

IRENEUSZ SADOWSKI, BOGDAN W. MACH*

Parametryzacja i kategoryzacja jednostek naukowych w roku 2013 jako praktyka ewaluacyjna i proces instytucjonalny – przypadek nauk humanistycznych i społecznych

1. Wprowadzenie

Filozofia i praktyka przyznawania kategorii (i związanej z nią dotacji) jednostkom naukowym budzi żywe kontrowersje w polskim środowisku naukowym. Najbardziej słyszalne głosy protestu wobec przyjętych rozwiązań płyną ze środowisk nauk społecznych i humanistycznych. Nie będziemy w tym artykule dyskutować podnoszonych przez te środowiska szczegółowych argumentów (to materia na odrębny tekst). Nie ma też chyba powodu, by zastanawiać się nad ogólnym problemem, czy w nauce powinna, czy też nie powinna, obowiązywać jakaś forma merytokracji – materialnego i symbolicznego, społecznie usankcjonowanego, nagradzania w proporcji do jakości i ilości osiągnięć naukowych¹. Naszym zasadniczym celem w tym tekście jest ukierunkowana na przyszłość interpretacja praktycznych rozwiązań instytucjonalnych i ewaluacyjnych, zastosowanych w parametrycznej ocenie i kategoryzacji jednostek naukowych w roku 2013.

Kompleksowa ocena działalności naukowej i badawczo-rozwojowej, której skutkiem jest kategoryzacja jednostek naukowych (JN) przez Ministerstwo Nauki i Szkolnictwa Wyższego (MNiSW), to przede wszystkim mechanizm służący dzieleniu części środków na naukę. Jest to zarazem oficjalny system atestacji i sygnalizacji znaczenia naukowego jednostek oraz system bodźców, który poprzez presję o charakterze finansowym i prestiżowym wskazuje preferowane przez państwo kierunki i strategie działania. Sposób prowadzenia państwowej polityki naukowej nie jest rzeczą jasną obojętną dla pozycji, jaką w nauce światowej mają dokonania rodzimych badaczy (zob. Allik 2013). Równocześnie ocena działalności naukowej i badawczo-rozwojowej to praktyka ewaluacyjna, a więc forma pomiaru, w toku której od cząstkowych wskaźników ilościowych przechodzimy do „zmiennej nominalnej”, przypisując instytucjom naukowym jedną z czterech

* Dr Ireneusz Sadowski, Instytut Studiów Politycznych PAN, e-mail: is@spoleczenstwo.pl; prof. dr hab. Bogdan W. Mach, Instytut Studiów Politycznych PAN

¹ Odrębną sprawą jest to, za pomocą jakiego mechanizmu (rynkowego, parametrycznego, eksperckiego) ma być generowany efekt merytokratycznego podziału środków (nagród).

wartości („kategorii”): A+, A, B albo C. Jako system sygnałów i bodźców proces oceny JN wymaga interpretacji w terminach jego potencjalnych skutków, natomiast jako praktyka ewaluacyjna wymaga on oceny pod kątem trafności i rzetelności pomiarowej. Oba te aspekty zamierzamy w tym opracowaniu podjąć, skupiając się na przebiegu i wynikach oceny w grupie nauk humanistycznych i społecznych (NHS).

Zakończoną niedawno ocenę JN charakteryzowało to, że był to proces bardziej złożony od ocen poprzednich, a w jego ramach zdefiniowano i obsadzono znacznie więcej ról. Udział licznego grona osób na różnych etapach rozłożył (choć, rzecz jasna, w nierównym stopniu) odpowiedzialność za ostateczne wyniki pomiędzy ustawodawcą, Ministerstwo Nauki i Szkolnictwa Wyższego (MNiSW), jednostki, które przygotowywały ankiety osiągnięć, Komitet Ewaluacji Jednostek Naukowych (KEJN), przewodniczących zespołów ewaluacyjnych, „szeregowych” ewaluatorów oraz OPI i ICI² (odpowiedzialne za teleinformatyczną stronę przedsięwzięcia). Jedni uczestnicy nie zawsze byli w stanie dostrzegać problemy, z którymi stykali się drudzy (zresztą przyjęte reguły celowo wykluczały możliwość centralnego sterowania procesem oceny), zróżnicowana była także dostępność informacji zwrotnej. Z tego powodu Jerzy Brzeziński apelował niedawno na łamach Forum Akademickiego, aby w imię przejrzystości reguł obowiązujących w polskiej nauce zabierać głos na temat przebiegu parametryzacji (zob. Brzeziński 2013). Niniejszy tekst stanowi jedną z odpowiedzi na ten apel.

2. Realne znaczenie kategorii naukowej

Dyskusję o kategoryzacji JN warto rozpocząć od określenia, jaka jest obecnie realna waga kategorii naukowej, którą otrzymuje JN. Najogólniej rzecz biorąc, w ramach określonej grupy jednostek środki publiczne przeznaczone na statutową działalność naukową JN dzielone są proporcjonalnie do liczebności kadry³, z uwzględnieniem współczynnika określającego kosztochłonność badań, rodzaju JN oraz współczynnika przypisanego poszczególnym kategoriom naukowym. Wartości tego ostatniego współczynnika dla poszczególnych kategorii wynoszą: 1,5 (A+), 1 (A), 0,75 (B), 0,4 (C) (zob. Załącznik nr 2 do Rozporządzenia w sprawie kryteriów..., Dz. U. z 2010 r., Nr 218, Poz. 1438).

Najłatwiej ocenić realne skutki kategoryzacji, jeżeli za podstawę odniesień weźmiemy najliczniejszą kategorię – B. Wówczas przyznanie kategorii A+ jest równoznaczne z podwojeniem otrzymywanej dotacji, natomiast przyznanie kategorii C – z analogiczną „karą”, czyli jej prawie dwukrotnym zmniejszeniem (pomijając znacznie poważniejszy

² Ośrodek Przetwarzania Informacji oraz Index Copernicus International – podmioty, które obsługiwały parametryzację od strony informatycznej.

³ Pomijamy tu inercyjny mechanizm przeniesienia, który uzależnia wysokość dotacji od jej wysokości w latach poprzednich.

aspekt – odebranie dotacji po 6 miesiącach⁴). Jednostki, którym przyznano kategorię A, uzyskują względem kategorii B znacznie niższą premię – w wysokości 33%. Oznacza to, że najbardziej poważny kontrast pod względem finansowym wiąże się ze znalezieniem się w kategoriach skrajnych: A+ i C, czym zdecydowana większość JN nie jest w ogóle zagrożona. Realnie rzecz biorąc, dla dużej części JN kategoryzacja jest jedynie grą o te dodatkowe 33%, ponieważ ani nie są zagrożone odebraniem dotacji, ani też nie mają szans na uzyskanie „plusa”.

Trzeba jednocześnie zwrócić uwagę, że ta 33-procentowa różnica jest odczuwalna najbardziej w przypadku instytutów PAN, w których wynagrodzenia są silnie uzależnione od dotacji na badania naukowe. Całkiem niedawno w artykule prasowym Jan Cieśliński (2014) słusznie zauważył, że z formalnego punktu widzenia pracownicy uczelni wyższych za swoją pracę naukową nie otrzymują wynagrodzenia. Zdecydowana większość środków, którymi rozporządzają szkoły wyższe, pochodzi bowiem z dotacji na zadania związane z kształceniem studentów. Nawet w przypadku uczelni największych, prowadzących zakrojone na szeroką skalę badania, takich jak Uniwersytet Warszawski, kwota dotacji MNiSW na działalność dydaktyczną znacząco przekracza łączną kwotę wszystkich dotacji MNiSW na badania naukowe (której część stanowi bazowa dotacja na działalność naukową). Dysproporcja między dotacją na dydaktykę a dotacją na badania naukowe jest w przypadku uczelni słabszych drastycznie większa. Dla nich środki na badania statutowe są nieznacznym dodatkiem do zasadniczej puli przekazywanej w ramach finansowania zadań dydaktycznych. Opierając się na sprawozdaniach rektorów z działania ich uczelni w roku 2012, łatwo wyliczyć, że Uniwersytet Warszawski otrzymał w przeliczeniu na jednego pracownika około 110 tys. zł z racji prowadzonej dydaktyki i nie więcej niż 84 tys. z racji prowadzonych badań, w przypadku Uniwersytetu w Białymstoku te kwoty wynosiły odpowiednio: 83 i 11 tys. zł, a w przypadku Uniwersytetu Kardynała Stefana Wyszyńskiego: 96 i 6 tys. zł.

Taki stan prawny rodzi pytanie o to, jak MNiSW definiuje zadania pracownika naukowo-dydaktycznego. Ile pracy powinien przeznaczać on na zadania związane z kształceniem (przygotowanie i prowadzenie zajęć, konsultacje itp.), a ile na zadania związane z nauką (prowadzenie badań, praca nad publikacjami itp.)? Kierując się przelicznikami z rozporządzenia o dotacji na badania statutowe, można by stwierdzić, że badania powinny konsumować mu około 40% czasu pracy⁵. Przez porównanie pensji pracownika PAN i pracownika uniwersytetu moglibyśmy twierdzić, że ekwiwalentna praca naukowa powinna temu ostatniemu zabierać około 2/3 czasu. Kierując się wielkością pensum

⁴ Zob. Ustawa o zasadach finansowania..., Dz. U. z 2010 r., Nr 96, Poz. 615.

⁵ Wielkość dotacji na badania statutowe jest w przypadku uczelni ustalana w oparciu o współczynnik równy 0,4, a w przypadku instytutów PAN – równy 1,0.

określonego w ustawie Prawo o szkolnictwie wyższym (Dz. U. z 2005 r., Nr 164, poz. 1365), jak to uczynili Antonowicz i Brzeziński (2013), moglibyśmy określić tę proporcję nawet na poziomie 80%. Trzymając się definicji określonej za pośrednictwem mechanizmu dzielenia środków, można by twierdzić, że w praktyce udział pracy badawczej waha się w granicach od 5% do najwyżej 20%⁶. Jak widać, sytuacja jest dosyć niejednoznaczna, a norma obyczajowa zdaje się znacząco odbiegać od normy prawnej.

Czy można powiedzieć, że MNiSW faktycznie zakłada, że w przeciętnym pracowniku uczelni jest tak mało naukowca? Otóż nie. Jak słusznie zauważył Ciesliński, dotacja dydaktyczna jest w istocie rodzajem dotacji naukowo-dydaktycznej. Środki przekazane w jej ramach mają w części służyć rozwojowi warsztatu naukowego, prowadzeniu badań, a także dydaktyce, opierającej się na zdobytym w ten sposób doświadczeniu. Komponenty kadrowy i badawczy mają w ramach środków na kształcenie łączną wagę równą 0,4⁷. Jeśli więc dla uproszczenia przyjmiemy, jak wyżej, że od pracownika uniwersytetu oczekuje się poświęcenia przeciętnie około 40% czasu na prowadzenie badań (w zależności od momentu rozwoju kariery, a więc stopnia ogólnego merytorycznego przygotowania), sytuację będzie można interpretować bardziej klarownie. Po pierwsze, proporcja 40% wydaje się wciąż odbiegać od środowiskowej normy. Po drugie, rzuca ona nowe światło na ogólne znaczenie obecnej parametryzacji i kategoryzacji.

Jeśli uznamy, że przynajmniej 40% dotacji na dydaktykę dotyczy w istocie finansowania badań prowadzonych przez pracowników naukowo-dydaktycznych, to trzeba jednocześnie zauważyć, że środki te nie są dzielone w sposób merytokratyczny, lecz w sposób – nazwijmy to umownie – egalitarny. Uwzględniony jest oczywiście przelicznik dotyczący posiadanego stopnia lub tytułu, ale nie są uwzględnione realne, bieżące efekty pracy naukowej. Gdyby składnik kadrowy dotacji na kształcenie potraktować jako uwzględniający jakąś formę założeń na temat wyników pracy, to założenia te określałyby *implicite* etatową pracę naukową profesora jako wartą 1,7 etatowej pracy naukowej doktora i 2,5 etatowej pracy naukowej magistra, ale z kolei jako dwa razy mniej cenną niż praca profesora zagranicznego. Dla tych ostatnich przypisano bowiem wagę 5,0 (co powoduje np., że uniwersytety z Polski wschodniej powinny zacząć masowo zatrudniać profesorów z Białorusi i Ukrainy – będzie to z pewnością opłacalne dla obu stron).

Realnie merytokratyczne zasady dystrybucji środków dotyczą jedynie dotacji *stricte* naukowych, z których większość rozdzielana jest w ramach konkursów prowadzonych przez Narodowe Centrum Badań i Rozwoju oraz Narodowe Centrum Nauki. Podmio-

⁶ Może on wynosić więcej w przypadku tych, którzy uzyskują środki w postępowaniach konkursowych, ale wtedy dodatkowe wynagrodzenia przewidziane są w granicy.

⁷ Gdyby szukać analogii z sytuacją instytutów badawczych, wypadałoby doliczyć także komponenty związane z prowadzeniem doktoratów i z posiadanymi uprawnieniami.

towa dotacja dla jednostek naukowych wynosi obecnie około 1 mld zł. Uzyskana przez JN kategoria ma znaczenie jedynie przy podziale owej puli.

