

AZ AMERIKAI FELSŐOKTATÁSI ÉRTÉKELÉSI TRENDEK ÁTTEKINTÉSE

Paul Ewald

*Department of Social and Behavioral Sciences
University of New England*

Bevezetés

Ebben a dolgozatban az amerikai felsőoktatásban honos értékelési gyakorlatokat kívánom bemutatni. Az amerikai felsőoktatás értékelési rendszerét kísérő figyelem több tekintetben előzmény nélkül való. Ennek számos, egyenként meglehetősen összetett oka van. Ezek közül három:

Először, a felsőoktatás kiszélesedése az intézmények és az egy bizonyos helyen tanuló diákok számának növekedését, ezáltal a képzési programok és akkreditációs eljárások még nagyobb sokszínűségét eredményezte. A huszadik század folyamán a relatív növekedés és a stagnálás periódusai ciklikusan ismétlődtek, és minden egyes ciklus a színvonallal, a célokkal és az elszámoltathatósággal kapcsolatos újabb kérdéseket hozott magával. A tesztelés és értékelés sokféle gyakorlatát ajánlották.

Másodsor, a felsőoktatás egyre inkább támaszkodik állami erőforrásokra, miután mind a társadalomban betöltött szerepe, mind a beiratkozott hallgatók száma nőtt. Mivel az egyes államok a költségvetési pénzekből egyre nagyobb részt juttatnak a felsőoktatásnak, a köz egyre jobban érvényesíti akaratát a törvényhozásban és követel elszámoltathatóságot. Sok államban ennek eredményeképpen az intézményes értékelést előíró állami utasítások születtek. Az államok a költségvetési ösztönzőket gyakran az általuk kiadott utasításoknak való megfeleléssel kötik össze. Az állam és az intézmények ilyen jellegű kapcsolatát később részletesen tárgyalom.

Végül az értékelés néhány formája lehetővé tette az oktatók számára, hogy tanterveiket átalakítsák, dialógusba bocsátkozzanak egymással a tanítás és tanulás problémaival kapcsolatban. Elsősorban két okból foglalkozunk a középfokú oktatás utáni, nem pedig az alapfokú, illetve a középfokú oktatással. Az oktatás színvonala és általában az alap- és középfokú (K 12) oktatás miatti aggodalmak a 1980-as években nagyon erőteljesen jelentkeztek. Ezért számos államban sok osztályában kötelező értékelést vezettek be standardizált objektív mérések felhasználásával. Az elmúlt tíz év folyamán a fejlődés ilyen alakulása nem sok teret hagyott az innovációnak, így az alap- és középfokú értékelési eljárások (bár van néhány figyelemre méltó kivétel) gyakorlatilag nem is változtak. Valószínű, hogy az értékelés elméletében, gyakorlatában és technológiájában bekövetkező majdani változások a felsőoktatásból fognak kiindulni és innét „szivárognak le” az alap-

és középfokú oktatásba. Az értékelési szaktudásnak a felsőoktatásban való koncentráltságán kívül ez specifikusan a tanárképzésben bekövetkezett változásoknak is lesz majd köszönhető. Általánosabban, ha az értékelési eljárásoknak ebben a dolgozatban vázolt fajtái még szélesebb körben elterjednek, akkor a leendő gyakorló tanárok, mint minden más középiskola után továbbtanuló diák, többféle értékelési alternatívával találkozhatnak, melyekből jó néhányat átültetnek saját oktatási gyakorlatukba. A középfokú és a középfok utáni oktatást a dolgozatban később részletesen tárgyaljuk.

A dolgozat négy részből áll. Az első rész háttéranyagot szolgáltat, felvázolja azokat az eseményeket, amelyek elvezettek az állami szerepvállalásig és a kötelező vizsgákig. A második részben az alkalmazott értékelési eljárások részletesebb ismertetésére kerül sor. A harmadik részben megpróbáljuk azonosítani a jelenlegi eljárások kihatásait. Az utolsó fejezet pedig azokat a kérdéseket vizsgálja, amelyek még megválaszolatlanok, illetve konfliktusos forrásai lehetnek, és amelyeket az értékelési eljárások továbbfejlesztésének biztosításához körül kell járni.

Háttér

Az értékelés fejlődése az amerikai felsőoktatásban mindig is az alap-, illetve középfokú oktatásban bekövetkezett változások és reformok nyomán ment végbe. Ez a fejlődés úgy folyt le, hogy az alsóbb osztályok információt adtak arról, mi várható a középfokú oktatást követően, továbbá olyan tanulságokkal szolgáltak, amelyekből a felsőoktatás is profitálhatott.

Az 1970-es évek elejétől kezdődően az alap- és középfokú oktatás szintjén több állam is elszámoltathatósági törvényeket hozott a helyi oktatási intézménystruktúra a színvonalának emelése és az oktatás fejlesztésének serkentése érdekében. A hangsúly az oktatási módszereken volt és nagyfokú helyi beleszólást tett lehetővé. Ez az 1970-es évek végére megváltozott. Mivel az adott irányú fejlődés nem volt kielégítő, az oktatáspolitikai kidolgozói először a kompetenciateszteket tették kötelezővé és a hangsúlyt a tudásanyagra helyezték. A minimum-kompetencia tesztelése kezdetben csak a középiskolásokat célozta meg, később azonban a kötelezően előírt követelményeket az alsóbb osztályokra is kiterjesztették. A 80-as évek közepén az állami törvényhozások már tantervi útmutatásokat is kötelezővé tettek, melyek a tesztek anyagához igazodtak. *Darling-Hammond* (1988) szerint ezeknek az egymást követő hullámoknak az eredménye a szűken meghatározott, tantervre épülő tesztelés lett. Ezeket a teszteredményeket ma már olyan célokra, olyan döntések meghozatalára használják, amelyek nem felelnek meg a tesztfejlesztők eredeti szándékának. Így a tanárok ma a tesztek megoldására tanítanak, ami csorbítja a helyi ellenőrzést és döntéshozatalt.

Az 1980-as évek közepén már egyre gyakrabban lehetett hallani az oktatási reformot sürgetők hangját (*Ewell*, 1991). 1984-ben és 85-ben négy, széles körben ismertté vált cikk jelent meg az elsődipomás képzésről. Az egyik ilyen felmérést a *Southern Educational Regional Board* (1985) végezte Access to Quality Undergraduate Education címmel. A felmérés az egyetemi tanulmányaikat megkezdő nagyszámú hallgatót vizsgált meg és alapvető készségbeli hiányosságokat mutatott ki, ami a középfokú oktatás súlyos problémáit jelezte. Ez a jelentés nyilvánvalóvá tette, hogy az alapvető készségek teszte-

lésére a felsőoktatásban is szükség van. Egy, a National Institute of Education által 1984-ben közreadott felmérésben (*Involvement in Learning*, 1984) a fejlődés három lehetséges útját emelte ki: magasabb követelmények, a diákok aktív részvétele és világos visszajelzés a diáknak saját teljesítményéről. A másik két felmérésben (*Integrity in the College Curriculum*, American Association of Colleges, 1985; *To Reclaim a Legacy*, Bennett, 1984) a hangsúly a tanterven, a tananyagban és a struktúrában volt. Ez a négy felmérés az oktatásban hatalommal rendelkező vezetők figyelmét az elszámoltathatóság és a fejlesztés kérdéseire irányította, továbbá kijelölte az amerikai felsőoktatás reformjának menetrendjét.

Ezeket a felméréseket elsősorban oktatási szakemberek publikálták oktatási szakembereknek és az oktatásban képviselt érdekcsoportoknak, vagyis egy belülről eredő reformmozdulást képviselnek (Ewell, 1991). 1986-ra azonban a National Governor's Association és az Education Commission of the States egyaránt közzétett egy olyan beszámolót, amely az elszámoltathatóságot sürgeti. Ezek a vizsgálatok azt sugallták, hogy a diákok teljesítménye és ezáltal maguknak a felsőoktatási intézményeknek a hallgatók képzésében mutatott teljesítménye olyan információ kellene, hogy legyen, amely a nagy nyilvánosság számára is hozzáférhető. Így a diákok és szüleik ennek tudatában hozhatnának döntést az oktatásba fektetett pénzeikről és az államok kormányai is felhasználhatnák a költségvetés elosztása során.

Az 1980-as évek végén érezhetővé vált, hogy az amerikai közvélemény nagyon is figyelemmel követi a felsőfokú oktatás színvonaláról folyó vitákat. 1987-ben két egyetemi tanár, Hirsch (University of Virginia), és Allan Bloom (University of Chicago) könyvet írt az amerikai elsődipomás képzésről és általában az egyetemi oktatásról. Ezek a tanulmányok egyszerre jelentek meg a New York Times sikerkönyvlistáján (The New York Times Best Seller List). Ez valószínűleg példa nélküli volt az amerikai népszerűsítő kultúrában. Bloom a *Closing of the American Mind* című könyvében azt állítja, hogy a felsőoktatási tantervek teljesen szétestek, és mivel nem tartalmazzák a klasszikus nyugati hagyományok legnagyobb műveit, megfosztják valamitől a ma egyetemeken tanuló generációkat. Hirsch *Cultural Literacy* című könyvében szintén azt fejt ki, hogy a felsőoktatásban tanuló diákok kevesebbet tudnak, mint az előző generációk. A közös tudásbázis, illetve a közös intellektuális nemzeti nyelv hiánya mára összevisszaságot eredményezett és már-már a kulturális örökség hiányával fenyeget. E tanulmányokat a felsőoktatás reformjáról szóló könyvek egész sora követte, de az ismertségüket egyik sem tudta már megközelíteni.

Ez a három elem – a fejlődés szükségességének megfogalmazása belülről, az elszámoltathatóság követelése kívülről, valamint a növekvő közfigyelem és érdeklődés, – a velük párhuzamosan jelentkező fejlődési irányokat is képviseli. Közös vonásuk, hogy a háttérben egy átalakulóban levő amerikai gazdaság áll, változó oktatási követelményekkel. Lawrence Cremin (1990) ezt a gazdasági helyzetet veszi kiindulópontnak. Cremin szerint ebben az évszázadban az amerikai gazdaság már túl van harmadik fő korszakváltásán. (1920-ban az iparban dolgozók száma először haladta meg a mezőgazdaságban dolgozók számát. Az 1950-es évek közepére az értelmiségi, igazgatási munkakörökben dolgozók száma túlszárnyalta a fizikai dolgozókéét. 1990-re pedig az amerikai munkaerőnek durván a fele az információfeldolgozás valamilyen ágában dolgozott). A gaz-

daság azok fejlődésben levő szektorainak mindegyike magas szintű kommunikációs, problémamegoldó és analitikai készségeket kíván, amelyeket pedig éppen a felsőoktatás hivatott megadni. A felsőoktatás iránti kereslet az USA-ban változatlanul sohanem látott méretű. Ez azonban egyre magasabb elvárásokkal és az elszámoltathatóság egyre növekvő igényével jár.

Az elvárások és a kereslet ilyen mértékű növekedése több intézményi változást eredményezett. *Ewell* (1989) az amerikai felsőoktatás történetében két éles váltást lát. Az egyik a növekedéstől a sokszínűsödés felé, a másik pedig a tantárgyközpontú egyetemi tanulástól a hivatásra való felkészítés felé mutat.

Mint azt korábban megjegyeztük, a növekedés állandó tényező az amerikai felsőoktatásban. Gondoljunk csak el, hogy 1910-ben a 19–24 évesek 2,9%-a volt beiratkozott vagy bentlakó diákja az egyetemeknek, illetve felsőoktatási intézményeknek (*Cremin*, 1990; 554–555. o.). Az 1920-as években az egyetemekre és főiskolákra való beiratkozások száma megkétszereződött, 1920 és 1940 között pedig bizonyos felsőoktatási intézmények, a junior college-ok (elő-egyetemek) száma megháromszorozódott (*Cremin*, 1990. 248–249. o.). Jelentős volt a növekedés a második világháború után is, amikor az 1944-es Serviceman's Readjustment Act illetve később a Truman Commission Report, Higher Education for American Democracy (*Cremin*, 1990. 251. o.) életbe lépett. 1950 és 1970 között a felsőfokú oktatási intézményekbe való beiratkozások száma megháromszorozódott. 1980-ban pedig az USA 3000 intézményében több mint 12 millió diák tanult. Az a növekedés és sokszínűsödés, amelyről *Ewell* beszél, csak a legújabb keletű változás. A hivatásközpontú oktatás azonban nem új az amerikai felsőoktatásban, már az 1890-es évekről *Cremin* (1988) a következőket mondja: „Gyakorlatilag minden fajta felsőoktatás szakmai jellegű, amennyiben diákokat a felnőtt munkaszerepekre készíti fel”. (243. o.)

