

OCENJEVANJE RAZISKOVALNIH PROGRAMOV NA PODROČJU VZGOJE IN IZOBRAŽEVANJA: ETIČNE DILEME IN METODOLOŠKI PROBLEMI

Zdenko Kodelja

Pedagoški inštitut, Ljubljana

Ocenjevanje raziskovalnih dosežkov je v sodobnih razvitih družbah ključnega pomena za načrtovanje in vodenje raziskovalne politike na nacionalni ravni. Bolj ko se je povečevalo vlaganje v znanstvene raziskave, večjo težo so takšnemu ocenjevanju pripisovale ustanove, ki so pristojne za načrtovanje, vodenje in financiranje razvoja znanosti v neki državi, saj so rezultati takega ocenjevanja po eni strani osnova za njihove odločitve o tem, kaj in koliko financirati oz. kako usmerjati razvoj raziskovalne dejavnosti, po drugi strani pa so lahko uporabljeni tudi kot argument za dokazovanje odgovorne uporabe proračunskega denarja. Način ocenjevanja raziskovalnih dosežkov se razlikuje ne samo od države do države, temveč tudi v različnih obdobjih znotraj ene in iste države. Kljub temu pa je mogoče v vsakem obdobju opaziti tudi neke skupne značilnosti takšnega ocenjevanja v različnih državah. Te podobnosti so običajno plod večjega ali manjšega posnemanja ocenjevalnega modela, ki v nekem obdobju velja za najboljšega. Danes je denimo mnogim državam skupna hkratna uporaba kvalitativnih in kvantitativnih oblik ocenjevanja. Vendar je videti, da dobiva kvantitativno ocenjevanje na podlagi bibliometričnih kazalcev vedno večjo težo. Ti kazalci so za ene le navidezno racionalno sredstvo za »vladanje znanosti in učenjakom ter za to, da dajejo utemeljitve znanstvenega videza birokratskim odločitvam«,¹ za druge pa sredstvo za vodenje pravične raziskovalne politike, ki razdeljuje finančna sredstva v skladu z raziskovalnimi dosežki znanstvenikov. Toda tudi v drugem primeru začnejo ti kazalci negativno vplivati na samo raziskovanje, ko postanejo tako pomembno merilo za ocenjevanje uspešnosti raziskovalcev, da sta od njih odvisna položaj in poklicna kariera raziskovalcev. Takrat namreč najvažnejši cilj znanstvenikov ni več toliko znanstveno odkritje kot tako, ampak čim več objavljati.² To je problem, s

katerim se že soočamo tudi pri nas, čeprav je vzpostavljen sistem ocenjevanja raziskovalnih dosežkov, ki upošteva tudi druge kriterije. Prisoten je bil tudi pri letošnjem panelnem ocenjevanju raziskovalnih programov, ki sicer spada med kvalitativne oblike ocenjevanja. Če ne pri vseh, pa je bilo tako vsaj pri ocenjevanju programov na področju vzgoje in izobraževanja. To pa zato, ker so imeli bibliometrični kazalci, ki so bili dani ocenjevalcem na razpolago kot eden od elementov za oblikovanje njihove ekspertne ocene o kvaliteti posameznih raziskovalnih programov, velik vpliv na njihove ocene in posledično seveda tudi na samo razvrstitev ocenjevanih programov. V nadaljevanju bom predstavil nekatere etične dileme in metodološke probleme, s katerimi sem se soočil kot eden od ocenjevalcev na tem panelu.³ Na njem so se pokazale nekatere težave, ki so bolj ali manj značilne za panel kot obliko kvalitativnega ocenjevanja,⁴ druge pa so bile verjetno značilne le za ta panel.⁵ Zato tega, kar je veljalo za panelno ocenjevanje programov na področju vzgoje in izobraževanja, ne moremo enostavno posplošiti na druge panele.

To, kar se je zgodilo na panelu, ki je potekal 26. junija 2008 na Pedagoškem inštitutu v Ljubljani, in tudi sama metodologija ocenjevanja raziskovalnih programov, pa sta gotovo nekaj, kar terja resen premislek, saj so posledice lahko usodne za razvoj raziskovanja na področju vzgoje in izobraževanja. Dejstvo je namreč, da so rezultati ocenjevanja vsaj za nekatere krivični in zato tudi nesprejemljivi. Prav tako je dejstvo, da so rezultati ocenjevanja že povzročili konflikte na medosebni ravni. Oboje je, milo rečeno, slabo. Zato je treba poskusiti čim bolj nepristransko presoditi, ali so občutki, da se je nekaterim zgodila krivica, upravičeni, in če so, kaj so vzroki za storjeno krivico. Ali je glavni vzrok pristransko in moralno sporno ravnanje tistih, ki smo na tem panelu sodelovali, ali pa je vzrok že v samem postopku ocenjevanja, ki ga je določila Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS)?