Reasumując – wydatki budżetu państwa na działalność dydaktyczną w uczelniach publicznych miały w roku 2013 roku wynieść ponad 7 mld zł (plan finansowy), zaś wydatki na naukę w postaci dotowania działalności statutowej – około 1 mld zł. Jeśli 40% z pierwszej sumy rozdziela się w sposób quasi-egalitarny, a drugą w oparciu o ocenę parametryczną, to można skonkludować, że niewiele ponad 1/4 środków przeznaczanych w trybie pozakonkursowym na bazowe, statutowe badania naukowe prowadzone przez JN dzielona jest w oparciu o kategoryzację. Jeśli uwzględnimy jeszcze konkursy NCBR i NCN, wówczas oddziaływanie oceny parametrycznej będzie dotyczyło zaledwie jednej na każde sześć złotych realnie wydawanych przez MNiSW na naukę. Kategoryzacja ma więc dziś głównie znaczenie prestiżowe, ale być może nawet pod tym względem jest ono mniejsze niż w przypadku uprawnień do nadawania stopni. Zważywszy na to, że ustawa przewiduje poszerzenie strumienia środków rozdzielanego przez NCN i NCBR, m.in. kosztem działalności statutowej⁸, znaczenie kategorii – przy zachowaniu obecnych reguł – będzie jeszcze małe.

W oparciu o to spostrzeżenie można wysnuć wnioski pesymistyczne lub postulatycznie pozytywne. Ograniczając się do tych drugich, należy stwierdzić, że chcąc zadbać o merytokratyczne dzielenie środków na naukę, MNiSW powinno wprowadzić do podziału dotacji dydaktycznej komponent oparty na przyznanej kategorii bądź też przekazać część tych środków na dotacje badawcze *par excellence* (badania statutowe lub postępowania konkursowe). Warto jednocześnie zwrócić uwagę na fakt, że jeśli interpretujemy dotację dydaktyczną jako w istocie naukowo-dydaktyczną, to wówczas w bilansie dwóch strumieni finansowania (na badania statutowe, na szkolnictwo wyższe) stroną realnie dyskryminowaną są instytuty badawcze, przede wszystkim instytuty PAN (otrzymują one więcej z tego pierwszego, ale z drugiego, który jest źródłem znacznie bardziej obfitym, nie otrzymują nic)⁹.

Istnieje jeszcze jedna implikacja ewentualnej zmiany mechanizmów dystrybucji środków na naukę. Zmiana taka pomogłaby na nowo zdefiniować istotę uniwersytetu. Przynajmniej od ubiegłego roku toczy się na ten temat w Polsce dyskusja, u której pod-

⁸ Środki finansowe przeznaczone corocznie przez Ustawę o zasadach finansowania nauki (Dz. U. z 2010 r., Nr 96, poz. 615, z późniejszymi zmianami) na finansowanie zadań NCBR i NCN mają do roku 2020 osiągnąć wartość nie mniejszą niż 50% środków finansowych na naukę z puli, na którą składają się przede wszystkim: działalność statutowa, działalność upowszechniająca naukę, finansowanie KEJN i KPN, nagrody i stypendia.

⁹ Rozwiązałoby to również sygnalizowany przez Antonowicza i Brzezińskiego problem porównywalności wyników między jednostkami uczelnianymi a instytutami badawczymi bez konieczności stosowania arbitralnych wag (Antonowicz, Brzeziński 2013: 57-58).

staw legła spowodowana niżem demograficznym finansowa presja na likwidację kierunków „nierentownych”. Jeśli mimo zmniejszającej się liczby studentów państwo zechciałoby utrzymać (albo zwiększyć) poziom finansowania uniwersytetów, powinno wyżej ocenić wagę ich funkcji naukowej (pozwalając m.in. na tworzenie etatów wyłącznie naukowych), a niżej wagę ich funkcji edukacyjnej. Wydaje się, że można tego dokonać jedynie przez modyfikację sposobu dzielenia środków na cele związane z realizacją tych funkcji. Gdyby się tak stało, byłby to wyraźny sygnał, że byt uczelni zależy nie tylko od skuteczności naboru, ale również od jakości prowadzonych badań. Uniwersytet – mówiąc metaforycznie – powróciłby wówczas spod rządów studenta pod rządy korporacji naukowej.

3. Ocena jednostek naukowych jako element polityki publicznej

Wiele wskazuje na to, że dzielenie środków w oparciu o pewną formę parametryzacji stanie się trwałym elementem polskiego systemu instytucjonalnej nauki. Podobne praktyki są od dawna wykorzystywane w krajach najwyżej rozwiniętych, choć również tam potrafią budzić kontrowersje¹⁰. Nie można wykluczyć, że w przyszłości dojdzie do modyfikacji zasad przyznawania kategorii naukowych – jednak zarówno z punktu widzenia państwa, jak i znacznej części środowiska naukowego, parametryzacja i kategoryzacja są rozwiązaniami pożądanymi. Trzeba przy tym wyraźnie zaznaczyć, że podział środków zgodnie z arbitralnie wyznaczonymi parametrami należy do porządku działań administracyjnych, a nie naukowych, stąd też zazwyczaj zawierać będzie wiele założeń merytorycznie wątpliwych (z punktu widzenia sposobów uprawiania nauki), lecz motywowanych ogólną polityką naukową państwa. W sytuacji gdy interes państwa i interes środowiska naukowego są zbieżne, pewna doza arbitralności nie stanowi problemu. Inaczej dzieje się, gdy dochodzi do rozbieżności pomiędzy tymi interesami.

3.1. Ewaluacja a polityka państwa w obszarze nauki

Cele państwa w obszarze polityki naukowej możemy wyczytać m.in. z rządowego dokumentu „Strategia rozwoju nauki do 2015 roku” (2008). Są nimi – parafrazując ów dokument – poprawienie pozycji polskiej nauki w kontekście porównań międzynarodowych, mocniejsze powiązanie z instytucjonalną nauką w Europie, silniejsze powiązanie rezultatów badań z edukacją, kulturą, gospodarką i innymi obszarami życia społecznego, a zwłaszcza poprawa zdolności do komercjalizacji wyników. Warto zauważyć, że niektóre z tych priorytetów są niezgodne z interesami niektórych dyscyplin. Z punktu widzenia wielu (większości?) humanistów komercjalizacja czy poprawa pozycji międzynarodowej

¹⁰ Jak pisali Dag Aksnes i Randi Elisabeth Tøxt (2004) w Norwegii zarzucano tym praktykom, że tworzą fałszywy obraz efektywności naukowej, oparte są na błędnych założeniach i wypaczonych danych, a także, że ich wyniki odpowiadają obiegowym opiniom, a zatem nie warto przeznaczać na nie środków publicznych.

mogą być traktowane jako niepożądane lub nieosiągalne, zwłaszcza jeśli humanistykę określi się jako dziedzinę, której wytwory skierowane są do rodzimego odbiorcy i zorientowane na sferę niematerialną (nie tylko w sensie ontologicznym, ale także finansowym). Nic więc dziwnego, że najbardziej zagorzałymi krytykami reform stali się przedstawiciele tej grupy dyscyplin, wspierani siłami dziedziny artystycznej¹¹.

Jak wspomnieliśmy we wstępie – każdy system parametrycznej oceny można interpretować jako system zachęt, bodźców służących ukierunkowaniu aktywności. W kontekście możliwej rozbieżności priorytetów jest to zatem jednocześnie sposób na „przymuszenie” środowiska naukowego do silniejszego podążania w kierunku wskazanym przez Ministerstwo. W obecnej parametryzacji ten kierunek został określony jako względnie jednolity dla wszystkich grup nauk (GN). Świadczy o tym fakt, że we wszystkich czterech GN przyjęto z grubsza takie same kryteria oceny efektów pracy naukowej: kryterium bibliometryczne (obejmujące element patentowy), kryterium potencjału naukowego (głównie kadrowego), kryterium efektów finansowych oraz kryterium „niematerialnych” osiągnięć o znaczeniu ogólnospołecznym. Opór przedstawicieli nauk humanistycznych i społecznych (NHS) budzi przede wszystkim kryterium I oraz kryterium III, chociaż temu ostatniemu nadano wagę równą zaledwie 0,05. Ważniejszy zarzut dotyczy tego, że sposób oceny sprawia wrażenie zorientowanego na ilość, a nie jakość osiągnięć badawczych. I jest tak niewątpliwie, ponieważ pierwsze kryterium, będące głównym komponentem oceny (jego waga w NHS wynosiła 0,65), opierało się na zliczaniu określonych rodzajów publikacji, bez udziału oceny eksperckiej. O ile w przypadku artykułów przybliżoną miarą wartości jest liczba punktów przyznanych czasopismu na podstawie oszacowania jego wpływu (*impact factor*), o tyle w przypadku monografii nie przewidziano merytorycznego różnicowania wartości. Stanowi to istotne obciążenie wyników zakończonej niedawno ewaluacji.

Standaryzacja kryteriów jest jednak pożądaną odpowiedzią ze względu na inną ważną kwestię – interdyscyplinarność jednostek naukowych oraz prowadzonych badań. Do oceny w ramach NHS zgłaszały się przecież jednostki, w których pracowali naukowo przedstawiciele innych dziedzin naukowych, a stopień wewnętrznego instytucjonalnego wymieszania różnych dyscyplin w NHS jest bardzo duży (z tendencją rosnącą). Zmusza to do ujednolicenia (przynajmniej w pewnym rozsądnym zakresie) zasad w obrębie GN,

¹¹ Można mieć wątpliwości, czy jednostki z dziedziny nauk o sztuce i twórczości artystycznej powinny być objęte tym samym systemem ewaluacji, co pozostałe instytucje naukowe. Antonowicz i Brzeziński (2013) sugerują nawet przeniesienie ich oceny pod kuratelę Ministerstwa Kultury i Dziedzictwa Narodowego. Ma to sens o tyle, że ocenie podlegają tu osiągnięcia artystyczne, a co ważniejsze – priorytety polityki kulturalnej mogą być inne niż polityki badawczo-rozwojowej. Problematyka ewaluacji jednostek z dziedziny nauk o sztuce i twórczości artystycznej wymaga odrębnego rozważenia. Nie zajmujemy się nią w tym tekście.

ale także pomiędzy GN. Alternatywne podejście, to jest wyznaczenie kompletnie odmiennych reguł dla poszczególnych GN, stworzyłoby więcej problemów, niż by rozwiązało. W takiej sytuacji trzeba by bowiem instytucjonalnie zadbać o hermetyczność pól badawczych i przedstawić akceptowalne przez środowiska uzasadnienia tej hermetyczności. Obecny system proponuje rozwiązanie pośrednie – ogólne kryteria oceny są wspólne, ale ich wagi i szczegółowa zawartość różnią się pomiędzy GN. To zróżnicowanie może być w przyszłej parametryzacji większe niż obecnie – w niektórych obszarach, jak sądzimy, powinno być większe.

Generalnie można przypuszczać, że istnieje przynajmniej kilka przesłanek, które przemawiają na rzecz trwałości wielu elementów obecnego systemu uniwersalnej i silnie zestandaryzowanej oceny. Podstawowym może być dążenie Ministerstwa do skutecznej realizacji priorytetów polityk publicznych i związana z tym obywatelska presja na to, by alokacja środków publicznych była transparentna, a odpowiedzialność za nią opierała się na zdolności do ich precyzyjnego rozliczenia. Wydaje się, że jest to współcześnie ważniejsza składowa społecznej legitymizacji, niż zapewnienie daleko idącej autonomii nawet wpływowym środowiskom zawodowym. Na rzecz złagodzenia oporu środowisk naukowych działa przy tym to, że zostało ono znacząco włączone w ustalanie uniwersalnych i zestandaryzowanych reguł gry. Konflikt o sposób prowadzenia oceny stał się więc w dużej mierze wewnętrzną sprawą naukowców. Nie zmienia to oczywiście faktu, że ostateczne decyzje co do kształtu oceny zostały podjęte przez Ministerstwo i w sensie prawnym Ministerstwo za nie odpowiada.

Warto pamiętać, że wśród niespełna 1000 jednostek naukowych poddawanych ocenie nauki humanistyczne i społeczne stanowią mniej więcej 30%, a nauki o sztuce i twórczości artystycznej ok. 10%. W tej sytuacji nauki o życiu oraz ścisłe i inżynierskie stają się i ze względu na swą liczebność, i ze względu na jasno formułowane priorytety rządowe „praktycznie stanowiącą większość”, a właściwe dla nich ilościowe i zestandaryzowane sposoby oceny mają największe szanse na stabilizację i powielanie się w przyszłości. Sposób oceny przyjęty w takim kontekście może się okazać korzystny z punktu widzenia części nauk społecznych, ale z pewnością może być traktowany jako „gwałt” przez większość humanistyki oraz nauk o sztuce i twórczości artystycznej.

Należy także zauważyć, że najsilniejsze jednostki naukowe, których status został wzmocniony oceną parametryczną, będą dążyły do zachowania istniejącego sposobu dzielenia środków. Liczba i słabość głównych „przeigranych” parametryzacji, czyli tych JN, które znalazły się w kategorii C, nie będzie zapewne w tej sytuacji przekładać się na powstawanie znaczących grup wpływu na rzecz radykalnej zmiany sposobu oceny JN. Również tylko niektóre z jednostek w kategorii B mogą być z istniejącego systemu niezadowolone na tyle, by aktywnie przeciw niemu protestować. Argumentem empirycznym jest też w pewnym sensie to, że modyfikację zasad parametryzacji w stosunku do

poprzednich jej odsłon, niezależnie od oczywistych różnic w „technologicznym oprzyrządowaniu”, trudno nazwać rewolucyjną. Można więc wnosić, że istnieje pewien „polityczny” konsens co do zasadniczych podstaw dzielenia środków na naukę. Liczba kategorii, wagi, szczegółowe kryteria – są to elementy wtórne wobec przekonania o dających się sparametryzować wynikach pracy naukowców i znaczącej roli bibliometrii.

