

Bíralunk, de hogyan?

Egy kritika margójára

A Magyar Tudomány 1998. 8. számának 988. lapján Zolnai László és Gácsi Zoltán tollából *Mérünk, de mit?* címmel. Egy formula margójára alcímmel megjelent egy közlemény, amely a szerzők szerint „...az értékelés általános, a »józan ész« által elvárt tulajdonságai” alapján bírálja az MTA intézeteinek munkáját 1996. évben értékelő személyek és bizottságok tevékenységét. Mint egyike a megbíráltaknak, indokoltnak tartom, hogy a szóban forgó cikkkel kapcsolatosan a következő észrevételek nyilvánosságot kapjanak.

Sajnálattal állapítom meg, hogy a bírálat a leglényegesebb kérdéseket, így a *tudományos teljesítmény mibenlétének meghatározását, annak számbavételi lehetőségeit, az egyes teljesítményelemek kijelölését, súlyozását* nyíltan nem vitatja, hanem látszólag csupán *módszertani kérdésekre* koncentrálna. Nyilván ennek is megvan a maga oka. Mint az értékelési módszerek egyik kidolgozója és az MTA Matematikai és Természettudományi Kuratóriuma által felkért Kutatáselemzési Munkabizottság elnöke, elmondhatom, hogy a Bizottságot nagyon nehéz feladat elé állították. Először is meg kellett határozni a különböző szakterületeken működő MTA-kutatóintézetek *tevékenységének* legfontosabb, kvantitatíve is értékelhető *elemeit*, majd ezek jellemzésére mutatószámokat kellett kidolgozni, végül egy *összegző mutatót* kellett javasolnia. Azt, hogy a Bizottság igyekezett pártatlanul és az értékelési eljárás szakmai követelményeit betartva eljárni, igazolják jelentéséből a következők: „A vizsgált természettudományi intézetek hozzávetőlegesen három *informatikai körbe* sorolhatók. Az intézetek döntő hányada abba a körbe tartozik, ahol a természettudományok alapszciplináinak általánosan, szokások, hagyományok követése révén elterjedt kommunikációs rendje érvényesül. Ezeknek az intézeteknek a publikációs tevékenységét értékelve, tudományos kutatási eredményességük is jellemezhető. (Ezek a fizikai, kémiai, matematikai területeken dolgozó intézetek).” — hozzátehetjük az élettudományok diszciplináit és legtöbb szakterületét is.

„Az említett intézetek olyan témákban dolgoznak, amelyek információs csatornáit:

- viszonylag jól reprezentáltak a Science Citation Index adatbázisában;
- a kutatók publikációs igényei belső normatív elvárások révén érvényesülnek (az eredményeket publikálni kell, célszerű a „jobb” folyóiratokban publikálni);
- jelentős részben feltáró jellegű tudományos alapkutatásokat folytatnak, következőképp: a SCI-ben lévő folyóiratok *hatástényező* adatai (átlagos idézettség) alkalmazhatók a folyóiratok minősítésére;

— a kapott idézetek száma többé-kevésbé jellemzi az eredmények *nemzetközi hatását*. (Nem színvonalát, minőségét, hanem *hatását*, ami természetesen összefügg az előbbi kritériumokkal is.)

Az intézetek *második körébe* azok rendelhetők, amelyek tevékenysége nagyobb részben nem az alapszciplinák közé, hanem az *alkalmazott tudományok* (és részben a fejlesztés,

ill. a magas színvonalú *szakmai szolgáltatások*) kategóriájába tartozik. Ez a megállapítás nem értékmegjelölés, hanem a tevékenység többiétől eltérő jellegének jelölésére szolgál.”

„A *harmadik körbe* tartozik...például...” a Földrajztudományi Intézet, amely csak részben sorolható a természettudományi kutatóintézetek csoportjába, hiszen tevékenységének tekintélyes hányada inkább a társadalomtudományok területére esik.

