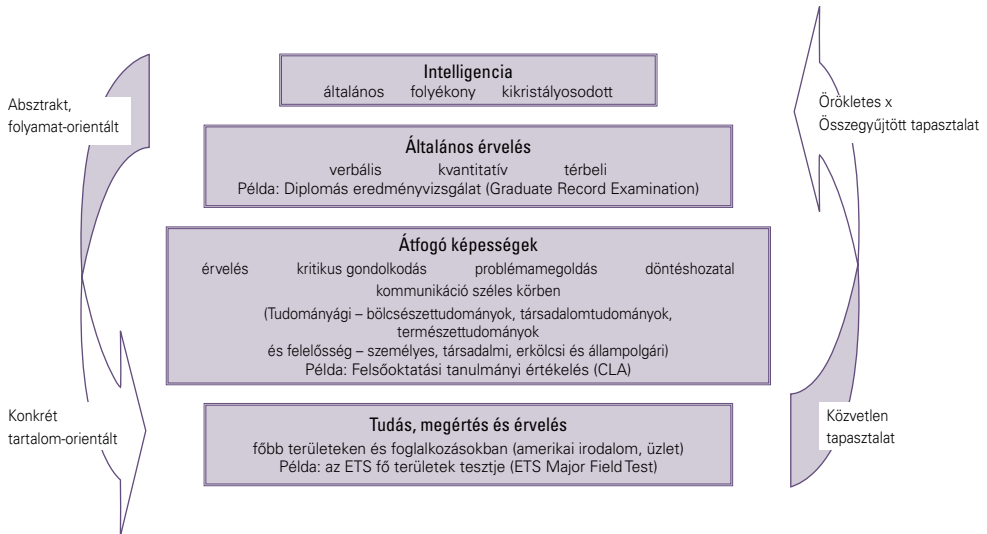
Az itt következőkben ismertetjük azokat a tanulmányi eredményeket, amiket mér, és azokat, amiket nem mér a CLA. Emellett tárgyaljuk a CLA szemléletmódját a tanulmányi eredmények értékelésében, amely – amellett, hogy hosszú és sikeres története van (*Shavelson*, 2007a,b) –

lényegesen különbözik az olyan tipikus általános oktatási¹ mérésektől, mint az ETS² hagyományosabb *Measure of Academic Proficiency and Progress* (MAPP) mérési módszere.

Amit a CLA mér, és amit nem

A tanulmányi eredmények igazán átfogó értékelése magába foglalhatná az 1. ábrán bemutatott készségek többségét, vagy akár mindegyikét. Nincsen olyan mérési program, amely képes lehetne mérni a középfokú szint feletti oktatás összes fontos eredményét, de ez nem zárja ki, hogy legalább néhány tényezőt értékeljen. Ezt az alapelvet szem előtt tartva a CLA-t úgy tervezték, hogy mérje a hallgatók kritikus gondolkodását, elemző érvelését, problémamegoldó-képességét és írásbeli kommunikációs kompetenciáit. A CLA (legalábbis kezdetben) ezekre a „átfogó képességekre” összpontosított, mert ezek minden szakon jelen vannak, és szerepelnek szinte minden felsőoktatási intézmény küldetésnyilatkozatában. Ezek a képességek idővel fejlődnek a tudományág-orientált oktatás, az általános oktatás és a hallgatók által kialakított és a felsőoktatásba magukkal hozott általános képességek közti kölcsönhatásban (Shavelson, 2007a).

1. ábra. A tanulmányi eredmények sematikus ábrája (Shavelson, 2007a)



A CLA program nem méri az általános intellektuális képességet, sem pedig az intelligenciát, és nem is törekszik erre. A különböző tudományágak megértését, azokhoz kapcsolódó tartalmi vagy műveleti tudást sem méri. Egyetértünk Dwyer, Millet és Payne (2006) véleményével, hogy hasznos lenne, ha rendelkezniénk eszközökkel ezen fontos középfok utáni képzési eredmények mérésére is. Kérdéses ugyanakkor, hogy valóban megvalósíthatók-e ilyen eljárások az intézmények közti összehasonlítások céljából, különösen akkor, ha a szakok eredményeit összesíteni kell, illetve a kétszakosságot is figyelembe kell venni. Mindazonáltal, ha létre lehet hozni, és

1 Némi határozatlansággal beszélünk általános oktatásról, mert az oktatók és a közvélemény jellemzően úgy vélik, hogy ez a kifejezés a felsőfokú képzés első két évfolyamára vonatkozik. Mi azonban a teljes egyetemi alapképzés során biztosított oktatásra utalunk.
 2 Educational Testing Service (<http://www.ets.org>), melyet 1947-ben alapítottak, és olyan világhírű teszteseteket fejleszt, szolgáltat, mint a TOEFL. (a szerk.)

egy közös skálára lehet kalibrálni ilyen méréseket, akkor valószínű, hogy azok a CLA tesztszeivel együtt (inkább, mint azok helyett) lennének használhatóak.

A CLA-ban alapvetően az intézmény (nem az egyes hallgató) az elemzési egység. Célja, hogy összegző értékelést nyújtson az intézmény (egészében szemlélt) képzési és egyéb programjai által *hozzáadott-értékről*, bizonyos fontos tanulmányi eredmények szempontjából. E mérések kitűzött célja, hogy jelzést adjanak a vezetőknek, tanszékeknek és hallgatóknak a fejlesztésre szoruló néhány kompetenciáról, a hallgatók által az intézményben elért teljesítményszintről. Kiemelten arról is, hogy ez a szint magasabb, alacsonyabb, vagy a körülbelül elvárható, tekintettel az intézménybe érkező hallgatók képességi szintjére.

A CLA önmagában nem határozza meg azokat az okokat, melyek következtében egy intézmény hallgatói a vártnál jobban vagy rosszabbul teljesítenek, arra sem tesz javaslatot, hogy milyen tantervi, vagy más változtatásokat kellene az iskolának megvalósítania hallgatói teljesítményének fokozása érdekében. A főiskolának saját tanszékeire, akkreditációs bizottságok felülvizsgálataira, helyi mérések adataira és egyéb forrásokra támaszkodva kell meghatároznia, hogy mit kell tennie az eredmények javításáért. Ugyanakkor – mivel a CLA mutatói az intézményi vezetői ciklusokon keresztül standardizálva vannak – segítségükkel vizsgálhatók (más mutatókkal együtt) az intézmény által bevezetett reformok és irányelvek hallgatói teljesítményre gyakorolt általános hatásai.³

Röviden, a CLA fő célja, hogy információt szolgáltatson, amely segít az egyetemeknek és főiskoláknak meghatározni, mennyit fejlődnek hallgatóik, és hogy ez a fejlődés összhangban áll-e más intézmények hasonló hallgatóinak fejlődésével. Egyes vezető felsőoktatási intézmények a CLA alkalmazásával vizsgálják az intézményen belüli alternatív oktatási programok eredményét is.

Ideális esetben a tanulási eredmények értékelése túlmutat a CLA *közvetlen tanulmányi mérésein*, és részben érinti az 1. ábrán bemutatott egyéb területeket is. A közvetlen mérések alatt azt értjük, hogy a hallgatóknak ténylegesen el kell végezni bizonyos kognitív kihívást jelentő feladatot, és e teljesítmény minőségét értékeljük numerikus skálán (Carroll, 1993). Ezzel szemben a *közvetett méréseket* úgy tervezik, hogy információt nyújtsanak az intézmény tanulmányi környezete szempontjából fontosnak tartott jellemzőkről. Például a Spellings-bizottság megjegyezte, hogy a Hallgatói Elköteleződés Országos Felmérése (*National Survey of Student Engagement*, NSSE) vizsgálja a „hallgatók részvételét és elköteleződését a tanulmányaik és fejlődésük elősegítésére tervezett programokban. A hallgatói részvétel mérését – a hallgatók által a foglalkozásokon és azon kívül tanulmányi tevékenységekre fordított idő és erőfeszítés, az oktatókkal való megbeszélésektől a kurzusokon nem kötelezően megjelölt könyvek olvasásáig” – úgy tervezték meg, „hogy *közvetítőként* szolgáljanak a hallgatók alapképzési tapasztalatainak értékéhez és minőségéhez” (22. old. kiemelés hozzáadva). A közvetlen és közvetett mérések közti különbségről (és az NSSE-ről) a cikkben még szót ejtünk.

Végül a felsőoktatás-politika szempontjából különösen fontosak a felsőoktatásba történő felvétel, a tanulmányok költségei, és a diplomaszerezési arány statisztikai mutatói. A tanulmányok hozzáférhetőségét és a tanulmányok átfogó értékelését ezekkel az indikátorokkal is elemezni kell.

Összességében a CLA csupán egy a hallgatók tanulásának sok fontos mutatója közül. Ideális esetben a tanulmányi eredmények más közvetlen méréseivel, a tanulmányok intézményi támo-

3 A CAE kutatói sok CLA-t alkalmazó intézménnyel társulva dolgoznak együtt, hogy meghatározzák és megvalósítsák a jobb CLA eredményekhez vezető legjobb gyakorlatokat. Ide tartozik a Council of Independent Colleges (Független Főintézmények Tanácsa), a University of Texas (a Texasi Egyetem) rendszere, a Higher Education Commission of West Virginia (Nyugat-Virginiai Felsőoktatási Bizottság), állami kutatóegyetemek és a Lumina longitudinális csoport.

gatásának közvetett eszközeivel és egyéb, a felsőoktatás szabályozását meghatározó politikai döntéshozók és a nyilvánosság számára fontos mutatókkal együtt alkalmazzák.