Mislim, da bi bilo narobe, če bi se tem vprašanjem poskušali izogniti z znanim in mnogokrat uporabljenim argumentom, da je pač pri takem ocenjevanju vedno tako, da so eni z oceno zadovoljni, drugi pa ne, in da se zato s tem sploh nima smisla ukvarjati. Menim tudi, da je bilo naše ravnanje v veliki meri odvisno od položaja, v katerega smo bili postavljeni v skladu s proceduro, ki jo je zahtevala predpisana metodologija ocenjevanja. Zato bom svoj pogled na dogajanje pred panelom in na njem začel prav z vprašanjem metodologije, čeprav ta ni popolnoma determinirala našega ravnanja. V zadnji instanci smo se vendarle sami svobodno odločili, kako bomo ravnali. Zato ne moremo preprosto prenesti odgovornosti za naše odločitve in ravnanje na druge, tj. na metodologijo ali na tiste, ki so jo določili. S tem pa nikakor ne mislim, da bi morali vso odgovornost ali morebitno krivdo sprejeti nase. Kajti metodologi-

ja celotnega postopka ocenjevanja programov je nedvomno močno vplivala tako na naše odločitve kakor tudi na posledice našega ravnanja.

Zame je ta metodologija ocenjevanja raziskovalnih programov sporna zaradi več razlogov:

Prvič zato, ker so bili v našem primeru predmet ocenjevanja raziskovalni programi tako s področja vzgoje in izobraževanja kakor tudi kineziologije.⁶ Sam menim, da gre za tako različne programe, da jih ni mogoče ocenjevati skupaj. Veliko vprašanje je tudi, ali kineziologija sploh spada v družboslovje.⁷ Večina predstavljenih raziskav v okviru kineziologije je namreč videti veliko bolj podobnih naravoslovnim ali medicinskim raziskavam kakor družboslovnim. Vsekakor pa nimajo skoraj nič skupnega z izobraževanjem. Zato sem uradnim osebam na ARRS večkrat povedal, da se mi zdi sporno, da bi vodje programov s področja izobraževanja ocenjevali tudi programe s področja kineziologije in nasprotno. Sam sem vztrajal, da programov iz kineziologije ne bom ocenjeval, ker se na to področje ne spoznam. Tako sem tudi ravnal v postopku ocenjevanja pred panelom in na njem. Vsako drugačno ravnanje bi bilo zame moralno nesprejemljivo. To sem tudi vsem javno povedal. Na panelu sem tudi predlagal, da kineziologi ocenjujejo kineziologe, raziskovalci s področja edukacijskih ved pa programe s področja vzgoje in izobraževanja. To je bila možna rešitev tega problema, ki bi bila, če bi jo na panelu sprejeli, sprejemljiva tudi za ARRS (tako mi je bilo vsaj rečeno na ARRS). Čeprav se je veliko udeležencev panela z mano strinjalo, da ni prav, da ocenjujemo drug druge, so razen enega, vsi to storili. To je bila seveda njihova pravica, še več, po metodologiji celo neke vrste dolžnost, čeprav moram reči, da so me na ARRS sicer pregovarjali, da ocenim tudi kineziološke programe, nihče pa me ni silil, da to proti svoji volji tudi storim. To pa ne pomeni, da tistim, ki so ravnali drugače kot sem sam, kar koli očitam. Vsakdo zase ve, na kaj se spozna, in vsakdo je ravnal tako, kot je presodil, da je prav. Ne zdi pa se mi prav, da metodologija ocenjevanja ne samo dopušča, ampak v primerih, kot je bil naš, tudi predvideva takšno ocenjevanje. Prepričan sem namreč, da ni mogoče korektno oceniti programa na področju, ki ga ne poznaš, pa čeprav se še tako trudiš biti nepristranski. Še na področju, ki ga precej dobro poznaš, se večkrat zgodi, da nekaterih stvari ne poznaš dovolj, da bi lahko kompetentno razsojal. Sam sem v takem primeru, ko je šlo za predlog ocenjevanja nekaterih tujih programov, ocenjevanje zavrnil, ker bi zaradi pomanjkljivega poznavanja problematike, ki je bila predmet raziskovanja, lahko naredil komu veliko krivico. Iz tega sklepam, da sedanja metodologija pri nas (in seveda tudi v državi, iz katere sem bil povabljen, da ocenim raziskovalne programe) ne onemogoča,

