

Felsőfokú Tanulási Eredmények Értékelése (CLA): Tények és hiedelmek¹

Stephen Klein – Roger Benjamin – Richard Shavelson – Roger Bolus

Az USA oktatási kormányzata (Secretary of Education) által felkért, A Jövő Felsőoktatása Bizottság szerint (lásd U.S. Department of Education, 2006) – mely Spellings-bizottság néven vált ismertté – a CLA „az egyik legátfogóbb országos kezdeményezés annak felmérésére, hogy a hallgatók valójában mekkora tudásra tesznek szert a különböző felsőoktatási intézményekben”, illetve a CLA „elősegíti a bizonyítékon alapuló értékelés kultúrájának terjedését a felsőoktatásban” (22. o.). A bizottság ezen túlmenően javasolta, hogy „a felsőoktatási intézmények mérjék hallgatóik tanulási eredményeit a Collegiate Learning Assessment által kínálthoz hasonló eljárások alkalmazásával, melyek a diákok tanulmányi fejlődését követik nyomon az egyes felsőoktatási intézményekben...” (23. o.). Az Amerikai Főiskolák és Egyetemek Szövetsége, egyetértésben az Amerikai Állami Főiskolák és Egyetemek Szövetségével (AACSB) (2006, 4. o.), arra jutott, hogy:

A hozzáadott értéket mérő közvetlen értékelés legjobb példája a Felsőfokú Tanulási Eredmények Értékelése (CLA), mely a RAND Hozzáadott Érték Értékelési Kezdeményezéséből (Value Added Assessment Initiative – VAAI) nőtt ki, és 2004 tavasza óta áll a főiskolák és egyetemek rendelkezésére. A teszt túlmutat a feleletválasztós formán, életszerű feladatok elvégzésével méri a teljesítményt, amelyekben a hallgatóknak többféle kiinduló dokumentumot kell elemezni, és írásban (például emlékeztető, ajánlás formájában) adják meg a választ.

Ha a hallgatói tanulási eredmények egy viszonylag újszerű értékelése ilyen dicséretet és figyelmet kap, jogos, hogy sokan górcső alá veszik. Megkérdézik, hogy mit mér, mik a szakmai értékei és mi a haszna. Különböző bizottságok tekintik át eljárásait, erősségeiről és gyengeségeiről – néha – versenytársak formáinak általában megalapozott, olykor kevésbé megalapozott véleményt. Banta és Pike (2007) például azt vitatta, hogy a CLA által alkalmazott, a hozzáadott érték elvén alapuló tanulásértékelés vajon helyes-e, Kuh (2006) pedig kételyeit fejezte ki a CLA megbízhatóságával és hasznosságával kapcsolatban.

¹ Jelen szöveg a Kiss P. (szerk., 2010) *Kompetenciamérés a felsőoktatásban* című, az Educatio Nonprofit Kft. által kiadott kötetben megjelent fordítás alapján készült, annak rövidített és további szöveggondozáson átesett változata. A szerzőkről: Stephen Klein és Roger Benjamin, Council for Aid to Education; Richard Shavelson, Stanford University; Roger Bolus, Research Solutions Group.

A továbbiakban ezért röviden bemutatjuk a CLA-program értékelésnél, pontozásnál érvényesített alapelveit, az általa alkalmazott mutatókat, illetve a CLA-módszer és más kiterjedt mérési programok közti hasonlóságokat és eltéréseket. Végül pedig a CLA-mérések néhány fontos statisztikai jellemzőjéről, többek között a megbízhatóságáról szólnunk.

A CLA értékelési szemlélete

Elsőként azt vizsgáljuk, hogy milyen tanulási eredményeket mér és mit nem mér a CLA. Emellett tárgyaljuk a CLA szemléletmódját a tanulási eredmények értékelésében, amely – amellett, hogy többéves, sikeres története van (Shavelson, 2007a,b) – lényegesen különbözik az általános képzés² olyan tipikus méréseitől, mint az ETS³ hagyományosabb, Measure of Academic Proficiency and Progress (MAPP) nevű mérési módszere.