Kolejny, powiązany argument dotyczy ustalania reguł gry w jej trakcie. Ostateczne zasady obecnej parametryzacji poznaliśmy niedługo przed jej zakończeniem, co jest sytuacją dalece niewłaściwą. By jej uniknąć w przyszłości, większości zainteresowanych stron może zależeć na utrzymaniu zasad wypracowanych w obecnej edycji. Przewidywalność zasad jest pożądana nawet z punktu widzenia słabszych jednostek, ponieważ zwiększa ich możliwości strategicznego planowania i budowania planów naprawczych.

Wydaje się w końcu, że nie ma sprzeczności pomiędzy pragmatyczną akceptacją reguł zaproponowanych przez doradczy wobec MNiSW Komitet Ewaluacji Jednostek Naukowych a wysokimi standardami pracy naukowej, tak jak widzi je samo środowisko naukowe – również w obrębie NHS. Przystosowanie się do narzuconej odgórnie logiki parametrycznej wymaga bowiem przede wszystkim zmian dotyczących sposobu publikacji wyników. Przykładowo, autor pracujący wiele lat nad swoim *opus magnum*, „zmuśzony” jest obecnie do tego, by co pewien czas przedstawiać środowisku swoje wyniki cząstkowe w postaci artykułów w dobrych czasopismach. Taki poziom „opresji” można prawdopodobnie uznać zarówno za znośny dla autora, jak i korzystny dla społeczności naukowej. Podobna implikacja praktyczna dotyczy promowania modelu publikacji wieloautorskich, czyli po prostu pracy zespołowej niesłusznie niedocenianej w pewnych przynajmniej obszarach NHS. Chodzi, na przykład o to, że współautorstwo tekstu wymusza rodzaj wewnątrzzespołowej recenzji – poszczególni autorzy stają się dla siebie nawzajem pierwszym źródłem oceny i krytyki, co podnosić może poziom pracy naukowej.

Powyższe argumenty mogą przemawiać za tym, że środowisko naukowe i państwowy mecenat skłonni będą do podtrzymania istniejących zasad. Nie zmienia to faktu, że warto podjąć dyskusję nad słabościami obecnej oceny – zwłaszcza tymi, które względnie łatwo usunąć. Na podstawie zebranych w aktualnej edycji doświadczeń chcielibyśmy przedyskutować jej reguły i przebieg.

3.2. Od projektu do realizacji

Wypada zacząć od kluczowego nowego elementu w systemie oceny jednostek, czyli od wspomnianego już Komitetu Ewaluacji Jednostek Naukowych (KEJN). Komitet jest ciałem doradczym MNiSW i składa się z czterech komisji do spraw czterech grup nauk: (a) humanistycznych i społecznych, (b) o sztuce i twórczości artystycznej, (c) ścisłych i inżynierskich oraz (d) o życiu. KEJN został powołany przez ministra nauki i szkolnictwa wyższego spośród kandydatów zgłoszonych przez środowiska naukowe. Jego podsta-

wowym zadaniem było przygotowanie zasad parametryzacji i kategoryzacji JN, przy czym liczba kategorii oraz obowiązek przeprowadzenia kategoryzacji odrębnie dla uczelni, PAN i instytutów badawczych stanowiły dla KEJN ustalone przez ustawodawcę *datum*. W praktyce prace KEJN skoncentrowane zostały na stworzeniu projektu Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym. Projekt ten, w drodze ustawowej procedury, stał się podstawą odpowiedniego Rozporządzenia MNiSW. W pracach nad projektem komisje KEJN były całkowicie autonomiczne, ale ich decyzje dotyczące projektu wymagały ostatecznej aprobaty przez KEJN w pełnym składzie.

W pełnym składzie KEJN zdecydował, że (a) parametryzacja wykorzystywać będzie porównywanie parami w grupach JN nazwanych Grupami Wspólnej Oceny (GWO) oraz że (b) porównania dotyczyć będą wyników osiąganych przez JN na czterech znanych z Rozporządzenia kryteriach. Ilość i składy GWO, szczegółowa kompozycja kryteriów, ich wkład (wagę) w ostateczny wynik porównań między jednostkami oraz ustalenie wartości kryteriów dla „jednostki referencyjnej” decydującej o kategoriach A, B, C pozostawiono komisjom, ale ich decyzje musiały zostać zaakceptowane przez KEJN w pełnym składzie. Podobnie postąpiono z kategorią A+, którą proponowały komisje, uwzględniając odpowiedni paragraf Rozporządzenia, ale zatwierdzało plenarne posiedzenie KEJN. Podkreślić jeszcze raz należy, że wszystkie działania KEJN, niezależnie od ich formalizacji, miały charakter doradczy – finalne decyzje podejmowało ministerstwo i to ono jest w sensie prawnym „właścicielem” tych decyzji.

Sformułowany w Rozporządzeniu system oceny oczekiwał na zasilenie danymi. Rolę ich dostarczyciela, jak się wydaje zupełnie słusznie, powierzono samym ocenianym. To na nich spoczął obowiązek przedstawienia stanu faktycznego. Rolą KEJN i powołanych przez Komitet zespołów ewaluacyjnych stała się więc przede wszystkim weryfikacja danych. Jest to fakt istotny z punktu widzenia etapu odwoławczego.

Informacje o efektach własnej działalności naukowej JN przedstawiały w postaci składającej się z 20 punktów i wielu podpunktów ankiety. Liczba zebranych w ten sposób danych wyraźnie przewyższała ilość informacji stanowiących podstawę oceny. Prowadziło to do pewnych nieporozumień – przykładowo niektóre JN odwoływały się od faktu, że w parametryzacji nie uwzględniono niektórych ich dokonań. Nie jest to zresztą jedyne świadectwo, że dokładne zasady ewaluacji nie stały się całkowicie jasne dla wszystkich jej uczestników. Niekiedy owo „niezrozumienie” zasad miało zapewne charakter intencjonalny i pragmatyczny, jak przy zgłaszaniu publikacji, które formalnie nie kwalifikowały się do oceny (albo wątpliwe było, czy powinny zostać zgłoszone).

Zasadę „sprawozdawczej maksymalizacji”, czyli domniemania naukowej wartości wszystkich efektów pracy, stosowały bardzo liczne JN, nawet te, które otrzymały finalnie kategorię A+. Obecna formuła oceny w większości przypadków nie przewiduje

bowiem negatywnych konsekwencji dla jednostek przedkładających nierzetelne dane. W publicznej dyskusji wielokrotnie podkreślano, że podstawowym środkiem kontroli procesu sprawozdawczego (i ewaluacyjnego) powinna być jawność ankiet lub kart oceny. W szczególności wydaje się całkowicie rozsądne, by przynajmniej dane o zgłoszonych publikacjach (kryterium I) i głównych osiągnięciach (kryterium IV) mogły być zupełnie jawne (takie są przecież z samej swej istoty, JN nie robi nic więcej, niż przedstawia je w sposób syntetyczny). To samo powinno dotyczyć przynajmniej liczby N, jeśli nie imiennej listy pracowników etatowych. Wzmocniłoby to proces wyłapywania błędów o element kontroli wzajemnej. Być może decydenci uznali, iż bezpieczniej jest uniknąć otwierania tej puszki Pandory – wolny dostęp do sprawozdań to zaproszenie do potencjalnej, ogólnopolskiej awantury (zwłaszcza że, podobnie jak w innych środowiskach naukowych, również w Polsce istnieją utrwalone międzyinstytucjonalne i personalne idiosynkrazje). Należałoby jednak ministerstwo zachęcić do dogłębnego rozważenia, czy priorytetem uczynić większą przejrzystość podziału środków, czy też „ogólną” dobrą atmosferę.

Prawdopodobnie najpoważniejszym zagrożeniem dla trafności oceny jest obecnie możliwość „handlu” publikacjami, czyli przekazywania praw do ich afiliowania. Do tego zagadnienia nawiązemy jeszcze w kolejnej części artykułu, ale już teraz warto podkreślić, że ta luka może być wykorzystywana na dwa sposoby. Pierwszy dotyczy książek i artykułów będących dziełem „wolnych strzelców”. Drugi dotyczy na przykład możliwości współpracy silnych jednostek naukowych, mających „zapas” publikacji, ze słabszymi od nich uczelniami, które zatrudniają kadre tych pierwszych w charakterze wykładowców. By ograniczyć możliwości takich manipulacji, MNiSW powinno bezwzględnie wymagać pisemnych deklaracji autora na etapie składania ankiety bądź też zezwolić na afiliowanie publikacji jedynie u swojego podstawowego pracodawcy.

Pewne słabości procesu oceny wynikały także z tempa wdrożenia reformy. Ośrodek Przetwarzania Informacji (OPI), odpowiedzialny za informatyczną stronę składania ankiet, poprawiał system już w trakcie jego działania. Pierwszy problem pojawił się, gdy udostępniono możliwość wprowadzania danych przez JN (i niektóre natychmiast zaczęły z tego korzystać), natomiast w formularzu zabrakło jednego pola (dotyczyło objętości w arkuszach wydawniczych). Mimo że błąd ten szybko skorygowano, w bieżącej ocenie brak informacji o liczbie arkuszy nie mógł być podstawą do odrzucenia publikacji. W praktyce ewaluatorzy musieli więc tę objętość sprawdzać sami.

Są w systemie informatycznym również takie elementy, za których słabość odpowiada Index Copernicus International (ICI). Chodzi przede wszystkim o funkcjonalność, która miała umożliwić wyłapywanie duplikatów (w tej samej ankiecie lub w dwóch różnych). Automatyczne wykrywanie duplikatów to narzędzie niezbędne, ponieważ trudno byłoby manualnie odszukać duplikaty w ankietach JN, liczących niekiedy po kilkaset pozycji. Problem polegał jednak na tym, że algorytm wyszukiwania zastosowany przez

ICI działał na zasadzie porównywania dokładnego łańcucha znaków. Oznacza to, że pojedyncza literówka (a nawet dodatkowa spacja!) sprawiała, że duplikat publikacji stawał się dla systemu niewidoczny. Tę wadę należy usunąć, stosując bardziej zaawansowany algorytm porównywania tytułów (choćby taki, jakie stosuje się przy wykrywaniu plagiatów).

Trzeba także podnieść i kwestię bardzo życzliwego stosunku MNiSW i KEJN do zawartości ankiet nadesłanych przez JN podlegające ocenie. Stosowano, i słusznie, zasadę *in dubio pro reo*. Jednak zbyt daleko posunięta wyrozumiałość w weryfikacji nadesłanych danych staje się niesprawiedliwa wobec jednostek solidnych, to znaczy tych, które po prostu dobrze przygotowały swoje sprawozdania. Tam, gdzie jednostki nie dopełniły obowiązku dostarczenia rzetelnych danych, które były od nich wymagane na etapie składania ankiety, one same ponoszą odpowiedzialność za niższą ocenę. Zgodnie z Rozporządzeniem ewaluacja przebiegała „na podstawie złożonej ankiety”. Żądanie rzetelności sprawozdawczej wymaga oczywiście, by także MNiSW, KEJN i współpracujące z nimi firmy informatyczne dopięły następną edycję oceny „na ostatni guzik”.

4. Ocena jednostek naukowych jako praktyka ewaluacyjna (pomiarowa)

Jedna z najważniejszych dyskusji na temat zasad parametryzacji dotyczyła tego, czy działalność naukową oceniać bardziej „ilościowo” (system bibliometryczny) czy bardziej „jakościowo” (system ekspercki). Zarówno samo stanowisko wypracowane w KEJN, jak i zabierane głosy (Antonowicz, Brzeziński 2013, Brzeziński 2013, Lewicka 2013) sugerują, że można mówić o konsensie wokół systemu mieszanego. W generalnej dyskusji podkreślane są wady zarówno czystego systemu eksperckiego, jak i systemu ściśle bibliometrycznego, więc rozwiązanie łączące te dwa mechanizmy ma na swój sposób zabezpieczać przed wyraźnymi wypaczeniami. Wypada jednak zauważyć, że zastosowane wagi dają prymat elementom ilościowym, ograniczając element *peer review* do 15% (IV kryterium w przypadku NHS i NSiTA). Jak twierdzi Maria Lewicka: „System ekspercki (...), choć pozwala na oceny jakościowe, obciążony jest grzechem subiektywizmu – jest nieodporny na świadome manipulacje i nieświadome zniekształcenia. Młode demokracje, w tym Polska, są na te zniekształcenia wyjątkowo podatne. System ekspercki jest też nieporównanie bardziej pracochłonny, absorbuje ogromną ilość czasu znacznej liczby tych pracowników, którzy tworzą eksperckie ciała oceniające produkcję kolegów” (2013: 2). Czy można zatem uznać, że obecnie przyjęta proporcja (85/15) między podejściami jest właściwa? I czy generalne kryteria oceny trafnie i rzetelnie opisują naukowe efekty pracy poszczególnych JN? Poniżej postaramy się przedyskutować te kwestie.