Ewell megfigyelése fontos az 1990-es évek szempontjából. A XX. század első évtizedei meghatározóak voltak az amerikai kutatóegyetemek megteremtésében és későbbi dominanciájának kialakulásában. A német egyetemi modell erőteljes hatására, a *Morrill* telekadományoknak, adóbevételeknek és a filantrópia által is biztosított bőséges anyagi forrásoknak köszönhetően az amerikai egyetemek a tudományos kutatásokat tudományágak szerint szerveződő tanszékekben szilárdították meg, melyek aztán az amerikai egyetem strukturális középpontjába kerültek. Ezen tudományágak befolyása a század folyamán csak megerősödött, az egész Amerikára kiterjedő szövetségek, társaságok, szakcéhok megalakulásával. Sokszor mondták már, hogy az amerikai egyetemi oktató elsődlegesen a tudományága iránt elkötelezett, anyaegyeteme iránt pedig csak távoli, másodlagos elkötelezettséget érez.

Ewell meglátása szerint a hivatásra felkészítés felé való váltáson belül két strukturális fejlődési irány lelhető fel. Az első a nyilvánosság felé való elszámoltathatóságra, a második pedig újra a hivatásfelkészítésre helyezett hangsúlyból következik. *Ewell* szerint a tudományág-központú szervezeti és ösztönzési struktúra összeegyeztethetlenséget jelent a mai diákok oktatási igényei, szükségletei között. Ezek a fejlődési irányok és szerkezeti váltások komoly kihívást jelentenek az amerikai felsőoktatásban és fontos következményei vannak az értékelés gyakorlata és felhasználása szempontjából is.

Az amerikai felsőoktatásban dolgozó vezetők tudatában vannak ezeknek a váltásoknak. Közülük például *Donald Stewart*, a College Board elnöke nyilvánosan is reagált

ezekre a kérdésekre. Az 1987-ben az American Association for Higher Education konferenciáján tartott beszédében *Stewart* több olyan kihívást nevezett meg, amivel a felsőoktatásnak szembe kell néznie. Például a hallgatók kellő számban való vonzása, kiválasztása, felvétele és megtartása a kisebbségi hallgatók és egyéb, alulreprezentált rétegek felsőoktatásba való áramoltatása, a hallgatók felzárkóztató, illetve fejlesztő csoportokba való besorolásának lehetősége, hatékony oktatási intézkedések megtervezése, a tanárok és a tanárképzés színvonalának emelése, a diákok teljesítményének megfelelő szintfelmérése, a „hozzáadott érték” dokumentálása az egyes intézményekbe járó diákok számára, a végzősök kompetenciájának igazolása. Ezek a kihívások, illetve célok az egész USA-ban érvényes oktatáspolitikai célkitűzéseit is mutatják. Ezeket egyre inkább lehet direkt módon az értékelési gyakorlathoz kötni. Mint azonban ahogy *Stewart* is rámutat, az értékelés felhasználásáról és értékéről többféle felfogás uralkodik. Az egyetemeken kívül az egyetemi problémák megoldását sokszor csak újabb és újabb tesztekben látják. Az egyetemeken belül azonban, különösen azok, akik az értékelés felé való elmozduláshoz tartozóknak tartják magukat, sokkal többféle tevékenységet tartanak az értékelési gyakorlat körébe tartozónak.

Az amerikai felsőoktatásban az értékelést újabban legalább három különböző szintre tartozónak képzelik el, amelyek mindegyikének más a célja, felépítése és az eszköze (*Cross*, 1990). Az első ezek közül az állami szintű elszámoltathatóság a törvényhozás, a kormányzók, az egyetemi és más felsőoktatásbeli tisztségviselők, igazgatótanácsok felé. A második az intézményes értékelés, ami rendszerint az egyetemi oktatóktól, illetve a vezetéstől indul ki. A harmadik pedig az oktatási célú értékelés, amelyet az oktatók maguk terveznek és saját céljaikra használnak. Az értékelést felhasználhatják még a diákok minősítésére és osztályzására, elsőrendű célja azonban hatást gyakorolni oktatás fejlődésére. Az államok szintjén az elszámoltathatóságra irányuló, illetve intézményes értékelésnek mindig el kell különülnie a diákok osztályzásától és minősítésétől.

Az oktatási célú értékeléssel bővebben fogunk foglalkozni az értékelés gyakorlatának keretében. Azt azonban itt is fontos megjegyezni, hogy mind az intézményi értékelésben, mind az állami szintű elszámoltathatóság szintjén van fejlődés. Érdekes szem előtt tartani (*Astin*, 1988), hogy a felsőoktatási intézményeket hagyományosan a következő kritériumok szerint ítélték és ítélik meg: (1) felvételizők és a felvettek színvonala (az input); (2) az anyagi források és a (3) a hírnév. *Astin* szerint azonban az ilyen információk keveset árulnak el az intézményben folyó oktatás hatékonyságáról és minőségéről. *Astin* korábbi, a National Merit Scholarship Corporation számára végzett munkái azt mutatják, hogy azok az intézmények, amelyek diákjai kevésbé jó előmenetelűek, kevesebb forrásra támaszkodhatnak és kisebb a hírnevük, sokkal hatékonyabbak voltak a később PhD fokozatot is szerzett diákok diplomáig való eljuttatásában, mint azok az iskolák, amelyek a fenti kritériumok szerint jobb megítélés alá estek. Ezekre az adatokra és egy nagyszabású, az egész USA-ra kiterjedt, a Cooperative Institutional Research Project keretében végzett értékelésre alapozva, hívta fel a figyelmet *Astin* arra, hogy szükség van olyan intézményi értékelésekre, amelyekkel nemcsak input adatokhoz lehet jutni, hanem a tanuláshoz kapcsolódó környezeti változók többszörös értékeit, illetve a többszörös outcome értékeket is fel lehet tárni. Longitudinálisan kombinálva, ezek az adatok biztosabb alapot adnak az intézményi hatékonyság kvalitatív megítéléséhez. Az output-ot az input függ-

vényében vizsgáló értékelési megközelítés *Astin* korai munkájából következett és a „hozzáadott érték” elnevezést kapta. Ez a longitudinális elő-utó teszt megközelítés most is használatos nagy mintájú felméréseknél. Az utóbbi évtizedben fordult csak a figyelem ezzel egyenlő mértékben a környezeti tanulási változók értékelése felé.

A tudástartalmat és a tananyagot figyelembe véve *Astin* és mások a kognitívon kívül az affektív és nem-kognitív outcome értékelése mellett érveltek. Sok intézmény, ha nem is a túlnyomó többség, ma is azt tartja, hogy céljuk olyan értékek fejlődésének elősegítése, mint az empátia, a tolerancia, a másokkal való törődés és az állampolgári érzés. Több érv is szól e tulajdonságok értékelése mellett. Legkevésbé persze az elszámoltathatóság, bár az intézmények részéről van olyan igény, ami erre alapot nyújthatna. Egy másik érv az, hogy ezeknek a tulajdonságoknak az értékelése nagyon értékes forrása lehetne olyan intézményi szinten nyilvántartott adatoknak, amelyek segítenének abban, hogy az intézmények képesek legyenek teljesítményüket növelni a hagyományosan affektív területek fejlesztésében. A harmadik, egy többé kevésbé teoretikus vélemény és ezzel *Astin* nem áll egyedül, hogy a hallgatói fejlődés effektív és kognitív területei közötti hagyományos megkülönböztetés önkényes és mesterséges. Holisztikusabb megközelítések szerint az egész személyiség számít. Talán felesleges felhívni a figyelmet arra, hogy az ilyesfajta értékelés csakis az intézményes hatékonyság javítását szolgálhatja és nem lehet alapja a hallgatókról alkotott minősítő ítéleteknek.

Ewell (1991) az 1980-as évek értékelési kísérleteire alapozva három, a felsőfokú oktatásban kialakuló értékelési trendet ír le: a „polcra levezető” tesztelési technológia helyett új eszközök kifejlesztését, a hallgatói fejlődés változó modelljét (az additív lineáris modell felől a nem lineáris komplex modell felé), valamint a teszteredmények alapján történő megítélés helyett a naturalisztikusabb megközelítések felé való elmozdulást. E trendek közül kettőről itt lesz szó bővebben, míg az eszközökről a tesztelés gyakorlatáról szóló részben.

A hallgatói fejlődést ma már nem lineárisnak, inkább sok összetevőjűnek tekintjük, ami nem teszi lehetővé az egy tényezőre visszavezetett magyarázatot. Ez olyan értékelési gyakorlatot jelent, amelyben inkább az ismételt, semmint egyszeri mérés, illetve tesztek és ellenőrző tesztek szükségesek a hallgató fejlődésének értékeléséhez. Az értékelt eredményeket mint kognitív és nem kognitív tulajdonságok komplex kombinációját holisztikusan modellezzük (*Ewell*, 1991, 17–23 o.).

Jól látható jelei vannak annak is, hogy elmozdulás történt a hagyományos tesztelési eljárásoktól a természetesebb értékelési formák felé. Ez, a kurzusokhoz ténylegesen jobban kötődő megközelítéseket, szemináriumi projekteket, „klinikai gyakorlatot” (internship), terepgyakorlatot (field placement), portfóliót (szintfelmérési anyagok gyűjteményét), szimulációkat, tudásdemonstrációkat, a felsőbb évfolyamokon pedig tényleges gyakorlatot (capstone experience) foglal magában. Legtöbbször több értékelővel és javítóval dolgoznak. Bár így az értékelő tevékenység reliabilitása némileg csökken, de jobban illeszkedik a tanítási folyamatba, hitelesebb képet ad arról, hogy a hallgatók mit tanultak meg, hogy mit tudnak kezdeni azzal, amit megtanultak. A hallgatói fejlődés elemei a naturalisztikusabb megközelítések és az eszközök kiválasztásának területén (ez utóbbiról később szólunk majd) sok olyan értékelési eljárásban láthatóak, amelyekről ebben a dolgozatban szó van.

1987-ben *Boyer, Ewell, Finney* és *Mingle* egy, mind az ötven államra kiterjedő felmérést végzett annak a kiderítésére, hogy hány állam ír elő kötelező értékelést és hogy hogyan látják az állam szerepét a felsőoktatásban. Abban az időben az államok kétharmadában éppen készülöben volt a formális értékelés. 1991-ben már az államok 82%-ában számoltak be formális értékelési tevékenységről (*Hutchings* és *Marchese*, 1991). *Boyer* és munkatársainak felmérésre válaszolva az államok szakértőinek többsége úgy gondolta, hogy az értékelések megtervezése és kivitelezése az egyes intézmények dolga, saját szerepüket a támogatásban, továbbá a lehetőségek megteremtésében látták. Ezzel szemben tíz állam az értékelési programok megtervezésében és lebonyolításában való aktív részvételben látta saját szerepét. Az állam aktívabb szerepét olyan tényezők határozzák meg, mint például az államnak az oktatásba való beleszólásával kapcsolatos hagyományok, illetve az állam rendelkezésére álló anyagi erőforrások (*Ewell* és *Boyer*, 1988). Több államban bőséges anyagi támogatást adtak az értékelési tevékenységhez, ami néhány esetben az intézménynek az utasításokra és a kezdeményezésre adott reagálásától függött. Az egyes államok kormányainak utasításain és kezdeményezésein túl az USA Művelődésügyi Minisztériuma (Department of Education) 1988-ban, a Felsőoktatási Akkreditációs Tanács (Council on Postsecondary Accreditation) pedig 1987-ben kezdte megkövetelni az akkreditációs ügynökségektől, hogy információt szerezzenek a tanulási eredményekkel kapcsolatos információk gyűjtését.