Amit a CLA mér, és amit nem

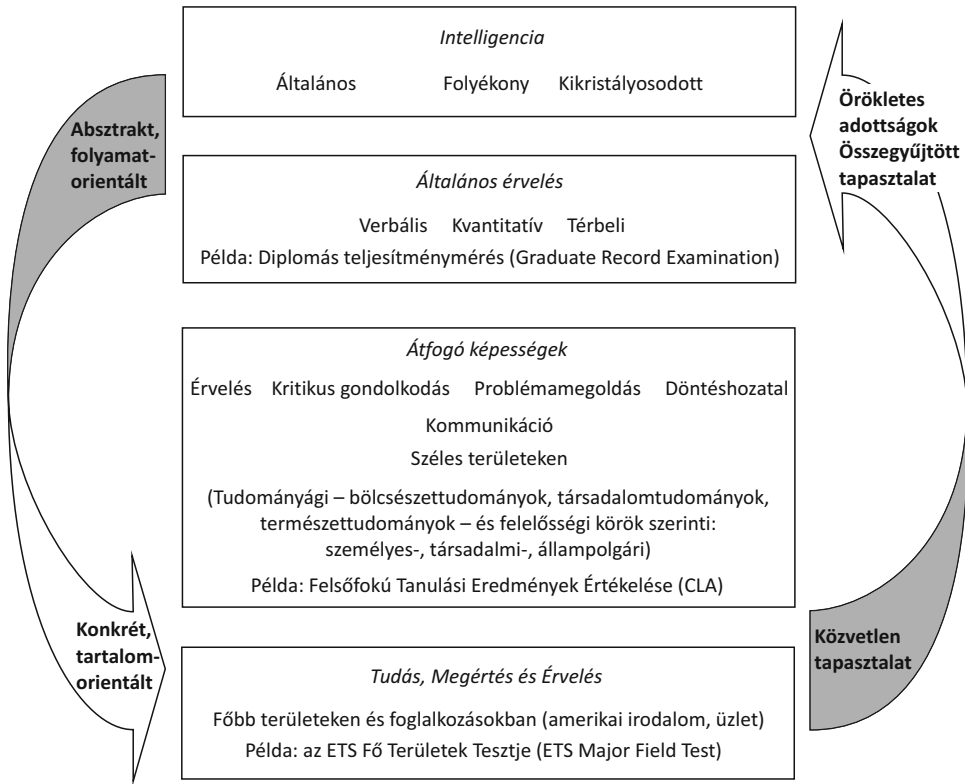
A tanulási eredmények igazán átfogó értékelése magába foglalná az *1. ábrán* bemutatott készségek többségét vagy akár mindegyikét, de nincsen olyan mérési eljárás, amely képes lenne a középfokú szint feletti oktatás összes fontos hatását mérni. Ez azonban nem zárja ki, hogy legalább néhány aspektust értékeljünk. Ezt az alapelvet szem előtt tartva a CLA-t úgy terveztük, hogy felmérje a hallgatók kritikus gondolkodását, elemző érvelését, problémamegoldó képességét és írásbeli kommunikációs kompetenciáit. A CLA (legalábbis kezdetben) ezekre az „átfogó képességekre” összpontosított, mert ezek minden szakon jelen vannak és szerepelnek szinte minden felsőoktatási intézmény küldetésnyilatkozatában. Ezek a képességek idővel, a szakspecifikus képzés, az általános képzés és a hallgatók által fejlesztett és magukkal hozott általános képességek közti kölcsönhatásban alakulnak ki (Shavelson, 2007a).

A CLA-program nem mér sem általános szellemi képességet, sem intelligenciát, és nem is törekszik erre. A CLA nem méri a különböző tudományágak megértését és az azokhoz kapcsolódó tartalmi vagy művelési tudást sem. Egyetértünk Dwyer, Millet és Payne (2006) véleményével, hogy hasznos lenne, ha rendelkeznének eszközökkel a 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

² Némi bizonytalanság övezi az általános képzés fogalmát, mert a felsőoktatásban 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 általános képzésre utalunk e címszóval.

³ Educational Testing Service (<http://www.ets.org>), melyet 1947-ben alapítottak, és olyan világhírű teszteket fejleszt és szolgáltat, mint a TOEFL (a szerk.).

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



hozni, és egy közös skálára lehet kalibrálni ilyen méréseket, akkor valószínű, hogy azok a CLA tesztszűzjeivel együtt (nem azok helyett) lennének használhatóak.