A CLA eszközei

A CLA önmagukban is értelmes, holisztikus, összetett feladatokkal méri a hallgatók kritikus gondolkodását, elemző érvelését, problémamegoldó-képességét és írásbeli kommunikációs képességeit, követve az 1930-as évekre visszanyúló progresszív oktatási mozgalom hagyományát (Shavelson, 2007. b.). A CLA egyes feladatai az írásos kommunikációs képességeket hangsúlyozzák, míg mások valóság-hű „munkából vett minta” teljesítményfeladatokat tartalmaznak. Minden feladat megfelel sok tudományos alapképzési szak és általános oktatási program hallgatói számára.

A CLA teljesítményfeladatai a Klein (1982, 1983) által a kaliforniai jogi szakvizsga számára fejlesztett eszközei, illetve a New Jersey Felsőoktatási Osztály által kialakított „Kritikus gondolkodási feladatok” (Ewell, 1994, Erwin és Sebrell, 2003) alapján készültek el. Ezzel a megközelítéssel az elvégzendő feladatok valóság-hű munkafeladatok, melyeket az oktatásban, munkában, politikában és mindennapi gyakorlatban előforduló tevékenységekről mintáztunk. A feladatok teljesítéséhez kritikus és elemző gondolkodásra van szükségük a tesztet végzőknek a kapott információról, és világosan, megfelelő ítéletalkotással kell megfogalmazniuk döntéseiket, értékelésüket, vagy javaslataikat (lásd McClelland, 1973).

Az információs technológia legújabb fejlődésének köszönhetően lehet a feladatok gazdagságát anélkül biztosítani, hogy a hallgatókat túlterhelnénk. Az értékelés minden feladatát egy interaktív internetes felületen adjuk meg, amely papírmentes, elektromos adminisztrációt és online pontozást és eredménykimutatást biztosít. Az esszé típusú feladatokra adott válaszokat jelenleg bírák pontozzák, akik átfogó képzést követően online osztályozzák a válaszokat. Így az egész rendszer papírmentes, ami jelentősen csökkenti a költségeket. Úgy véljük, hogy az elkövetkező néhány éven belül a teljesítményfeladatokat is számítógépes szoftver fogja majd pontozni.⁴

A 2. ábra egy teljesítményfeladatot mutat be. E feladatban a hallgatók – egy kisrepülőgépek számára elektronikus navigációs és kommunikációs berendezéseket gyártó vállalatnál – a DynaTech-nél dolgoznak, és főnökük megkérte őket, hogy értékeljenek egy repülőgépet (a „SwiftAir 235-ös”) megvásárlása mellett, illetve ellene szóló érveket a vállalat számára. A vásárlással kapcsolatban aggályok merültek fel egy SwiftAir 235-ös közelmúltbéli balesete miatt. A hallgatók a helyzet értékeléséhez könyvtári információkat kapnak a SwiftAir 235-ösről, különösen és általában a repülőgép-balesetekről. Az információ részben releváns és megbízható, részben azonban nem. A probléma részét képezi, hogy a hallgatók mely információkat használjanak fel, és melyeket hagyjanak figyelmen kívül. A hallgatók a több információforrás integrálásával megoldást találnak a problémára, döntést hoznak, és javaslatot tesznek.

A hallgatók valóság-hű módon adnak választ, például emlékeztetőt írnak a főnöküknek, elemezve a különböző megoldási lehetőségek mellett és ellen szóló érveket, előre jelezve a lehetséges problémákat, és azok megoldásait, javaslatot téve, hogy mit tegyen a vállalat, és felsorakoztatva a véleményüket, illetve javaslataikat alátámasztó bizonyítékokat. A teljesítmény pontozásában vizsgálják és értékelik a megfelelően alátámasztott alternatív megoldásokat a

4 A közelmúltban készített elővizsgálat szerint a kézzel és géppel készített összpontszámok közti összefüggés ugyanazon a teljesítményfeladaton ($r = 0,86$) hasonló a két ember által készített pontszám összefüggéseivel. A két pontozási módszer hasonló átlagokat és standard derivációkat is hozott.

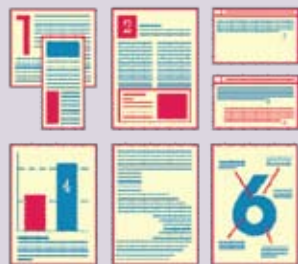
problémára, illetve az alternatív megoldási módokat.⁵ E feladat esetében néhány hallgató elég lényeglátó módon felismerte, hogy nem kívánatos következményei lehetnek, ha a DynaTech saját műszereivel felszerelt repülőgépe zuhanna le.

2. ábra. A CLA teljesítményfeladatának formátuma

Ön Pat Williams, a DynaTech (amely egy precíziós elektronikai műszereket és navigációs berendezéseket gyártó vállalat) elnökének asszisztense. Sally Evans, a DynaTech értékesítési csapatának egyik tagja azt javasolta, hogy a DynaTech vásároljon egy kis magánrepülőgépet (egy SwiftAir 235-öst), amelyet ő, és az értékesítési csapat többi tagja használhatna az ügyfelek látogatására. Pat már épp jóváhagyta volna a beszerzést, amikor baleset történt egy SwiftAir 235-össel. Az alábbi dokumentáció áll rendelkezésre:

1. Újságcikkek a balesetről.
2. Szövetségi Baleseti Jelentés a repülés közben összetört egymotoros repülőgépekről.
3. Pat önnek írt e-mailje, és Sally Patnek írt e-mailje.
4. Táblázatok a SwiftAir teljesítmény-jellemzőiről.
5. Az Amatőr pilóta c. folyóirat cikke, amelyben összehasonlítják a SwiftAir 235-öt hasonló repülőgépekkel.
6. Képek és leírás a SwiftAir 180-as és 235-ös modellekről.

Kérjük, készítsen emlékeztetőt, melyben több kérdést tárgyal, köztük azokat az adatokat, melyek alátámasztják vagy cáfolják, hogy a SwiftAir 235 szárnytípusa több repülés közben történő balesethez vezet, a balesethez esetleg hozzájáruló egyéb, figyelmet érdemlő tényezőket, valamint az ön összefoglaló javaslatát arról, hogy a DynaTech megvásárolja-e a repülőgépet, vagy ne.



A CLA program emellett két esszékérdés-típust is alkalmaz. A 30 perces „szedj ízekre egy érvelést” típus bemutat egy érvelést, és azt kéri a hallgatóktól, hogy kritizálják meg, elemezve a szerző érveinek érvényességét, ne pedig csak egyszerűen egyetértve, vagy szembehelyezkedve a szerző álláspontjával (*példát lásd a 3. ábrán*).

A 45 perces érveléstípus bemutat egy nézőpontot egy általános érdeklődésre számot tartó témában, és azt kéri tőlük, hogy válaszoljanak rá bármely szemszög(ek)ből. Az egyik ilyen téma: „Manapság a különböző specialisták rendkívül túlértékeltek. Több általánosan tájékozott emberre van szükség – olyan emberekre, akik széles körű áttekintést tudnak adni.” A hallgatóknak releváns érveket és példákat kell felhozniuk véleményük megvilágítására és igazolására. A „szedj ízekre egy érvelést”, és az „érvelj” feladatokra adott válaszokat számítógép pontozza.

⁵ Minden bíráló által végzett pontozás név nélkül történik. A bírák semmilyen információval nem rendelkeznek az értékelt válaszokat adó hallgatók tulajdonságairól. A kézírás sem okoz problémát, mert minden hallgató gépeli a válaszokat.

3. ábra. Példa a 30 perces „Szedj ízekre egy érvelést” feladatra

A University of Clairat általában a világ legjobb egyetemei között tartják számon, oktatóik hírneve miatt, amit elsősorban bizonyos tanszéki munkatársak kiterjedt kutatói és publikációs munkája alapoz meg.

Számos tanszéki oktató nemzetközileg elismert szaktekinvély. Például az angol tanszék számos tagja rendszeresen meghívást kap, hogy tanítson más országok egyetemein.

Ráadásul a fizika tanszék két friss diplomását fizikai Nobel-díjra jelölték. Emellett a diplomaszerezést követően a hallgatók 75%-a talál állást.

A tanszékei hírneve alapján a University of Claira nyilvánvalóan jó választás lehet bárkinek, aki minőségi képzést keres.