Ezen adatok és arányok ellenére még mindig nehéz pontosan meghatározni az értékelési tevékenység szintjét az amerikai felsőoktatási intézményekben. Rendelkezésünkre áll néhány anekdotikus jelzés is. 1987-ben az American Association of Higher Education létrehozta az AAEH Assessment Forum-ot. Ez a projekt szolgáltatásokat, konzultációkat, anyagi forrásokat valamint tanulmányírási támogatást nyújt és évenkénti értékelési konferenciát rendez. Az első konferencia 1987-ben 700 érdeklődött vonzott, ez a szám 1991-re megduplázódott. Az Assessment Forum szerint a felsőoktatási intézmények 90%-a értékeli valamilyen formában, de ez keveset árul el ezeknek a törekvéseknek a minőségéről, mélységéről, illetve sikerességéről. *Peter Ewell* valószínűleg több látogatást tett az egyetemeken az értékeléssel kapcsolatban, mint bárki más az USA-ban. Az ő nem hivatalos becslése alapján az egyetemek 15%-án folyik olyan jellegű értékelés, ami valószínűleg hatással van az alsóbb évfolyamokon folyó oktatás hatékonyságára. *Ewell* szerint, egy másik 15%, vagy ellenáll az értékelésben való részvételnek, vagy teljesen tanácstalan azzal kapcsolatban, hogy mit kellene csinálnia. A középben levő nagy csoport pedig lehetőség szerint megpróbálja végrehajtani az utasításokat (*Hutchings* és *Marchese*, 1991).

Értékelési eljárások

Az értékelési eljárások megközelítése

Az USA-ban használatos értékelési eljárásokról valaki azt jegyezte meg, hogy ezek az eljárások részben az európai gyakorlat újrafelfedezéséből és a standard objektív tesztek kritikájából állnak és csak részben újítások (*Darling és Hammond, 1988*). Ez a megfigyelés nagyon hasznos a jelenlegi gyakorlat tárgyalásához. Az USA oktatási szakemberei minden szinten, alternatív megoldásokat keresve, továbbra is elégedetlenségüknek és aggodalmuknak adnak hangot a standard tesztek oktatási gyakorlatban teljesen félresiklott szerepével kapcsolatban. Az újítások nagy része nem más, mint újrafelfedezés, bár vannak tényleges újítások is, egyelőre nehéz belátni, hogy ezek életképesek.

Az „amerikai gyakorlat” bármely területen való áttekintésének kísérlete során hamar szembetaláljuk magunkat a sokszínűség tényével. Nem meglepő tehát, hogy az értékelés is sokféleképpen, sok forma, tipológia kategória szerint tagozódhat. Mivel az értékelés változatlanul kritikus helyet foglal el a felsőoktatásban, ezek az osztályozások és leírások a kétségtelen konvergenciákon és divergenciákon keresztül végül nagyobb áttekinthetőséghez fognak vezetni.

Az értékelés legtagabb értelmű osztályozása a tevékenység funkciójából indul ki és az elszámoltathatóságra irányuló, illetve a fejlődést mérő tesztekkel különböztet meg. Hagyományos terminológiával, ezek a formák tipikusan az összegző (szummatív), illetve segítő (formatív) értékelést jelentik. Ezeknek a funkcióknak a finom módosítása, illetve kiterjesztése látható az értékelés *Cross* által definiált hármastételében. *Cross* legfőbb meglátása az, hogy ez a három terület – az állami szintű összevethetőség, intézményes értékelés és oktatási célú értékelés – különböző módokon kezd összeolvadni és kiegészíti egymást.

Az értékelés alapvető taxonómiáját *Ewell* (1991) egy kétszer kettős mátrixként vázolja fel, egyéni és csoport, illetve fejlődés és demonstráció dimenziókkal. Az egyéni-fejlesztési rubrikába például egy olyan alapkészség-teszt található, amit besorolás, korrepetálás majd újratestelés követhet. Az egyéni-demonstráció elé sorolt értékelések minősítési, illetve szelektálási célokra a legértékesebbek. A felsőoktatásban az utóbbi időkben a csoport-fejlesztésre helyeződött a hangsúly, ez az a terület, amely a legtöbb újítási lehetőséget nyújtja. Ide sorolható például az általános oktatási eredmények felmérése, illetve azok az eszközök, amelyek egy-egy szak hatékonyságát mérik. A leggyorsabban valószínűleg a mátrix csoport-demonstráció cellája változik, mivel egyre inkább a hallgatócsoportok összehasonlításának irányában mozdul el, illetve a tandíjfizető nyilvánosság számára elszámoltathatóságot biztosít. Az akkreditációs ügynököket éppen a ilyen csoport-demonstráció értékelés érdekli a legjobban.

Adminisztratív nézőpontból, *Rossman és El-Khawas* (1987) korlátozott és átfogó megközelítéseket különböztet meg. A korlátozott megközelítések a tanuló teljesítményének egyetlen aspektusát, esetleg egy bizonyos tanítási programot vesznek célba. Erre példa többek között a *junior capstone projekt* vagy a hallgatók memóriájának fejlesztését célzó értékelések. Az átfogó megközelítések több komponenset is figyelembe vesznek és többszörös méréssel, több időpontban dolgoznak. Ennek példái a fokozatszáró, illetve a

felvételi standardizált tesztek. Egy másik példa lehet a szakmai kompetenciára épülő tanterv. Az értékelésnek és lebonyolításnak ezek a különböző felfogásai egyáltalán nem összeegyeztethetetlenek, és legtöbbször éppen kiegészítik egymást.

Az értékelési eljárások formái

A legtöbb értékelési program alapját változatlanul a standard tesztek képezik, melyeket továbbra is nagyfokú ambivalencia övez. Ezekből az eljárásokból való kiábrándultság volt a legfontosabb ösztönzője annak, hogy nagyon sok oktatási szakember vehetett részt az alternatív értékelési megközelítések kidolgozásában. Bár a pszichometrikusok gyakran hangoztatják, hogy az objektív tesztformák magasabb szintű kognitív készségek értékelésére is alkalmasak, ezt igazán nem sok empirikus tény igazolja. *Darling-Hammond* (1988) egyszerűen azt állítja, hogy van valami furcsa abban, hogy standard feleletválasztós kérdéseket használnak a felsőoktatásban. A standard feleletválasztós tesztek problémái mindenki által jól ismertek és még az alapvető tanárképzési szakirodalom is a kimúlásukat jósolja, amint más alternatívák jelennek meg (például *Gage-Berliner* 1991. 613–617 o.). Ha azonban gondosan áttekintjük a problémát, világossá válik, hogy a feleletválasztós tesztekkel szembeni aggodalmak inkább azok kizárólagos vagy rossz célra való felhasználásával és az eredmények nem megfelelő interpretálásával kapcsolatosak. A feleletválasztós tesztek kritikáinak fő célpontja az USA-ban a többi forma rovására való elterjedtsége. Az ambivalencia abból ered, hogy az oktatási szakemberek nagyon is tudatában vannak e tesztelési formák előnyeinek (a gyors, könnyű és olcsó adminisztrálhatóságnak). Ezek a tulajdonságok, amelyek különösen a tapasztalat hiánya esetén nyilvánvalóak, az egyetemi szakemberek számára meg éppenhogy jól jönnek. Nekik ugyanis az idő elsőrendű fontosságú. Ezek a szakemberek ugyan megnevezik a feleletválasztós teszteléssel folytatott értékelés problémáit és buktatóit, de azon az állásponton vannak, hogy megfelelő előkészítéssel ezek a problémák kiküszöbölhetők. A legfontosabb annak a figyelembevételére, hogy a feleletválasztós tesztelés kizárólagos alkalmazása nagyon gyenge értékelési programot ad. Az a felfogás viszont, hogy a standard tesztek részét képezhetnék egy többszörös mérési megközelítési formának, egyre inkább elfogadottá válik (*Kean*, 1993).

Az eszközök tekintetében több olyan fejlődési irány is van, amely megérdemli, hogy foglalkozzunk vele. Az 1980-as évek elején és közepén megnövekedett az értékelés iránt, minek következtében újra hadrendbe állították a polcról levezhető standard tesztek (például a Graduate Record Field Exam, CLEP vizsgák, stb). 1989-re már négy, az általános oktatási eredményeket tesztelő vizsga volt kipróbálás alatt, illetve használatban. Ezek közül három mátrixos felépítésű. Az ACT College Outcomes Measures Project (ACT-COMP) például három tartalmi és három folyamati kategória mátrixából áll, így mind a tudást, mind pedig a készségeket méri. Az ACT-COMP egyéb tulajdonságai között említést érdemelnek még a sokrétű problémamegoldó formátum, a csoportszintből becsült eredmények és a teszt-ellenőrző teszt különbségeinek értékei a tantervi hatás kimutatására. A másik két mátrixos felépítésű vizsga az Educational Testing Service *Academic Profile* nevű tesztje és a College Basic Subjects Examination (College Base). Az ilyen természetű vizsgák közül a negyedik az ACT-COLLEGE Assessment of Academic

Proficiencias (CAAP), amely egyben a leghagyományosabb is. Ezt a vizsgát a felsőoktatási intézményekben félidős vizsgaként szokták alkalmazni. Öt negyvenperces feleletválasztós tesztet tartalmaz az íráskészség, matematika, szövegértés, kritikus gondolkodás és a tudományos érvelés területén. Bármilyen más kombinációban igen rugalmasan kivitelezhető és norma-orientált.

1990-ben indították el a talán legújított erejű nagyszabású értékelést a New Jersey General Intellectual Skills Assessment-et, röviden GIS-t. *Ewell* szerint a GIS olyan utat jelent, amely a jövő standard általános célú értékelés számára töri az utat. A GIS szintén egy mátrixra épül, s négy készségterület az információgyűjtés, az információelemzés, az információbemutató, illetve a kvantitatív elemzés alapján fejlesztették ki. Ezen készségek mindegyike egyúttal a három tudományterület (társadalomtudományok, természettudományok, bölcsészettudományok) szerinti csoportosításban is szerepel. Az eredmények mindhárom esetben készségek szerint csoportosítva is kombinálhatók. A vizsga száz százalékban teljesítménycentrikus. Minden diák kap egy feladatot a készség-tudásanyag mátrixból és minden egyes feladat értékelése körülbelül 90 percet vesz igénybe. Az 1988/89-es próbatesztelés azt mutatta, hogy a taxonómiát minimum hét feladat fedi le megfelelően és 200 diák szükséges ahhoz, hogy használható csoporteredményeket kapjunk. Az egyetemi oktatók munkacsoportokban az Educational Testing Service (ETS) kifejlesztett a core pontozási (core scoring) technikával pontozzák a feladatokat. A pontozási útmutató alapján minden feladatra meghatározzák a minimálisan elegendő pontértékeket. A próbatesztek általában alátámasztották a GIS kivitelezhetőségét nagyszabású tesztelés esetén, továbbá kimutatták, hogy (a feleletválasztós teszteléssel összehasonlítva) sokkal inkább alkalmazható az oktatóra és a diákokra, mint a felsőoktatási intézmény szintjének a képviselőire. A GIS-szel további vizsgálatokat és kísérleteket is folytatnak intézményi szinten, például spirális mátrix szerkesztéssel, core judgmental pontozással és intakt osztálytermi mintavételi stratégiákkal (*Ewell*, 1991. 15–17. o.).

Az értékelés teljesítménycentrikus megközelítését *Marchese* (1987) Assessment Center Method-ként mutatja be (az értékelési központ módszere). *Marchese* szerint ez az értékelési gyakorlat az 1930-as években Németországban, Angliában és az USA-ban honos egyetemi kutatásból nőtt ki. Alapvető kiindulási pontja az, hogy bármely emberi teljesítmény túl komplex ahhoz, hogy egy írásbeli vizsgával (paper and pencil exam) értékelni lehessen, az ugyanis egyszerűen nem adhat elég információt. Az Assessment Center Method szimulációs és egyéb olyan feladatokkal dolgozik, amelyek közvetlen megfigyelést tesznek lehetővé. Szakértők minősítik a teljesítményt és megfigyeléseik alapján szelekciós döntéseket hoznak. Ezt a módszert használták a brit tisztképzésben megfelelő férfiak kiválasztáskor, és az USA-ban a második világháború alatt a kémek toborzásánál. Az 1950-es évek közepén az American Telephone and Telegraph értékelési központokat hozott létre vezetőinek minősítésére, ezt a gyakorlatot ma több mint kétezer amerikai nagyvállalat követi. Az elmúlt években a különböző módosított assessment center-szerű megközelítéseket több egyetemi program is magáévá tette.