da program ocenjuje tudi nekdo, ki ni ekspert za specifično tematiko, ki je predmet raziskovanja v takem programu. Vem, da ni mogoče vedno dobiti idealnega ocenjevalca, toda obstajati bi morali ne samo jasno in natančno določeni minimalni kriteriji, ki še omogočajo kompetentno ocenjevanje, ampak tudi postopek, ki bi zagotavljal, da ocenjevalci te kriterije izpolnjujejo. Prav tako bi vsakdo od udeležencev panela moral imeti možnost preveriti, ali jih res izpolnjujejo. Če obstoječa metodologija tega ne zagotavlja, bi jo bilo treba spremeniti. Kajti obstoječa metodologija kvečjemu omogoča posamezniku, ki zaradi etičnih razlogov noče ocenjevati nečesa, kar ne pozna, da tega ne stori. To pa ne zadošča, saj ne preprečuje, da drugi ocenijo program tega istega posameznika, tudi če se na tematiko, ki je predmet raziskave v tem programu, ne spoznajo. To pa je zame nesprejemljivo in krivično. Poleg tega taka ocena ne more veljati za oceno ekspertov, kar naj bi (tako je bilo mogoče vsaj razumeti intenco ocenjevanja, ki ga izvajajo eksperti in ne uradniki ARRS) bila. Kot ekspertna ocena pa taka ocena ne more veljati zato, ker je nujni pogoj zanjo ravno to, da ocenjevalci niso samo eksperti na svojem področju, ampak morajo biti eksperti prav za področje, ki ga ocenjujejo. V nasprotnem primeru ocena ne more biti ekspertna ocena. Če jo kdo tako razume in jo kot tako zagovarja z argumentom, da je taka ocena ekspertna ocena zato, ker so jo dali eksperti, je to posledica logične napake v dokazovanju imenovane »argumentum ad verecundiam«.⁸ Če pa kdo meni, da za korektno in kompetentno ocenjevanje ni potrebno, da ocenjevalec take pogoje izpolnjuje, ker ocenjuje le na podlagi objektivnih kazalcev uspešnosti raziskovanja, ki so zbrani v tabelah, ki jih pripravijo na ARRS, potem ni pravega razloga, da to počnemo raziskovalci oz. vodje programov, saj bi to lahko opravili uradniki na sami agenciji.

Drugi razlog, zaradi katerega je zame obstoječa metodologija sporna, je dejstvo, da je moral vsak vodja programa oceniti preostale programe. To je sporno, tudi če ocenjujemo le tiste programe, ki obravnavajo tematiko, katero vsaj do določene mere poznamo: recimo to, da smo ocenjevali programe na področju vzgoje in izobraževanja. Sporno je najprej zato, ker smo ocenjevali programe, v katerih sodelujejo raziskovalci, ki jih ne samo večinoma osebno poznamo, ampak smo z nekaterimi tudi sodelovali in še vedno sodelujemo v različnih projektih, strokovnih združenjih itd. Zato nepristranskosti in s tem tudi pravičnosti ocenjevanja ni mogoče zagotoviti zgolj s tem, da nihče ne ocenjuje »svojega« programa. Pravzaprav bi se v tem položaju, ko pogoji za nepristransko ocenjevanje niso zagotovljeni, naša pristranskost ali nepristranskost celo jasneje pokazala, če bi ocenjevali tudi »svoj« program. Zato bi bilo najbrž bolje, če bi vodje programov ocenjevali vse programe, tudi »svojega«, pod pogojem, da bi bile ocene javne in v primeru ugotovljene uporabe raz-

ličnih kriterijev ocenjevanja »svojega« in drugih programov tudi korigirane na panelu. Sedaj pa je tako, da je »svojemu« programu škodil tisti, ki je pri ocenjevanju uporabljal milejši kriterij kot drugi ocenjevalci, ki so ocenjevali »njegov« program. Zato bi bilo treba (na to sem opozarjal že v komentarju k ocenam programov pred panelom in na samem panelu) v takih primerih oceno korigirati. Prav tako bi postopek ocenjevanja moral vključevati možnost, da se v skladu s predpisanim postopkom korigira oceno takrat, ko se ocene ocenjevalcev za neki program ali del programa med seboj pomembno razlikujejo.

Etična dilema, s katero smo bili ocenjevalci pri tem soočeni, pa je, ali takrat, ko veš, da predpisani postopek ocenjevanja raziskovalnih programov ne zagotavlja nepristranskega ocenjevanja, sprejeti vlogo ocenjevalca ali ne. Kakor koli se namreč odločiš, ima tvoja odločitev lahko negativne posledice. Kljub temu pa smo vsi programe ocenili. Oceno smo podali, čeprav smo dobro vedeli, da to ni prav. Zakaj so drugi tako ravnali, ne vem, sam pa v tem primeru (v nasprotju s tistim, ko je šlo za ocenjevanje programov s področja kineziologije kot nečesa, česar ne poznam) nisem vztrajal, da programov s področja vzgoje in izobraževanja ne bom ocenil. Vendar ne zato, ker bi mislil, da je kriva metodologija, ki predpisuje tako ocenjevanje, in ne ocenjevalci, ki ravnajo v skladu s predpisanim postopkom, ampak zato, ker sem menil, da pri ocenjevanju še vedno lahko ravnam v skladu z enim temeljnih etičnih načel: *primum non nocere*, tj. predvsem ne škoditi. To pa zato, ker je ravno dejstvo, da je v vsakem ocenjevanem programu vsaj nekdo, ki ga poznam, ali smo celo v prijateljskih odnosih, od mene terjalo, da vsak program ocenim, kolikor je mogoče nepristransko. Koliko mi je to uspelo, je seveda drugo vprašanje.