Az utóbb említett két informatikai kör publikálási és hivatkozási szokásai eltérnek az első körétől, s ezért az általános mutatószámok itt nem, vagy csak kevésbé használhatók értékelés céljaira. Az értékelés elvégzéséhez ilyenkor a szakértői (bizottsági) értékelésnek van nagyobb szerepe.”

„Az egyes teljesítményelemek súlyozását természetesen máshogyan is el lehet végezni, a Bizottság csupán egy lehetséges megoldást igyekezett felvázolni. Az egyes teljesítményelemek súlyát az *összteljesítményben* tulajdonképpen az értékelést felhasználónak kellene tudománypolitikai döntések nyomán megadnia, természetesen figyelembe véve az értékelés módszertani szempontjait is.” „Reméljük, hogy akkor, ha ez megtörténik, az értékelés eredményeinek felhasználására is sor kerül, mert öncélúan értékelni legalább olyan helytelen, mint egyáltalán nem értékelni.” — áll a Bizottság jelentésében.

**

Az említettekhez a következőket kell még hozzátennem.

Tudományos szervezetek eredményeinek kvantitatív értékelése *szakmai tevékenység*, amelynek végzéséhez feltétlenül szükséges mind az értékelendő szakterület, mind az elméleti és az alkalmazott tudománymetria ismeretanyagának egy bizonyos színvonalon való tudása. Nyilvánvaló, hogy az értékelési eljárás során az értékelőknek nem csak a szakmai, de az etikai követelményeknek is meg kell felelniük. Meg vagyok győződve arról, hogy a különböző szinteken, fórumokon a szóban forgó akadémiai értékelési módszereket érdemben alakítók etikai vétséget nem követtek el. Ezért csak rosszindulatú feltételezésnek minősíthetem a bírálók következő megjegyzését: „Az eredmény láttán érdemes eljátszani a gondolattal: vajon milyen érdekek és célok tették ezt a felmérést olyanná, amilyen lett?” — idézet a 993. oldalról. Az általam vezetett bizottság tagjaival kapcsolatosan joggal elvárható annyi szakmai hozzáértés és tisztesség feltételezése mind az értékelés megrendelőitől, mind maguktól az értékeltektől, hogy elfogadják, igyekeztünk a ránk bízott feladatot legjobb tudásunk és lelkiismeretünk szerint elvégezni. A Bizottság nem volt elfogult egyik intézet vagy szakterület kárára vagy javára sem. Sajnálom, hogy a bírálók által képviselt intézet az adott értékelési követelmények között nem tudott a legjobbak közé kerülni. Éppen ezért a bírálóknak inkább azzal a gondolattal lenne érdemes eljátszania, hogyan lehetséges az általuk képviselt intézethez hasonló kutatási területen működő másik két intézet kiváló helyezése. Megjegyzendő az előzőekhez, hogy a bírálók munkahelye a kritizált értékelési eljárás 1. változata szerint (a Bizottság jelentésének 3. táblázata) a 19 intézet rangsorában a 7. helyet foglalta el. Célszerű, ha az értékelés módszereit arra alkalmas módon alkalmazásuk előtt az értékeltekkel megvitatják. Azt azonban naivitás lenne feltételezni, hogy az értékelést megrendelő csak olyan mutatószámok használatát támogatná, amelyek — akár indokoltan, akár indokolatlanul is — de mindenki számára előnyösek lennének.

A bíráló cikk 993. oldalán található összefoglaló rész 1. pontjába foglalt következtetés (A módszer a tudományos kutatás általánosan pozitív jellegűnek elfogadott produktumainak létrehozását bünteti.) képtelen és helytelen, mert az alkalmazott értékelő módszerek egyike sem büntette a bírálók által „pozitív jellegűnek elfogadott”-nak nevezett produktumok létrehozását.