Egy másik olyan értékelési alternatíva, ami a standard tesztek gyenge pontjainak felderítéséből nőtt ki, a portfóliós értékelés. A képzőművészetekben hosszú hagyománya van a portfólióknak, sok jelenlegi gyakorlat is a művészetektől kölcsönöz és használ fel

vizsgákhoz különböző eljárásokat. Ebben az értelemben a portfólió értékelés bármely olyan, a hallgató bizonyos időszakban végzett munkája alapján koordinált mintavételi eljárást jelent, amelyet bírálatra, minősítésre be lehet nyújtani, továbbá a diák kompetenciájának, teljesítményének és hozzáértésének bemutatására szolgál. Sok olyan intézmény, ahol régóta használják a portfóliós értékelést, jelentős haladást ért el a pontozási eljárások fejlesztésében és a reliabilitás javításban. Mivel produktum-centrikus és az autentikusságot hangsúlyozza, a validitás bizonyos típusait éppen a portfóliós értékeléseknél lehet tetten érni (Forrest, 1990).

Egy másik olyan tevékenység, amely az európai gyakorlatok újrafelfedezése címszó alá illik, a külső vizsgáztatók igénybevétele. Ez Európában széleskörben elterjedt és az amerikai felsőoktatásban is használatos volt körülbelül a XIX. század közepéig. A tanzsékek létrejöttével azonban, a tantervkészítés fokozatosan helyi szintre került, bevezették a választható tantárgyak rendszerét, a külső vizsgáztatók meghívása átadta helyét a tanár által kifejlesztett teszteknek és minősítő eljárásoknak (Brubacher és Rudy, 1958). A külső vizsgáztató közreműködésének előnyei között a legfontosabb, hogy különválnak az oktatás és a minősítés funkciója, a tanárok és a minősítők együttműködése, nő a validitás és lehetővé válik a magasabbrendű kognitív készségek értékelése (Fong, 1987).

Az oktatási célú értékelés – másképpen visszacsatolás – kategóriájába széles skálán helyet foglaló értékelési tevékenységek tartoznak (Cross, 1989; Cross és Angelo 1988). Cross (1989) az oktatási értékelés különböző megközelítési módjait egy olyan hipotézis formájában adja közre, amely abból indul ki, hogy az értékelési tevékenység akkor a leghasznosabb, ha folyamatos, ha annyira közel áll a tanítás és tanulás folyamatához amennyire csak lehetséges, továbbá ha diagnosztikus visszacsatolást biztosít mind a tanár, mind pedig a diák felé. Az oktatás visszacsatolási lehetőségeit Cross a tanárok és diákok, illetve a visszacsatolás adói és fogadói kétszer kettes mátrixában vázolja fel. Az oktatási értékelés leggyakoribb formája az, amikor a tanár nyújt visszacsatolást (feedback) a diáknak. Az amerikai felsőoktatásban azonban ma már ennek fordítottja – diák-tanár feedback – is teljesen elfogadott gyakorlat. Ez a két forma különböző megítélések alapja, egyrészt az érdemjegyeké, másrészt az előléptetéseké, az oktatói kinevezéseké. A másik két kvadráns a tanárok közötti konzultációt és a hallgatók egymásra vonatkozó értékelését, illetve a hallgatók együttműködő tanulási formáit tartalmazza. Cross áttekinti az egyes kvadránsokkal kapcsolatos kutatásokat és azt tárgyalja, hogy hogyan lehet ezeket a formákat úgy használni, hogy a tanteremben folyó tanítást és a tanulást fejleszthessék.

A fentebb leírt különböző gyakorlatok és eljárások egyedül vagy egymással kombinálva is előfordulhatnak. Céljuk lehet a kognitív vagy nem-kognitív fejlődés mérése, az általános oktatás vagy csak a humán tárgyak eredményeinek értékelése, tanulmányi irányok meghatározása. Használhatók besorolásra, átsorolásra, visszacsatolásra, bizonyítványkiadásra, minősítésre. Formájuk lehet tesztelés-újratesztelés, folyamatos értékelés, egyszeri értékelés. Szinte bármilyen kombináció létezhet és valószínűleg létezik is. Néhány konkrét példát a későbbiekben leírunk majd. Bár semmiképpen sem a teljesség igényével, a jelenleg használatos értékelési programok áttekintésére törekszünk. A következő példákat úgy választottuk ki, hogy jól illusztrálják a legmarkánsabb programokat és

eljárásokat, az intézményeknél most jelentkező problémákat, valamint azokat az intézményeket, amelyek érintettek az ilyenfajta értékelésben.

Példák az értékelési gyakorlatokra

Az értékelési szakirodalomban az egyik legtöbbet említett intézmény a wisconsini *Alverno College*, amely jó példája az Assessment Center Method-nak és sok más korábban leírt fogalomnak. Az *Alverno* tanterve nyolc hallgatói képességre épül, melyek a következők: kommunikáció, elemzés, problémamegoldás, értékmegítélés, társadalmi érintkezés, a környezetért való felelősségvállalás, a mai világ dolgaiba való bekapcsolódás és az esztétikai reagálás. E felfogás szerint ezek olyan integrált képességekről van szó, amelyek nem értékelhetők külön-külön, vagy nem teljesítményre épülő környezetben. A képességek mindegyikét hat egymásra épülő szinttel írják le. A diákokat rendszeresen külső vizsgáztatók mérik teljesítménycentrikus formában. Az *Alverno* 1973 óta dolgozik ezekkel a módszerekkel, megközelítésekkel és megkezdődött a képzési-program, illetve intézményi szintű értékelés (*Paskow és Francis, 1990*).

A külső vizsgáztatókat folyamatosan alkalmazó vizsgarendszer valószínűleg legjobb intézményi példája az USA-ban *Swarthmore*. Ezt a rendszert 1923-ban vezették be az iskola kiemelkedő diákjai számára. A módszer lényeg, hogy harmadév alatt a diákok tanulási tervet készítenek, amibe az utolsó éves szemináriumokat veszik bele. Ezeket a szemináriumokat olyan átfogó vizsgák zárják le, amelyeket más egyetemek tanárai készítenek és értékelnek. Az az időtartam, amelyet ez a program eddig megélt, már maga a siker jele, s az oktatók is nagyra értékelik, mivel elősegíti együttműködésüket és munkájuk szakmai ellenőrzését is (*Fong, 1987*).

Az ilyen jellegű értékelésre a harmadik példa a *Baldwin-Wallace College*, ami szintén bölcsészettudományi jellegű intézmény. A *Baldwin-Wallace* egy módosított Assessment Center módszert a felkészített külső vizsgáztatók rendszerével kombinálja. A diákok olyan, hivatáshoz kapcsolódó készségeit értékelik, mint vezetői készség, szóbeli kommunikáció és problémamegoldás. A megközelítés célja, az iskola és a munkabárállás közötti átmenet megkönnyítése (*Paskow és Francis, 1990*).

Ez a három példa jól mutatja a teljesítménycentrikus szimulációk és a külső vizsgáztatók rugalmas alkalmazhatóságát. *Swarthmore* azért használ külső vizsgáztatókat, hogy a különböző tantárgyakban minősítse és ellenőrizze (validate) a diákok tanulmányi haladását. *Baldwin* inkább a szakmai kompetencia szempontjából fontos készségekre koncentrálna, különösen a manager- és a tanárképzés területén. Az *Alverno* pedig a diákok fejlődésének összetett leírásából kiindulva kísérli meg a tanulás és a tanulást elősegítő képességek közötti kapcsolat megerősítését.

Az amerikai felsőoktatásban nagy érdeklődés övezte és övezi a portfólió értékelést is (*Forrest, 1990; Eckstein és Noah 1993, 193, 242 o.*). A portfólió vonzó tulajdonságai közé tartozik, hogy az oktatás folyamatára épít és így nem bontja meg azt és a komplex készségek értékelését is magában foglalja. Az oktatók szerint ez elősegíti az oktatási folyamat megújítását, a diákok pedig olyan produktumokkal rendelkeznek, ami mutatja, hogy mit tudnak és mit tudnak kezdeni ezzel a tudással. A portfólió iránti kezdeti lelkesedést azonban gyakran lelohasztják annak látható korlátai. Újabban a *William and Mary*

College kísérletezett a portfólió értékelési móddal. A programban 50 hallgató vett részt. A diákok munkáit több értékelő gyűjtötte és olvasta, specifikus pontozási kritériumokat használva. Így találták, hogy ezzel a módszerrel jól értékelhető az íráskészség és a kritikus gondolkodás, más általános képességek mérésére azonban nem igazán alkalmas. A *William and Mary College* általános oktatási tantervében az íráskészség értékelésére azóta is a portfólió módszert használják (*College of William and Mary*, 1988; *Ewell* 1991).

Az értékelési programok kifejlesztésében, különösen a felvételi utáni tesztelés területén az Egyesült Államokban mindig is nagyon aktívak voltak az alsóbb-szintű felsőoktatási intézmények (junior college). Mivel ezekben a felsőoktatási intézményekben a legtöbb esetben nincsen semmiféle felvételi követelmény nagyon változó képességű és teljesítményű hallgatói populációval rendelkeznek. Emiatt aztán gyakran céltáblái az állami törvényhozásnak és minőségkontrollra irányuló mechanizmusokat kívánnak tőlük. Egy ilyen átfogó tesztprogram jó példája a *Miami-Dade Community College*. Ez azért is figyelemre méltó példa, mert annak ellenére, hogy állami szintű utasítások először 1985-ben születtek, a program néhány eleme 1960-ból származik. Emellett, a *Miami-Dade*-nek évente közel 75 ezer rendes hallgatója és 40 ezer egyéb hallgatója van, ezáltal a nagyszabású vizsgáztatási gyakorlatnak is szép példáját adja (*Paskow és Francis*, 1990). Minden olyan diáknak, aki valamilyen kurzust fel akar venni a *Miami Dade*-en, standard felvételi szintű teszt-sorozatot kell megoldania. A teszteredmények döntenek el, hogy az illető milyen kurzust vehet fel. Ahhoz pedig, hogy egy tanulmányi programot teljesítsen valaki, illetve megkapja két év után az associate degree-t, továbbá ahhoz, hogy felsőbb-szintű kurzusokra beiratkozhasson sikeres, záróvizsga jellegű standard vizsgát kell tennie. A felvételi szintű vizsgát mind a diákok, mind pedig a tanárok nagyon kedvezően ítélik meg, mivel a diákok olyan kurzusokba kerülnek, amelyben szakmailag jól érzik magukat. A záróvizsga jellegű tesztelés kritikája azonban vegyes. A legfontosabb ellenvetés az, hogy az etnikai kisebbséghez tartozó diákok jelentős része nem tudja teljesíteni ezt a vizsgát és így az megakadályozza őket tanulmányaik folytatásában, illetve befejezésében. *Robert McCabe* (1988) a *Miami-Dade* rektora, kétségbeejtőnek tartja a záróvizsga hatását az etnikai kisebbségekre nézve. A kritikák rámutatnak arra is, hogy vannak olyan diákok, akinek bár nem sikerült ez a vizsga, de mégis folytatták tanulmányait (még az állami utasítás előtt) magasabb szinten sikeresen teljesítettek, ez pedig igen csak megkérdőjelezi a vizsga előrejelző validitását. A *Miami-Dade* ugyanakkor felvételi szinten sikeresebb programot képvisel a nyitott egyetemi felvételi rendszerénél. A fokozatszáró vizsgával kapcsolatos problémák nem korlátozódnak a *Miami-Dade*-re, viszont jó példái annak, hogy a szigorúan vett állami utasítások milyen következményekkel járhatnak. Ez a bizonyos vizsga, a College-Level Academic Skills Test, (röviden CLAST) Floridában az egész államra kiterjedően kötelező. Néha „feljövő junior”-ként beszélnek róla, mivel az összes junior (harmadik év előtt álló) diákot átvilágítja. Az etnikai kisebbségekkel kapcsolatos hatása mellett további kritika vele szemben, hogy megköti a tantervet és az oktatást, csorbítja a helyi intézményi ellenőrzés szerepét, állami szinten központosítja a felsőoktatást és a sokszínűség ellen hat (*McCabe*, 1988). A következőkben néhány példa következik arra, hogy hogyan reagálnak az intézmények az állami utasításokra.