Tretji razlog, zaradi katerega ocenjujem sedanjo metodologijo ocenjevanja raziskovalnih programov kot problematično, so kriteriji za ocenjevanje odličnosti. S tem ne trdim, da veljavni kriteriji niso pomembni. So pomembni. Menim pa, da niso zadostni. Vsaj ne na področju raziskovanja vzgoje in izobraževanja. Uporaba zgolj teh kriterijev (SCI, SSCI, objave pod kategorijo A1⁹) lahko vodi v oblikovanje izkrivljene podobe znanstvene odličnosti raziskav na področju vzgoje in izobraževanja. Ti kriteriji so namreč veliko primernejši za naravoslovje in njemu sorodne vede kot za velik del družboslovja. Prenašanje naravoslovnega modela na celotno družboslovje je, kot je znano, problematično.¹⁰ Uporaba istih kriterijev za področja, ki ne spadajo v isto kategorijo, pa je po definiciji krivična. Zato bi po mojem mnenju metodologija morala upoštevati tudi specifiko posameznih ved in področij raziskovanja. Pri tem ne gre za poskus izogibanja oceni znanstvene odličnosti na podlagi rigoroznih in mednarodno primerljivih kriterijev. Gre za to, da metodologija nekaterih pomembnih kriterijev ne upošteva.¹¹ Recimo citatov v znanstvenih monografijah, ki so

izšle pri uglednih tujih oz. mednarodnih založbah, ali pa navedbo nekega dela v knjigi ali poglavju knjige, katerega avtor je na svojem področju celo avtoriteta v svetovnem merilu. Poleg tega se ne upošteva, da nekatere zelo pomembne revije na družboslovnem področju sploh niso vključene med tiste, v katerih citati štejejo kot SSCI. Upoštevati je treba tudi, da sam citat še ne pomeni nujno, da je citirano delo tako izvrstno, da mu lahko avtomatsko pripišemo vrednost znanstvene odličnosti. Lahko je na primer citirano kot paradigmatški primer popolnoma zgrešene teorije, interpretacije itd.¹² Zgolj iz podatka o številu citatov tudi ne vemo, ali se citirano delo nanaša na področje, ki je predmet raziskovanja v okviru določenega raziskovalnega programa, ali ne. Prav tako ni razvidno, kdo je neko delo citiral. Če gre za domačega avtorja, se lahko zgodi, da gre za objavo zgolj prevedenega ali nekoliko predelanega besedila, ki je že bilo objavljeno v slovenščini. V takem primeru je tako, da citat istega dela v slovenski reviji ne šteje nič, v mednarodni pa šteje ne samo kot citat, ampak se iz tega tudi sklepa na znanstveno odličnost citiranega dela. Zato bi ocenjevalci morali imeti na razpolago dodatne informacije, na podlagi katerih bi lahko natančneje ocenili znanstveno odličnost določenega raziskovalnega programa.

Če metodologija ocenjevanja raziskovalnih programov v tem pogledu ne bo spremenjena, bo to lahko imelo dolgoročno zelo hude negativne posledice za raziskovanje vzgoje in izobraževanja v Sloveniji. Ena od takih posledic, ki jo je mogoče z veliko verjetnostjo predvideti, je, da se bodo že na naslednjem razpisu oblikovale raziskovalne skupine povsem drugače, recimo tako, da se bo vključevalo vanje zadostno število naravoslovcev z velikim številom citatov in objav v revijah, ki spadajo v kategorijo A1".¹³ To pa bo pomenilo nujno redukcijo raziskovalne tematike na samo nekaj z naravoslovjem bolj ali manj povezanih tem, številne teme, pomembne tako za razvoj edukacijskih ved kakor tudi za razvoj celotnega področja vzgoje in izobraževanja v Sloveniji, pa preprosto ne bodo predmet raziskovanja. Vsaj ne v okviru programskega financiranja. Veliko vprašanje je tudi, koliko bo uporaba predvsem teh kriterijev za oceno kakovosti oz. odličnosti raziskovalnih dosežkov resnično prispevala k dvigu kakovosti raziskovanja na celotnem področju vzgoje in izobraževanja. S tem pa, še enkrat ponavljam, ne trdim, da so obstoječi kriteriji neprimerni, niti ne gre za manifestacijo obrambnega mehanizma, znanega pod imenom »kislo grozdje«. Skratka, s temi pripombami nočem zmanjševati pomena dosežkov raziskovalnega programa, ki ima po teh kriterijih najboljše dosežke med ocenjenimi programi, češ da ti kriteriji, zato ker jih drugi nismo sposobni v enaki meri doseči, za raziskovanje vzgoje in izobraževanja niso pomembni. Da ne gre za to, priča dejstvo, da sem ta program glede na obstoječe kriterije ocenil z visoko oceno tako v ocenjevalnem postopku pred panelom kakor na

samem panelu. Prepričan pa sem, da ti kriteriji niso zadostni. Moja kritika je torej namenjena kriterijem in ne enemu od ocenjevanih programov.