A bírálat alapvető hibája, hogy a bírálók *nem azt a formulát alkalmazták, amelyet az értékelők használtak*. A bírálók ugyanis a következő képletet (1989. o.) írják fel:

1. képlet

$$P_k = \sum_{i=1}^n W_i \cdot \frac{a_{i,k}}{\sum_{j=1}^m a_{i,j}}$$

ahol:

- P_k : egy adott intézet bruttó teljesítménymutatója,
 W_i : az i -edik teljesítményelem súlyfaktora,
 n : az összes teljesítményelem száma,
 $a_{i,k}$: a k -adik értékelte intézet i -edik teljesítményelem szerinti mutatószáma,
 $a_{i,j}$: a j -edik intézet i -edik teljesítményelem szerinti mutatószáma,
 m : az összes intézet száma.

Megjegyzendő, hogy a bírálók által bírálatukat alátámasztandóan használt s az értékelőknek tulajdonított 1. képletnek számos pontatlansága van, ezeket e helyütt nem elemzem. A képlet azt a képtelenséget sugallja, hogy minél nagyobb értékű adott intézet adott teljesítménymutatója ($a_{i,k}$) azaz: minél jobban szerepel egy intézet az adott versenyszámban, az ($a_{i,k}/\sum a_{i,j}$) hányados annál nagyobb, így ha ezt az adatot egy számból (W_i -ből, azaz a megfelelő súlyfaktorból) kivonjuk, nyilvánvalóan annál kisebb értéket kapunk.

Nem tudom elképzelni, honnan származik az 1. formula, de a Bizottság sem számszerű, sem szöveges formában ilyen kiszámítási módról soha nem tárgyalt, nemhogy ilyet alkalmazott volna. Elkészerítő, ha a bírálók úgy gondolják, hogy a Bizottság, amelynek tagjai között két matematikus (egyik akadémikus) is volt, az említett butaságot találta volna ki, amelyet tetéz, hogy e képtelenséget feltételezik a kuratóriumokról és az AKT-ről is, amelyek megtárgyalták a Bizottság jelentését. Az említett formulával szemben a Bizottság a 2. képletet alkalmazta az általa ajánlott egyik módszer (lásd Scientometrics, 41 (1998) 185–200) szerint:

$$(BT)_k = \frac{W(M_1)}{K} M_{1,k} + \frac{W(M_2)}{K} M_{2,k} + \dots$$

$$\sum_{i=1} M_{1,k} \quad \sum_{i=1} M_{2,k}$$

2. képlet

ahol:

- $(BT)_k$: a k -adik intézet Bruttó Teljesítménymutatója,
 $W(M_1)$: az M_1 mutatószámmal jellemzett tevékenység súlyfaktora,
 K : az intézetek száma összesen,
 $M_{1,k}$: a k -adik intézet M_1 mutatószáma.

Megjegyzendő, hogy a Bizottság más módszereket is alkalmazott a Bruttó Teljesítmény kiszámítására.

A Bizottság jelentését az MTA Matematikai és Természettudományi Kuratóriuma az intézetek értékelésének alapjául elfogadta. A későbbi anyagi következmények alapjául azonban nem az említett Bizottság által készített értékelés eredményei szolgáltak. Az Akadémiai Kutatóhelyek Tanácsa ugyanis mind az alkalmazott a teljesítményelemek számát, mind a súlyozást alapvetően megváltoztatta.

A módosítás csak a következő elemeket (zárójelben azok súlya) vette figyelembe:

- összes tudományos publikáció száma/kutatók száma (8),
- SCI-publikációk száma/kutatók száma (16),
- a folyóiratcikkek (normált) hatástényezőinek összege/kutatók száma (24),
- idézetek száma/kutatók száma (32),
- pályázati bevételek összege/kutatók száma (15),
- PhD-disszertációk száma/kutatók száma (5).

Az AKT-által megadott módszer szerint az adatokat a jelen cikk szerzője számította ki a következők szerint. Az intézetek egyes mutatószámait átlagoltam. Kiszámítottam az átlagadat (\bar{x}) szórását (SD).