4.1. Bibliometria czy ocena ekspercka?

Na świecie zebrano już wiele doświadczeń w kwestii walorów i słabych stron obu rozwiązań. Badania empiryczne pokazują generalnie, że oceny eksperckie są w więk-

szym bądź mniejszym stopniu wypaczone – dzieje się tak w różnych państwach i nie należy oczekiwać, że w Polsce mogłoby być inaczej. Stephen Cole, Jonathan Cole i Gary Simon (1981) na podstawie badań dotyczących oceny projektów grantowych składanych do National Science Foundation ustalili, że metoda ta redukowała losowość decyzji o przyznaniu grantu jedynie o połowę. Poddano w nim analizie oceny eksperckie w ramach starannie zaplanowanego eksperymentu, a wnioski opierały się na dekompozycji wariacji. Dowodziła ona, że zróżnicowanie ocen tego samego projektu przez różnych ekspertów było bardzo duże (i to niezależnie od dziedziny naukowej). Szansa uzyskania grantu w co najmniej tym samym stopniu zależała od jakości wniosku, jak od doboru recenzentów.

Z drugiej strony także wskaźniki oparte na cytowaniach podatne są na wpływ czynników innych niż sama wartość opublikowanych wyników. Częstość odwołań zależy m.in. od środowiskowej pozycji autorów, ich liczby i afiliacji, formy publikacji, dlatego nie zawsze stanowi precyzyjną miarę jakości i wagi osiągnięcia naukowego. Badanie dotyczące cytawalności publikacji biomedycznych przeprowadzone przez badaczy związanych z brytyjskim Wellcome Trust (Allen, Jones, Dolby, Lynn, Walport 2009) pokazało na przykład, że potencjalny wkład wyników do rozwoju wiedzy, oceniany w momencie ukazania się artykułu, pozwalał tylko częściowo przewidzieć, jak popularny stawał się on w kolejnych latach (korelacja na poziomie 0,45). Karol Życzkowski i Jakub Zakrzewski, komentując bibliometryczne wskaźniki dotyczące ich własnych publikacji w kontekście rzeczywistej wartości naukowej tekstów, stwierdzili, że „na wszelkie dane bibliometryczne należy patrzeć z wielką dozą krytycyzmu”, a wnioskowanie w oparciu o nie „powinno być prowadzone jedynie w jedną stronę: mała liczba publikacji i cytowań prac danego badacza sugeruje, choć nie implikuje, że jego dorobek nie jest znaczący” (Życzkowski, Zakrzewski 2012: 3).

Tam, gdzie bibliometryczna i ekspercka ocena jednostek naukowych stosowane były obok siebie, stwierdzano rzecz jasna dodatnią korelację między ich wynikami, ale często o stosunkowo niskiej sile. Analiza wyników przeprowadzonej w Norwegii ewaluacji jednostek naukowych pokazała, że związek ten wynosił 0,34, co sugeruje pewien problem z ogólną rzetelnością (Aksnes, Taxt 2004). Co istotne – wskaźnik bibliometryczny dobrze przewidywał ocenę ekspercką wśród jednostek, które afiliowały wiele tekstów w prestiżowych czasopismach, ale już znacznie gorzej wśród jednostek pod tym względem słabszych. Taka sytuacja stwarza miejsce na dwie interpretacje – albo wskaźniki publikacyjne są ślepe na pewien typ wartościowej pracy naukowej, albo też oceny eksperckie ulegają w znacznej mierze wpływowi czynników pozamerytorycznych (albo jedno i drugie). Autorzy wspomnianego studium konkludowali, że rekomendowane jest branie pod uwagę obu kryteriów, a sytuacje dużej rozbieżności powinny zostać poddane dodatkowemu sprawdzeniu. Podobne sugestie odnajdziemy w tekście Brytyjczyka,

Henka Moeda (2007), pod wiele mówiącym tytułem: „*The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review*”. Autor wymienił trzy przyczyny, które powodują, że ocena ekspercka nie jest w pełni trafną metodą ewaluacji: oceniający często ulegają presjom politycznym i środowiskowym, są obciążeni specyfiką własnej specjalizacji, a w końcu – zamiast dokonywać oszacowania na podstawie dostarczonych informacji, często wyrażają po prostu panujące w środowisku przekonania. Moed podkreślał jednocześnie, że metody bibliometryczne najslabiej sprawdzają się w humanistyce, tu więc zastosowanie *peer review*, mimo jego wad, jest rekomendowane.

Jak pokazują wspomniane badania, miary bibliometryczne oferują większą rzetelność i porównywalność, ale nie są w pełni trafne. Z kolei „miękką” ocenę ekspercka jest jedynie przybliżeniem, może podnosić ogólną trafność wyniku, ale sama cechuje się niską rzetelnością, więc bez suplementu w postaci jakichś „twardych” miar jest narażona na wypaczenia. Paradoksalnie oceny eksperckie są najbardziej cenione w sytuacjach, kiedy są zgodne ze wskaźnikami bibliometrycznymi (oczekuje się wysokiej korelacji). W praktyce jednak ich przydatność polega przede wszystkim na tym, że umożliwiają identyfikację słabszych JN, które dzięki pewnym „manipulacjom” osiągnęły w metrykach publikacyjnych zawyżone wartości. Innymi słowy – metoda *peer review* jest w ewaluacji JN potrzebna najbardziej do weryfikacji „twardych” wskaźników, a nie ich suplementowania subiektywną opinią opartą na zupełnie innych kryteriach. Samo uśrednianie dwóch ocen jest zatem dość powierzchowną formą poprawy trafności i rzetelności ewaluacji. A tak się właśnie dzieje w obecnej formule oceny parametrycznej – ekspertom nie zlecono oceny monografii stanowiących niezwykle ważny element produkcji naukowej w NHS (kluczowego pierwszego kryterium), za to nakazano im ocenę niejednoznacznie zdefiniowanych osiągnięć ogólnospołecznych, na dodatek w oparciu o „autopromocyjne” deklaracje JN. Doświadczenia zebrane w innych krajach podpowiadają, że powinno być odwrotnie.

4.2. Sposób przyznawania kategorii

Przejdźmy do omówienia mechanizmu nadawania jednostkom kategorii. Jak wiadomo, zostały one przyznane na podstawie czterech kryteriów: publikacji, potencjału, efektów materialnych oraz innych kluczowych osiągnięć¹². Problemem o charakterze metodologicznym było to, że wartości składowych mogły silnie zależeć od tych cech JN, które nie były bezpośrednio związane z jakością prowadzonych badań, np. od ich wiel-

¹² Poszczególnym kryteriom przyznano wagi wynoszące w NHS odpowiednio: 0,65; 0,15; 0,05; 0,15. Owe mnożniki różniły się pomiędzy GN, a w ich obrębie dodatkowo między rodzajami instytucji. W NHS tego wewnętrznego zróżnicowania nie zastosowano.

kości. Parametry kryterium I i III wyznaczone zostały jako ilorazy mające w mianowniku przeciętny stan kadrowy w okresie oceny. Oznacza to, że owe kryteria potencjalnie premiowały jednostki niewielkie, z kolei kryterium II premiowało jednostki duże, przede wszystkim wydziały uniwersyteckie, posiadające kilka rodzajów uprawnień do nadawania stopni. Ponieważ z punktu widzenia JN łatwiej jest manipulować liczbą etatów (zwłaszcza w kierunku ich zmniejszania) niż posiadanymi uprawnieniami, rozwiązanie takie wydaje się korzystne zwłaszcza dla małych instytutów i niepublicznych szkół wyższych. Jednak kompensowanie tego wypaczenia za pomocą kryterium potencjału nie wydaje się ani eleganckie (skąd waga 0,15? dlaczego np. nie 0,3 albo 0,05 – jak w naukach o życiu?), ani do końca skuteczne.

Drugi kluczowy element systemu oceny to podział ogółu JN na grupy wspólnej oceny (GWO). Warto zauważyć, że proces kategoryzacji jednostek naukowych prowadzony jest wedle modelu rywalizacyjnego, a nie modelu „obiektywnego kryterium”, dlatego tak istotne jest to, kto z kim jest porównywany. Z punktu widzenia systemu zachęt (bodźców) jest to sytuacja dla państwowego mecenatu dobra, ponieważ każde nawet najlepszym jednostkom ciągle oglądać się za plecy. Komponent wzajemnych porównań wydaje się zresztą praktycznie niemożliwy do uniknięcia – w Wielkiej Brytanii, gdzie formalnie zadekretowano ocenę JN w oparciu o pewne „obiektywne” warunki brzegowe, a nie wobec siebie nawzajem, w praktyce i tak ma miejsce to drugie (zob. Moed 2007). Zarówno tam, jak i w Polsce o kategorii decyduje nie tylko kondycja danej jednostki, ale również kondycja wszystkich pozostałych w obrębie GWO. Warto też zauważyć, że w praktyce MNiSW i KEJN mogą (i powinny) kierować się również informacjami zebranymi w poprzednich edycjach kompleksowej oceny. Wydaje się przykładowo rozsądne, by jednostki referencyjne wyznaczyć w 2017 roku na poziomie nie niższym niż w edycji bieżącej. W tym sensie kryteria oceny będą się stopniowo obiektywizować. To zresztą kolejny argument na rzecz zachowania obecnych kryteriów.

Poważnym problemem pozostaje jednak określenie składu GWO, zwłaszcza ze względu na różne typy instytucji w obrębie jednej dyscypliny oraz wielodyscyplinarność jednostek uczelnianych, co Antonowicz i Brzeziński określili mianem „ewaluacyjnego węzła gordyjskiego” (zob. tychże 2013: 58). W naukach o życiu (NZ) oraz naukach ścisłych i inżynierskich (NSI) konieczność odrębnego traktowania jednostek uczelnianych, jednostek PAN i pozostałych instytucji badawczych wynika przede wszystkim z tego, że wobec każdego z trzech typów nieco inaczej sformułowane są wymagania dotyczące efektów pracy (różne wagi). W NHS oraz naukach o sztuce i twórczości artystycznej (NSTA) takich różnic nie ma, więc można by się zgodzić ze wspomnianymi autorami, że niewiele stoi na przeszkodzie, by potraktować je łącznie. Ich rozdzielenie stanowi obecnie ukłon w kierunku uczelni, które oprócz badań ponoszą ciężar kształcenia i wolałyby uniknąć konkurencji z instytutami PAN. W praktyce jednak rywa-

lizacja ta ma znaczenie jedynie w przypadku kategorii A+, a przyglądając się wynikom zakończonej niedawno parametryzacji, wypada skonstatować, że wydziały najsilniejszych polskich uniwersytetów bynajmniej nie stoją w takich porównaniach na straconej pozycji. Połączenie GWO należących do tej samej dyscypliny być może nieco rozszerzyłoby konkurencję i zwiększyło przejrzystość decyzji dotyczących nadawania najwyższej kategorii, ale nie jest to kwestia szczególnie paląca – ostatecznie wobec wszystkich i tak zastosowano te same praktycznie jednostki referencyjne¹³.

Większe znaczenie praktyczne ma w NHS kwestia heterogeniczności GWO. Jak słusznie zauważyli Antonowicz i Brzeziński (2013), w tym kontekście elementem gry między JN staje się nie tylko jakość badań, ale także formalne zaszeregowanie. Są przecież w Polsce JN, które zostały postawione w sytuacji zmuszającej do strategicznego rozważenia, w jakiej roli bardziej opłaca im się występować. Chodzi nie tylko o różnice kosztochłonności, ale również przynależności do GWO, a także wzbogacania wewnętrznej struktury organizacyjnej o zakłady i katedry, których pracownicy prowadzą produkcję naukową w dyscyplinach, w których łatwiej jest zdobywać punkty (na przykład zasobne wydziały prawa mogą przygarniać filozofów czy politologów). Przywoływani tu autorzy zaproponowali trzy rozwiązania tego problemu, jednak każde z nich związane jest z pewnymi trudnościami. Najlepsze z nich wydaje się rozwiązanie pierwsze, to znaczy wykluczenie publikacji, które ewidentnie należą do innej GN, przy zachowaniu szczególnej wrażliwości na prace interdyscyplinarne. Pozwala ono przeciąć „gordyjski węzeł” za pomocą ockhamowskiej brzytwy, to znaczy wykorzystując zastosowaną w parametryzacji szkatułkową strukturę podziału jednostek (GWO osadzone są w GN). By nie spisywać na straty części dorobku pracowników dużych jednostek uczelnianych, wystarczyłoby zezwolić na przekazanie publikacji do dorobku innej jednostki, ale jedynie w obrębie tej samej uczelni (dołączając pisemne potwierdzenie), co i tak jest obecnie możliwe. Jednostki, decydując się na złożoną strukturę, muszą po prostu być świadome, że ich dorobek będzie oceniany z perspektywy głównej dyscypliny i dopasowywać do tego kierunki badań (a nie odwrotnie – żądać dostosowania reguł gry do własnej specyfiki). W tym sensie jest to także bodziec do rozsądnej restrukturyzacji uczelni¹⁴. O los prac interdyscyplinarnych nie należy mieć przesadnych obaw – skoro ewaluatorzy stanęli przed trudnym (przynajmniej w NHS) zadaniem wskazania, która praca jest naukowa, a która nie, to z o wiele łatwiejszym zadaniem, polegającym na stwierdzeniu, czy praca w jakimś stopniu dotyczy danej dziedziny, też powinni sobie poradzić. W kontekście

¹³ Przykładowo, w konstrukcji jednostek referencyjnych dla instytutów PAN w HS2SP uwzględniano parametry łącznego rozkładu jednostek uczelnianych (HS1SP).