Minden számítógép által adott „nyers” pontszám átszámításra kerül egy közös mérési skálára. Ennek következtében az egyik feladaton egy adott pontszámot kapott válaszok körülbelül azonos szintű relatív jártasságot tükröznek, mint egy másik feladatra adott, azonos pontszámot elért válaszok. Ez a jellegzetesség lehetővé teszi a különböző feladatok pontszámainak egyesítését és átlagolását, az intézményi átlagok kiszámításához. Emellett a skála SAT-hoz (*Scholastic Aptitude Test*⁶) kapcsolásával a CLA program képes a pontszámok időbeli összevetésére, például a belépő elsőévesek és a diplomaszerezés előtt álló végzős hallgatók közti fejlődés mérésére.⁷

A CLA megközelítése abból indul ki, hogy a kritikus gondolkodás, az elemző érvelés, a problémamegoldó-képesség és az írásbeli kommunikációs képességek elválaszthatatlanul és összetett módon egybefonódnak a feladat- és válaszkövetelményekben. A teljesítményfeladatok hitelességének megőrzése érdekében különösen fontos, hogy azok a valódi tevékenységi területekről származzanak. A feladatok elemekre bontása és a kritikus gondolkodás, az elemző érvelés, a problémamegoldó-képesség és a kommunikáció önálló mérése lehetetlen lenne e feladatok összetettsége és egységessége miatt. Ehhez fel kéne adni életszerű munkafeladat-jellegűket, és az így létrehozott mesterséges részfeladatok szétszabdálnák a valós összetettséget. Mindazonáltal a MAPP és más hasonló mérések ezt teszik. Az ő értékelési megközelítésük a Carnegie Alapítvány úttörő munkájából ered, amely a 20. század elején mérte az egyetemi és főiskolai hallgatók tanulmányait, és amelyet nagymértékben befolyásoltak a behavioristák, elsősorban E. L. Thorndike (*Shavelson*, 2007. b.). Ezek az értékelések sokdimenziós „fogalmi szerkezetből” indulnak ki – ahogy a CLA is – de azután megpróbálják azt alkotóelemeire darabolni – kritikus gondolkodás, elemző érvelés és kommunikáció. Ezután az egyéni tesztfeladatokat, részteszteket vagy teszteket úgy szerkesztik meg, hogy lefedjék az egyes elemeket. Végül az elemeket újra összerakják, hogy kialakuljon az összpontszám.

6 A SAT egy a felsőoktatási felvételen széles körben használt teszt az Egyesült Államokban, melyet ma a College Board, egy nonprofit szervezet fejleszt és működtet, korábban az ETS gondozta. (a szerk.)

7 A SAT helyett inkább az ACT-t (American College Testing) választó hallgatók ACT pontszámait átváltjuk a SAT skálára. Ezt a két vizsga közti, a legtöbb felsőfokú felvételi eljárásában alkalmazott szabványos átváltási rendszerrel végezzük. A CLA elemzései és a jelen cikkben szereplő eredmények nem tesznek különbséget az eredeti SAT pontszámok, illetve az ACT-ből átváltott pontszámok között.

Ez a „fogalmi szerkesztésorientált” megközelítés (ellentétben a CLA munkát mintázó megközelítéssel) feltételezi, hogy a részek összeillesztésével pontos képet kaphatunk az egészről. E stratégia elsődleges előnyei, hogy külön pontszámok nyerhetőek az egyes szerkezeti elemekre, és legalábbis elméletben pontosabban különíti el a problémákat. A legnagyobb hátránya – és egyben az egyik fő oka annak, hogy a CLA nem alkalmazza –, hogy gyakran mesterséges, és figyelmen kívül hagyja azt a ténytet, hogy az egész általában sokkal több, mint a részek összessége: a biztonságos autóvezetés képessége sokkal több, mint a közlekedési szabályok ismerete, a műszerek leolvasásának képessége, a kormányzás, gyorsítás és fékezés.

Bár a CLA szemlélete alapvetően eltér a MAPP által alkalmazottól, és más széles körben használt feleletválasztós tesztekől, néhány más jellemzőjüket alkalmazza. Például a CLA felméréseket szabványosított feltételek között adják ki az intézmények között, és a nyers pontszámokat az adminisztrációk között összehasonlítható skálapontokká alakítják. A CLA program ezen jellemzői lehetővé teszik az intézmények számára, hogy megvizsgálják, hallgatóik milyen jól teljesítenek az idő multával, és más intézmények hasonló helyzetű hallgatóihoz képest.

Banta és Pike (2006) kijelenti, hogy „a felsőoktatásban sokkal ígéretesebb útját biztosítja a növekedés és fejlődés értékelésének, ha a hallgatók általános és tudományág-specifikus tudását és képességeit vizsgáljuk a szakterületükön, mint ha a leggyakrabban alkalmazott, pusztán általános képességeket mérő tesztek veszzük alapul”. Egyetértünk abban, hogy a hallgatói eredmények értékelése fejleszthető a tudományág specifikus eredményeinek értékelésével, de ez nem jelentheti az olyan átfogó képességek értékelésének elutasítását, amelyeket a felsőoktatási intézmények kimondottan fejleszteni próbálnak hallgatóiknál, és amelyeket a munkaadók értékelnek. Ráadásul a felvett elsőévesek és a diplomát szerző végzős hallgatók teljesítményében mérhető átlagos fejlődés a CLA-n nagyobb, mint egy szórásnyi. Ez mindenképpen hatalmas hatásméretet, ami azt jelzi, hogy a CLA érzékeny a felsőfokú oktatás következtében kialakult változásokra.

A CLA további megkülönböztető jegyei

Mint fentebb jeleztük, a CLA szemlélete eltér a legtöbb nagy tesztprogramétól. Például minden CLA mérést online végzünk el, minden kérdés nyílt (nincsenek feleletválasztós kérdések), és minden hallgató válaszait számítógép vagy képzett bírák pontozzák, szintén online. Emellett a feladatokat véletlenszerűen kapják a hallgatók, így az adott hallgató a teljes kérdéshalmaznak csak egy kis részére válaszol, de minden feladat eljut egy-egy iskolába.⁸ Ez a „mátrixos mintavételi” stratégia nagymértékben csökkenti a tesztelés egy hallgatóra jutó terhét, miközben továbbra is ugyanazt a kérdéshalmazt használja minden intézményben.

Az általános és középiskolai szinten széles körben használt tesztprogramokkal ellentétben a CLA csak egy az intézmény diákjaiból vett *mintát* vizsgál. Ráadásuk a hallgatók jellemzően önkéntesen vesznek részt ebben a programban, ezt nem követelik meg tőlük. Bár természetesen kívánatos lenne minden hallgatót vizsgálni, vagy egy valóban véletlenszerű mintavételt alkalmazni minden iskolában, a felsőoktatási intézmények többségében azonban e megoldások egyike sem valósítható meg. A MAPP és az NSSE esetében is felmerül ez a probléma.

A főiskolák és egyetemek számos módszert alkalmaznak, hogy ösztönözzék hallgatóikat a CLA teszt elvégzésére, például könyvutalvánnyal, kurzusfelvételnél előnyökkel, és különböző nyereményekkel. Mindazonáltal eddig nem volt kimutatható semmilyen értelmezhető összefüggés az ösztönzés típusa és a részvételi arány, vagy a hallgatók motivációja között, hogy

8 A CLA programban egynél több alkalommal részt vevő hallgatók – például elsőévesként, később pedig végzősként – automatikusan különböző feladatokat kapnak. A program jelenleg nyolc különböző teljesítményfeladatot alkalmaz és minden esszé jellegű kérdésből is több változatot.

jól teljesítsenek a teszten. A teljesítménytesztet megíró hallgatók kitöltenek egy utó-értékelési kérdőívet, amely megkérdezi, hogy mennyire igyekeztek jó eredményt elérni a feladaton. Az adatok elemzése szerint, miután a SAT pontszám hatását kontrolláltuk, az erőfeszítésre vonatkozó saját beszámolóik az intézményi átlagok varianciájának mindössze mintegy öt százalékáról adnak számot, ami túl kevés, hogy jelentős hatással legyen az eredményekre.⁹

A hallgatók megnevezték a CLA megírásának elsődleges motivációs tényezőjét is. Válaszaikat az alábbi három csoportba rendeztük: (1) Személyes fejlődés (erősségeim és gyengéim megértése, vagy hogyan viszonyul a pontszámom a többi diákéhoz); (2) Ellenszolgáltatás (javadalmazás, ajándékutalvány, nyeresémhúzás nagy földijjal); és (3) Követelmény (a kurzus vagy az egész intézmény értékelési tevékenységébe illesztve). A SAT pontszámaik hatásának kontrollálása után az, hogy melyik típusú tényezőt említik, a hallgatói CLA pontszámok varianciájának mindössze egy százalékát befolyásolta.

Eredmények bemutatása a hozzáadott-érték felől

A CLA program egyik legfontosabb ismérve az eredmények bemutatása abból a szempontból, hogy az intézmény hallgatói jobban, rosszabbul vagy körülbelül ugyanúgy teljesítettek, mint az az induláskor mért kompetenciaszintjük ismeretében elvárható. A program azt is vizsgálja, hogy az átlagos hallgatói teljesítmény fejlődése az iskolába történő beiratkozás és a diploma-szerzés között megfelel-e a más felsőoktatási intézményben tanuló, hasonló helyzetben lévő hallgatókéinak. A program tehát tájékoztatja az intézményeket, hogy a hallgatóik fejlődése megfelel-e a más intézményekben elérhető fejlődésnek (pl. *Raudenbush*, 2004). Így a CLA program betartja azt az elvet, hogy a középfok utáni értékelési programoknak a hallgatói tanulmányi fejlődés mérésére és az ahhoz való hozzájárulásra kell összpontosítaniuk.¹⁰

A fejlődés értékelésének „hozzáadott-érték” alapú megközelítését azért vettük át, mert (a) egy intézmény vizsgált hallgatói bizonyos mértékben jobb vagy rosszabb készségekkel rendelkezhetnek, mint társaik, (b) egy intézmény hallgatóinak belépő tulajdonságai egyik évben eltérőek lehetnek a másik évben részt vevők tulajdonságaitól, és (c) az intézmények között jelentős különbségek vannak belépő hallgatóik kompetenciaszintjében.¹¹ Ezek a nehézségek eleve kizárják az intézmények közti átlagpontszámok egyszerű összehasonlításának lehetőségét. Így ahhoz, hogy ésszerű és praktikus módszert biztosítsunk a fejlődés mérésére az intézmények között és egy iskolán belül az idő múlásával, először a vizsgálati programban valójában részt vevő hallgatók belépő kompetenciáikhoz kell igazítani a pontozást minden intézményben.