A CLA-ban alapvetően nem az egyes hallgató, hanem az intézmény az elemzési egység. Célja, hogy összegző értékelést nyújtson az intézmény (egészében szemlélt) képzési és egyéb programjai által *hozzáadott értékről*, bizonyos fontos tanulási eredmények szempontjából. E mérések kitűzött célja, hogy jelzést adjanak a vezetőknek, tanszékeknek és hallgatóknak a fejlesztésre szoruló néhány kompetenciáról és a hallgatók által az intézményben elért általános teljesítményszintről. Kiemelten arról is, hogy ez a szint magasabb-e, alacsonyabb-e, mint ami az intézménybe érkező hallgatók képességszintje alapján várható.

A CLA önmagában nem határozza meg azokat az okokat, amelyek következtében egy intézmény hallgatói a vártnál jobban vagy rosszabbul teljesítenek, arra sem tesz javaslatot, hogy milyen tantervi vagy más változtatásokat kellene megvalósítani a hallgatók teljesítményének növelése érdekében. A felsőfokú intézménynek saját tanszékeire, akkreditációs bizottságainak 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 át standardizálva vannak – segítségükkel vizsgálhatók (más

mutatókkal együtt) az intézmény által bevezetett reformok és irányelvek hallgatói teljesítményre gyakorolt általános hatásai.⁴

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 mérésén*, és részben érinti az *1. ábrán* bemutatott egyéb területeket is. A közvetlen mérésen azt értjük, hogy a hallgatóknak ténylegesen el kell végezni bizonyos kognitív kihívást jelentő feladatokat, 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. A Spellings-bizottság megjegyezte például, hogy a Hallgatói Elköteleződés Országos Felmérése (National Survey of Student Engagement – NSSE) vizsgálja a „diákok részvételét a tanulmányaik és fejlődésük elősegítésére tervezett programokban, és az ezek melletti elköteleződésüket. A hallgatói részvétel mutatóit, melyek a tanórák alatt és azokon kívül a hallgatók tanulmányokra fordított idejét és erőfeszítéseit fejezik ki (az oktatókkal való megbeszélésektől a nem kötelező szakirodalmolvasásáig)”, úgy tervezték meg, „hogy *becslésként* szolgáljanak a hallgatók alapképzési tapasztalatainak értékéhez és minőségéhez” (US Department of Education 2006, 22. o., kiemelés hozzáadva). A közvetlen és közvetett mérések közti különbségről (és az NSSE-ről) a cikk folyamán még szót ejtünk.

Végül a felsőoktatás-politika szempontjából különösen fontos tényezők a felsőoktatásba történő felvétel, a tanulmányok költségei és a diplomaszerzési 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 számos fontos mutatója közül. Ideális esetben a tanulási eredmények egyéb közvetlen mérési módszereivel, a tanulmányok intézményi támogatásának közvetett eszközeivel és más, 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

⁴ A CAE kutatói sok a CLA-t alkalmazó intézménnyel társulva dolgoznak együtt, hogy meghatározzák és megvalósítsák a jobb CLA-eredményekhez vezető legjobb gyakorlatokat. Idetartozik a Council of Independent Colleges (Független Főintézmények Tanácsa), a University of Texas (Texasai Egyetem) rendszere, a Higher Education Commission of West Virginia (Nyugat-virginiai Felsőoktatási Bizottság), állami kutatóegyetemek és a Lumina longitudinális csoport.

képességeit, követve az 1930-as évekre visszanyúló progresszív oktatási mozgalom hagyományát (Shavelson, 2007b). A CLA egyes feladatai az írásos kommunikációs képességeket hangsúlyozzák, míg mások valóság-hű, „munkából vett” teljesítményfeladatok. Minden feladat megfelel a különböző egyetemi alapképzési szakok és általános oktatási programok számára.

A CLA teljesítményfeladatai a Klein (1982, 1983) által a kaliforniai jogi szakvizsga számára fejlesztett eszközei, illetve a New Jersey Felsőoktatási Hivatal által kialakított „Kritikus gondolkodási feladatok” (Ewell, 1994; Erwin–Sebrell, 2003) alapján készültek. Ebben a megközelítésben az elvégzendő feladatok valóság-hű munkafeladatok, melyeket az oktatásban, munkában, irányításban és mindennapi gyakorlatban előforduló tevékenységekről mintáztunk. A feladatok teljesítéséhez kritikus és elemző gondolkodásra van szükségük a válaszadónak. A kapott információk alapján világosan, megfelelő igazolást adva kell megfogalmazniuk döntéseiket, értékelésüket vagy javasolataikat (lásd McClelland, 1973).