¹⁴ Szkoły wyższe o egzotycznej strukturze, łączące np. architekturę i pielęgniarstwo, zmuszone zostaną do przyjęcia struktury wydziałowej, może to jednak okazać się dla nich w ostatecznym rozrachunku korzystne.

wymienionych kwestii warto rozważyć jeszcze sprawę samej istoty podziału na GWO. Nie jest ona przecież tożsama z podziałem na dyscypliny. Wynika raczej z założenia, że JN, a nawet poszczególni pracownicy w ich obrębie, muszą cechować się wspólną specyfiką badań, ale przede wszystkim zbliżoną możliwością zdobywania punktów za publikacje i pozyskiwania środków. Istnienie GWO opiera się więc nie tyle na przesłankach dotyczących przedmiotu czy tradycji, ale dotyczących trafności oceny – lub inaczej rzecz ujmując – jej sprawiedliwości.

Z tego punktu widzenia uzasadnienie ma raczej dzielenie grup na mniejsze (oczywiście w granicach rozsądku), a nie ich łączenie. Weźmy przykładowo dość problematyczną, liczącą niemal sto jednostek grupę HS1SP. Na tle pozostałych GWO w NHS zasadne byłoby nadanie jej zbiorczej etykiety „inne”, ponieważ obejmuje jednostki prowadzące badania w obszarze: psychologii, socjologii, nauki o polityce, pedagogiki, kulturoznawstwa, nauki o bezpieczeństwie, nauki o poznaniu i komunikacji społecznej itp. Istnieją podstawy, by twierdzić, że możliwości zdobywania punktów są w tych dyscyplinach niejednakowe, dotyczy to np. psychologii porównywanej z innymi naukami w HS1SP. Jak stwierdzili w swoim raporcie holenderscy badacze, Peter Van den Besseelaar i Loet Leydesdorff (2007), metryki publikacyjne stanowią najbardziej trafny wskaźnik „produkcji” naukowej, jeśli stosuje się je w obrębie spójnej dyscypliny. Do ilustracji tej prawidłowości wybrali właśnie psychologię i socjologię, stwierdzając: *„Clearly, psychology journals have on average higher impact factors, and this indicates that psychologists have longer reference lists per paper than sociologist do. Also what counts as a top journal in psychology is different from what would count as top in sociology, if we take the impact factor of the journal (IF) as a criterion”* (Van den Besseelaar, Leydesdorff 2007: 29). Rzeczywiście – jeśli spróbujemy na ministerialnej liście A odnaleźć czasopisma zawierające frazę „psychol-” (licząc się z różnymi formami nazw), to znajdziemy ich grubo ponad 250, natomiast wpisując „socio-” otrzymamy 57 wyników i kolejne 87, jeśli użyjemy frazy „social”.

W przypadku tego typu GWO KEJN powinien poważnie rozważyć podział np. na grupę obejmującą pedagogikę, psychologię i kognitywistykę (dyscypliny tradycyjnie bliskie, skupione na funkcjonowaniu jednostki ludzkiej, z możliwym odrębnym potraktowaniem pedagogiki, bądź szerzej nauk o osobowej edukacji) oraz grupę nauk społecznych w wąskim sensie, takich jak socjologia, politologia czy kulturoznawstwo (dyscypliny zainteresowane poziomem mezo- i makrospołecznym – systemami społecznymi, kulturą, instytucjami itp.). Trzeba zauważyć, że ze względu na przedmiot badania i możliwości publikacyjne bardziej zbliżona do tych ostatnich wydaje się np. historia, która traktowana jest przecież odrębnie. Ostateczny argument dotyczący demarkacji GWO musi jednak dotyczyć tego, czy jedna z dyscyplin nie jest w jakiś sposób uprzywilejowana już w przedbiegach. Wiele wskazuje natomiast na to, że w przeszłości najwyższe oceny

mogą być zarezerwowane wyłącznie dla jednostek psychologicznych (widać to już teraz, patrząc na czołówkę HS1SP; w HS2SP najlepszy okazał się co prawda instytut zajmujący się filozofią i socjologią, ale w punktach za publikacje wyprzedził go instytut zajmujący się psychologią).

Warto poczynić także uwagę na temat samej konstrukcji systemu kompleksowej oceny. Szkatułkowa struktura klasyfikacji jednostek powinna stanowić podstawę dla funkcjonalnego zróżnicowania zasad parametryzacji w poszczególnych podzbiorach. Obecnie tego typu różnice dotyczą jedynie czterech głównych GN, natomiast wydaje się, że ze względu na trafność werdyktu warto zezwolić na dostosowanie reguł przynajmniej na trzech poziomach. Na pierwszym, ogólnym poziomie przyjęto zestaw uniwersalnych kryteriów oceny. Na drugim poziomie, w obrębie GN, zdefiniowana została szczegółowa treść tych czterech kryteriów – podano specyficzne wskaźniki dla NHS, NZ, NSI oraz NSTA. Na tym poziomie powinny być także rozstrzygane decyzje dotyczące przyznawania kategorii A+ (jest to bowiem decyzja strategiczna). Wydaje się natomiast, że pozostałe regulacje można by określać już na trzecim poziomie, tzn. dla poszczególnych GWO. Chodzi w szczególności o parametry takie, jak wartości wag, poziom limitu na monografie i sposób dzielenia punktów za publikacje wieloautorskie (od którego autora, w jakiej proporcji). Z punktu widzenia całego procesu byłoby to korzystne o tyle, że specyfika dyscypliny określana byłaby w obrębie jej przedstawicieli (zwiększając trafność oceny, przynajmniej trafność fasadową), natomiast zachowana zostałaby względna porównywalność i określone konkretnymi parametrami priorytety ogólne. Instytucjonalnie rzecz biorąc, zwiększyłby się nieco zakres samorządności (a więc i legitymizacji systemu), nie naruszając przy tym jego podstaw.

4.3. Kryterium osiągnięć naukowych i twórczych

Osobnego omówienia wymagają poszczególne kryteria i parametry, ponieważ każdy z nich ma swoje specyficzne znaczenie z punktu widzenia kategoryzacji. Pierwsze kryterium, „osiągnięć naukowych i twórczych”, to w przypadku NHS właściwie wskaźnik bibliometryczny¹⁵. Przypisano mu podstawowe znaczenie i najwyższą wagę. Każda z JN była oceniana na podstawie $3N$ ¹⁶ swoich najwyżej punktowanych publikacji. W praktyce wypadało w ankiecie zgłosić ich trochę więcej, ponieważ w miejsce tych pozycji, które z różnych powodów zostały odrzucone przez zespoły eksperckie, uwzględniane były kolejne ze zgłoszonej listy (traktowane jako rezerwowe). Wydaje się, że bezpieczna rezerwa powinna wynosić nie mniej niż 20%. Łączna liczba punktów za uwzględnione publikacje dzielona była przez N , dając w efekcie parametr określający przeciętną pro-

¹⁵ W przypadku towaroznawstwa brano pod uwagę również patenty.

¹⁶ Gdzie N to przeciętna liczba etatów w okresie oceny.

dukcję publikacyjną *per capita*. Wśród 3*N* „najlepszych” (w sensie punktowym) publikacji co najwyżej 40% mogły stanowić zwarte monografie. W praktyce tworzy to narzucony odgórnie model publikacyjny, wedle którego „produkcja” publikacyjna statystycznego pracownika naukowego lub naukowo-dydaktycznego powinna w ciągu czterech lat wynosić 1,8 artykułu i 1,2 książki. Model ten można realizować na różne sposoby, również przez podział ról – na przykład jedni w ramach jednostki produkują po 2 książki, inni po 4 artykuły. Wydaje się, że sama ilość oczekiwanych wytworów nie jest przesadnie wysoka, zwłaszcza że monografie nie musiały liczyć więcej niż 6 arkuszy wydawniczych, a napisanie jednego artykułu rocznie nie powinno przekraczać możliwości nawet tych, którzy są mocno obłożeni dydaktyką.

Rozwiązanie to miało swoje dobre i złe strony. Właściwą decyzją wydaje się na przykład uwzględnienie liczby 3*N* publikacji. Włoskie doświadczenia pokazują, że zbyt niski limit na sprawozdawane książki i artykuły może powodować, że w systemie oceny pojawi się dodatkowy czynnik obniżający rzetelność (Abramo, D’Angelo 2009). Zbyt wiele zależy wówczas od uwzględnienia bądź pominięcia pojedynczej pozycji. Ponadto trudno w takim przypadku mówić o jakimś konkretnym modelu publikacyjnym, a ten stanowi przecież pewien wyznacznik ogólnych oczekiwań względem kadry naukowej.

Przejdźmy do wad. Po pierwsze, nie został wyznaczony dolny limit *N*. Najmniejsza wartość dotycząca przeciętnego stanu kadrowego, jaka została podana w tej edycji oceny, wyniosła 0,7. Oznacza to, że punkty za publikacje nie były w takim przypadku dzielone, ale mnożone¹⁷. Jest to bardzo źle skrojony bodziec, zachęca do przesadnego odchudzania etatowych kadr, swoistej „instytucjonalnej anoreksji”. MNiSW mogłoby zadbać o zdefiniowanie pożądanego instytucjonalnego modelu JN, ustalając na jakimś poziomie minimalną liczbę kadr (równe na przykład 10). W praktyce chodziłoby o wprowadzenie na potrzeby określania wartości kryterium I prostej zasady, że jeśli *N* jest mniejsze niż te przykładowe 10, wówczas nadajemy mu wartość 10. Jest to wskazane ze względu na interes samych pracowników nauki, jak i państwa, ponieważ likwiduje lukę pozwalającą na działanie „wirtualnych” JN. Ta ostatnia musi być przecież definiowana przede wszystkim jako zespół badawczy, a obecnie dopuszcza się, by zespół badawczy mieścił się w obrębie jednej osoby (dwie osoby mogą stworzyć trzy zespoły po 0,7 etatowego pracownika). Nie oznacza to też automatycznie, że instytuty i prywatne szkoły wyższe muszą utrzymywać taką akurat kadrę dydaktyczną – mniej liczna grupa będzie musiała jednak pracować bardziej efektywnie (publikując – trzymając się wciąż przykładu – 30 tekstów) lub zrezygnować z ubiegania się o kategorię naukową.

¹⁷ Obecnie jednostka, która zatrudniałaby przez cztery lata jedną osobę na pół etatu, a ta osoba opublikowałaby jedną książkę, otrzymałaby automatycznie 40 pkt w pierwszym kryterium.

Jeszcze większe obciążenie parametryzacji może stanowić wspomniana już możliwość zgłaszania publikacji osób, które nie były pracownikami JN. Brzeziński (2013) nazywa go wprost „patogennym”, ponieważ umożliwia „kupowanie” afiliacji książek i artykułów. Diagnoza ta wydaje się słuszna, ponieważ już w tej edycji problem ów się pojawiał, a biorąc pod uwagę zdolność instytucji do dopasowywania się do narzuconych warunków prawnych, w kolejnych edycjach przybierze na sile. Nie zawsze jest też jednoznaczne, jak traktować publikacje, które autorzy afiliują równocześnie za granicą i w Polsce. Co z książkami osób, które funkcjonują jako „wolni strzelcy” nauki? W naukach humanistycznych i społecznych może za takich uchodzić np. wielu publicystów. „Kupienie” afiliacji nie mówi w praktyce nic o realnym potencjale JN. Dlatego też przedstawiona przez Brzezińskiego propozycja rozwiązania tego problemu wydaje się trafna – o afiliacji powinno decydować podstawowe miejsce pracy, przy czym powinno się także uwzględnić dorobek doktorantów (wystarczy w tym celu rozszerzyć imienny wykaz zawarty w ankiecie jednostki o osoby, których studia lub przewody nie zostały jeszcze zakończone). Relatywnie wysoki limit (20%) na prace osób, które nie były zatrudnione w JN powoduje, że mamy do czynienia nie z systemem oceny JN, ale systemem oceny luźnych koalicji instytucji i naukowców. Jeśli MNiSW nie zależy na promowaniu tego drugiego modelu, powinno wyraźnie postawić na model kadrowy, czyli brać pod uwagę jedynie pracowników na pierwszym etacie.

Elementem bardzo obciążającym trafność pierwszego kryterium jest w NHS fakt, że wszystkie zwarte monografie traktowane są jako równoważne i otrzymują po 20 (względnie 25) punktów. Nadaje to specyficzne znaczenie całemu parametrowi. Jego trzy podstawowe komponenty to właśnie zwarte monografie, artykuły w czasopismach naukowych oraz rozdziały w monografiach zbiorowych, które w praktyce traktowane są jak artykuły (z limitem 5 tekstów na książkę). Każdy z nich ma osobną interpretację w kontekście karty oceny. Punkty za artykuły w czasopismach z ministerialnych list wydają się być względnie trafną miarą jakości naukowej. Punkty za rozdziały, przydzielane równo – po 4 za sztukę, to miara mało trafna, ale w praktyce ma funkcję „wypełniacza”. Takie teksty są uwzględniane w ocenie jedynie wtedy, gdy zostanie na nie miejsce, a w najlepszych JN rzadko tak się dzieje (miejsca na liście publikacji wypełniają tam w pierwszej kolejności artykuły w czasopismach za 5 punktów lub więcej). Trafność oceny zaburzą natomiast zwarte monografie. Określenie ich limitu i proste zliczanie powoduje, że pierwsze kryterium określa głównie to, jak wiele książek, w tym elektronicznych (byleby miały numer ISBN i objętość 6 arkuszy) jest w stanie wygenerować dana JN. Jest to więc przede wszystkim miara sprawności własnego wydawnictwa i zdolności pracowników do tworzenia tekstu (niezależnie od jego jakości). Zważywszy na to, jak niewiele artykułów przedstawicieli NHS znalazło się w czasopismach powyżej 20 punktów (a nawet 25, jeśli weźmiemy pod uwagę języki kongresowe), waga tego kompo-

mentu kryterium I jest szczególnie duża. W efekcie główny składnik łącznej oceny staje się miarą efektywności publikacyjnej, a nie naukowej.