9 Amikor az elemzés egysége a kar (school), a SAT önmagában a teljesítményteszt-pontszámok varianciájának mintegy 70%-át magyarázza. A SAT és az erőfeszítésre vonatkozó saját beszámolóik további mintegy 3–7%-t magyaráznak együtt a mintától függően (pl. elsőévesek vagy végzős hallgatók). Az erőfeszítésre vonatkozó saját beszámolókat természetesen befolyásolhatja a hallgatók benyomása arról, hogy mennyire jól oldották meg a feladatot, például azt mondják, hogy nem igyekeztek nagyon, ha úgy érzik, nem teljesítettek jól. Azzal is lehet érvelni, hogy a CLA inkább a tipikus, mint a maximum teljesítmény mérésére szolgál (*Cronbach*, 1990.) Ha így van, akkor a hallgatói motivációt tekinthetjük olyan eredménynek, hogy egyes intézmények sikeresebben nevelik diákjaikat arra, hogy legyenek büszkék munkájukra, és mindig minden tőlük telhetőt tegyenek meg.

10 Ez szemben áll az általános- és középiskolai szintű No Child Left Behind (NCLB = Egy gyermek sem marad le) program filozófiájával, amely egy intézmény adott szintű jártasságot elérő diákjainak százalékára összpontosít (a fejlődés mértékétől függetlenül).

11 Valójában egyes intézmények belépő hallgatói olyan jó képességekkel rendelkeznek, hogy jobb teljesítményt érnek el, mint más intézmények végzősei. Az értékelés hozzáadott-érték megközelítésének alkalmazásával azonban a viszonylag kevésbé jó képességű belépő osztályokkal rendelkező intézmények is annyi hallgatói tanulmányi fejlődést tudnak mutatni, mint a legjobb képességű elsőéves osztályokkal rendelkező intézmények. Ráadásul a hallgatói pontszámok jóval a lehetséges maximum pontszám alatt vannak még a legmagasabb pontszámokat elérő hallgatókkal rendelkező felsőoktatási intézményekben is. Más szóval a plafonhatások nem jelentenek gondot, mert mindenki számára rengeteg lehetőség van a növekedésre.

Ez úgy történik, hogy kiszámolják az intézmény valós átlag CLA pontszáma és a hallgatóitól elvárt átlag közti különbséget.¹² Az elvárt átlagot az átlagos CLA és SAT pontszámok közti erős empirikus együttjárásból származtatjuk a CLA programban részt vevő összes iskolában (pl. *Raudenbush*, 2004). Ezután minden felsőoktatási intézménynél kivonjuk az elsőéves hallgatók várt átlagpontszámát a végzős hallgatók várt átlagpontszámából. Ezzel a hozzáadott-érték megközelítések többségéhez hasonlóan (*Wainer*, 2004) egy egyszerű összegző modellt feltételezünk: azt, hogy a négyéves képzés (és tapasztalat) hatása egyenlő növekményt ad hozzá a hallgatók eredményeihez. Az e célra használt regressziós egyenletek az elsőéves és a végzős hallgatók esetében közel azonos meredekségűek, de meglehetősen eltérő metszéspontúak (a részleteket lásd a CLA Test Report-ban: www.cae.org).¹³

Való igaz, hogy előfordulhatnak problémák a hozzáadott-érték (és egyéb) tanulás-értékelési mutatókkal, ha a módszerek mögött meghúzódó feltételezéseket jelentős mértékben megsértik (*Banta és Pike*, 2007, *Braun*, 2005). Például a CLA megközelítése feltételezi, hogy a CLA és a SAT pontszámok közti kapcsolat az egy iskolában vizsgált hallgatók mintái között lineáris, amit a kognitív mérések közti hasonló kapcsolat hosszú története is alátámaszt (a humán kognitív mérések olvasható áttekintését lásd: *Martinez*, 2002). A megközelítés azt is feltételezi, hogy a vizsgált hallgatói mintában a CLA és a SAT pontszámok között talált kapcsolat általánosítható az adott intézmény összes hallgatójára az adott változók között. Az elemzés továbbá azt is feltételezi, hogy a CLA programban részt vevő intézmények mintája általában reprezentatív az intézmények országos körére.

Amint azt a CLA teszt beszámolóiban megjegyeztük (CLA, 2006), empirikus bizonyítékok igazolják, hogy e két feltétel teljesül. Például, ha az intézmény az elemzési egység, akkor a CLA-t megíró elsőéves hallgatók átlagos SAT pontszáma és az intézmény összes elsőéves hallgatójának átlag SAT pontszáma között a korreláció 0,96 (az IPED-ek szerint). Tehát úgy tűnik, az intézmények nem válogatják ki a legjobb hallgatókat a CLA-ra.

Emellett szoros összefüggést lehet felfedezni az Education Trust IPED adatbázisában szereplő mintegy 1400 intézmény jellemzői és a CLA programban részt vevő több mint 100 intézmény tulajdonságai között.¹⁴ Például, mint az 1. táblázatban látható, a 2005–2006-os tanévben az átlagos négy év utáni diplomaszerezés aránya átlagosan a két intézményi csoportban 36% és 38% volt, és mindkét csoport azonos 3,5-ös Baron szelektivitási osztályozást kapott.

12 A hozzáadott-érték becslésének szemléletét e cikkben nem tudjuk egészében áttekinteni, a kérdés átfogó tárgyalását lásd a *Journal of Educational and Behavioral Statistics* című folyóirat különszámában (2004, vol. 29, No. 1).

13 A CLA program követte a SAT módszerét az egyéni hallgatók pontszámának 1600-as maximumában, azaz a legmagasabb teljesítmény és esszé pontszám itt is 1600. Ez aggodalmakat váltott ki az esetleges mesterséges plafonhatások miatt a CLA esetében. Elemzésünk szerint az elsőévesek kevesebb mint 1%-a, illetve a végzős hallgatók kevesebb mint 3%-a éri el a korlátot jelentő maximumot. Így nem utalt semmi a jelentős plafonhatásra. Az eredmények szerint csak néhány intézmény javított besorolásán (például a „elvárt szint feletti”-ről „elvárt szint feletti”-re), amikor megszüntették a korlátot, annak ellenére, hogy nem változtattunk a regressziós egyenleteken, vagy a standard jövőbeli pontszám-számításokon, hogy elkerüljük az ilyen irányú aggályokat.

14 A College Results Online (Online Főiskolai Eredmények) nem tartalmazza az Egyesült Államok összes 4 éves, IV. fokozatra jogosult felsőoktatási intézményét. Csak az alábbi kritériumoknak megfelelő intézményeket tartalmazza: 1) Az állami, vagy magán nonprofit szektorba tartoznak. Ez kizárja a profitorientált 4 éves intézményeket, például a University of Phoenix-et. 2) Bejelentették a Graduate Rate Survey (Végzettség Arányának Felmérése) adatokat a 2004-es adatgyűjtéskor. 3) Szelektivitási osztályozást kaptak a Baron's Profiles of American Colleges (Baron Amerikai Főiskolák Bemutatása) 2005. évi kiadásában. E szigorítások mintegy 1400 intézményre korlátozzák az elemzés alá kerülő 4 éves intézmények körét. Ennek megfelelően a statisztikai elemzések eredményei és az intézmények körének leírásai eltérhet az összes 4 éves intézményre vonatkozó hasonló leírások eredményeitől. Mivel azonban a tárgyalt intézményi kör vette fel a 4 éves nem profitorientált, GRS cohort-ba tartozó bachelor programok összes hallgatóinak 96%-át, az ilyen variancia valószínűleg csekély.

1. táblázat. *Négyéves intézmények a CLA-ban és országosan az intézmények legfontosabb jellemzői szerint*

Intézmény jellemzője	Országos	CLA
Állami	36%	42%
Történelmileg fekete főiskola, vagy egyetem (HBCU) százalék	6%	10%
Pell támogatásban részesülő alapképzésben részt vevők átlagos százaléka	33%	32%
Négyéves képzésben diplomát szerzők átlagos aránya	36%	38%
Hatéves képzésben diplomát szerzők átlagos aránya	52%	55%
Elsőévesek átlagos megtartási aránya	75%	77%
Átlagos Baron szelektivitási osztály	3,5	3,5
Becsült átlagos közepes elsőéves SAT pontszám	1061	1079
FTE alapképzésben részt vevő hallgatók átlagos száma (kerekített)	4500	6160
Hallgatókhoz kapcsolódó átlagos kiadás/FTE hallgató (kerekített)	12 230 \$	11 820 \$

Forrás: College Results Online adatbázis, melyet az Education Trust kezel, amely az Egyesült Államok legtöbb 4 éves, IV. fokozatra jogosult felsőoktatási intézményét tartalmazza. Az IPED-ekből és más forrásból előállított adatokat az Education Trust engedélyével közöljük. Részleteket lásd: www.collegeresults.org/aboutthedata.aspx. Mivel nem minden intézmény jelentette be a táblázat összes adatát, az átlagok és százalékok esetenként kis mértékben eltérő nevezőkön alapulnak.