A virginiai *State Council for Higher Education* (Felsőoktatási Tanács) úgy reagált egy bizonyos minőségi oktatásról szóló törvényhozási döntésre, hogy 1987-ben elfogadott egy dokumentumot *Guidelines for Student Assessment* címmel (*Hutchings és Marchese, 1991*). Ez az útmutató meglehetősen szabadságot hagyott az egyes intézményeknek és később, mint az állami előírás iskolapéldáját emlegették. A *University of Virginia*, amit *Thomas Jefferson* alapított és amit a legjobb egyetemek között tartanak számon Amerikában, mint minden más virginiai felsőoktatási intézmény, szintén arra kényszerült, hogy megfeleljen ezeknek az előírásoknak. Ezen az egyetemen évente 13 ezer diák verseng az betölthető 2600 helyért. Az igen engedékeny állami előírások és az egyetem kiválósága ellenére az értékelési kísérletek, illetve az egyetem és a tanács közötti viszony már a kezdetektől problematikusnak bizonyult. Az egyetem az utasítás kiadása után nagyratörő tervet fogadott el. Ennek központi elemei voltak a diákok egyetemi tapasztalatainak időszakos megismerésére szolgáló elbeszélgetések, a diákok fejlődésének nyomkövetése, a főszak eredményeinek mérése és a végzetek szakmai fejlődésének nyomkövetése. A tervet a Tanács elfogadta. Az elfogadás után az egyetem 35 ezer dollár kiegészítő támogatást igényelt elképzeléseinek megvalósításához a Tanács azonban alig több mint háromezer dollárt adott erre a célra.

Az értékelési programban dolgozó oktatók végül is nagyon elégedetlenek lettek az állami utasítással, mondván, hogy az állam nem mondta meg pontosan mit ért értékelésen és hogy nem áll rendelkezésre olyan megfelelő szakirodalom, ami segítené őket erőfeszítéseikben, továbbá, hogy nem jelölték meg a kijavítandó problémákat. Az egyetemi tanács szintén visszautasította azt az elképzelést, hogy oktatási szakemberek meg tudnának egyezni az oktatás kívánatos eredményeiben (*Hutchings és Marchese 1991. 52. o.*). A Felsőoktatási Tanács elfogadta ugyan az első évekről adott beszámolókat, de nagyon kritikusan viszonyult az átfogalmazott tanrendekhez, lecsökkentett tervekhez, a tudományos kutatás megtervezéséhez és a csekély anyagi támogatással kapcsolatos panaszokhoz. 1988/89-ben az egyetem új koordinátort vett fel a hallgatók értékelésének munkálataihoz és a munkát kiterjesztette az első diplomáig tartó négy év hallgatóinak longitudinális vizsgálatára. Kifejlesztettek egy 15 oldalas kérdőívet és kiosztották a felvett diákok egynegyede között. A diákok egyharmada elbeszélgetésen vett részt. Emellett öt tanszék hozzájárult a szakjukat érintő hármas próbaértékeléshez. A végzett diákok nyomkövetésének a tervét egyelőre felfüggesztették. Az évvégi beszámolót megküldték a Felsőoktatási Tanácsnak. A tanács válaszul kilencoldalas kritikát küldött a 37 oldalas évvégi beszámolóra, amit ellenséges hangú levélváltás követett és a kapcsolat a két fél között egyre romlott. Az egyetem folytatta a longitudinális vizsgálatot és az állami utasítások általános betartását is (hiszen ilyen esetben az állami ösztönző juttatások forognak kockán). A *University of Virginia* esete több kérdést is jól illusztrál. Az állami hivatalnokok és az egyetemi oktatók nagyon gyakran nagyon különbözően látják az értékelési tevékenység formáit és szerepét. Amikor ezek az eltérő vélemények ütköznek, a valószínű eredmény az, hogy az intézmény az utasítás betűjéhez igen, de a szellemiségéhez nem lesz hű. Néha ez is elég. Az azonban valószínűtlen, hogy bármelyik fél ideálisnak tartson egy ilyen állapotot.

Tennessee államban, ezzel ellentétes példaként, szorosabb együttműködést mutat az állam és az egyetem. Tennessee állam már 1979-ben egy teljesítmény-támogatásos rendszert kezdeményezett. Eszerint az állami támogatás egy részét a teljesítménycentrikus kritériumok határozzák meg: az akkreditált programok száma, a diákok teljesítménye, a végzett hallgatók oktatással szembeni elégedettsége és az oktatás javítását célzó tesztelés tervezete. Bár az ennek való megfelelés teljességgel önkéntes, az egyetemek költségvetésének több mint 5%-a a teljesítménykritérium szerint kerülhet odaítélésre (*Hutching és Marchese, 1991*). A University of Tennessee-Knoxville (UT-K) számára például ez akár évi 5 millió dollárt is jelenthet. Ez az egyetem volt a legnagyobb intézményi felhasználója az ACT-College Outcomes Measure Project (ACT-COMP) által közreadott tesztnek. Minden évben a felvett hallgatók felével és minden végzős hallgatóval megoldatják. 1988 óta az UT-K emellett az Educational Testing Service Academic Profile tesztjét, az ACT Collegiate Assessment of Academic Proficiencies és a College Basic Academic Subjects Exam tesztjeit is folyamatosan próbatesztezik (*Paskow és Francis, 1990*), hogy lássák, az oktatási eredmények mérésében melyik bizonyul jobbnak ACT-COMP-nál. A főszaconkénti értékeléshez körülbelül 50 oktató által kifejlesztett, illetve tanszékenként tervezett értékelési eszközök is használatban vannak. A UT-K értékelési program nagyságrendje igazán lenyűgöző, ami az egyetemi vezetés elkötelezettségének is köszönhető. Az egyetem így nagymértékű anyagi támogatásban is részesül, az értékelés eredményeit pedig a hosszú távú tervezéshez használják fel.

Végül az utolsó két példa olyan kiemelkedően erős egyetemeket idéz, amelyek egyéni értékelési tevékenységet folytatnak. Mind a *Harvardnak* mind pedig a *Stanfordnak* sajátos programja van.

A *Harvard* értékelési programja, bármely más ilyen programnál jobban, az intézményen belüli kutatás formájában folyik, igencsak hagyományos módszerű mintavétellel, tervezéssel és elemző technikákkal. Az ilyen természetű kutatás ösztönzése céljából 1986-ban létrehozták a Harvard Seminar on Assessment-et (Harvard Értékelési Műhely), mely már a harmadik évben száz résztvevőt vonzott. Az eredményekről két beszámolót adtak ki és több mint egy tucat projektje volt folyamatban. A tanulmányozott témák között volt például a nemek közötti különbségek a felsőoktatásban, a tanácsadás gyakorlata, vitacsoportot alkalmazó módszerek, a középiskolai felkészítés és a felsőfokú tanulmányok sikere, valamint különböző, az egyetemi tanulmányokkal kapcsolatos eredmények (*Paskow és Francis 1990*). Ezeknek a kutatóműhelyeknek a fő jellemzői a következők: egyaránt résztvesznek bennük az oktatók és az egyetemi, intézményi vezetők, teljesen önkéntes alapon szerveződik, a kutatási feladatokat olyanok tervezik meg, akik maguk is a tanulmányozandó tanítási-tanulási környezetben vannak, és a hangsúly a már meglévő programok megújításán és fejlesztésén van.

A Stanford University Teacher Assessment Project-et (Stanford University Tanárértékelési Projekt) a National Board for Professional Teaching Standard -dal (a Szakképzési Színvonal Országos Tanácsa) karöltve, a Carnegie Corporation anyagi támogatásával végzi (*Paskow és Francis, 1990*). A projekt célja a tanárok és a diákok értékelése a tanárképzésben a tárgyi tudásra, a tanártól megkívánható gyakorlati készségekre és stratégiákra irányuljon. A programigazgató, *Lee Shulman* meggyőzően érvel a kiegyensúlyozott módszer-tartalom megközelítés mellett mind a tanárképzésben, mind pedig az

egyetemi oktatók értékelésében (*Shulman, 1987*). A projekt mind a hagyományos, mind pedig az Assessment Center értékelési megközelítéseihez alkalmazza. Az értékelendő személyeket megkéri például arra, hogy vizsgálják meg szakterületük tanácskönyveit, vagy vegyenek részt a közös tervezési tevékenységben más tanárokkal, illetve minősítsék a többiek és maguk teljesítményét a tartalom és a pedagógia oldaláról. Ezeket az anyagokat összerakják és portfólióként betekintésre, továbbá értékelési célokra megőrzik. Ezek az eljárások szándék szerint a tanári diplomaeredményekhez és a tanárképzés tanterveinek fejlesztéséhez is kapcsolódnának.

Az itt áttekintett példák közül több általános következtetést is levonhatunk. Nagyon sok felsőoktatási intézményben, illetve néhány, példaként felhozott program (Alverno Swarthmore, UT-K) esetében az értékelési tevékenység egyáltalán nem újdonság, jó néhány esetben tizenöt vagy annál több éve dolgoznak saját értékelési eljárásokon. Az értékelési eljárások változatossága erős helyi jellegzetességeket mutat. Az amerikai oktatás története – és az amerikai karaktere is talán – az idők során az egységesség ellenében bontakozott ki, ellenállt a központosításnak. Néhány esetben, amikor az egységesség és a központosítás jelen van, csak csekély siker tapasztalható (Miami-Dade Community College, University of Virginia). Az erősebb programok, mint például a *Harvardé*, szigorúan az önkéntes részvételen alapulnak.

Bár az értékelési tevékenység nagyon sok esetben nem új keletű, két aspektusa azonban az. Először az állami törvényhozás és az oktatási bizottságok nagy befolyása, néhány esetben ellenőrzési tevékenysége az állam felsőoktatási intézményeiben folyó értékelési tevékenységnek mind a kiterjedésére, mind pedig a formájára vonatkozólag. Ezt a tényezőt a szakirodalom nemegyszer az intézményi szintű értékelési programok legfőbb motivációs tényezőjének tartja és valószínűleg a jövőben sem lesz kevésbé jelentős. Másodszor, míg a tényleges értékelési eljárások közül sok valóban nem új, az oktatók számára azonban mégiscsak újak. Sokak véleménye szerint a felsőoktatásban a probléma a szak tudás tekintetében éppen az ellenkezője az alap és középfokú oktatásban jelentkezőnek. Az alsóbb osztályokban a tanárok pedagógiaiul nagyon képzettek, de rendszeresen bírálják őket tárgyi tudásuk hiányosságai miatt. A felsőoktatásban tanítók viszont saját területüknek, tudományáguknak szakemberei, de nincs képzettségük arra vonatkozóan, hogy hogyan kell tanítani. Ez vezetett ahhoz a paradox helyzethez, hogy jelenleg a legtöbb elemi iskolai tanár lényegesen többet tud az értékelésről és a tesztelésről, mint az átlag egyetemi professzor (*McCabe, 1988*). Következésképpen, a felsőoktatást érő bírálatok nagy része, csakúgy, mint az oktatók érdeklődése az értékelés iránt, éppen ebből az aránytalanságból fakad. Az értékelés néhány esetben éppen annak a lehetőségét adja meg az oktatók számára, hogy tanárként fejlődhessenek.

Az értékelés hatása

A tesztelési és értékelési eljárások halmazati hatása igen paradox módon jelentkezik az USA-ban. Egyrészt, a végzős középiskolásokra váró vizsgák száma minimális a fejlett országokban élő társaikhoz viszonyítva (*Eckstein és Noah 1993*). Másrészt pedig, a folyamatos értékelés és a nemrégiben kötelezővé tett alap- és középfokú iskolai tesztek következtében az amerikai diákoknak kell a világon a legtöbb tesztet kitölteniük, hiszen

évente több mint százmillió standard tesztet íratnak Amerika-szerte (Neill, 1992). Mivel a felsőoktatási értékelési eljárások megváltoztatását célzó állami utasítások és ösztönzők csak az 1980-as évek vége óta vannak jelen, túl korai lenne még ezek hatásának pontos természetét és nagyságrendjét megítélni. Vannak azonban jelei annak, hogy ezt a hatást nemsokára, vagy a közel jövőben legalább négy területen lehet majd érezni a középfokú és az azt követő oktatás közötti kapcsolatban, a tanárképzésben, a felsőoktatásban oktatók felkészítésében, illetve a pszichometriában és a tesztfejlesztésben. A következőkben egyenként foglalkozunk ezekkel.