Četrty razlog, zaradi katerega menim, da je obstoječa metodologija ocenjevanja raziskovalnih programov problematična, je dejstvo, da ne preprečuje v zadostni meri precej različnih ocen enega in istega raziskovalnega programa. To pa zato, ker navodila o tem, kakšne rezultate mora program doseči, da je ocenjen recimo z oceno pet, niso dovolj natančna in tudi ne upoštevajo specifike posameznih znanstvenih disciplin ali področij raziskovanja. Če je na področju naravoslovja morda doseženo dovolj široko soglasje o tem, kaj šteje za vrhunski znanstveni dosežek v svetovnem merilu, pa za področje vzgoje in izobraževanja, kolikor vem, takega soglasja ni. Zato bi morali ocenjevalci raziskovalnih programov s tega področja pred ocenjevanjem na panelu (ali najpozneje na samem panelu) vsaj približno uskladiti svoje poglede na to, kaj eventualno šteti kot vrhunske znanstvene dosežke v svetovnem merilu. Podobno velja za oceno dosežkov na mednarodni in pravzaprav tudi na nacionalni ravni. Poleg tega menim, da korektne ocene ni mogoče oblikovati zgolj na podlagi bibliometričnih kazalcev in kratke, le nekaj strani obsegajoče vsebinske predstavitev najboljših dosežkov dosedanjega programa in strukture novega programa. Vsaj tisto, kar so posamezni obstoječi programi sami navedli kot njihove najboljše dosežke (denimo knjiga, poglavje v knjigi, članek), bi moralo biti v celoti, tako kot je v nekaterih državah, predmet natančne recenzije (peer-review) vsaj dveh priznanih strokovnjakov za tematiko (kar mora biti razvidno tudi iz njihovih referenc), ki je bila predmet raziskovanja. Če se oceni pomembno razlikujeta, pa je treba dati besedilo v presojo še vsaj tretjemu recenzentu.¹⁴ Sedanja metodologija, ki daje prednost kvantitativnemu ocenjevanju znanstvenih dosežkov (ki temelji predvsem na bibliometričnih kazalcih) pred kvalitativnim (recenzije oz. *peer review*), sicer prispeva k večji objektivnosti ocen, vendar na račun njihove zanesljivosti, čeprav je res, da je včasih tudi zanesljivost recenzij vprašljiva.¹⁵ Toda to velja tudi za recenzije člankov, ki so objavljene v revijah – tudi tistih najuglednejših. Zato je zgolj na podlagi tega, da je neki članek objavljen v ugledni reviji, zmotno sklepati, da zaradi tega tudi ta objavljeni članek nujno spada v kategorijo znanstvene odličnosti.¹⁶ Poleg tega ima to, da se bibliometričnim kazalcem pripisuje prevelika teža v ocenjevalnem postopku, od katerega je odvisno financiranje raziskav in kariere raziskovalcev – kar seveda še zdaleč ni samo slovenska posebnost – še druge negativne posledice. Med njimi velja omeniti vsaj tisto, na katero opozarja Peter Lawrence, ugleden biolog z univerze v Cambridgeu, namreč, da najvažnejši cilj znanstvenikov niso več znanstvena odkritja, ampak čim večje število objav v uglednih znanstvenih revijah

z visokim faktorjem vpliva. Tako je število tovrstnih objav, ki je bilo najprej zgolj merilo – in kot takšno le sredstvo za ocenjevanje uspešnosti raziskovanja – postalo prioritetni cilj tudi za same znanstvenike. Ti se zato namesto z znanstvenimi problemi in njihovimi rešitvami vse bolj ukvarjajo s strategijami objavljanja člankov v uglednih znanstvenih revijah in z lovom na citate.¹⁷

Peti razlog, zaradi katerega se mi zdi obstoječa metodologija ocenjevanja raziskovalnih programov pomanjkljiva, pa zadeva tako ocenjevanje kakor tudi posledice tega ocenjevanja. Najprej je problem v tem, da se ocenjujejo skupaj obstoječi in novi programi, čeprav novih programov po določenih kriterijih sploh ne moremo oceniti (denimo vpliv raziskovalnih dosežkov programa za razvoj znanosti ali družbe), saj program, ki ga še ni, takih vplivov ne more imeti. Vprašanje je tudi, ali so primerljivi individualni dosežki raziskovalcev, ki so bili doseženi zunaj programa, in dosežki raziskovalcev, ki so rezultat raziskovalnega dela v okviru nekega že obstoječega programa. Če so primerljivi, potem je to videti nekako tako, kot da je vseeno, ali se financira raziskave posameznikov ali programskih skupin. Drugi problem pa je v ne dovolj jasni razlagi tega, kako se bodo financirali novi programi. Iz dosedanjih razlag pristojnih na ARRS je bilo pričakovati, da se bodo dobri novi programi financirali bodisi iz novih sredstev bodisi iz sredstev, ki se bodo sprostila z zmanjševanjem ali ukinjanjem financiranja slabih raziskovalnih programov. Ker očitno novih sredstev ni, se bodo novi programi financirali iz drugega omenjenega vira. Ključno vprašanje pri tem je, kaj bo okvir za določitev slabih programov, ki bodo ukinjeni. Ali bodo izbrani izmed vseh najslabše ocenjenih programov, ki jih je doslej financirala ARRS, ali izmed najslabše ocenjenih programov znotraj določenega področja (denimo družboslovja), ali celo zgolj znotraj posameznega panela. V prvem in drugem primeru je težava v tem, da so bili kriteriji ocenjevanja lahko na posameznih panelih različni. Tako se lahko zgodi, da bo ukinjen program, ki ne bi bil ukinjen, če bi bil ocenjen v skladu s kriteriji, ki so bili uporabljeni na nekem drugem panelu. V tretjem primeru pa je problem, če se zgodi tako, kot se je na našem panelu, ko so vsi programi dobro ocenjeni in torej ni programa, ki bi ga lahko ukinili zato, ker je slab. Ukiniti dober program, pa čeprav je na lestvici dobrih programov na zadnjem mestu, bi bilo krivično in v nasprotju z dosedanjo razlago ARRS, kako bo financirala nove programe.