A hozzáadott-érték mutatói feltételezik, hogy a kiválasztási elfogultság kontroll alatt van – azaz a végzős hallgatók teljesítményének összehasonlítása a mérceként tekintett intézményekben „pártatlan” és a végzős hallgatók közötti minden, az értékelésen nyújtott teljesítményükön kívüli különbség kontrollált. Ezt véletlenszerű kiválasztás nélkül egyszerűen lehetetlen megvalósítani, és a véletlenszerű kiválasztást megközelíteni próbáló súlyozási eljárások rendkívül nehezen lennének használhatóak a hozzáadott-érték, vagy más pontozási rendszerekben (pl. *Braun*, 2005). Az IPED adatainak elemzése mindazonáltal azt mutatta, hogy a CLA pontszámok számos intézményi jellemzővel (köztük a méret, Carnegie-osztályozás, kisebbségi százalék, állami-, illetve magánszektor, és a Baron szelektivitási mutató) nem jártak együtt szignifikánsan, ha az intézmény átlagos SAT pontszámának hatását kontrolláltuk.

Így, bár a tökéletestől messze áll, ez az eredmény azt mutatja, hogy a CLA pontszámokat valószínűleg nem alakítják oda nem tartozó tényezők (lásd *Raudenbush*, 2004, „A” típusú intézményi hatások).¹⁵

Végül nem tartható az a feltételezés, hogy kizárólag az intézmény, vagy egy intézményen belüli program lenne a hallgatók CLA-n, vagy bármely más tanulmányi értékelésen mutatott teljesítmény-változásának egyetlen „oka”. Bár a hallgatók életére egyik legjelentősebb hatással a felsőfokú tanulmányaik vannak, az egyéb tényezők (például az érés) is közrejátszhattak a felvétel és a diplomaszerezés közötti fejlődésben. Egyszerűen nem tudjuk meghatározni, hogy e növe-

¹⁵ Ezek az eredmények a 2005 őszi elsőéveseket vizsgáló 103 iskolával, valamint a 2006 tavaszán végzős hallgatókat tesztelő 91 iskolával lefolytatott regressziós modelleken alapulnak, ahol minden intézmény rendelkezett IPEDs adatokkal, és legalább 25 hallgatóval minden vizsgálatban. Az e modellel nyert eredmények jelzik, hogy a SAT önmagában a CLA pontszámok varianciájának 74%-át magyarázta az elsőéveseknél, míg a SAT és IPEDs változók kombinációja 75%-ra adott magyarázatot. Tehát az emelkedés mindössze 1% volt. A megfelelő százalékok a végzős hallgatók esetében 76, illetve 78 százalék. Ráadásul a várt értékek közti korreláció minden modellből 0,98 volt az elsőéveseknél, és 0,97 a végzősöknél. Röviden, az IPEDs változók beemelése a modellbe csekély, vagy semmilyen hatást nem gyakorolt a várt, vagy a hozzáadott-érték pontszámokra. Ez az eredmény jelzi, hogy semmilyen előnnyel nem járna, ha a CLA eredményeket intézménytípusonként mutatnánk be.

kedés milyen mértékben tulajdonítható a felsőfokú tapasztalatoknak, illetve más tényezőknek. Mivel azonban a CLA-n mutatott növekedés nagysága az elsőévesek és a végzős évfolyamok között általában annyival nagyobb a más méréseken tapasztalt mértéknél, hogy gyanúnk szerint a fejlődés jelentős részben a felsőfokú képzésben biztosított tanulmányi lehetőségeknek köszönhető.¹⁶

A fent tárgyalt szempontokat is figyelembe véve, választani kell: alapul vegyük, vagy ne vegyük alapul a hozzáadott-érték eredményeket. A CLA az ilyen pontszámok alkalmazása mellett áll ki, de mindig a fent említett nehézségeket mérlegelve, mert a hozzáadott-érték megközelítés ezen ellentmondások felismerése mellett fontos módszert biztosít az intézmények számára saját teljesítményük értékelésére. Ráadásul a CLA program hangsúlyozza, hogy a hallgatók tanulásáról kialakított mutatói nem az egyetlen információípust jelentik, amit a hallgatói és intézményi teljesítmény vizsgálatára és fejlesztésére fel kellene használni. Sőt, a CLA mutatók mindenképpen az intézmény vagy a képzési program teljesítményéről szóló információk egyik lehetséges forrása csupán. A CLA eredményeket más hallgatói tanulmányi mérések eredményeivel együtt kell felhasználni.

Banta és Pike (2006) az ETS elnöke és vezérigazgatója, *Kurt Landgraf* kijelentésére alapozták kritikájukat a hozzáadott-érték mutatók CLA általi használatával kapcsolatban. E szerint „A hozzáadott-érték modellekből származó eredmények nem szolgálhatnak következményekkel járó döntések alapjául. Bármely tisztességes és érvényes oktatói értékelési rendszerben szükség van más mérésekre is.” Ezt nem vitatjuk. A hozzáadott-érték eredményeket más mutatókkal együtt kell alkalmazni, de nincs ok a kizárásukra. Egész egyszerűen a CLA által mért kompetenciák a felsőoktatás létfontosságú részét képezik és a hozzáadott-érték megközelítés a legmegfelelőbb módszer a mérésekre. Ráadásul *Landgraf* utalt a hozzáadott-érték pontszámok használatára az oktatókkal kapcsolatos fontos személyi döntéseknél, amit a CLA nem javasol, és a CLA programban részt vevő intézmények nem alkalmazzák az eredményeket erre, vagy hasonló célokra.

A CLA és az NSSE hasonlóságai és különbségei

A Spellings-bizottság az intézmények számára megfontolandó mérési módszerekként említette a CLA-t, az MAPP-t és az NSSE-t. A CLA és a MAPP megközelítése közti különbségeket már ismertettük, melyek közvetlen tanulmányi mérések. Tehát most a CLA és az NSSE program közti fontos hasonlóságok és különbségek leírása következik, miközben összehasonlítjuk a közvetett és közvetlen tanulmányi méréseket.

A CLA és NSSE program egyaránt a részvételhez hozzájáruló hallgatók mintájára korlátozódik. A CLA a hozzáadott-érték modell fent ismertetett alkalmazásával kezeli ezt a megszorítást. Az NSSE más módszert követ. Véletlenszerű mintát választ egy intézmény hallgatóiból, de a hallgatók közül sokan nem vesznek részt, illetve fejezik be a vizsgálatot. Ezért az NSSE program önkéntes mintákkal kénytelen dolgozni, éppúgy mint a CLA.

Azonban a CLA-val ellentétben az NSSE nem súlyozza az eredményeit a válaszadók releváns háttér-jellemzőit figyelembe véve, ami értelmezési nehézségekhez vezet. Például, ha egy intézményben a hallgatók, akik az NSSE-n azt mondják, gyakran olvasnak a kurzusokon nem kijelölt könyveket, megtörténhet, hogy még jóval felsőfokú tanulmányaik megkezdése előtt olvastak ilyeneket. Egy másik intézmény hallgatói, akik azt mondják, hogy olvasnak a kurzusokon nem kijelölt könyveket, talán éppen akkor tettek így először.

¹⁶ A CLA program jelenleg egy empirikus kutatást végez, az érésnek az eredményekre gyakorolt hatásairól.

Kiderül, hogy az NSSE felismerni sem képes ezeket a különbségeket, nem hogy hatékonyan ki tudná azokat igazítani. Az intézmények egyszerű összehasonlítása a tekintetben, hogy hallgatóik mennyire mondják magukat „elkötelezettnek” a képzési folyamat iránt, ezért analóg az intézmények átlagos CLA pontszámainak összehasonlításával, a hallgatóknak a felsőfokú tanulmányaik megkezdése előtti átlagos képességszintjében mutatkozó különbségek kontrollálása nélkül. Ráadásul az egyik intézményben részt vett hallgatók az NSSE-n „elkötelezettebbek” lehetnek, míg azok, akik egy másik intézményben vesznek részt a vizsgálatban, kevésbé elkötelezettek lehetnek társaiknál. Az NSSE program nem tudja kizárni és nem képes kiigazítani sem ezeket a különbségeket.

A CLA (és más közvetlen tanulmányi mérések) kérdéseket vagy feladatokat ad, melyekre a hallgatók válaszai ismert pontossági szabványok szerint értékelhetők. Ezzel szemben az NSSE-re adott hallgatói válaszok pontossága nem vizsgálható. Az NSSE mintegy 85 kérdést tartalmaz arról, hogy a hallgatók milyen gyakran vesznek részt különböző tevékenységekben. E kérdések egy része konkrét és a válaszok tartalma bemérhető. Például, egy kérdés azt kérdezi, hány 20 oldalas, vagy azt meghaladó dolgozatot vagy beszámolót írt a hallgató. A választási lehetőségek: 1–4, 5–10, 11–20 és 20-nál több. Ez a kérdés egyértelmű, a válaszlehetőségek világosak és félreérthetetlenek. Az NSSE kérdéseinek többsége azonban nem rendelkezik ezen tulajdonságokkal. Például, az egyik tétel azt kéri a hallgatóktól, hogy számoljon be arról, milyen gyakran „használt elektronikus médiumot (levelezési lista, chat csoport, internet, azonnali üzenet stb.) egy feladat megbeszélésére vagy elvégzésére.” A válaszlehetőségek: „nagyon gyakran”, „gyakran”, „néha”, és „soha”. Válaszlehetőségek más kérdésekre: „nagyon”, „meglehetősen”, „némi-ig”, és „nagyon kicsit”. Azonban, ami az egyik hallgatónak „nagyon gyakran” (vagy „nagyon”) az lehet „néha”, (vagy „nagyon kicsit”) a másiknak, azaz lehet, hogy mindkét hallgató azonos gyakorisággal vesz részt az adott tevékenységben.