Az amerikai középfokú és az azt követő oktatás kapcsolatát elemző problémafelvetések a XVII. század végéig nyúlnak vissza (Brubacher és Rudy, 1958). Ezt megelőzően a klasszikus tanulmányok megkezdése előtt latinból és görögből felvételi vizsgáztatták diákokat. A XVIII. század közepén egyes intézmények már aritmetikából is vizsgáztattak, a XIX. század elején pedig néhány intézmény földrajzi, történelmi, angol nyelvtani és irodalmi ismereteket is elvárt a hallgatóktól. Ez a tantervi kiszélesedés nehezítette a középiskolák számára azt, hogy pontosan tudják, miből készítsék fel diákjaikat. Ezt a nehézséget tovább súlyosbította a felsőfokú oktatási intézmények közötti egyeztetés hiánya, illetve a középiskola utáni továbbtanulás iránti igény növekedése. A XIX. század utolsó évtizedeitől kezdődően többféle intézkedést próbáltak foganatosítani a közép- és a felsőfok közötti átmenet megkönnyítésére. E próbálkozások közé tartozott például az, hogy az egyetemi oktatók feladata középiskolák rendszeres látogatása, az ott folyó oktatás, tantervek megfigyelése és az ezekről való vélemény alkotás (például Michiganben). Ugyancsak ilyen kísérlet volt az egyetem által jóváhagyott vizsgák (regents exams) és a hozzájuk tartozó tanmenetek bevezetése a középiskolákban (például New Yorkban). Az Association of Colleges and Secondary Schools (Felsőfokú és Középfokú Oktatási Intézmények Szövetsége) az 1890-es években megpróbálta bevezetni az egységes felvételi vizsga rendszerét, az első világháború után pedig kötelezővé tették a képességvizsgát (aptitude test). Hasonló törekvés volt a Carnegie Unit alkalmazása, a standardizálási gyakorlat időbeli megközelítése a junior college-nak a megjelenése és széles körben való elterjedése. A junior college tulajdonképpen átmenetet képez a középiskolai és az egyetemi szintű tanulmányok között (Brubacher és Rudy, 1959. 235–257 o.).

A középfokú és az azt követő oktatás közötti átmenet természetesen továbbra is kihívást jelent az oktatási szakemberek számára. Minden egyes korszak valami olyan fogás kérdést vet fel, amit nem lehet kikerülni. A felsőoktatási értékelési projektek és törekvések ennek szellemében készülnek és újító szándékú programokat ajánlanak (Boyer és munkatársai, 1987). Néhány ígéretes példa látható erre Ohio-ban, Texasban, Michiganban, Indianában. Az Ohio-beli Early Assessment of High School Juniors (középfokú oktatás harmadéveseit megcélzó korai értékelés) címet viselő program olyan eszközökkel méri a középfokú oktatás harmadik évében tanulókat matematikából és íráskészségből, amelyeket az Ohio State University és a Youngstown University fejlesztett ki. A felmérést, az egyetemre való jelentkezést megelőzően a hiányosságok felismerése és felzárkóztatás követi. Az indianai program szintén a középiskola alsóbb éveit célozza meg és a diákok fejlődését végig, a középiskolai tanulmányaikon keresztül, egészen az egyetem elkezdéséig követi nyomon. A diákok felméréséről készített eredménylapokat, beszámolókat a szülők és a felkészülési tanácsadó szakemberek számára is hozzáférhető-

vé teszik. Törekvéseik közé tartozik még visszajelzés az iskoláknak az egyetem előkészítő jellegű tantervi programjaiknak a hatékonyságáról (*Boyer és munkatársai, 1987*).

A valószínűsíthető hatás egy másik területe a tanárképzés. A *Lee Shulman* által vezetett stanfordi projekt a legjobb példája a tanárképzésben használatos megreformált értékeléseknek, s hatása valószínűleg a stanfordi egyetemen kívül máshol is érezhető lesz. *Shulman* (1987) szerint a jelenlegi tanárképzés teljesen a módszerrel kapcsolatos készségekre épül (process skills) és nem fordít figyelmet a tudásanyagra. Ennek bizonyítékként rámutat arra, hogy a tanárminősítésnél használatos értékelési eszközök és módszerek vizsgálatakor kiderült, a mérésnél nincs különbség abban, hogy alsótagozatos matematika vagy tizenkettedik (középiskola utolsó évfolyama) osztálybeli biológia tanításáról van-e szó. A mérés kizárólag a készségekre épül, néhány általános készséget hangsúlyozva és független a tanítandó osztály évfolyamától és a tananyagtól. *Shulman* úgy látja, hogy a tanítási módszereket nem lehet a tanítás tartalmától elválasztani és ennek alátámasztására a tanítási transzfer irodalmára hivatkozik. Az erre vonatkozó irodalom szerint kevés érv szól amellett, hogy a módszerrel kapcsolatos készségek egyik ismereti területről a másikkra transzferálódjanak. Az USA-ban a tanárképzésben éles határ húzódik a pedagógia tanszékek, valamint a természet- és bölcsészettudományi tanszékek között. Az előbbieket pedagógiát és módszertani készségeket tanítanak, a tananyag tanítását pedig a bölcsész és a természettudós oktatókra hagyják. *Shulman* álláspontja szerint mind a tantervben, mind pedig az értékelési gyakorlatban változtatásokra van szükség. *Shulman* és kollégái által kifejlesztett értékelési eljárások mindegyike kombinálja a tananyagot és a módszert. Az állami utasítások és az elszámoltathatóság felől közelítve a tanárképzés mindig a figyelem középpontjában volt. 1987-ben *Boyer* és munkatársai 50 államban felmérést készítettek arról, hogy mekkora mértékű az állam beavatkozása az értékelésbe. Ebből kiderült, hogy az értékelési gyakorlatnak ez az egyik fő területe. Bár a felmérésben nem is kérdeztek rá konkrétan a tanárképzésre, kilenc államban jelezték, hogy a tanárképzésben tesztelésre irányuló kezdeményezések vannak, három másik államban pedig próbatesztek folynak. A leggyakrabban alkalmazott forma ebben az időben a standard alapkészség teszt, ezek közül is a Pre-Professional Skills Test (PPST) volt, amit az Educational Testing Service adott ki. *Shulman* munkássága kifejezetten arra irányult, hogy minden állam, egyetem, főiskola és minősítő intézmény részére hozzáférhetővé tegyen egy alternatív tananyag-módszer értékelési könyvtárat. Jelenleg mindenki a standard tesztekkel használja, mivel nincsen elérhető alternatív forma. A képzési programok pedig ezekre a tesztekre készítene fel, így a tanárképzésben egy általános, a tananyagot figyelmen kívül hagyó megközelítés állandósul.

Egy másik olyan terület, ahol a tanítás mindennapi gyakorlata javulhatna, az egyetemeken és főiskolákon folyó oktatás. Az 1984/85-ben kiadott reformbeszámolóban megjelenő, a felsőoktatást érő bírálatok legnagyobb része éppen a felsőoktatásban végzett oktatást és annak az egyetemi feladatok között való alárendeltségét érte. A kutató egyetemeknek a század elején való előretörése és a felsőoktatás ezt követő bürokratizálódása miatt a tanítás csak egy lett az oktatói feladatok között (*Cremín, 1988*). *Jacque Barzun* (1993), a Columbia egyetem volt dékánja és igazgatója írta 1967-ben: „A tanítástól való eltávolodás kétségtelen tény, de ... kultúrához kötődő tény. Másképpen kifejezve ez az új egyetem egy új társadalom produktuma, melynek mellékterméke az oktatástól való eltá-

volódás” (54. o.). Az 1970-es években a felsőoktatási tandíjak növekedésével az amerikai közvélemény és az oktatók maguk is lassan kezdték megkérdőjelezni a publikált kutatások értékét és minőségét. Követelni kezdték, hogy több figyelmet kapjon az első diploma előtti oktatási tevékenység. Tekintettel arra, hogy az egyetemi tanszékek valóban nem a tanítást tartják elsőrendű fontosságúnak, az értékelési tevékenység célja az ezen való változtatás lehet. Ennek ellenére mindig is voltak kiváló oktatók a felsőoktatásban, voltak és vannak olyan intézmények is, amelyek anyagi forrásaik és energiáik legnagyobb részét a minőségi oktatásra fordítják. Ezekben az esetekben az értékelés célja éppen ennek a ténynek és az elszámoltathatóságnak a demonstrálása.

Az oktatók autonómiájának és az egyetemi szabadságnak a fényében az értékelésben tapasztalható oktatói részvétel dinamikája is érdemel néhány szót. *Hutchings* (1987) az egyetemi oktatói szakszervezetnek az állami utasításokra történt reagálásokról egy lehetséges példáját adja. Az idézett példában a szakszervezeti képviselők az egyetemi szervezeti vezetők és az állam képviselői közötti tárgyalások olyan értékelési alapelveket eredményeztek, mint például az értékelés célja a tanulás fejlesztése, az oktatók szabadon fejleszthetik ki a saját oktatási programjukhoz szabott mérőeszközöket, az értékelés eredményeit nem használják fel tanulási programokba való felvételhez, illetve azok lezárásának ellenőrzésére, az eredményeket csak intézményen belül dolgozzák fel, de nem használják a tanárok, az oktatók, a tanszékek, a karok vagy az intézmények bármilyen jellegű összehasonlítására. Az ehhez hasonló elvek az egyetemi vezetők, oktatók, a kívülről utasítók és esetleg a szakszervezetek közötti hatékony, informális, néhány esetben teljesen hivatalos érvényű szerződéseké válhatnak. Ezek mögött az alapelvek mögött meghúzódó engedmények alapvetően az oktatók bevonásának áráként foghatók fel. Az egységesség és az összehasonlíthatóság itt áldozatul esik egy, az oktatók által irányított cselekvésközpontú megközelítésnek. Ez az oktatói szabadság kérdéskörével függ össze és sok intézményben ez az egyetlen irány, amely felé az értékelés fejlődhet. Akkor, amikor az ilyen alapelvekben megegyezés történik, rendszerint két feltételnek kell teljesülnie. Először is kell egy olyan intézményi struktúra, amely nyilvánvalóan elismeri és tartalmazza a magas színvonalú oktatást és hajlandó erre anyagi forrásokat is áldozni. Másodszor pedig maguknak az oktatóknak is el kell gondolkodniuk saját tanítási gyakorlatukon. Ahol ez a két feltétel megvan, ott az értékelési tevékenység a szakmai fejlődés eszközévé válhat. Az oktatási célú értékelés használatához és megismeréséhez a természetes ösztönzőerő magától adódik olyan fiatal oktatók esetében, akik másoddiplómájukon/szakképesítésükön dolgoznak (graduate training) és tanítási tapasztalat, illetve tanári előképzettség nélkül töltenek be oktatói állásokat. Az általam átnézett értékelési szakirodalomban a felsőoktatásban végzett értékelési tevékenység leggyakrabban idézett hatása az, hogy olyan munkatársi légkört teremt és tart fenn, amelyben az oktatók pedagógiai kérdéseket, problémákat vitathatnak meg egymással. Ezzel ellentétes hatás (ami az intézményi szintű ösztönzők és elismerések szempontjából fontos), hogy szakmai, egyetemi előmeneteli szempontból a rendszeres értékelés többet kíván meg az oktatóktól, mint amit egyébként az egyetemek általában. Az értékelés megtervezése, kifejlesztése, végrehajtása, adatainak felhasználása időigényes feladategyüttes, ami azt kívánja meg az oktatóktól, hogy elgondolkojanak a tudás, tanítás, tanulás problémáin. Azt, hogy végül is a felsőoktatásban az oktatás, és ezzel összefüggésben a tanulás fejlődik-e

az értékelési gyakorlat eredményeképpen, az fogja meghatározni, hogy az egyes intézmények hogyan kezelik az alapelvek felállításának, az ösztönző és premizálórendszereknek és az oktató munkatársak egyéni érzékenységének kérdését.