Az információs technológia fejlődésének köszönhetően a feladatok gazdagságát anélkül lehet biztosítani, hogy a hallgatókat túlterheljék. Az értékelés minden feladatát egy interaktív internetes felületen adjuk meg, amely elektronikus adatfelvételt, illetve online értékelést és eredményközlést biztosít. Az esszé típusú feladatokra adott válaszokat jelenleg bírák pontozzák, akik átfogó képzést követően online pontozzák a válaszokat. Így az egész rendszer papírmentes, ami jelentősen csökkenti a költségeket. Úgy véljük, hogy néhány éven belül a teljesítményfeladatokat is számítógépes szoftver fogja majd pontozni.⁵

[...]

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 Test⁶) 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 külön-külön a kritikus gondolkodás,

⁵ 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.

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 eljárásban 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.

az elemző érvelés, a problémamegoldó képesség és a kommunikáció mérése lehetetlen lenne e feladatok összetettsége és egységessége miatt. Ehhez fel kéne adni életszerű munkafeladat-jellegüket. Az így létrehozott mesterséges részfeladatok szétszabdalják a valós helyzetek összetettségét, még akkor is, ha 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, 2007b). Ezek az értékelések sokdimenziós „fogalmi szerkezetből” indulnak ki – ahogy a CLA is –, de azután megpróbálják ezt 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.

Ez a „fogalmi szerkezetből” kiinduló megközelítés (ellentétben a CLA valós munkatevékenységből kiinduló megközelítésével) feltételezi, hogy a részek összeillesztésével pontos képet kaphatunk az egészről. E stratégia elsődleges előnyei, hogy külön pontszámok nyerhetők 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 tényt, 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, a gyorsítás és a fékezés.

Bár a CLA szemlélete alapvetően eltér a MAPP által alkalmazottól és más széles körben használt feleletválasztós tesztektől, mégis van néhány közös vonásuk. Például a CLA-felméréseket is szabványosított feltételek között adjuk ki a különböző intézményeknek, és a nyers pontszámokat a vezetői ciklusokat összehasonlítani képes skálapontokká alakítjuk. A CLA-program ezen jellemzői lehetővé teszik az intézmények számára, hogy megvizsgálják, hallgatóik milyen jól teljesítenek az idő múltával és más intézmények hasonló helyzetű hallgatóihoz képest.

Banta és Pike (2006) kijelenti, hogy „a felsőoktatásban a növekedés, fejlődés értékelésének sokkal ígéretesebb útja, ha a hallgatók általános és tudományág-specifikus tudását és képességeit vizsgáljuk a szakterületükön, mint ha a leggyakrabban alkalmazott, pusztán általános képességeket mérő teszteket vesszük alapul”. Egyetértünk velük abban, hogy a hallgatói eredmények értékelése pontosabb a tudományág-specifikus eredmények értékelésével, de ez nem jelentheti az olyan átfogó képességek felmérésének elutasítását, melyeket a felsőoktatási intézmények kimondottan fejleszteni igyekeznek, és a munkaadók is értékelnek. Ráadásul a felvett elsőévesek és a diplomát szerző végzős hallgatók teljesítményének különbségében mérhető átlagos fejlődés a CLA-n nagyobb, mint egy szórásnyi. Ez mindenképpen jelentős hatásméret, ami azt jelzi, hogy a CLA érzékeny a felsőfokú oktatás következtében kialakult változásokra.

A CLA megkülönböztető jegyei

Mint fentebb jeleztük, a CLA szemlélete eltér a legtöbb nagy tesztprogramétól. 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 tanuló a teljes kérdéshalmaznak csak

egy kis részére válaszol, de minden feladat megjelenik egy adott intézményben.⁸ Ez a „mátrixos mintavételi” stratégia nagymértékben csökkenti a tesztelés egy hallgatóra jutó terhet, 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. A hallgatók jellemzően önkéntesen vesznek részt ebben a programban, nem kötelező a válaszadás. Bár természetesen kívánatos lenne minden hallgatót vizsgálni vagy valóban véletlenszerű mintavételt alkalmazni minden iskolában, a felsőoktatási intézmények többségében azonban e megoldások egyike sem valósítható meg. A MAPP és az NSSE is küzd ezzel a problémával.