To je seveda le hiter in zato nedvomno pomanjkljiv poskus refleksije etičnih dilem in metodoloških problemov, povezanih z ocenjevanjem raziskovalnih programov na področju vzgoje in izobraževanja, ki je potekalo letos poleti. Prepričan sem, da nastali položaj zahteva resno obravnavo in poskus iskanja rešitev, ki bodo kljub legitimnim različnim interesom raziskovalcev vendarle preprečile, da se bo znanstveno življenje na tem

področju nadaljevalo kot *bellum omnium contra omnes*.¹⁸ Kajti če se bo to zgodilo, ne bo dobro ne za posameznike ne za raziskovalne ustanove na področju vzgoje in izobraževanja. Še več, dolgoročno bo to lahko imelo usodne posledice za raziskovanje vzgoje in izobraževanja v Sloveniji.

Opombe

- [1] P. Bourdieu (2004: 50–51).
- [2] P. Lawrence (2008).
- [3] O tem sem govoril že na posvetu o posledicah panelnega ocenjevanja raziskovalnih programov, ki je bil na Pedagoškem inštitutu nekaj dni po panelu, in v zapisu, ki sem ga najprej posredoval raziskovalnemu svetu za družboslovje pri Javni agenciji za raziskovalno dejavnost Republike Slovenije (ARRS), nato pa še tistim, ki so bili v okviru ARRS pristojni za evalvacijo izvedbe vseh panelov.
- [4] L. Langfeldt (2004: 51–62; 2006: 31–41).
- [5] Ena od značilnosti tega panela je bila, da je šlo za ocenjevanje raziskovalnih programov znotraj nekega interdisciplinarnega področja in ne znotraj neke znanstvene discipline. Ekspertno ocenjevanje interdisciplinarno zasnovanih programov pa se sooča s posebnimi metodološkimi problemi (G. Landel, 2004), tudi če drugih problemov, ki bodo predmet analize v nadaljevanju, ne bi bilo.
- [6] V Angliji so izobraževanje, psihologija in vede, ki preučujejo šport, res združeni pod enega od skupnih glavnih panelov, vendar se kot subpaneli obravnavajo in ocenjujejo vsak zase (*Research Assessment Exercise*, 2008). To pa je ravno nasprotno kot v Sloveniji.
- [7] Tu puščam ob strani vprašanje, ali tudi vse raziskave, ki so se ocenjevale v okviru sklopa »izobraževanje«, res spadajo v družboslovje. Ali recimo filozofija vzgoje in zgodovina šolstva spadata bolj v družboslovje ali humanistiko? Odgovor je odvisen od tega, ali se ti vedi obravnavata kot filozofska in zgodovinska disciplina ali ne. Pravzaprav je podobna težava tudi s pedagogiko, katere raziskovalni program je pri nas uvrščen v družboslovje, v Italiji pa je bil na primer uvrščen v humanistiko (skupaj z zgodovinskimi, filozofskimi in psihološkimi vedami) in ne v skupino družbenih in političnih ved (A. Barbara, A. Costantini, M. Costantini, I, 2006/4, 2007/1: 3). Sama razvrstitev teh disciplin z vidika ocenjevanja raziskovalnih rezultatov niti ne bi bila toliko pomembna, če se merila ocenjevanja v družboslovju in humanistiki ne bi razlikovala. Vendar se.
- [8] J. Woods and D. Walton (1974: 135–153; J. Woods and D. Walton, 1982; D. Walton, 1985; D. Walton, 1989: 59–73).
- [9] SCI je okrajšava za izraz »Science Citation Index«, SSCI je kratica za »Social Sciences Citation Index«, A1" pa oznaka za knjigo, ki je izšla pri mednarodni založbi, ali za članke v zgornjih pet odstotkih revij na SCI področjih, ali za članke v reviji s faktorjem vpliva (IF), ki je večji od pet, ali v družboslovju za članke v zgornji polovici revij SSCI.