Azok, akik a felsőoktatási értékelésről írnak, általában meglehetősen kritikus álláspontot képviselnek a hagyományos pszichometria elméletével és hatásaival kapcsolatban. A bírálatok célpontja szerzőről szerzőre más, és összességükben igen széles alapon nyugszanak. Egyformán bírálják a tesztfelhasználókat és a tesztfelkészítőket, de a céljuk általában nem a pszichometria-elmélet elvetése, hanem más irányba való terelése, ahol egyéb célokat szolgálhatna. Rá kell mutatni arra is, hogy ezeknek a bírálatoknak a nagy része már régóta ismert és a mérési, értékelési eljárások valós problémáit vetik fel. Ezzel a bírálók is tisztában vannak. Álláspontjuk szerint az az érv, hogy ez az elmélet alkalmatlan ezekkel a hiányosságokkal való megbirkózásra, nem elégséges ma már a hagyományos eljárások védelmében való érvelésre. Néhány olyan terület, amellyel kapcsolatban aggodalmak, kérdések és a nagyobb elméleti előrehaladás igénye fogalmazódott meg, a következők:

- Szükség van a természetes környezetben végzett mérések pontosságának növelésére; a tanulók eredményeinek longitudinális megközelítéssel való modellálására és értékelésére, valamint sokkal nagyobb gondot kell fordítani a validitásra (*Ewell, 1991*).
- A standard pontozási (scoring) eljárások hatalmas információvesztéssel járnak főként a tanulásra és a tanításra vonatkozóan. A norma-orientált pontozási eljárások pedig nem adnak elégséges visszacsatolást sem a hallgatóknak sem az oktatóknak, ellene hatnak a jó diák-tanár kapcsolatnak, és az igazodást bátorítják a szellemi függetlenség szemben (*Austin, 1988*). A mérési eljárások visszahatnak a viselkedésre is. A standard eljárások a komplex gondolkodási tevékenységek kerülését, valamint nem elégséges olvasási, írási és problémamegoldási készségeket eredményezhetnek (*Darling-Hammond, 1988*).
- Azt sem helyénvaló feltételezni, hogy az egyes tesztfeladatok (itemek) önálló tudáselemek, amelyek egymástól teljesen elkülönülve alkotnak univerzumot és elég belőlük véletlen eljárásokkal mintát venni. Önmagában egy teszt reliabilitásának (megbízhatóságának) és validitásának a kimutatása nem elég ok arra, hogy azt a tesztet használjuk. Egy teszt akkor jó, ha azt méri, amit a hallgatókkal el akarunk sajátítani, a validitás és a megbízhatóság csak ez után következik (*Shulman, 1987*).
- A hagyományos tesztek annyit árulnak el, hogy a diák tanult, de azt nem, hogy hogyan tanult, így tanítási eszközként nem alkalmazhatóak (*Mentkowski, Astin, Ewell, és Moran 1991*).
- Ennél elsőbb bírálatok is léteznek: szűklátókörűnek, félrevezetőnek, primitívnek jellemzik a hagyományos pszichometriai tesztek (*Novak és Ridley 1987*). Néhányan úgy tartják, hogy az ok-okozati interferenciák iránti hagyományos igény a társadalom- és viselkedéstudományi paradigmákban gátolja az emberi tanulás, fejlődés és viselkedés megértését, végül pedig a pozitívizmus fényében bírálják gyenge pontjait (*Mentkowski, Astin, Ewell, Moran, 1991*). Ezek a bírálatok általában azzal a dorgálással járnak együtt, miszerint az sem védhető ki az egyetemi ok-

tatók és értékelést végzők részéről, hogy ha semmit sem tudnak a hagyományos tesztelméletről és annak gyakorlatáról. Hiányosságai ellenére ugyanis ez az alapja a továbblépésnek.

Shulman (1987) úgy látja, hogy az oktatóknak az értékelési eljárások és technikák kifejlesztése mellett az ezzel foglalkozó tudományos kutatásban is el kellene mélyedniük, főleg annak a tudásnak a növekedésével kapcsolatban, amit mérni vélnek. Bár a kritikák legnagyobb részben nem fogalmazzák meg a hagyományos pszichometria elvetését, mindenképpen elvetik azokat az eljárásokat, amelyek ennek nyomán létrejöttek. Amit itt ki kell emelnünk az az, hogy a felsőoktatási értékelés szélesedésével és sokszínűsödésével, a tesztfejlesztőknek egyre jobban informált és egyre igényesebb felhasználókat kell kielégíteni. Emellett, a felsőoktatásban dolgozók sokkal jobb hatalmi pozícióban vannak – az alsóbb szinteken tanító kollégáikhoz képest – ahhoz, hogy nyomást gyakoroljanak a pszichometrikusokra és tesztfejlesztőkre annak érdekében, hogy azok a tudományos kutatás ilyen irányait is figyelembe vegyék.

Konklúzió: Négyféle dilemma és a jövő

Az amerikai felsőoktatásban az értékelés, ahogy eddig bemutattuk és tárgyaltuk, valamiféle mozgalomként van jelen, folyamatosan mutatva a lendületesség és a terjeszkedés jeleit. A törvényi ösztönzők és a közvélemény támogatása következtében nem is valószínű, hogy ez a közeljövőben megváltozna. A mozgalomban résztvevő egyetemi munkatársak és az értékelést támogató felsőoktatási szövetségek határozottan állást foglalnak ezen az oktatási szinten folyó különböző értékelési tevékenységek helyénvalósága és célja is mellett. Mozgalmi jelleg miatt azonban megvan annak a kockázata, hogy egyre jobban ideologizálódik és inkább a retorika, mint sem az eredmények lesznek meghatározóak (*Ewell*, 1989). Ebben a pillanatban több véleményütközés és dilemma is világosan látható, továbbá az is nyilvánvaló, hogy a valójában létrejövő változások, attól is függenek, hogy ezeket az ellentéteket hogyan lehet feloldani.

Ésszerűség és politika

Az egyetemi oktatók büszkék arra, hogy rendkívül racionális emberek, az ésszerűség talaján pedig szinte lehetetlen az értékelés ellen érveket felhozni. Tágabb értelemben, az értékelés nem más, mint információgyűjtés azokról a tevékenységekről, amelyekben mindannyiunk közös érdekében részt veszünk (*Ewell*, 1989). Az amerikai felsőoktatásra vonatkoztatva, általában elfogadottak azok az érvek, amelyek a miért és hogyan mellett szólnak. Ráadásul a tanulás elősegítésének szempontjából az értékelés szószólóit közvetlenül is érinti az értékelés használhatósága és gyakorlati haszna. A legutóbbi időkben például a Harvard Assessment Seminars életrehozója és kollégái egy kitűnő könyvet állítottak össze felsőoktatási dolgozók számára a kutatástervezésről a (*Light, Singer és Willet*, 1990). A könyv egyik célja annak a bemutatása volt, hogy a pedagógiai és a társadalomtudományi kutatás nem szükségszerűen ezoterikus tevékenység, hanem inkább olyasmi, amiben minden tájékozott oktató résztvehet. Fontos látni, hogy maga a Harvard Assessment Seminars is viszonylagosan elvont és komplex kutatási tervekkel és elemzé-

sekkel indult, de azóta egyszerűsítette megközelítéseit (Ewell, 1989), vonzóvá vált a gyakorlati használhatóságra helyezett hangsúly. Másrészt viszont politikai és/vagy praktikus szempontból is lehet az értékelés ellen érvelni. Lényegében olyan aggodalmakról van szó mint például az, hogy az információ nem megfelelően kerül felhasználásra, hogy csorbítja az oktatók függetlenségét, vagy az oktatókkal szemben támasztott követelmények túlzottak, nem reálisak. Ezek az aggodalmak legtöbbször konkrétan egy dologra vonatkoznak és egy-egy helyhez kötődnek. Az hogy hogyan, milyen mértékben lehet összebékíteni a politikai indíttatású és praktikus szempontokat az ésszerűekkel talán az első kihívás az értékelés széleskörben történő elterjedéséhez vezető úton.

A hagyományos a megreformált tesztelmélet kapcsolata

Második kihívást az jelenti, hogy mennyire képesek az értékelés hívei valamint a méréselmélettel foglalkozók és tesztfejlesztők közös álláspontra jutni. Mint az az előző részből látható volt, sokan szembehelyezkednek a hagyományos tesztelmélet számos alapvető feltételezésével. Ennél is kibékíthetlenebbek azonban azok az érték-konfliktusok, amelyek a gyakorlattal és használattal kapcsolatos döntésekben játszanak szerepet. A felsőoktatási értékelési megközelítések és eljárások fejlődése az elmúlt tíz évben lényegében a hagyományos mérési területen kívül eső, önálló kezdeményezések eredménye volt. Az értékelésben résztvevők számára a kritikus kérdés itt a validitás. Az az állítás, hogy a reliabilitás (megbízhatóság) becsült értékei határt szabnak a validitásnak és releváns következményei vannak a tesztkészítők számára, a felsőoktatásban működő értékelők legtöbbjé számára elfogadhatatlan. Az tesztkészítők legfontosabb célkitűzése az éppen mért szerkezet minél pontosabb meghatározása és annak az összefüggésnek a megértése, amiben a mérés történik, továbbá annak a tisztán látása, hogy az eredményeket hogyan fogják majd felhasználni. Ezek tulajdonképpen mind validitási problémák (Mentkowski és munkatársai, 1991).

Az oktatás és a kutatás viszonya

A harmadik dilemma az, ami mindenütt jelen van a felsőoktatásban és igen jelentős akadályt jelenthet az eljárások megváltozásának széleskörű elterjedésében. Ugyanabban az időben, amikor Barzum a „tanítástól való eltávolodást” írta le (1968), Jencks és Reisman (1968) olyan a körülményekről számolt be, amelyek alsóbbrendű tevékenységé fokozzák le a tanítást. Szerintük a tanítás nem olyan hivatás, mint a kutatás. Ha valaki sikeres tanár, nem jelent nagyobb presztízst, elismerést vagy befolyást, mintha csak közép-szerű tanár lenne. Leírtak néhány módszert arra vonatkozóan, hogy azok akik erre a pályára akarnak lépni, hogyan szerezhethének „klinikai gyakorlat”, illetve műhelyszerű formában tanári képzettséget. Akkoriban még kevés intézmény kínált ilyen képzési lehetőséget, és ez mind a mai napig így van. Legújabbban is csak Keokane (1993) a Duke University rektora jelölte meg a kutatóegyetem egyik kötelezettségeként a posztgraduális hallgatók tanári pályára történő felkészítését. A következőket írja: „A legtöbb posztgraduális hallgatónak csak halvány elképzelése van arról, mit jelent előadásorozatot összeállítani, tanmenetet készíteni, egy kurzus teljes szellemi irányultságáért felelősséget vál-

lalni. Ennek eredményeképpen, az oktatói pálya első éveit a legtöbb egyetemi oktató számára rendkívüli nehézségekkel járnak, pedig a képzés néhány ponton történő nagyon egyszerű megváltoztatásával ezek egy részét ki lehetne küszöbölni”. (114. o.)

Ennek ellenére intézményi, sőt kulturális szempontból is hatalmas tehetlenségi erő hat az olyan törekvésekkel szemben, amelyek a felsőfokú oktatási tevékenységet hivatás- és más tevékenységekkel egyenrangúvá szeretnék emelni. Az értékelés újabb keletű központi szerepének a legbiztatóbb aspektusa éppen az, hogy a tanítás szakmai előrelépésének formája lehet. Optimális körülmények között az értékelés egy olyan folyamat része lehet, amelynek az elején a magas szintű oktatással kapcsolatos elvárások fogalmazódnak meg, a végén pedig olyan szakmai ösztönzők és elismerés következik, amit a kitűnő oktatási tevékenységért adnak. Néhány figyelemreméltó kivételtől eltekintve (ezeket az USA-ban leginkább az elit egyetemeken találhatjuk) ennek az elterjedésére még várni kell. A lehetőség azonban már létezik és ha kihasználják, az amerikai felsőoktatás módszeres fejlődését eredményezheti.