Az i -edik intézet (x_i) ezek után a következő pontszámokat kaphatta:

<i>besorolás</i>	<i>pontszám</i>
$x_1 \geq x + SD$	4
$x \leq x_1 < x + SD$	3
$x - SD \leq x_1 < x$	2
$x_1 < x - SD$	1

Az egyes intézeteknek az adott teljesítménymutató szerint elért pontszámát szoroztam a megadott kategóriára vonatkozó súlyfaktoriall, majd az egyes szorzatokat intézetenként összegeztem. Így alakult ki az intézeteknek egy bizonyos sorrendje. A sorrend alapján az intézeteket három, ill. négy kategóriába csoportosították, amely besorolás volt az alapja bizonyos támogatások szétosztását érintő pénzügyi intézkedéseknek. Megjegyzendő, hogy az alkalmazott eljárás az értékelés során elkövetett esetleges módszertani hibákat jelentős mértékben csökkentette.

A különféle értékelési eljárások értelemszerűen nem adhattak teljesen azonos eredményeket, hiszen más és más mutatószámok más és más súllyal szerepeltek. Az azonban megfontolandó, hogy az intézetek bruttó pontszámára a Bizottság (a 2. képlet szerint) és az AKT módszer szerinti sorrendek között számított Spearman-féle rangkorrelációs koefficiens értéke: 0,864.

A bírálókkal szemben megállapíthatjuk, hogy bármely teljesítménynek elismert tevékenység révén hozott legkisebb mértékű eredmény is *növelte a bruttó mutatót*.

Minden értékelés orientál. A Bizottság s ezzel összhangban az AKT szándéka nyilvánvaló volt — a súlyfaktorok ezt egyértelműen mutatják: a *nemzetközi hatást kifejtő tudományos publikációs tevékenységet* kell értékelni és elismerni. A nagyon különböző szakterületek teljesítményeinek összehasonlítása kétségtelenül nehéz és csak *kellő szakértelemmel, megbízható adatbázisra építve* szabad végrehajtani, de még mindig kisebb kockázattal és veszteséggel jár, mint az értékelés elhanyagolása, az elismerés és támogatás lobbyzás révén történő érdemtelen osztogatása.

Az értékelés: kellemetlen lehet egyesek számára, főként, ha számokban fejezik ki a teljesítményt. Az viszont igaz, hogy a helytelen értékelés árthat, és hogy a tudományos kutatás értékelési eljárásai még nem kiforrottak. Kevés az ehhez a szakmához értő szakember, és sokszor hiányoznak a megbízható adatok is.

A Bizottság értékelési módszerében a bírálók az egyik legnagyobb hiányosságnak azt tartják, hogy az értékelés a tudományterületek „eltérő relatív tökeellátottságát” nem vette figyelembe. Valóban így van, de a tudományos kutatók világversenye kíméletlen, abszolút verseny abban az értelemben, hogy nincsen tekintettel a versenyzők egyéni, szervezeti, anyagi stb. körülményeire. Az értékelés eredményeinek alapján viszont az *illetékes tudománypolitikusoknak és irányítóknak* az értékeltek speciális körülményeire való figyelemmel kell levonni a megfelelő következtetéseket és megtennie a szükséges intézkedéseket.

Számos közleményre hivatkozhatnánk, amelyek mind a tudománymetria módszereinek ésszerű és célszerű alkalmazása mellett szólnak. E helyütt egyetlen dolgozatra utalok: C. Oppenheim angol kutatóhelyek egy kutatóra jutó idézeteinek számát és a Universities Funding Council Research Assessment Exercise elnevezésű, országos szintű szakértői értékelés pontokban kifejezett adatait hasonlította össze. A két adatsor Spearman-féle rangkorrelációs tényezőjét: 0,82-nek találta (Journal of Documentation, 51 (1995) 18—27).

Remélem, hogy a tudományos teljesítmény értékeléséről mindazokkal, akiket ez a téma érdekel vagy hivatalból illet, a jövőben kölcsönös jószándékot, szakmai tudást vagy etikuss viselkedést feltételező eszmecseréket folytathatunk.

Vinkler Péter