A főiskolák és egyetemek számos módszert alkalmaznak, hogy ösztönözzék hallgatóikat a CLA-teszt elvégzésére, például könyvtulványt, kurzusfelvételnél előnyöket, különböző nyereségeket ajánlanak. Mindazonáltal eddig nem volt kimutatható semmilyen értelmezhető összefüggés az ösztönzés típusa és a részvételi arány vagy a hallgatók motivációja között, hogy jól teljesítsenek a teszten. A teljesítménytesztet megíró hallgatók kitöltenek egy utólagos 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 5%-áról adnak számot, ami túl kevés, hogy jelentős hatással legyen az eredményekre.⁹

A hallgatóktól megkérdeztük azt is, hogy főként mi motiválta őket a CLA megírására. Válaszaikat az alábbi három csoportba rendeztük: (1) személyes fejlődés (erősségeim és gyengeségeim megértése, vagy hogyan viszonyul a pontszámom a többi diákéhoz); (2) ellenszolgáltatás (javadalmazás, ajándékutalvány, nyereseményhúzás nagy fődíjjal); és (3) az intézményi követelmények teljesítése (a kurzus vagy az egész intézmény értékelési tevékenységébe illeszkedve). A SAT-pontszámaik hatásának kontrollálása után az, hogy melyik típusú tényezőt említik, a hallgatói CLA-pontszámok varianciájának mindössze 1%-át befolyásolta.

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.

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 továbbá mintegy 3–7%-ot 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, mely szerint egyes intézmények sikeresebben nevelik diákjaikat arra, hogy legyenek büszkék munkájukra, és mindig minden tőlük telhetőt tegyenek meg.

Eredmények bemutatása a hozzáadott érték felől

A CLA-program egyik legfontosabb mutatója, hogy az intézmény hallgatói jobban, rosszabbul vagy körülbelül ugyanúgy teljesítettek, mint ahogy az induláskor mért kompetenciaszintjük ismeretében elvárható lenne. A program azt is vizsgálja, hogy az átlagos hallgatói teljesítmény fejlődése az iskolába történő beiratkozás és a diplomaszerezés között megfelel-e a más felsőoktatási intézményben tanuló, hasonló helyzetben lévő hallgatókéknak. A program tehát tájékoztatja az intézményeket, hogy a hallgatóik fejlődése megfelel-e a más intézményekben elérhető fejlődésnek (pl. Raudenbush, 2004). A CLA azt az elvet követi, hogy a hallgatók tanulmányi fejlődésének mérésére és az ahhoz való intézményi hozzájárulásra kell összpontosítani az értékelési programoknak a középfok utáni képzési szinten.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 kompetenciáihoz kell igazítani a pontozást minden intézményben.

Ez technikailag úgy történik, hogy kiszámoljuk 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

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.

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.).

elsőéves és a végzős hallgatók esetében közel azonos meredekségűek, de meglehetősen eltérő metszéspontúak (a részleteket lásd a CLA Test Reportban: www.cae.org).¹³

Való igaz, hogy előfordulhatnak problémák a hozzáadott érték (és egyéb) tanulásiértékelési mutatókkal, ha a módszerek mögött meghúzódó feltételezéseket jelentős mértékben megsértik (Banta–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 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 Az 1. táblázatban látható például, hogy a 2005–2006-os tanévben a 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.

¹³ A CLA-program követte a SAT módszerét az egyéni hallgatók pontszámának 1600-as maximumában, azaz a legmagasabb teljesítmény- és esszépontszám itt is 1600. Ez aggodalmat 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 szintű”-ről „elvárt szint feletti”-re), amikor megszüntették a korlátot, annak ellenére, hogy nem változtattunk a regressziós egyenleteken vagy a standard jövőbeli pontszámszámításokon, hogy elkerüljük az ilyen irányú aggályokat.