Egységesség és összevethetőség szemben az egyéni felhasználással

Ez az a dilemma, ami mindig is jelen volt és az értékelési tevékenységek különböző fajtáival, azok felhasználásával kapcsolatban folyamatosan figyelmet kell rá fordítani. Azok a feltételek, amelyek mellett az intézmények és az egyes oktatók bekapcsolódhatnak a tanítás, a tanulás és a programok színvonalának értékelésébe, általában teljesen vagy részben ellene hatnak a hallgatói teljesítmények egységes és összevethető mérésének. Ezt a kérdést az előzőekben többször is tárgyaltam, itt már nem szükséges róla külön szólni. Talán csak annyit jegyezni meg, hogy mindig voltak és lesznek is olyanok, akiknek ez az „alku” nem elfogadható. Olyan pontról van szó, ami mindig részletes tárgyalásokat fog megkövetelni az érdekelt felektől.

Ezt az áttekintő tanulmányt a magyar olvasóknak készítettem, olyan időszakban, amikor a magyar felsőoktatás is átalakulóban van. Talán helyénvaló lenne a dolgozat végén az előbbiekről lehetséges helyi jelentőségéről szólni.

1992-ben a Rektori Konferencia, a Felsőoktatási Egyeztető Tanács és a Magyar Ekivalencia és Információs Központ kiadott egy dokumentumot *Hungarian Higher Education 1992* címmel. Ebben a dokumentumban szerepelt egy javaslat, „A felsőoktatás fejlesztése Magyarországon 2000-ig: tézisek és javaslatok” címmel. Valószínűleg más vázlatok és javaslatok is megfogalmazódtak, ám ezek áttekintése túlnőne ennek a tanulmánynak a keretein. A szóban forgó javaslat azonban rámutat azokra a kihívásokra, amelyekkel az elkövetkezendő években a magyarországi felsőoktatásnak szembe kell néznie. Vegyük például a következőket: az általános és középiskolai tanárképzést egységesíteni fogják és az egyetemek tevékenységi körébe utalják (1.6 bekezdés); a központi támogatás hatékonysági és eredményességi mutatók alapján kerül felosztásra (1.9); a legjobb oktatók megtartásának problémája (2.1); az értékelés új rendszerének kidolgozása (2.5) a felvételi vizsgák felváltása egy, a hallgató kezdeti teljesítményén alapuló szelekciós rendszerrel (3.4); nagyobb lemorzsolódási arány, részbizonyítványok kiadása azoknak, akik a diploma megszerzése előtt hagyják abba tanulmányaikat (3.5); a posztgraduális hallgatók bevonása az oktatásba (4.1.2.a) a minőségi és hatékonysági mu-

tatók új ellenőrző rendszerének bevezetése, akkreditációs és minősítő intézmények felállítása, illetve módszerek kidolgozása arra, hogy ezeknek a mutatóknak az értékeit hogyan érvényesítsék (4.1.7).

A fentiek mindegyikének az értékelési eljárásokra is vannak következményei. Ide tartozik a tanítás, a tanulás, a programok és intézmények értékelése a köz felé való elszámoltathatóság és a színvonal tartása céljából. Sok értékelési és minősítési eljárás, módszer még hiányzik, s ezeket viszonylag rövid idő alatt kell majd kifejleszteni és a különböző célokra kiválasztani. Az egyetemi oktatókra és vezetőkre, valamint a köztisztviselőkre marad annak eldöntése, hogy a fejlődési folyamat által követett politikát szem előtt tartva milyen értékelési formákat és eljárásokat válasszanak. Magyarországon most olyan kérdések vannak napirenden, mint centralizáció vagy decentralizáció, helyi autonómia és ellenőrzés, valamint a szűkös anyagi források elosztásának a kérdése. A problémák tehát általánosságban ugyanazok mint az amerikai intézmények esetében, de a megoldás teljes mértékben azok kezében van, akik maguk is közvetlenül érintettek. Ebben a dolgozatban a magam szerény eszközeivel talán sikerült olyan információkat nyújtani, amelyek segítenek felhívni az érdekeltek figyelmét a különböző lehetséges értékelési megközelítésekben rejlő csapdákra és buktatókra, de ugyanakkor rámutatnak azok előnyeikre, erőnyeikre is.

Fordította: Kiss Anikó és B. Németh Mária

Irodalom

- Association of American Colleges (1885): *Integrity in the College Curriculum: A Report to the Academic Community*, Association of American Colleges, Washington D. C.
- Astin, A. W. (1988): *Assessment and Human Values: Confessions of a Reformed Number Cruncher*. (előadás) American Association for Higher Education Assessment Conference, June 8–11, Chicago.
- Barzum, J. (1993): *The American University: How it Runs, Where It's Going*. Second edition. University of Chicago Press, Chicago.
- Bennett, W. J. (1984): *To Reclaim a Legacy: A Report on the Humanities in Higher Education*. National Endowment for the Humanities, Washington D. C.
- Bloom, A. (1987): *The Closing of the American Mind: How Higher Education Has Failed Democracy and Impoverished the Souls of Today's Students*. Simon and Schuster, New York.
- Boyer, C. M., Ewell, P. T., Finney, J. E. és Mingle, J. R. (1987): *Assessment and Outcomes Measurement: A View From the States*. AAHE Bulletin, March.
- Brubacher, J. S., és Rudy, W. (1958): *Higher Education in Transition. An American History: 1636–1956*. Harper and Row, New York.
- College of William and Mary, Office of the Provost (1989): *Assessment of Undergraduate Liberal Education: A Report to the State Council of Higher Education of the State of Virginia*, Office of the Associate Provost, College of William and Mary Williamsburg, Virginia.
- Cremin, L. A. (1988): *American Education. The Metropolitan Experience: 1876–1980* Harper and Row, New York.

Az amerikai felsőoktatási értékelési trendek áttekintése

- Cremin, L. A. (1990): *Popular Education and its Discontents*. Harper and Row, New York.
- Cross, K. P. és Angello, T. A. (1988): *Classroom Assessment Techniques: A Handbook for Faculty*. National Center for the Improvement of Postsecondary Teaching and Learning, University of Michigan, Ann Arbor.
- Cross, K. P. (1989): *Feedback in the the Classroom: Making Assessment Matter*. AAHE Assessment Forum, Association for Higher Education, Washington D. C.
- Cross K. P. (1990): Streams of Thought About Assessment (előadás) American Association for Higher Education Assessment, Conference, June 27–30, Washington D. C.
- Darling-Hammond, L. (1988): Assessment and Incentives: The Medium is the Message. In: *Three Presentations From the Third AAHE Conference on Assessment in Higher Education* No. 6. 1988, Chicago American Association for Higher Education Assessment Forum, Washington D. C.
- Eckstein, M. A. és Noah, H. J. (1993): *Secondary School Examinations: International Perspectives on Policies and Practice*. Yale University Press, New Haven.
- Education Commission of the States (1986): *Transforming the State Role in Undergraduate Education: Time for a Different View*. Education Commission for the States, Denver.
- Ewell, P. T. és Boyer, C. M. (1988): Acting out State Mandated Assessment: Evidence from Five States, *Change*, **20**. 4. sz. 40–47.
- Ewell, P. T. (1991): To Capture the Ineffable: New Forms of Assessment in Higher Education. In *Reprise 1991: Reprints of Two Papers Treating Assessment's History and Implementation*. American Association for Higher Education, Washington D. C. 1–45.
- Ewell, P. T. (1989): Hearts and Minds: Some Reflections on the Ideologies of Assessment. In: *Three Presentations from the Fourth AAHE Conference on Assessment in Higher Education*. American Association for Higher Education Assessment Forum, Washington D. C. 1–26.
- Fong, B. (1987): *The External Examiner Approach to Assessment*. AAHE Assessment Forum, Second National Conference on Assessment in Higher Education, No 6. Denver. American Association for Higher Education, Washington D. C. 14–17.
- Forrest, A. (1990): *Time Will Tell: Portfolio Assisted Assessment of General Education*. AAHE Assessment Forum, American Association for Higher Education, Washington D. C.
- Gage, N. L. és Berliner, D. C. (1991): *Educational Psychology*, Fifth edition. Houghton Mifflin, Boston.
- Hefferman, J. M., Hutchings, P. és Marchese, T. J. (1988): Standard Tests and the Purposes of Assessment. *American Association for Higher Education Assessment Forum*, Washington D. C.
- Hirsch, E. D. (1987): *Cultural Literacy: What Every American Needs to Know*. Houghton Mifflin, New York.
- Hungarian Rectors' Conference (1992): *Hungarian Higher Education 1992*. Coordination Office for Higher Education and Hungarian Equivalence and Information Centre, Budapest.
- Hutchings, P. (1987): Six Stories: *Implementing Successful Assessment*. (előadás) AAHE Second Conference on Assessment, Denver.
- Hutchings, P. és Marchese T. J. (1991): Watching Assessment: Questions, Stories and Prospects. In *Reprise 1991: Reprints of Two Papers Treating Assessment's History and Implementation* American Association for Higher Education, Washington D. C. 47–71.
- Jencks, C. és Riesman, D (1968): *The Academic Revolution*. Doubleday. New York.
- Keane, M. H. (1993): Getting It Right: Authentic Assessment and the True Multiple Measures Approach *Education Week*, October, 27–29.
- Keohane, N. O. (1993): The Mission of the Research University. *Daedalus*, fall, 101–125.
- Light, R. J., Singer J. D. és Willet, J. B. (1990): *By Design: Planning Research on Higher Education*. Harvard University Press, Cambridge, Ma.
- Marchese, T. J. (1987): Third Down, Ten Yards to Go. *AAHE Bulletin* **40**. 4. sz. 3–6.

Paul Ewald

- Mentkowski, M., Astin, A. W., Ewell, P. T. és Moran, T. (1991): *Catching up Theory with Practice: Conceptual Frameworks of Assessment*. American Association for Higher Education Assessment Forum, Washington D. C.
- National Governors' Association (1986): *Time for Results: The Governors' 1986 Report on Education* National Governors' Association, Washington D. C.
- National Institute of Education, Study Group on the Conditions of Excellence in American Higher Education (1986): *Involvement in Learning: Realizing the Potential of American Higher Education*. U.S. Government Printing Office, Washington D. C.
- Neill, M. (1992): *Assessment and the 'Educational Impact Statement' Education Week*, September 23.
- Novak, J. D. és Ridley, D. R. (1987): *Assessing Student Learning in Light of How Students Learn*. (előadás) AAHE Assessment Forum, American Association for Higher Education, Washington D. C.
- Paskow, J. és Francis, E. (1990): *Assessment Programs and Projects: A Directory*. American Association for Higher Education, Washington D. C.
- Rossmann, T. E. és El-Khawass, E. (1987): *Thinking about Assessment: Perspectives for Presidents and Chief Academic Officers*. American Council on Education and the American Association for Higher Education, Washington D. C.
- Shulman, L. S. (1987): *Assessing Content and Process: Challenges for the New Assessments*. (előadás) AAHE Assessment Forum Second National Conference on Assessment in Higher Education, June 14–17, Denver.
- Southern Regional Education Board (1985): *Access to Quality Undergraduate Education: A Report to the Southern Regional Education Board by its Commission for Educational Quality*, Southern Educational Regional Board, Atlanta.
- Stewart, D. M. (1987): *The Ethics of Assessment* (előadás). AAHE Assessment Forum Second National Conference on Assessment in Higher Education, June 14–17, Denver.

ABSTRACT

PAUL EWALD: A REVIEW OF CURRENT TRENDS IN ASSESSMENT IN
AMERICAN HIGHER EDUCATION

This article presents an overview of assessment practices in contemporary American higher education. Assessment is described as a tool used to improve educational practice, and as a means of establishing accountability. The article is divided into four sections. In the first, general background information is presented. This includes some historical context, philosophical and theoretical issues, and reform efforts in the mid-1980s that led to mandates for accountability. In the second section a description of the range of assessment practices in use is presented along with examples of how some specific institutions are making use of them. In the third section an attempt is made to identify some of the major effects that assessment appears to be having. The final section delineates our dilemmas that need to be addressed or resolved in order for assessment to move forward in the United States at the post-secondary level. In the concluding comments the author suggests that in the coming years Hungarian higher education may be confronting some of the same issues and questions that have given rise to the assessment movement in the U.S. in recent years.

English version of the paper can be obtained from the author.

MAGYAR PEDAGÓGIA **94**. Number 3–4. 179–205. (1994)

Levelezési cím / Address for correspondence: Paul D. Ewald, Department of Social and Behavioral Sciences, University of New England, 11 Hills Beach Road, Biddeford, Maine, 04005. U.S.A.