Takie potraktowanie monografii jest bodźcem do wydawania byle czego, dlatego ze względu na sens całego przedsięwzięcia należy dokonać zmiany zasad dotyczących kryterium publikacyjnego. Jako że monografie są jednym z najważniejszych efektów pracy naukowej w NHS, a zatem muszą być one uwzględniane, trzeba zgodzić się z Antonowiczem i Brzezińskim (2013), że jedynym rozsądnym rozwiązaniem staje się *peer review*. Co ważne – jest to bardziej właściwy element do wykorzystania oceny eksperckiej niż ocena pozostałych (ogólnospołecznych) efektów działalności naukowej (kryterium IV), ponieważ dotyczy naukowej treści, a nie abstrakcyjnej siły oddziaływania. Można to sobie wyobrazić na kilka sposobów. Na przykład taki, że jednostki podchodzące do oceny zobligowane będą do przesłania KEJN kilku swoich najważniejszych książek (można uwzględnić również prace zbiorowe), a te zostałyby rozdzielone i przekazane do oceny powołanym przez KEJN recenzentom (każdy recenzent oceniałby tylko jedną książkę z danej jednostki – taki mechanizm znacząco zwiększa rzetelność oceny). W odróżnieniu od obecnych zasad dotyczących kryterium IV powinny też zostać szczegółowo zdefiniowane zasady takiej recenzji (kryteria oceny monografii, takie jak: ocena warsztatu badawczego, ocena jakości/jasności tekstu, spójności tematycznej, ocena rozmiaru prac niezbędnych do zebrania materiału, ocena wagi dokonania dla dyscypliny, osobno w kraju i za granicą – każde kryterium z osobną wagą). Przy takim rozwiązaniu trzeba by też wyraźnie zróżnicować liczbę punktów za monografie recenzowane przez KEJN i pozostałe monografie uwzględnione w ocenie. Pierwsze mogłyby być oceniane na skali 0-100, a pozostałe otrzymywałyby domyślnie np. 12/15 punktów¹⁸. Element *peer review* wymagałby zapewne powiększenia liczby ekspertów, ale to również mogłoby się okazać w ogólnym rozrachunku korzystne (zmniejszyłaby się liczba innych zadań, zwłaszcza weryfikacyjnych, natomiast zwiększyłby się ogólny udział środowiska).

Dwie inne problematyczne kwestie dotyczą samego limitu na monografie zwarte i sposobu traktowania publikacji wieloautorskich. Wydaje się, że w pierwszym przypadku należałoby limit powiązać z konkretną GWO, a nie GN (jak to się dzieje obecnie). W NHS istnieją wyraźne wewnętrzne różnice dotyczące oczekiwanej proporcji wyników publikowanych w formie monografii. Najmniejszego limitu chcieliby zapewne psychologowie, nieco większego socjologowie, jeszcze większego historycy, a największego

¹⁸ Wymagałoby to także obniżenia limitu na punktowane rozdziały w monografiach zbiorowych do trzech (tylko tyle osobnych rozdziałów z jednej książki mogłaby zgłosić konkretna JN). Byłoby to korzystne o tyle, że zniechęcałoby np. do wydawania słabych książek zakładowych. Ich generalna ranga nie zostałaby jednak obniżona, ponieważ te najlepsze miałyby szansę zostać zgłoszone do oceny eksperckiej jako kluczowe monografie (trzeba by oczywiście ustalić warunek afiliacji, np. większość redaktorów oraz przynajmniej 50% rozdziałów z afiliującej jednostki).

zapewne filozofowie. Oczekiwania te zależą przede wszystkim od stopnia, w jakim sam tekst stanowi owoc pracy twórczej. Warto zwrócić uwagę, że współczynniki kosztochłonności są w części oparte na analogicznym założeniu – praca jednych silniej związana jest z dodatkowym kosztem badań (zwłaszcza empirycznych), a innych słabiej. Skoro uznano ten fakt w innym elemencie podziału środków, tym bardziej powinno się to uznać przy kwalifikacji i ocenie „produktów”. To GWO powinna wyznaczać limit na monografię, najlepiej w granicach 10-50% (wyższy limit groziłby, że nie mierzymy już efektów naukowych, ale publikacyjne).

Nieporozumieniem jest przejście od NZ i NSI sposobu dzielenia punktów za publikacje wieloautorskie (nie mówimy tu o książkach zbiorowych). Specyfika badań tamtych dziedzin powoduje, że zdarzają się prace sygnowane przez setki autorów. Jednak przyjęcie w NHS, że punkty za tekst będący efektem pracy dziesięciu autorów nie podlegają dzieleniu¹⁹, a powyżej tej liczby dzielone są co najwyżej przez 2, jest sytuacją kuriozalną. Warto zauważyć, że artykuły i książki otrzymują w tym drugim przypadku połowę punktów, ale jednocześnie zajmują jedynie połowę publikacyjnego „slotu”. W praktyce nie jest to więc względnie niższa ocena publikacji, tylko traktowanie jej jako „mniejszej jednostki publikacyjnej”. Jeśli już tak robimy, to dlaczego nie rozważyć ogólnie traktowania artykułów jako „mniejszych”, a książek jako „większych jednostek publikacyjnych”? Gdyby przyjąć zasadę dzielenia publikacji na jednostki publikacyjne różnej wielkości, wówczas można bardziej elegancko rozwiązać kwestię limitu na monografie. Wystarczy na przykład przyjąć, że książka za 20 punktów zajmuje pięć „slotów”, a rozdział i artykuł „konsumują” tylko po jednym „slocie”. Można wówczas stworzyć również więcej kategorii jednostki publikacyjnej – przykładowo zamiast usuwać z oceny recenzje, przeglądy literatury itp., można by przyjąć, że uzyskują one pół punktu, ale jednocześnie zajmują pół „slotu”. Podobnie z tłumaczeniami, opracowaniami dokumentów, map itp. W ten sposób pozwolilibyśmy JN nawet na zgłoszenie samych monografii, mimo że nie byłoby to równoznaczne z uzyskaniem przez nią przewagi w punktach. Oczywiście takie rozwiązanie trzeba suplementować wspomnianą ekspercką recenzją wybranych książek, by wskaźnikowo ocenić – *pars pro toto* – względną wartość dorobku jednostki. Ta część „produkcji” naukowej, która jest w praktyce ewaluacyjnej traktowana „hurtowo”, powinna być rzeczywiście przeliczana na jednakowe jednostki, a nie tak, jak to się dzieje obecnie, gdy jedne formy wydawnicze są cenione wyżej niż inne.

Warto wspomnieć być może i o kwestii premii za książki w językach kongresowych. Premia ta zmniejszyła się w ostatniej edycji oceny (w porównaniu z poprzednią) ze 100% do 25%. Sama idea premiowania takich publikacji jest całkowicie zgodna z najważniejszymi priorytetami wyznaczonymi przez MNiSW (umiędzynarodowienie), dlatego

¹⁹ Gdy każdy wskazuje odmienną afiliację.

wspomniany spadek może nieco dziwić. Z drugiej strony o randze naukowej JN nie może decydować zdolność do tłumaczenia na języki obce. Ze względu na te dwie kwestie premia w obecnej wielkości wydaje się rozsądnym kompromisem, można jedynie zastanawiać się, czy publikowanie w języku angielskim, niemieckim, francuskim, hiszpańskim, rosyjskim i włoskim wiąże się z tym samym, czy też różnym stopniem potencjalnego umiędzynarodowienia. Być może za ten pierwszy, będący *lingua franca* współczesnej nauki, wypadałoby przyznać premię nieco wyższą.

4.4. Pozostałe kryteria oceny

Drugie kryterium – „potencjału naukowego” – dotyczy przede wszystkim siły kadrowej. To od niej zależy bowiem liczba i rodzaj posiadanych uprawnień, a także liczba nadawanych stopni (główne składniki parametru). Jak już wspomnieliśmy, jest to rodzaj rekompensaty dla dużych JN za to, że trudniej im manipulować mianownikiem parametru publikacyjnego i materialnego. Nieco niezręczne może wydawać się jednak to, że jako „efekt” czy też „produkt” działalności badawczej traktuje się uprawnienia czy stopnie. Ponadto idzie o te same elementy, które uwzględnione są przy podziale dotacji dla szkół wyższych (tzw. dotacji dydaktycznej). Oznacza to, że komponent ów jest niejako podwójnie premiowany przy dzieleniu środków będących w dyspozycji MNiSW. Czy należałoby więc to kryterium usunąć z oceny? Wydaje się, że potrzebna jest bardziej kompleksowa korekta systemu dzielenia dotacji. Z merytorycznego punktu widzenia da się bowiem wskazać nie trzy, a cztery dystynktywne strumienie²⁰ – różniące się nie tylko przeznaczeniem, ale także pożądanym sposobem dystrybucji.

Pierwszy strumień dotyczyłby wyłącznie działalności dydaktycznej – na pierwszym i drugim stopniu (byłaby to *sensu stricte* dotacja na szkolnictwo wyższe). Drugi strumień dzielony byłby w trybie konkursu projektów (NCBR, NCN). Trzeci strumień dotyczyłby utrzymania potencjału, czyli dzielony byłby z uwzględnieniem stanu kadrowego, uprawnień i nadanych/uzyskanych stopni (powinien uwzględniać koszt promocji, byłaby to więc dotacja celowa – przynajmniej w części). Czwarty strumień, na statutowe badania naukowe, powinien mieć charakter ściśle merytokratyczny, a więc powinien być dzielony wyłącznie ze względu na jakość prac badawczych. Przy jego podziale brane powinny być pod uwagę jedynie realne „produkty” (a więc obecne kryterium I, III i ewentualnie IV), ale konsekwencją kategoryzacji powinna być jeszcze bardziej zróżnicowana dystrybucja dotacji. W ten sposób MNiSW realizowałoby cztery odrębne zadania – tym samym mogłoby je realizować w sposób bardziej przejrzysty i skuteczniej kontrolowany. Inna możliwość to połączenie trzeciego i czwartego ze wspomnianych strumieni (na utrzymanie potencjału i badania statutowe), które wymagałoby po prostu przesunięcia

²⁰ Nie licząc strumieni „technicznych”, np. na restrukturyzację czy utrzymanie aparatury.

części środków z obecnej dotacji dydaktycznej i objęcia jej algorytmem uwzględniającym kategorię naukową. Takie rozwiązanie stałoby się jednocześnie szansą (w dłuższej perspektywie czasowej oraz zakładając odpowiednią alokację środków pomiędzy strumienie) na to, by silniej uzależnić posiadanie uprawnień od jakości prowadzonych badań. Przy obecnym algorytmie podziału środków potencjał naukowy uzależnia się od wielkości naboru na studia I stopnia, a jakość wyników naukowych ocenia się na podstawie posiadanych uprawnień.

Również trzecie kryterium, „materialnych efektów działalności naukowej”, wymaga pewnego komentarza. Przynajmniej w niektórych GWO w obrębie NHS środki materialne, o których tu mowa, są w praktyce prawie wyłącznie pochodną przyznanych JN grantów²¹. Mamy tu nieco dwuznaczną sytuację. Z jednej strony środki jednego strumienia (badania statutowe) są dzielone w oparciu o to, jak zostały podzielone inne środki MNiSW (konkursy grantowe). Z drugiej strony proponuje się w ten sposób wykorzystanie dokonanej już przez ekspertów NCN i NCBR oceny projektów badawczych, czyli korzysta się z gotowego wskaźnika „pomysłowości”. Wskaźnik ten jest informacją o tym, na ile pracownicy poszczególnych JN są w stanie przekonać swoimi pomysłami własne środowisko i dokonać tym samym postępu wiedzy w określonym przez siebie kierunku. Mamy tu więc rodzaj bezkosztowego wykorzystania dobrej jakościowo informacji pochodzącej z *peer review* (jeśli wierzyć dobrym opiniom o procesie recenzentkim w NCN), ale równocześnie problem nakładania się strumieni finansowych. Czy da się wyjść poza tę dwuznaczność? Wydaje się, że tak – wystarczy wziąć w ocenie pod uwagę liczbę grantów, a nie sumy, na jakie opiewają, „pomysłowość” zostaje wtedy niejako oderwana od kosztu tych pomysłów. Takie rozwiązanie proponuje również Brzeziński (2013), słusznie uważając je za dodatkowy bodziec do prowadzenia samodzielnych badań przez młodszą kadre. Byłby to zresztą dość istotny wskaźnik dotyczący ogólnego efektu naukowego działalności JN, stanowiący rodzaj przeciwwagi dla kryterium publikacyjnego. Z pewnością należałoby mu przypisać znacznie wyższą wagę niż 0,05 (być może przynosząc go zarazem do innego kryterium, posiadającego już wyższą wagę²²).