Az NSSE hallgatólagosan feltételezi, hogy a válaszlehetőségek ilyen értelmezésbeli különbségei véletlenszerűen oszlanak meg a hallgatók között. Ha azonban ezek nem véletlenszerűek, hanem a különböző intézményekbe járó hallgatók háttértulajdonságaival, és tapasztalataival együttjárnak, akkor ez komoly árnyékot vethet az NSSE eredményeinek értelmezésére. Ez a CLA, vagy más közvetlen mérések esetében nem jelent problémát.

A CLA és az NSSE talán legnagyobb különbsége az, hogy a CLA a hallgatói tanulmányok közvetlen mérése, míg az NSSE legjobb esetben is csak közvetett mérés. Ez a különbségtétel egyezik azzal, amikor a baseball játékos hazafutáshoz vezető ütési képességét értékelik, ahelyett, hogy mérnék a játékos által valószínűleg ütött hazafutások számát, illetve a játékos erejét, szem-kéz koordinációját, fizikai felépítését, és más vélhetően az erre való képességéhez kapcsolódó tulajdonságait. Az előbbi, közvetett megközelítés, amelyet az NSSE követ, indokolt lehet, ha ezen tulajdonságok kombinációja szorosan kapcsolódik az ütött hazafutások számához. Azonban kevés empirikus bizonyíték áll rendelkezésre, amely arra mutat, hogy az NSSE-re adott hallgatói válaszok akár csak közepes mértékben is együttjárjanak a jegyekkel, a CLA pontszámokkal vagy a CLA hozzáadott-érték pontszámokkal (*Klein, Kuh, Chun, Hamilton és Shavelson, 2005*).

E megfontolások alapján valószínűleg jobb, ha az NSSE-t annak fokmérőjeként tekintjük, hogy a hallgatók milyen mértékben vesznek részt önmagukban is fontos tevékenységekben, ahelyett, hogy a tanulmányi eredmény mérésének helyettesítőjeként fognánk fel. Ez a javaslat egybevág az ETS Culture of Evidence beszámolójában foglaltakkal (*Dwyer, Millett és Payne, 2006*).

A CLA statisztikai jellemzői

Banta és Pike (2006) és Kuh (2006) kérdéseket vetett fel a CLA pontszámai, és különösen a hozzáadott-érték („reziduális”) pontszámok megbízhatóságával kapcsolatban. E fejezetben áttekintjük a CLA pontszámok néhány, a kérdés tekintetében releváns statisztikai tulajdonságát. Különös hangsúllyal vizsgáljuk a kézzel és számítógéppel adott pontszámok összehasonlíthatóságát, hogy a különböző bírák mennyire megegyező pontszámot adnak egy adott válaszra; hogy az intézményi átlagokat és a reziduális pontszámokat milyen mértékben befolyásolja, hogy az elemzés egysége az intézmény; illetve, hogy a hallgatók CLA pontszámai milyen mértékben függenek a hallgató demográfiai jellemzőitől.

Klein (megjelenés előtt) 0,78 korrelációt talált a kézi és gépi átlagpontok között az „érvelj” és a „szedj ízekre egy érvelést” típusú feladatokban. Ez közel van a két bíráló közti 0,8–0,85 korrelációhoz. Emellett a válaszok pontozásához használt módszer (azaz: kézi, szemben a gépi pontozással) csekély, vagy semmilyen hatást sem gyakorolt az eredmények más mérésekkel, mint a SAT pontszámokkal vagy a jegyekkel való összefüggésére (lásd a 7. táblázatot, Klein et al., 2005). Ezek az eredmények egybevágóan mások hasonló feladatokra vonatkozó beszámolóival (pl. Powers et al., 2000, 2001).

Minden teljesítményteszt-választ kézzel pontoznak. A bírák semmilyen információval nem rendelkeznek a hallgatók tulajdonságairól, akiknek a válaszait pontozzák. A kézzel pontozott válaszok 10 százalékos véletlenszerű mintáját automatikusan és függetlenül ismét pontozza egy második bíráló, a bírák közti konzisztencia ellenőrzése és értékelése érdekében. A 90 perces teljesítményfeladatokra adott két kézi pontszám átlagos korrelációja 0,80, ami hasonló, mint a jogi szakvizsga esszéválaszainak osztályozásakor kapott érték (Klein és Bolus, 1983).

Kuh (2006) kritizálta a CLA-t az egyéni hallgatói pontszámok intézményi szintű összesítéséről. Azt hangsúlyozta különösen, hogy ebben az esetben „A hallgatói pontszámok hibamennyisége többszöröződik, és további hibát okoz az eredményekben, ami megnehezíti a megalapozott értelmezést” (9. old.). A méréselmélet valójában épp az ellenkezőjét jósolja (lásd Brennan, 1995); azaz a pontszámok nemhogy kevésbé, de sokkal inkább megbízhatóvá válnak, ha az eredményeket intézményi szinten összesítik. Ha így van, akkor a megbízhatóság növekedése a mérések magasabb szintű korrelációjához vezetne (feltéve, hogy számottevő változatosság van a különböző intézmények pontszámai között).¹⁷

Kuh aggregációs hatásokról alkotott véleményének vizsgálatához kiszámítottuk a SAT és a CLA pontszámok korrelációját a hallgatót véve az elemzés alapegységének, és ismét az iskolát véve összehasonlítási egységként. Ezt az elemzést 2005 ősze óta 113 iskolában több mint 10 000 elsőéves, és 2006 tavasza óta 90 iskolában több mint 4000 végzős hallgató adatai alapján végeztük. A legtöbb esetben egy adott hallgató egy teljesítményteszt-kérdést vagy a két esszé típus kombinációját oldotta meg. Így az intézmény teljes pontszáma egyszerűen a teljesítménytesztek és az esszé pontszámok átlaga volt.

A 2. táblázat világosan mutatja, hogy Kuh várakozásával ellentétben, de a pszichomterikus elmélettel összhangban a korrelációk mintegy 0,35-tel magasabbak (és háromszor annyi varianciát magyaráznak), amikor az elemzés egysége a hallgató helyett a felsőfokú intézmény. Ez a hatalmas növekedés az intézményi szintű pontszámok sokkal nagyobb megbízhatóságából ered.

17 Az intézményi szintű CLA és SAT pontszámok között nagy különbségek vannak, az intézményi szintű pontszámok 900–1400 között mozognak.

2. táblázat. A CLA pontszámok és a SAT pontszámok korrelációja osztályonként (elsőévesek, szemben a végzősökkel), és elemzési egységként (hallgató szemben az iskolával)¹⁸

CLA pontszám	Egység = Hallgatók		Egység = Intézmény	
	Elsőéves	Végzős	Elsőéves	Végzős
Teljesítmény	0,56	0,54	0,91	0,88
Esszé	0,44	0,46	0,79	0,83
Összes	NA	NA	0,88	0,88

Kuh kifejezte aggályait a CLA hozzáadott-érték (reziduális) pontszámainak használatával kapcsolatban is. Azt mondta, hogy az ilyen pontszámok használata „különösen problematikus az intézményi teljesítmény összehasonlításakor a tesztpontszámmal kapcsolatos hiba (pontatlanság) mennyisége miatt, az olyan kiszámíthatatlan tényezők, mint a teszthelyzet, a hallgató kedélyállapota és sok egyéb következtében. Ez azt jelenti, hogy van bizonyos mennyiségű ismeretlen hiba a teszteredményekben, mert a hallgató egyik nap elért „igazi” pontszáma eltérhet a következő napon nyújtott teljesítményétől” (9. old.). Kuh (2009) így folytatja:

„Ezek nem pusztán elméleti tünődések. Az NSSE kísérletezett saját hozzáadott-érték változatával, hogy kiszámolja a várható intézményi szintű pontszámokat. Alapos vizsgálatot követően... a továbbiakban nem biztosítjuk ezt az információt, mert félrevezető lehet, ha az intézmények közötti különbségek kifejezésére használják.”

Ellenőrizve, hogy Kuh problémái az NSSE-vel érvényesek-e a CLA-ra, meghatároztuk az összes olyan intézményt, ahol legalább 80 hallgató elvégzett minimum egy teljesítményfeladatot vagy mindkét esszétípust. Ezután két véletlenszerű (helyettesítés nélküli), 40 hallgatóból álló mintát vettünk minden iskolából, és A és B mintának neveztük el őket. Végül az iskolát az elemzés egységként alkalmazva kiszámítottuk az A és B minta átlagos CLA pontszámai, illetve az A és B hozzáadott-érték (reziduális) pontszámai közti korrelációt. Ezt az elemzést külön végeztük el abban a 62 iskolában, amely a fent ismertetett mintavételi követelményeknek megfelelő számú elsőévest vizsgált 2005 őszén, illetve abban a 44 iskolában is, amely megfelelő számú végzős hallgatót vizsgált 2006 tavaszán.