¹⁴ A College Results Online (Online Főiskolai Eredmények) nem tartalmazza az Egyesült Államok összes négyéves, IV. fokozatra jogosult felsőoktatási intézményét. Csak az alábbi kritériumoknak megfelelő intézményeket tartalmazza: 1. Az állami vagy magán nonprofit szektorba tartoznak. Ez kizárja a profitorientált négyéves intézményeket, például a University of Phoenixet. 2. Bejelentették a Graduate Rate Survey (Végzetek 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 négyéves intézmények körét. Ennek megfelelően a statisztikai elemzések eredményei és az intézmények körének leírásai eltérhetnek az összes négyéves intézményre vonatkozó hasonló leírások eredményeitől. Mivel azonban a tárgyalt intézményi kör vette fel a négyéves nem profitorientált, GRS cohortba 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 | 1 061 | 1 079 |
| FTE-alapképzésben részt vevő hallgatók átlagos száma (kerekített) | 4 500 | 6 160 |
| Hallgatókhoz kapcsolódó átlagos kiadás per FTE-hallgató (kerekített) | 12 230 \$ | 11 820 \$ |

Forrás: College Results Online adatbázis, melyet az Education Trust kezel, amely az Egyesült Államok legtöbb négyéves 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 kismértékben eltérő nevezőkön alapulnak

A hozzáadott érték mutatói feltételezik, hogy a kiválasztási torzítás kontroll alatt van – azaz a végzős hallgatók teljesítményének összehasonlítása a mércének tekintett intézményekben „pártatlan” és a végzős hallgatók közti 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 használhatók a hozzáadottérték- 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

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 modellekkel 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%. 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 hozzáadottérték-pontszámokra. Ez az eredmény jelzi, hogy semmilyen előnnyel nem járna, ha a CLA-eredményeket intézménytípusonként mutatnánk be.

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 teljesítményváltozásának egyetlen „oka”. Bár a hallgatók életére a felsőfokú tanulmányaik jelentős hatással vannak, az egyéb tényezők (például az érés) is közrejátszhattak a felvétel és a diplomaszerezés közti fejlődésben. Egyszerűen nem tudjuk meghatározni, hogy e növekedés 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, gyanúnk szerint a fejlődés jelentős részben a felsőfokú képzésben biztosított tanulmányi lehetőségeknek köszönhető.¹⁶

A fent tárgyalt szempontokat is figyelembe véve, választani kell: építsünk-e a hozzáadott érték eredményeire? A CLA az alkalmazásuk mellett áll ki, a fent említett nehézségeket mindig 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 információtypust 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) Kurt Landgraf (az ETS elnöke és vezérigazgatója) kijelentésére alapozták kritikájukat a hozzáadottérték-mutatók CLA általi használatával kapcsolatban. Eszerint: „A hozzáadott-érték modellekből származó eredmények nem szolgálhatnak súlyos következményekkel járó döntések alapjául. Bármely tisztességes és érvényes oktatói értékelési rendszerben szükség van más mérésekre is.” Ezt nem vitatjuk. A hozzáadottérték-eredményeket más mutatókkal együtt kell alkalmazni, nincs viszont ok a kizárásukra sem. Egyszerűen fogalmazva, a CLA által mért kompetenciák a felsőfokú képzés kulcsfontosságú összetevői és a hozzáadott-érték-megközelítés a legmegfelelőbb módszer a mérésekre. Ezen túl Landgraf utalt a hozzáadottérték-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 is közvetlen tanulmányi mérés, közöttük már bemutattuk a különbséget. Most a CLA- és az NSSE-program közti fontos hasonlóságok és különbségek leírása következik, miközben összehasonlítjuk a közvetett és közvetlen tanulmányi méréseket.

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

A CLA- és NSSE-program egyaránt a hallgatók önkéntes részvételen alapuló mintájára korlátozódik. A CLA a hozzáadottérték-modell fent ismertetett alkalmazásával kezeli ezt a megszorítást. Az NSSE más úton jár. Véletlenszerű mintát választ egy intézmény hallgatóiból, ám a hallgatók közül sokan nem vesznek részt, illetve hagyják félbe a vizsgálatot. Ezért az NSSE-program önkéntes mintákkal kénytelen dolgozni, éppúgy, mint a CLA.

Azonban a CLA-val ellentétben az NSSE nem illeszti az eredményeit a válaszadók releváns háttérjellemzőihez, ami értelmezési nehézségekhez vezet. Például ha egy intézményben a hallgatók, akik az NSSE-n azt mondják, gyakran olvasnak a kurzusokon fel nem adott könyveket, megtörténhet, hogy még jóval felsőfokú tanulmányaik megkezdése előtt olvastak ilyeneket. Egy másik intézményben pedig talán először olvasnak nem kötelező könyveket az így válaszolók.