²¹ Nie ma się co oszukiwać, że badania prowadzone w dyscyplinach wielu GWO (poza wyjątkami, jak np. ekonomia) dadzą się łatwo skomercjalizować. Nawet jeśli kompetencje niektórych pracowników wzbudzą zainteresowanie biznesu, to dla obu stron bardziej korzystna okazuje się często inna forma współpracy – omijająca dodatkowe „opodatkowanie” (wynikające na przykład z wewnątrzuniwersyteckich kosztów administracyjnych). Jednostki naukowe w NHS nie dysponują na ogół zasobami sprzętowymi czy infrastrukturalnymi, które zapewniłyby ich udział w takiej współpracy.

²² W grę wchodzi zarówno kryterium I (można np. przyznawać za granty od 10 do 50 pkt, w zależności od programu), jak i kryterium II (takie przesunięcie zostało dokonane w ramach NZ). Wydaje się, że sama liczba dobrych projektów mówi jednak raczej o „osiągnięciach naukowych i twórczych” niż wąsko rozumianym „potencjale naukowym”.

Najbardziej problematyczne z punktu widzenia rzetelności ewaluacji wydaje się ostatnie kryterium, obejmujące „pozostałe efekty działalności naukowej” (a jak doprecyzowuje załącznik do Rozporządzenia z 13 lipca 2012 r. – chodzi o osiągnięcia „o znaczeniu ogólnospołecznym lub gospodarczym”). Dokładne zasady dotyczące tego elementu, określanego czasem jako ocena za „wrażenia artystyczne”, mogły różnić się między poszczególnymi GWO. Na przykład wytyczne w grupie nauk społecznych nakazywały brać pod uwagę przede wszystkim zasięg i oryginalność osiągnięć naukowych. Ze względu na uznaniowość tego drugiego elementu w praktyce kluczowy okazał się często stopień „umiędzynarodowienia” wymienionych działań. Nie było to do końca zgodne z literą aspektów wyszczególnionych w załączonej do Rozporządzenia z 13 lipca 2012 r. Karty Kompleksowej Oceny JN (KKO), ale powiedzmy też otwarcie, że zamieszczone tam enumeratywne określenie osiągnięć ogólnospołecznych, jednakowe dla wszystkich GN, nie było ani rozwiązaniem praktycznym, ani ścisłym, ani trafnym.

Największy problem wystąpił po stronie samych JN, które wypełniając ankietę w bardzo zróżnicowany sposób zinterpretowały sens IV kryterium. Na przykład – skoro jeden z pięciu elementów wymienionych w KKO mówił o upowszechnianiu wiedzy, wiele jednostek zachęconych zostało do wpisania różnych, organizowanych lokalnie *eventów* popularnonaukowych, ale takie osiągnięcia, jako mające niewielki zasięg, oceniane były nisko. Niektóre szkoły wpisywały osiągnięcia kształceniowe, a nie naukowe. Inne podawały po jednym osiągnięciu dla każdego z pięciu wymienionych w KKO przykładów. Ilustracje można mnożyć, ale wszystko sprowadza się do mało precyzyjnie sformułowanych zasad. Samo dodanie tego kryterium do algorytmu kompleksowej oceny sprawia wrażenie, jakby było efektem niedoskonałego kompromisu w dwóch sprawach. Pierwszą jest uwzględnienie jakiejś formy oceny eksperckiej (czyli postulowane przez wielu „zmiękczenie” ewaluacji), a drugą danie jednostkom swobody w samodzielnym definiowaniu dokonań (w zasadzie mogą się chwalić, czym chcą).

Również po stronie ewaluatorów taka formuła okazała się niezręczna. Osiągnięcia oceniane były nie pojedynczo, ale łącznie, na skali 0-100. Ze względu na słabe sprecyzowanie zadania, jego rzetelne wykonanie stawało się dużym wyzwaniem, a rezultat musiał w jakiejś mierze opierać się na „wrażeniach artystycznych”, przepuszczonych przez filtr pozytywnych lub negatywnych uprzedzeń wynikających z opinii środowiskowej. Ignorowanie tej opinii i wiedzy o realnym statusie danej jednostki oznaczałoby zresztą, że ocena opierałaby się wyłącznie na sile „autopromocyjnej” treści zawartej w ankiecie, a więc byłaby jeszcze bardziej przypadkowa.

Werdykt ekspercki wydawany na tak szerokiej skali punktowej okazał się pomysłem niedobrym. Mamy tu bowiem znacznie więcej poziomów zmiennej niż obiektów, którym są one przypisywane. Z punktu widzenia poznawczego ocena tego rodzaju była generalnie dość skomplikowanym zadaniem. Prowadzone w psychologii i socjologii ekspery-

menty metodologiczne pokazują, że zbyt długa skala utrudnia rzetelną ocenę i w praktyce powoduje uruchomienie jakiejś formy oszacowania upraszczającego (zob. Preston, Colman 2000, Tourangeau, Rips, Rasinki 2000). Fakt, że poszczególni ludzie (nieważne, czy są to respondenci w badaniach sondażowych czy ewaluatorzy posiadający stopnie naukowe) stosują w praktyce nieco inne mechanizmy uproszczeń, powoduje, że efektywne skale szacunkowej oceny nie powinny liczyć więcej niż 11 punktów (często twierdzi się, że i to jest zbyt duża ilość). Dodawanie kolejnych punktów skali nie zwiększa precyzji dokonywanej przy jej pomocy oceny. Dorzuca za to porcję losowej wariancji. Z tego względu precyzyjna ocena na skali 0-100 z naukowego punktu widzenia nie ma większego sensu (chyba że założymy nieskromnie, że naukowcy z zasady dysponują większą „rozdzielczością” kognitywną²³). Zresztą, po co jest potrzebna tak szczegółowa informacja? Skoro ma ona służyć przyznaniu jednej z 3 kategorii (o A+ decyduje osobny zestaw warunków), to wystarczy, by każdy z ekspertów decydował, czy dane osiągnięcia jego zdaniem kwalifikują JN do grupy A, B czy C.

Część wariancji ocen wynikała także z tego, że ewaluatorzy zawsze cechują się zróżnicowanym poziomem subiektywnych standardów (surowości ocen). Jak pokazały cytowane wcześniej badania, rezultat recenzji jest w części konsekwencją tego, kto został powołany do recenzowania. Niesystematyczne obserwacje poczynione w trakcie parametryzacji sugerują na przykład, że oceny mogły być skorelowane ze statusem JN, z której pochodził ekspert, a także z reprezentowaną przez niego dyscypliną. Nie mamy dostępu do danych, które pozwalałyby zbadać te hipotezy, ale MNiSW może to przecież uczynić. Wskazują na to pewne poszlaki. Wyliczając współczynniki zmienności²⁴ dla poszczególnych GWO w NHS, otrzymamy:

- HS1HS (historia, $n = 20$): $V = 0,29$,
- HS1FB (filologia, $n = 29$): $V = 0,42$,
- HS1FT (filozofia i teologia, $n = 16$): $V = 0,53$,
- HS1EK (ekonomia i zarządzanie, $n = 93$): $V = 0,59$,
- HS1SP (nauki społeczne, $n = 93$): $V = 0,65$,
- HS1PR (prawo, $n = 25$): $V = 0,70$.

Wyższy współczynnik zmienności świadczy o większym zróżnicowaniu wartości (niezależnie od liczby obiektów i wielkości średniej). Jak widać, w omawianym kryterium najbardziej zbliżone oceny otrzymały jednostki historyczne, a najbardziej zróżnicowane – jednostki prawnicze. Międzygrupowa wariancja tego współczynnika wynika częściowo zapewne z realnego zróżnicowania osiągnięć w danej grupie, jednak trudno domniemywać, by wynikała jedynie z niego. Także sama liczebność GWO nie tłumaczy zróżnicowania

²³ Można to łatwo zweryfikować sprawdzając, czy noty będące wielokrotnością 5 lub 10 były przez oceniających wykorzystywane relatywnie częściej niż pozostałe.

²⁴ Iloraz odchylenia standardowego i średniej.

wania ocen (najmniejsza i największa grupa cechowały się akurat najbardziej przeciętną dyspersją). Większe znaczenie miała obecność szkół niepublicznych, ale i ona nie wyjaśniała wszystkiego. Domniemanie, że „efekt ewaluatora” wystąpił, wydaje się zasadne, a weryfikację tej hipotezy można przeprowadzić na danych z systemu informatycznego, stosując metodologię analogiczną do tej Cole’a, Cole’a i Simona (1981).

Trzeba postawić również pytanie o to, jakim bodźcem jest kryterium IV w obecnie przyjętej formie. Jaka strategia działania może być preferowana przez JN ze względu na ekspercką ocenę osiągnięć o znaczeniu ogólnospołecznym? Wydaje się, że wymaga to od JN dobrego *publicity*, renomy budowanej różnymi metodami, nie tylko poprzez same wyniki badań. Zachęca do rozwijania i utrzymywania sieci kontaktów i wzajemnego wsparcia. Jest dobrym bodźcem do tego, by dbać o środowiskowy wizerunek i dobre relacje. Tych spraw próżno jednak szukać wśród priorytetów, które dla JN formułuje MNiSW.

W ogólnym rozrachunku trzeba stwierdzić, że wartości kryterium IV wydają się w znaczącej mierze przypadkowe (tj. oparte na splocie odmiennych uwarunkowań) i nie stanowią dla JN dobrego naukowego drogowskazu. Można wyobrazić sobie pewne kroki naprawcze. Oprócz wspomnianego już przejścia na nominalną skalę ocen należałoby doprecyzować (i to w nowej wersji rozporządzenia, a nie w ustalanych *post factum* zasadach roboczych) sposób oceny tych osiągnięć. Skoro i tak jest to kryterium o charakterze „jakościowym”, to najlepiej byłoby poprosić bezpośrednio: o opisanie kilku najważniejszych, poczynionych przez pracowników jednostki ustaleń naukowych (odkryć, opracowań, syntez itp. – wkład w wiedzę należy zdefiniować szeroko). Konsekwencje takiej formuły byłyby wielorakie. Po pierwsze, stanowiłoby to bezpośrednie odwołanie do „jakości działalności naukowej”, a nie jej „ogólnospołecznego i gospodarczego znaczenia” (chyba że przez pracę naukową rozumiemy też popularyzatorstwo, ale to raczej domena edukacji naukowej – finansowanej odrębnie). Po drugie, zmusiłoby to JN do oceny, czy ich działalność faktycznie wpływa na stan wiedzy w danej dyscyplinie i zmobilizowało do konkretnych kroków w kierunku tworzenia nowej wiedzy. „Produkcja” naukowa o charakterze wtórnym, ale sygnowana znaczącymi nazwiskami, często ma większe znaczenie statusowe niż nowatorskie dokonanie mniej znanych badaczy (które docenia się często wiele dekad później, a czasem nawet pośmiertnie). Należy też pamiętać, że ocena ekspercka ma w założeniu kompensować grzechy bibliometrii, polegające przede wszystkim na tym, że premiuje popularność. Lepiej więc oceniać „nowy wkład w wiedzę” niż „ogólne znaczenie” czy „zasięg”. Również ekspertom łatwiej będzie oceniać wagę konkretnego ustalenia, niż np. wagę serii konferencji lub seminariów z udziałem kilku zagranicznych badaczy. W końcu po trzecie – byłby to jednocześnie sposób na rozpowszechnianie informacji o nowych dokonaniach naukowych. Dyfuzja odkryć jest ważnym, a często powolnym procesem, które KEJN i MNiSW mogą efektywnie (choć „przy okazji”) wesprzeć.

4.5. Realne znaczenie parametrów²⁵

Jak już wcześniej wspomniano, realna waga każdego z czterech kryteriów zależała od statystycznego rozkładu jego wartości. Przykładowo, gdyby wszystkie JN otrzymały dokładnie tę samą wartość parametru publikacyjnego, wówczas jego faktyczna waga wyniosłaby 0, więc o przyznaniu kategorii decydowałyby pozostałe kryteria. Ocena sumaryczna uzyskiwana była na podstawie porównań parami dokonywanych osobno w każdym kryterium. Uwzględniane były przy tym zasady: nierozróżnialności, pełnego przewyższenia i częściowego przewyższenia. Uzależniało to wynik porównań od parametrów rozkładu, takich jak dyspersja czy kurtoza. Jak pokazuje rycina 1, rozkłady wartości parametrów były w poszczególnych GWO dość zróżnicowane, co sprawiało, że posiadały nieco odmienne znaczenie dla oceny ogólnej. W przypadku kryterium I najmniejszym rozrzutem wartości cechowała się grupa filozoficzno-teologiczna ($V=0,14$), zdecydowanie największym zaś grupa nauk społecznych ($V=0,41$). Ta ostatnia wartość świadczy prawdopodobnie nie tylko o zróżnicowaniu jakości pracy naukowej, ale również o tym, że heterogeniczność grupy wynika z pewnych wewnętrznych różnic merytorycznych – w końcu równie liczna grupa HS1EK, obejmująca także szkoły prywatne, ale bardziej jednorodna „gatunkowo”, cechowała się wyraźnie mniejszą dyspersją ($V=0,27$). Jest to kolejna przesłanka sugerująca, że jednostki znajdujące się w HS1SP powinno się oceniać w osobnych grupach.