- [10] Enako velja za ne dovolj diferencirano obravnavo naravoslovja in humanistike (O. Boulnois, 2008).
- [11] O nekaterih pomanjkljivostih kvantitativnega ocenjevanja znanstvene in strokovne uspešnosti raziskovalcev v Sloveniji gl. v: F. Mali, J. Jug (2006).
- [12] Na to opozarja tudi Bourdieu, ko pravi, da je »treba poiskati *limite* metode, ki se opira na dosledno kvantitativne kriterije in ki ne pozna zelo različnih modalitet in funkcij reference (ki gre lahko tako daleč, da v oklepaj postavi razliko med pozitivnim in negativnim citiranjem)« (Bourdieu, 2004).
- [13] Poseben problem predstavlja razvrščanje revij v posamezne kategorije. Tudi to ni specifičen slovenski problem. V zadnjem času poteka ostra polemika o klasifikaciji humanističnih revij v Franciji (v kategorije A-B-C, z namenom ocenjevanja raziskovalcev, inštitutov in univerz), kjer se je zato AERES (Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur) odločila, da vsaj za enkrat ne bo klasificirala revij s področja francoske in primerjalne književnosti (Jourde, 2008, besedilo je pod drugim naslovom in brez referenc izšlo tudi v decembrski številki slovenske izdaje tega časopisa, str. 40). Toda problem s tem ni bil rešen. Prav nasprotno, Olivier Boulnois, specialist za filozofijo srednjega veka in dekan znamenite *l'Ecole pratique des hautes études*, je z analizo razvrstitve revij na področju srednjeveške filozofije pokazal na problematičnost take klasifikacije. Sprašuje se, zakaj je recimo temeljna revija na tem področju, *Archives d'histoire doctrinale et littéraire du Moyen Age*, v kateri je objava članka sen vsakega medievalista, uvrščena le v kategorijo B. Zakaj je njen italijanski ekvivalent, revija *Documenti e Studi sulla tradizione filosofica medievale*, uvrščena celo v kategorijo C? V isto kategorijo je ravno tako nerazumljivo uvrščena tudi izvrstna revija *Quaestio*. Prav tako se sprašuje, zakaj so glavne francoske filozofske revije (*Archives de Philosophie*, *Etudes philosophiques*, *Philosophie*, *Revue de Métaphysique et de Morale*) uvrščene samo v kategorijo B, medtem ko so vse njim ekvivalentne anglosaške revije (*Philosophical Quarterly*, *Philosophical Review*, *Philosophical Studies*) uvrščene v najvišjo kategorijo A. Poleg tega je ugotovil, da revija *Revue des sciences théologiques et philosophiques*, ki je bila uvrščena v kategorijo B, sploh ne obstaja, obstoječa revija z zelo podobnim naslovom *Revue des Sciences philosophiques et théologiques* pa je bila uvrščena v kategorijo A. Nič bolj razumljiva ni trikratna uvrstitev revije *Medieval Philosophy and Theology*, ki je znotraj filozofije uvrščena enkrat v kategorijo A, drugič v kategorijo B, tretjič pa znotraj zgodovine spet v A. Vprašanje je tudi, zakaj je revija *Revue de Philosophie Ancienne* znotraj filozofije enkrat uvrščena v kategorijo A, drugič pa v B. Naivnež bi lahko verjel, pravi Boulnois, da bi bil v njej objavljen filozofski članek vrednoten, kot da je bil objavljen v reviji, ki spada v kategorijo B, zgodovinski pa, kot da bi bil objavljen v reviji, ki spada v kategorijo A. Toda po katerem kriteriju naj se potem vrednotijo članki, ki prečijo obe vedi, ali tisti, ki spadajo v zgodovino idej (političnih, teoloških, moralnih itd.) in so najbolj novatorski v okviru srednjeveške filozofije, se sprašuje Boulnois, ki klasifikaciji tudi očita, da ne vključuje revij, kot je recimo poljska revija *Studia Mediewistyczne*,

v kateri je izšlo veliko ključnih člankov za to področje. Poleg tega je zanj metodološka napaka, če na podlagi kategorije, v katero je neka revija razvrščena, sklepamo o kakovosti članka. Kot argument v prid tej trditvi navaja, da je bilo npr. od vseh Foucaultovih člankov, zbranih v štirih knjigah z naslovom *Dits et Ecrits*, le kakih pet objavljenih v revijah, ki so razvrščene v kategorijo A (O. Boulnois, »Entre rires et larmes. L'évaluation automatisée en sciences humaines«).