Kiderül, hogy az NSSE felismerni sem képes ezeket a különbségeket, nemhogy hatékonyan ki tudná azokat igazítani. Az intézmények egyszerű összehasonlítása hallgatói „elkötelezettség” tekintetében olyan, mintha az intézmények átlagos CLA-pontszámait a felsőfokú tanulmányok megkezdése előtti átlagos képességszintben mutatkozó különbségek kontrollálása nélkül vetnének egybe. 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 sem kizárni, sem kiigazítani 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. Az NSSE kérdéseinek többsége azonban nem ilyen. Például az egyik tétel azt kéri a válaszadótól, hogy számoljon be arról, milyen gyakran „használt elektronikus médiumot (levelezési lista, chatcsoport, internet, azonnali üzenet stb.) egy feladat megbeszélésére vagy elvégzésére”. A válaszlehetőségek: „nagyon gyakran”, „gyakran”, „néha” és „soha”. Válaszlehetőségek más kérdésekre: „nagyon”, „meglehetősen”, „némi” 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 együtt járnak, akkor ez komoly árnyékot vethet az NSSE eredményeinek értelmezésére. Ugyanakkor 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 különbség olyan, mintha a baseballjátékos hazafutáshoz vezető ütési képességét a játékos erején, szem-kéz koordinációján, fizikai felépítésén és más, vélhetően ahhoz kapcsolódó tulajdonságain keresztül

értékelnék, ahelyett, hogy mérnék a játékos által valójában ütött hazafutások számát. A közvetett megközelítés, amelyet az NSSE követ, indokolt lehet, ha a mért járulékos tulajdonságok kombinációja szorosan kapcsolódik az ütött hazafutások számához. Azonban kevés empirikus bizonyíték áll rendelkezésre arra, hogy az NSSE-re adott hallgatói válaszok akár csak közepes mértékben is együtt járnának a jegyekkel, a CLA-pontszámokkal vagy a CLA hozzáadottérték-pontszámokkal (Klein et al., 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 et al., 2006).

[...]

Összefoglalás és konklúziók

A Spellings-bizottság megerősítette a tanulásértékelés 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. Az első kritikák részben tényeken, részben hiedelmeken alapuló állításokat fogalmaztak meg, ezért jelen tanulmányunkban kísérletet tettünk arra, hogy rámutassunk: *a)* 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 felméréseire úgy tekint, mint egyre (bár fontosra) a felsőoktatási mutatók átfogó rendszerében. Más méréseknek, például a szakterületen vagy társadalmi, erkölcsi és személyes kérdésekben nyújtott teljesítménynek is helyt kell kapniuk egy valóban átfogó rendszerben, együtt a szakpolitikai célokhoz mérve vizsgált felvételi arányokkal, képzéssel és diplomaszerezéssel.

A CLA jelentősége abban áll, hogy olyan átfogó képességekre összpontosít, amelyeket a főiskolák és egyetemek általánosan fejleszteni kívánnak hallgatóikban: kritikus gondolkodás, elemző érvelés, problémamegoldás és 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 nyitott kérdésekre támaszkodik, és az összetett feladatot nem darabolja fel az azt alkotó kompetenciákra, hogy különálló képességpontszámokat adjon. Ez azért van így, mert a CLA szerint ezzel kiemelnénk az értékelést a valóságból, és az így mért részek összeadásával nem ragadnánk meg az egész összetettségét. A CLA tehát egy valóságű teljesítményen és írásbeli feladatokon alapuló, egészlegesen kialakított értékelési módszer. Az új technológiák előretörése – köztük az internetes lebonyolítás és pontozás, valamint a természetes nyelvet feldolgozó szoftver elérhetősége a pontozáshoz – megvalósíthatóvá teszi az értékelés költségei és időigénye tekintetében 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 megnyerésének nehézségeivel a mérésekben, ami viszont azt jelenti, hogy a részt vevő hallgatók mintája nem feltétlenül reprezentatív. A CLA a hozzáadottérték-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 kapcsolata 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 elvi korlátokat. Az oktatási értékelés valós világában a kérdés azonban úgy merül fel, hogy milyen megközelítésekkel leszünk képesek a szakpolitikai célkitűzések szempontjából releváns kapcsolatokat és mintázatokat leginkább azonosítani a változók között úgy, hogy az oktatásban és a pedagógiafejlesztésben 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. Emellett gyakorlati protokollokat kell adniuk, melyek segítik az oktatói közösséget, akik számára a megközelítéseket kialakították, hogy kezeljék a mintával és kiválasztási elfoglaltsággal, illetve az eredmények félreértésével és a téves következtetésekkel kapcsolatos aggályokat. Végül világossá kell tennünk megközelítéseink minden oldalát annak érdekében, hogy a fejlesztésére vonatkozó javaslatok fogalmazódhassanak meg, melyek megkülönböztetik az oktatási teljesítménymérés általános módszertani korlátait és az egyes mérési stratégiákhoz kötődő sajátosságokat, melyek ezzel együtt is lehetnek kevésbé vagy inkább hatékonyak.