¹⁸ A 3. táblázatban látható magas intézményi szintű korrelációkat időnként félreértelmezik, hogy a CLA csak egy másik SAT. Ha létezik a SAT – folytatódik az érvelés –, akkor miért van szükség a CLA-ra? A magas korrelációk azonban nem jelentik azt, hogy a két teszt ugyanazt méri, azaz ugyanazt a gondolkodást, vagy érvelést, vagy „kognitív feldolgozást”. Gondoljuk át a következőket: Ha a CLA-ra készítenénk fel a hallgatókat, akkor a tankönyv nagyban különbözne a Princeton Review által használt tankönyvtől, melynek segítségével felkészítik a hallgatókat a SAT tesztre. Azaz, ha a felsőoktatásban oktatók a CLA feladatokra készítenék fel a hallgatókat, akkor azt a fajta kritikus gondolkodást, elemző érvelést és kommunikációs képességeket tanítanák, amelyeket főiskoláik küldetés nyilatkozata szerint tanítanak. Vagy máshogy fogalmazva, még ha a CLA nagymértékben korrelál a SAT-tal, nem helyettesítenénk egy hallgató SAT pontszámát az ő jegyével, vagy teljesítményével mondjuk egy Amerika történelem kurzuson – a SAT ugyanis éppúgy nem mond semmit a hallgató történelemtudásáról és érveléséről, mint a hallgató CLA teszten mutatott érveléséről.

A 3. táblázat bemutatja, hogy az átlagos CLA pontszámok és a reziduális pontszámok kellően megbízhatóak az intézményekkel kapcsolatosan meghozandó irányítói döntések megalapozására akkor is, amikor iskolánként csak 40 hallgatót vizsgálunk (és a hallgatók többsége a teljes három órás CLA vizsgálatnak csak a felét végezte el).¹⁹ Az is figyelemreméltó, hogy a 3. táblázatban szereplő együttthatók alulbecslik az intézményi átlagok valós megbízhatósági szintjét, mert egyik minta sem tartalmazott az egyes intézményekben vizsgált diákok létszámának felénél többet.

3. táblázat. Az A és B minták intézményi átlagai és reziduális pontszámai közti korreláció osztályonként

Pontszám típusa	Elsőévesek (N = 62 intézmény)	Végzősök (N = 44 intézmény)
Átlagos összpontszám	0,94	0,86
Reziduális pontszám	0,77	0,70

Az elsőévesek és a végzősök közti hozzáadott-érték (reziduális) pontszámok különbségének intézményi szintű megbízhatóságának vizsgálatához hasonló stratégiát alkalmaztunk, mint a 3. táblázat megszerkesztéséhez. Az elemzés kezdetén meghatároztuk mind a 91 olyan felsőoktatási intézményt, ahol 2006 őszén legalább 10 elsőévest és 2006 tavaszán legalább 10 végzős hallgatót vizsgáltak. Ezután az elsőévesek egyik felét a főiskolán véletlenszerűen az A mintába, a másik felét pedig a B mintába soroltuk. Majd a végzősök egyik felét szintén véletlenszerűen az A mintába, a másik felét pedig a B mintába soroltuk. Végül minden mintán belül kiszámítottuk az évfolyamok közti intézményi átlagos hozzáadott-érték (reziduális) pontszámok különbségét. Például, ha az elsőévesek átlagos hozzáadott-érték pontszáma egy felsőoktatási intézmény A mintájában 0,20 szórásnnyival a várt *alatt* volt, és a végzősök átlaga az intézmény A mintájában 0,15 szórásnnyival a várt *fölött* volt, akkor az A minta e két osztálya közti nettó különbség 0,35 szórásnnyi volt. Ezt követően elvégeztük a megfelelő számításokat a B mintán is.

A két reziduális különbségérték közti korreláció a 91 intézményben 0,63 volt, ami elég magas az intézmények mintáiban szereplő kisszámú hallgató (azaz egy főiskolán mintánként mindössze 5 azonos évfolyamra járó hallgató) és a különbségértékek (különösen reziduálisok esetében) megbízhatóságával kapcsolatos elméleti aggályok mellett.

A faji és etnikai csoportokat, illetve a nemeket összehasonlítva, az átlagos tesztpontszámok és jegyek különbségei gyakran vetnek fel az esetleges előítéletességgel kapcsolatos kérdéseket. Az ilyen aggályok feltárására két regressziós egyenletet alkottunk a hallgatók CLA teljesítmény-teszt pontszámának előrejelzésére. Az egyik egyenlet kizárólag a SAT pontszámot használta fel az előrejelzéshez. A másik egyenlet használta a hallgató SAT pontszámát, illetve figyelembe vette a faji/etnikai csoportot, a nemet és azt, hogy otthon az angol volt-e az elsődlegesen beszélt nyelv. Készült egy másik egyenletpár is ugyanezekkel a független változókkal a CLA esszé összpontszámai előrejelzésére. Ezt az elemzést körülbelül 10 000 olyan elsőévessel végeztük, akik 2005 őszén CLA tesztet írtak, illetve mintegy 4000 végzős hallgatóval, akik 2006 tavaszán írták meg a CLA-t.

A 4. táblázat azt mutatja, hogy a hallgatók demográfiai tulajdonságainak felvétele a regressziós modellbe csekély, vagy semmilyen hatást nem gyakorolt az osztály, vagy bármely típusú CLA teszt pontszám előrejelzési pontosságára. Az eredmény alátámasztja azt az elméletet, hogy a CLA mérései nem kedveznek egyik társadalmi csoportnak sem.

¹⁹ A CLA program azt ajánlja, hogy az intézmények osztályonként 100 hallgatót vizsgáljanak a teljes teszttel, de elfogadja, hogy sok intézmény nem képes elérni ezt a célt.

4. táblázat. A CLA pontszámok százalékos varianciája kizárólag a SAT pontszámok magyarázatával, és a SAT pontszámok, valamint a demográfiai változók kombinációjával (amikor a hallgató az elemzés egysége)

CLA pontszám és osztály	Csak SAT	SAT és demográfia	Különbség
Esszé			
elsőéves	20	22	2
végzős	23	25	2
Teljesítményteszt			
elsőéves	32	33	1
végzős	30	30	0

Összefoglalás és konklúziók

A Spellings-bizottság megerősítette a tanuláértékelés jelentőségét a felsőoktatási intézmények fejlesztése és értékelése céljából egyaránt. Ezzel együtt a CLA-t a tanulás mérésére alkalmas módszerek jó példajaként említette, és ez a publicitás jogosan helyezte ezt az értékelési programot alapos vizsgálat alá. E vizsgálat kezdeti beszámolóí részben tényeken, részben hiedelmeken alapuló állításokat fogalmaztak meg. Ezért jelen tanulmányunkban kísérletet tettünk arra, hogy rámutassunk: (a) hogy a CLA mit mér, és mit nem mér, (b) milyen megbízhatóan méri, amit állítása szerint mér, és (c) hogyan lehet megkülönböztetni más közvetlen (pl. MAPP) és közvetett (pl. NSSE) hallgatói tanulmányi mérésektől.

A CLA program a jelenlegi értékelést úgy tekinti, mint a felsőoktatási mutatók átfogó rendszerének egy nagyon fontos, lényegi részét. Más méréseknek, például egy tudományágon belüli teljesítmény, a társadalmi, erkölcsi, és személyes helyzetek, vagy a politikai szempontból releváns méréseknek is helyt kell kapniuk egy valóban átfogó rendszerben.

A CLA jelentősége abban áll, hogy azokra az átfogó képességekre összpontosít, amelyeket a főiskolák és egyetemek szinte általánosan fejleszteni kívánnak hallgatóikban: a kritikus gondolkodás, az elemző érvelés, a problémamegoldás, és a kommunikációs képességek. Újszerűségét az adja, hogy holisztikus, „valóságű” feladatokkal vizsgálja azt, amit az oktatók, hallgatók és a nyilvánosság egyaránt a felsőfokú képzés fontos eredményének tart. Emellett és más közvetlen hallgatói tanulási mérésekkel ellentétben, a CLA teljes mértékben nyílt mérésekre támaszkodik, és az összetett feladatot nem darabolja fel az azt alkotó kompetenciákra, hogy különálló képesség-pontszámokat adjon. Ez azért van így, mert a CLA szerint ezzel kiemelnénk az értékelést a valóságból, és egy ilyen értékelés részeinek összesítése nem ragadja meg az egész összetettségét. Tehát a CLA egy valóságű teljesítményen és írásbeli feladatokon alapuló, egészen kialakított értékelési módszer. Az új technológiák előretörése – köztük az internetes lebonyolítás és pontozás, valamint a természetes nyelvet feldolgozó szoftver elérhetősége a pontozáshoz – megvalósíthatóvá teszik az értékelést (költségeit és időigényét tekintve) ott is, ahol ez a múltban nem volt lehetséges.