- [14] E. Reale, A. B. Costantini, A. Costantini (2007: 220).
- [15] Analize ocen dveh ocenjevalcev, ki ocenjujeta iste raziskovalne dosežke, kažejo, da so razlike v ocenah včasih zelo velike. Tako se dogaja tudi, da en ocenjevalec oceni neki znanstveni dosežek ali raziskovalni program z najvišjo oceno, drugi pa z najnižjo. Sploh pa te analize, ki so bile opravljene v različnih državah, opozarjajo, da je koeficient korelacije med ocenami dveh ocenjevalcev dokaj nizek (cf. A. Barbara, A. Costantini, M. Costantini (2006: 8–9)).
- [16] Lep dokaz za to je npr. razvpita »afera Sokal«, ko je ugledna ameriška družboslovna revija *Social Texts* leta 1996 objavila članek ameriškega fizika Sokala z naslovom »Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity«, v katerem je bilo, ker je šlo za avtorjevo potegavščino, veliko navideznih referenc ali pa referenc, ki so bile na napačnih mestih (o drugih razsežnostih te afere gl.: B. Rotar (2004)).
- [17] P. Lawrence (2007: 583–585). To pa ima še druge negativne posledice. Znanstveniki zgubljajo ogromno časa in energije za razmišljanje o tem, kako napisati članek, ki se bo zdel urednikom uglednih revij zanimiv za objavo; izogibajo se pisanju o novih temah in izbirajo najpopularnejše teme, ker je večja verjetnost, da bodo objavljene; raziskovalnih izsledkov ne objavljajo v celoti, ampak jih razdelijo na čim več delov, da lahko objavijo več člankov (čeprav bralec zaradi tega ne vidi celote raziskovanega problema); ker je v najuglednejših revijah prostor za članek zelo omejen, prikazujejo svoje izsledke zelo zgoščeno, svoje ugotovitve poenostavljajo, ne citirajo izvirnih del, ampak recenzije teh del, ker tako prihranijo prostor; zanemarjajo ali skrivajo podatke, ki so v nasprotju s tistimi ugotovitvami, o katerih pišejo (da ne bi bil članek videti nekoherenten), ali pa so v nasprotju z izsledki uglednih kolegov, ki bi lahko bili recenzenti njihovega članka; povečuje se število avtorjev posameznih člankov, ker se s tem povečuje možnost, da pridejo do reference, itd. Tako poudarjanje bibliometričnih kazalcev pa poleg tega vodi tudi v vse večjo tekmovalnost med znanstveniki, ki izloča izvrstne, a manj bojevitve znanstvenike (večinoma ženske). Paradoks je tudi v tem, da bi nemara bili, če bi jih merili s sedanjimi merili, izločeni celo nekateri dobitniki Nobelove nagrade, recimo taki, kot je Ed Lewis, ki jo je dobil leta 1995 za medicino, saj je objavljajal malo in neredno (ibid., P. Lawrence, 2008/8: 2–3).
- [18] Cf. P. Bourdieu (2004: 89).

Literatura

- Barbara, A., Costantini, A., Costantini M. (2006). Metodologie di valutazione della ricerca pubblica in Europa: L'area delle scienze politiche e sociali nel processo del VTR Italiano ISIG (Trimestrale di sociologia Internazionale), *L'Europa dei sociologi Italiani*, I, 2006/4; 2007/1.
- Boulnois, O. (2008). *Entre rires et larmes. L'évaluation automatisée en sciences humaines, Evaluation de la recherche en SHS*, 22 oct. 2008, <http://evaluation.hypotheses.org>.
- Bourdieu, P. (2004). *Znanost o znanosti in refleksivnost*. Ljubljana: Liberalna akademija.
- Landel, G. (2004). Conclave in the Tower of Babel: How Peers Review Interdisciplinary Research, http://repp.anu.edu.au/papers/2004_TASAIinterdisc.pdf
- Langfeldt, L. (2004). Judging quality. Expert panels evaluationg research: decision-making and sources of bias, *Research Evaluation* 2004/1.
- Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessment, *Research evaluation* 2006/1.
- Lawrence, P. (2008). Lost in publication: how measurement harms science, *Ethics in Science and Environmental Politics* 2008/8.
- Lawrence, P. (2007). The Mismeasurement of Science, *Current Biology* 2007/15.
- Mali, F., Jug, J. (2006). Odprta vprašanja in dileme kvantitativnega in kvalitativnega ocenjevanja v znanosti, *Organizacija znanja* 2006/4.
- Reale, E., Costantini, A. B., A. Costantini (2007). Peer review for the evaluation of academic research: lessons from the Italian experience, *Research Evaluation* 2007/3.
- Research Assessment Exercise 2008*, <http://www.rae.ac.uk/pubs/2006/01/docs/kall.pdf>.
- Rotar, B. (2004). Kdo bo dajal naloge znanosti?, spremna beseda k: P. Bourdieu, *Znanost o znanosti in refleksivnost*.
- Sokal, A. (1996). Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity, *Social Texts*, 1996/46–47.
- Walton, D. (1985). *Agruer's Position*. Connecticut and London: Greenwood Press.
- Walton, D. (1989). Reasoned Use of Expertise in Argumentation, *Argumentation* 1989/3.
- Woods, J. and Walton, D. (1974). Argumentum ad verecundiam, *Philosophy and Rhetoric* 1974/7.
- Woods, J. and Walton, D. (1982). *Argument: The Logic of the Fallacies*. Toronto: McGraw-Hill.