A CLA jelzi az intézménynek, hogy hogyan teljesít a mércének tekintett intézményekhez képest. Az egyetemeknek és főiskoláknak e jelzés mellett további információkra, például intézményi értékelésekre és IPED-adatokra is szükségük 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 reziduálisok megbízhatósága szintén megfelelő (előbbi 0,80, utóbbi 0,70 fölötti). Az átlagos reziduálisokban a vizsgált tanulói csoportok között mutatkozó különbségek megbízhatósága az iskolán belül szintén megfelel a mérési célokat is figyelembe véve.

Összegzésül elmondható, hogy a CLA megbízhatóan mér bizonyos fontos tanulmányi eredményeket a felsőoktatási alapképzésben. Megközelítése eltér a többi közvetlen tanulmányi

felméréstől, hiszen holisztikus, nyílt végű, valószerű feladatokat használ az értékelésben, és a hallgatók a való életben is előforduló helyzetekhez hasonlóan adják meg válaszaikat. Ez a megközelítés eltér azoktól a közvetlen tanulmányi mérésektől, amelyek megadott válaszokkal (feleletválasztós) szerkesztett tesztek alkalmaznak, és az összetett feladatokat részfeladatokra bontják, majd ezeket összegezve próbálják megragadni a jelenség egészét. Az élet nem feleletválasztós kérdésekbe csomagolva tárul elénk: nem hisszük, hogy az egész a részek pusztán összege. Ezenfelül 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 valóságghű kognitív feladattal kerül szembe. A divatos állításokkal ellentétben a közvetlen és a közvetett hallgatói tanulmányi értékeléseket egyaránt befolyásolhatja, hogy mely hallgatók vesznek részt, illetve nem vesznek részt a felmérésben, de csak a közvetlen értékelések rendelkeznek megvalósítható statisztikai módszerekkel a probléma kezelésére. Végül a CLA-mérések statisztikai jellemzői megfelelnek az általánosan elfogadott minőségi szabványoknak, vagy felülmúlják azokat. Ezt megállapítva azonban mindig van lehetőség bármely értékelési rendszer kutatására és fejlesztésére, és a CLA törekszik ilyen kutatási-fejlesztési terv megvalósítására.

Hivatkozások

- AAC&U (2005): *Liberal education outcomes*. Association of American Colleges and Universities, Washington, D. C.
- AASCU (Spring, 2006): *Value-added Assessment. Perspectives*. American Association of State Colleges and Universities, Washington, D. C.
- Banta, T. W., Pike, G. R. (2007): Revisiting the blind alley of value added. *Assessment Update*, Jan-Feb. 19. (1).
- Braun, H. J. (2005): *Using student progress to evaluate teachers: A primer on value added models*. Educational Testing Service, New Jersey.
- 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*. Cambridge University Press, New York.
- Cronbach, L. J. (1990): *Essential of psychological and educational testing*. 5. kiadás. Harper Collins, New York.
- Dwyer, C. A., Millett, C. M., Payne, D. G. (2006): *A Culture of Evidence: Postsecondary assessment and learning outcomes*. Educational Testing Service, Princeton, N. J.
- 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*. Educational Testing Service, Princeton, NJ.

- 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: Nolan, D., Speed, T. (eds.): *Festschrift for David Freedman*. Institute for Mathematical Statistics, Beachwood, OH.
- Kuh, G. (2006): *Director's Message in: Engaged Learning: Fostering Success for All Students*. National Survey of Student Engagement. Bloomington, Indiana.
- Landgraf, K. (2005): *Cover letter accompanying the distribution of Braun (2005) report.*
- Martinez, M. E. (2000): *Education as the cultivation of intelligence*. Erlbaum, Mahwah, NJ.
- 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). Educational Testing Service, Princeton, NJ.
- 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). Educational Testing Service, Princeton, NJ.
- 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.