A CLA és más értékelési programok az ezeket használó főiskolákkal és egyetemekkel együtt mind szembesülnek a hallgatók együttműködésének nehézségeivel a mérésekben, ami viszont azt jelenti, hogy a részt vevő hallgatók mintája nem feltétlenül reprezentatív. A CLA a hozzáadott-érték megközelítéssel kezeli ezt a problémát, azaz a felsőfokú teljesítménypontszámokat hozzáigazítja a részt vevő hallgatók belépési szintű képességeiben mért különbségekhez (a SAT pontszámaik alapján). Ez a megközelítés feltételezi, hogy a CLA és a SAT pontszámok kapcsola-

ta az egy iskolában vizsgált hallgatók mintájában megfeleltethető ezen változók kapcsolatának az adott intézmény összes hallgatója körében. Az elemzés azt is feltételezi, hogy a CLA programban részt vevő intézmények mintája összességében reprezentatív az intézmények országos megoszlására is. Empirikus bizonyítékok mutatják, hogy mindkét fenti feltétel teljesül (CLA, 2006). Bár ez a megközelítés nem optimális és a szelektivitás mindig aggodalmat kelt, de praktikusán kezelhető, s az így szerzett információk értékére tekintettel véleményünk szerint jobb ezt választani, mint hogy nem korrigáljuk az inputot, vagy kérdéses érvényességű közvetett mérésekre hagyatkozunk.

A fent ismertetett nehézségek mutatják az oktatási értékelés bizonyosságával kapcsolatos korlátokat. Az oktatási értékelés valós világában a kérdés azonban az, hogyan kell a megközelítéseket fejleszteni, hogy maximalizáljuk az irányelvek szempontjából releváns kapcsolatok és minták azonosításának valószínűségét a változók között úgy, hogy az oktatásban és a pedagógiai fejlesztésben hasznos információt biztosítsunk az oktatók számára. A különböző megközelítéseket csak akkor lehet hitelesen megítélni, ha megmutatják, hogyan kezelik e korlátok következményeit az alkalmazott értékelési módszerek, az elemzési egység tekintetében, illetve abból a szempontból, hogy milyen elemzéseket végeznek, és azok eredményeit hogyan értelmezik és alkalmazzák. A különböző megközelítéseknek emellett megvalósítási protokollokat kell adniuk, melyek segítik az oktatói közösséget, akik számára a megközelítéseket kialakították, hogy kezeljék a mintával és kiválasztási elfogultsággal, illetve az eredmények félreértésével és a téves állításokkal kapcsolatos aggályokat. Végül fontos, hogy a megközelítés minden oldalát világossá tegyük, hogy a minden egyéni stratégiából (melyek közt vannak jobbak és rosszabbak) származó oktatási értékelés módszertani korlátai közti különbségtételen alapuló fejlesztési javaslatokat kapjunk.

A CLA jelzi az intézménynek, hogyan teljesít a mércének tekintett intézményekhez képest. Az egyetemeknek és főiskoláknak e jelzés mellett további információkra, például intézményi értékelésekre és IPED adatokra is szüksége van, hogy meghatározzák az esetleges kihívásokat, amelyeket kezelni kell. Számos intézmény a bevezetett reformok, vagy új programok hatásait értékeli a CLA segítségével, hogy lássák, működik-e a „kísérlet”.

Kérdések merültek fel a CLA technikai jellemzőit illetően, így a feladatok számítógéppel történő pontozása egyes vélemények szerint nem megfelelő, illetve a CLA pontszámai megbízhatatlanok. Az itt és máshol bemutatott empirikus adatok ellentmondanak ezen állításoknak. Az emberi és gépi pontozás megbízhatósága közel egyező; az intézményi átlagok és reziduális pontszámok megbízhatósága szintén megfelelő (előbbi 0,80 fölötti, utóbbi 0,70). Az átlagos reziduális pontszámok évfolyamok közötti különbségének megbízhatósága az iskolán belül szintén megfelel a célokra, amelyekre a mérést használják.

Összegzésül elmondható, hogy a CLA megbízhatóan mér bizonyos fontos tanulmányi eredményeket a felsőoktatási alapképzésben. Megközelítése eltér a többi közvetlen tanulmányi méréstől, hiszen holisztikus, nyílt végű, valószerű feladatokat mintáz az értékelésben, és a hallgatók a való életben is előforduló helyzetekhez hasonlóan alkotják meg válaszaikat. Ez a megközelítés eltér az egyéb közvetlen tanulmányi mérésektől, például azoktól, amelyek adott válaszokkal (feleletválasztós) szerkesztett tesztek alkalmaznak, és az összetett feladatokat részfeladatokra bontják, majd ezeket összegezve próbálják megragadni az egészet. Az élet nem feleletválasztós kérdésekbe csomagolva tárul elénk: nem hisszük, hogy az egész a részek pusztá összege. Ezen felül a CLA és más közvetlen tanulmányi mérések meglehetősen elütnek az olyan közvetett mérésektől, mint az NSSE. Nem a hallgatók saját beszámolóira támaszkodnak a teljesítményről, hanem közvetlenül mérik, hogy mire képes a hallgató, amikor valóságghú kognitív

feladattal kerül szembe. A divatos állításokkal ellentétben a közvetlen és a közvetett hallgatói tanulmányi méréseket egyaránt befolyásolhatja az, hogy mely hallgatók vesznek részt, illetve nem vesznek részt az értékelésben, de csak a közvetlen értékelések rendelkeznek megvalósítható statisztikai módszerekkel e probléma kezelésére. Végül a CLA mérések statisztikai jellemzői megfelelnek az általánosan elfogadott technikai minőségi szabványoknak, vagy felülműlják azokat. Ezt megállapítva, azonban mindig van lehetőség bármely értékelési rendszer kutatására és fejlesztésére, és a CLA törekszik ilyen kutatási-fejlesztési terv megvalósítására.

Irodalom

- AAC&U. (2005): *Liberal education outcomes*. Washington, DC, Association of American Colleges and Universities
- AASCU (Spring 2006): *Value-added Assessment. Perspectives*. Washington, DC, American Association of State Colleges and Universities
- Banta, T. W. – Pike, G. R. (Jan–Feb. 2007): Revisiting the blind alley of value added. *Assessment Update*, 19 (1), pp.
- Braun, H. J. (2005): *Using student progress to evaluate teachers: A primer on value added models*. New Jersey, Educational Testing Service
- Brennan, R. L. (1995): The conventional wisdom about group mean scores. *Journal of Educational Measurement*, 32 (4), 385–396.
- CLA (2006): Lásd: www.cae.org/cla
- Carroll, J. B. (1993): *Human cognitive abilities: A survey of factor-analytic studies*. New York, Cambridge University Press
- Cronbach, L. J. (1990): *Essential of psychological and educational testing*. 5. kiadás. New York, Harper Collins
- Dwyer, C. A. – Millett, C. M. – Payne, D. G. (2006): *A Culture of Evidence: Postsecondary assessment and learning outcomes*. Princeton, N. J.: Educational Testing Service.
- Erwin, D. – Sebrell, K. W. (2003): Assessment of critical thinking: ETS's tasks in critical thinking. *The Journal of General Education*, 52 (1), 50–70.
- Ewell, P. T. (1994): *A policy guide for assessment: Making good use of the Tasks in Critical Thinking*. Princeton, NJ: Educational Testing Service
- Klein, S. – Kuh, G. – Chun, M. – Hamilton, L. – Shavelson, R. (2005): An approach to measuring cognitive outcomes across higher-education institutions. *Research in Higher Education*, 46, #3, 251–276.
- Klein, S. – Bolus, R. (1982): *An analysis of the relationship between clinical skills and bar examination results*. Report prepared for the Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners
- Klein, S. (1983): Relationship of bar examinations to performance tests of lawyering skills. Paper presented to the American Educational Research Association, Montreal, April. (reprinted in *Professional Education Researcher Notes*, 1982, 4, 10–11.)
- Klein, S. (Megjelenés előtt): Characteristics of hand and machine-assigned scores to college students' answers to open-ended tasks. In *Festschrift for David Freedman*, D. Nolan and T. Speed, editors: Beachwood, OH. Institute for Mathematical Statistics
- Kuh, G. (2006): *Director's Message in: Engaged Learning: Fostering Success for All Students*. Bloomington, Indiana, National Survey of Student Engagement
- Landgraf, K. (2005): *Cover letter accompanying the distribution of Braun (2005) report*

- Martinez, M. E. (2000): *Education as the cultivation of intelligence*. Mahwah, NJ. Erlbaum
- McClelland, D. C. (1973): Testing for competence rather than for "intelligence." *American Psychologist*, 28 (1), 1–14.
- Powers, D. – Burstein, J. – Chodorow, M. – Fowles, M. – Kukich, K. (2000): *Comparing the validity of automated and human essay scoring* (GRE No. 98-08a, ETS RR-00-10). Princeton, NJ: Educational Testing Service
- Powers, D. – Burstein, J. – Chodorow, M. – Fowles, M. – Kukich, K. (2001): *Stumping erater: Challenging the validity of automated scoring*. (GRE No. 98-08Pb, ETS RR-01-03). Princeton, NJ: Educational Testing Service
- Raudenbush, S. W. (2004): What are value-added models estimating and what does this imply for statistical practice. *Journal of Educational and Behavioral Statistics*, 29, 121–129.
- Shavelson, R. J. (2007a): Assessing student learning responsibly: From history to an audacious proposal. *Change*
- Shavelson, R. J. (2007b): *Student learning assessment: From history to an audacious proposal*. AAC&U.
- U.S. Department of Education (2006): *A test of leadership: Charting the Future of U.S. Higher Education*. Washington, D.C.
- Wainer, H. (Ed.) (2004): Added-value assessment special issue. *Journal of Educational and Behavioral Statistics*, 29, whole