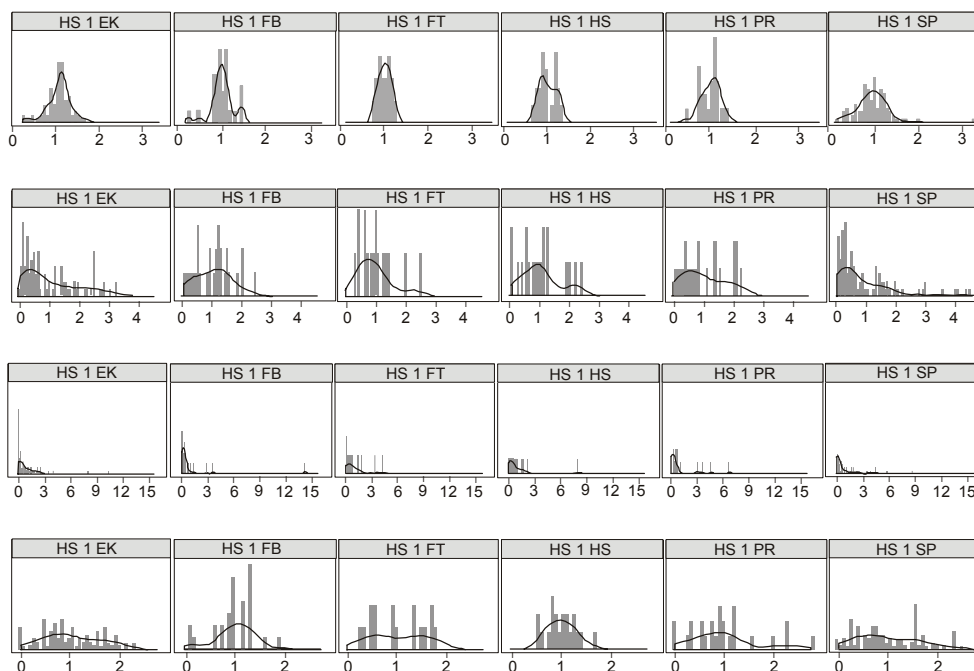
Ze wspomnianego wykresu wynika także, że pierwsze kryterium cechuje się znacznie mniejszym rozrzutem wartości niż pozostałe. Względnie rzadziej mogło dochodzić tu więc do sytuacji pełnego przewyższenia, przez co faktyczna waga tego kryterium mogła być nieco niższa, niż waga przypisana jej formalnie. Najwyższymi współczynnikami zmienności cechowało się kryterium III (od $V=1,25$ w HS1FT do $V=2,67$ w HS1EK), a następnie kryterium II (od $V=0,58$ w HS1FT do $V=1,11$ w HS1SP).

Warto spróbować w metodyczny sposób zmierzyć się z pytaniem o to, w jak dużym stopniu same dystrybuanty kształtują ostateczny wynik oceny. By na to odpowiedzieć, należy sprawdzić, jak w zależności od rozkładu zmienia się wynik porównywania parami. Jak kształtuje się on, gdy parametr ma niewielką dyspersję (jak w kryterium I), jak, gdy zróżnicowanie jest duże (jak w kryterium II i IV), a jak, kiedy na dodatek cechuje się wyjątkowo dużą asymetrią (jak w kryterium III)?

Tabela 1 prezentuje wyniki analiz symulacyjnych, dotyczących kilku dystrybucji hipotetycznych oraz – dla porównania – rozkładów empirycznych w HS1EK. Na potrzeby symulacji stworzona została grupa 100 wirtualnych jednostek, którym przypisywane były na zmianę parametry o różnym stopniu dyspersji, asymetrii (skośność) oraz koncentracji (kurtoza). Wszystkie parametry zostały wycelowane na średniej (jest ona

²⁵ Wykorzystane w tej części dane pochodzą z wyników ogłoszonych jeszcze przed fazą odwołań.

podstawą ustalania minimalnej wartości progu nierozróżnialności i progu pełnego przewyższenia).



Ryc. 1. Rozkład kryteriów I-IV (w kolejnych wierszach) w zależności od GWO (w kolumnach) Wartości na osi X odpowiadają wielokrotności średniej danego parametru w GWO

Następnie wyliczona została liczba punktów, jakie przy takim rozkładzie otrzymałaby JN, która znajdowałaby się na miejscach 1., 25. oraz 50. danego rozkładu. Parametrom nie nadano żadnych wag (a więc wszystkie mają niejako wagę 1). W przypadku porównawczych danych empirycznych postępowanie było analogiczne, przy czym ostateczny wynik został przeskalowany tak, aby uwzględnić odmienną liczbę porównań parami.

Pierwszy z zaprezentowanych parametrów („normalny A”) cechował się rozkładem Gaussowskim o bardzo niewielkiej dyspersji²⁶. Skutkuje to względnie niewielkimi różnicami punktacji. Kolejny parametr („normalny B”) posiadał ośmiokrotnie większą dyspersję, co automatycznie zmieniło konsekwencje punktowe. Warto zauważyć, że empiryczny rozkład kryterium w ekonomicznej GWO (kolumna F) stanowił ze względu na rozrzut wartości przypadek pośredni między wspomnianymi dystrybucjami.

²⁶ Lekko ujemna kurtoza wynika z ograniczenia liczby jednostek do 100.

Tabela 1. Liczba punktów JN^a (wybrane pozycje rankingowe) w zależności od rodzaju rozkładu wartości parametru^b

Parametr	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
Typ rozkładu	Normalny A	Normalny B	Monotoniczny	Sześcienny	Pareto	HS1EK: kryt. I	HS1EK: kryt. II	HS1EK: kryt. III	HS1EK: kryt. IV
Średnia	1	1	1	1	1	1	1	1	1
Min.	0,88	0,02	0,01	0,00	0,00	0,19	0,00	0,00	0,00
Max.	1,12	1,98	1,99	3,94	41,44	1,66	3,27	10,39	2,23
Odch. std. (V) ^b	0,05	0,40	0,58	1,14	4,45	0,28	0,92	1,70	0,59
Skośność	0	0	0	0	8,1	-0,6	0,9	3,5	0,2
Kurtoza	-0,3	-0,3	-1,2	0,0	70,8	1,3	-0,4	14,9	-0,8
Liczba pkt JN na pozycji 1.	14,3	94,1	83,1	93,7	99,0	93,2 ^c	94,3 ^c	98,9 ^c	91,8 ^c
Liczba pkt JN na pozycji 25.	0,9	41,1	46,1	49,1	27,4	22,9 ^c	44,9 ^c	48,6 ^c	42,8 ^c
Liczba pkt JN na pozycji 50.	-1,9	-7,9	-3,5	-3,3	-33,0	-1,6 ^c	-8,9 ^c	-10,7 ^c	-10,0 ^c

^a Wyniki symulacji hipotetycznej grupy 100 jednostek.

^b Dla wygody interpretacyjnej wszystkie zmienne wycelowano na średniej równej 1. W tym wypadku odchylenie standardowe odpowiada współczynnikowi zmienności V.

^c W przypadku danych porównawczych z HS1EK wynik przeskalowano tak, by był porównywalny (bez wag kryterialnych).

Trzeci typ rozkładu (monotoniczny) cechował się równymi interwałami między kolejnymi jednostkami, z czego wynikała większa dyspersja, ale też większa platykurtyczność rozkładu. Taka dystrybucja parametru zmniejszała nieco premię najlepszej jednostki i w praktyce przypominała rozkład kryterium IV w HS1EK (kolumna I). Sześcienn rozkładu monotonicznego (kolumna D) miał oczywiście znacznie większą dyspersję niż dwa poprzednie, ale z uwagi na zachowaną symetrię i zerową platykurtyczność nie stawał najlepszej jednostki w sytuacji korzystniejszej, niż miało to miejsce w przypadku rozkładu normalnego B. Wyraźną różnicę obserwujemy dopiero przy rozkładzie Pareto²⁷

²⁷ Jest to rozkład odpowiadający bardzo silnej, ale obserwowanej w warunkach naturalnych nierówności. Często dotyczy np. zróżnicowania w liczbie kontaktów czy w poziomie bogactwa. Potocznie opisywany jest jako sytuacja, w której grupa stanowiąca liczebnie mniejszość posiada absolutną większość zasobów (np. 20% jednostek posiada 80% określonego zasobu).

($\alpha = 0,5$), gdzie silna jest leptokurtyczność i asymetria rozkładu. Wówczas różnica między najsilniejszą jednostką a tą, która znajduje się w jednej czwartej rozkładu, wyniosła ponad 70 punktów, a dystans do jednostki medianowej przekroczył 130 punktów. Nieco podobne, choć wyraźnie mniejsze różnice dostrzec możemy w przypadku kryterium III (kolumna H). Wartości kurtozy i skośności, jakie obserwujemy w rozkładzie Pareto (kolumna E), mogą wydać się przykładem skrajnego, mało realistycznego zróżnicowania osiągnięć między jednostkami, ale sytuacja taka jest możliwa – zauważmy, że te same statystyki trzeciego kryterium były w niektórych grupach wyraźnie wyższe niż w HS1EK (np. w HS1FB wyniosły odpowiednio 32,3 oraz 4,7, podobnie było w HS1SP).

Jakie są konsekwencje takich różnic rozkładu dla ostatecznej oceny? Gdyby kryterium publikacyjne miało rozkład normalny A, kryterium kadrowe rozkład sześcienny (D), kryterium materialne rozkład Pareto (E), a kryterium osiągnięć eksperckich rozkład monotoniczny (C), wówczas poszczególne składowe oceny dla najsilniejszej jednostki w każdym z tych kryteriów wyniosłyby odpowiednio (po uwzględnieniu ważenia): 9,3; 14,1; 4,95; 12,5, a dla jednostek w trzecim kwartylu odpowiednio: 0,6; 7,4; 1,4; 6,9. Jak widać, pierwsze miejsce w rankingu potencjału i ocenie eksperckiej dawałoby wówczas więcej punktów niż publikacje. Nawet kryterium osiągnięć materialnych dawałoby jego liderowi więcej, niż znalezienie się na 25. miejscu (na 100) pod względem punktów za publikacje. Rozpatrzmy jeszcze jedną sytuację – wszystkie kryteria punktowane są zgodnie z rozkładem Pareto. Wówczas 25. miejsce w parametrze bibliometrycznym daje 17,8 punktu, natomiast pierwsze miejsce w pozostałych daje łącznie 34,7 punktu. Przeprowadzone symulacje w zestawieniu z danymi empirycznymi dowodzą, że kryteria II-IV powinny być przedmiotem szczególnej uwagi wśród jednostek aspirujących do uzyskania wysokiej kategorii.

Wszystkie te uwagi nie stanowią krytyki pod adresem przyjętej metody wyznaczania oceny końcowej. Pokazują jednak, że realna waga poszczególnych parametrów może być inna niż waga nadana im *a priori*. Wiele zależy od rozkładu i konkretnego miejsca, jakie zajmuje w nim JN. Wydaje się jedynie, że przy wyznaczaniu progów lepiej byłoby posłużyć się miarą bardziej odporną niż średnia – to znaczy medianą. Skrajnie wysokie bądź niskie wartości miałyby wówczas mniejszy wpływ na ustalenie obszaru nierozróżnialności (obecnie jest to 10% średniej). Trzeba tak uczynić, zwłaszcza jeśli kryteria II-IV mają z założenia raczej rozstrzygać w przypadku „remisów publikacyjnych”, niż stanowić samodzielną podstawę do znaczącego awansu w rankingu. Obecnie sytuacja, w której jedna JN jest w stanie całkowicie zdystansować pozostałe pod względem uprawnień lub efektów materialnych, może być dla niej korzystna nie tylko ze względu na uzyskiwane przez nią samą punkty, ale również ze względu na to, że punktacja drugiej pod tym względem jednostki zależeć będzie od pułapu średniej (jej wynik łatwiej znajdzie się w obszarze nierozróżnialności z kolejnymi, niżej ocenianymi jednostkami).

Warto przyrzeć się także korelacjom pomiędzy parametrami w obrębie poszczególnych GWO (tabela 2). Możemy z nich odczytać, czy parametry – jako miary efektywności naukowej – wzajemnie się „potwierdzają” czy raczej „uzupełniają”. Jak widać, sytuacja zależy od grupy, przykładowo wśród jednostek ekonomicznych kryterium III jest praktycznie niezależne od pozostałych (zdolność do zdobywania środków grantowych nie zależy od potencjału czy publikacji). Z kolei w grupie filozoficzno-teologicznej suma grantów (i innych pozyskanych środków) dość ściśle zależy od potencjału kadrowego (inaczej niż we wszystkich pozostałych GWO).

Tabela 2. Wartości korelacji między kryteriami I-IV w NHS^a

HS1EK					HS1FB					HS1FT				
	K1	K2	K3	K4		K1	K2	K3	K4		K1	K2	K3	K4
K1	1				K1	1				K1	1			
K2	0,48	1			K2	0,23	1			K2	0,1	1		
K3	0,03	0	1		K3	0,45	-0,1	1		K3	0,26	0,66	1	
K4	0,55	0,8	0,1	1	K4	0,65	0,52	0,35	1	K4	0,23	0,49	0,21	1
HS1HS					HS1PR					HS1SP				
	K1	K2	K3	K4		K1	K2	K3	K4		K1	K2	K3	K4
K1	1				K1	1				K1	1			
K2	0,42	1			K2	0,42	1			K2	0,27	1		
K3	0,6	0	1		K3	0,2	0,29	1		K3	0,23	0	1	
K4	0,58	0,77	0,36	1	K4	0,41	0,78	0,49	1	K4	0,46	0,57	0,32	1

^a Oznaczono korelacje powyżej 0,6 i poniżej 0,1

W kontekście omawianego wcześniej zagadnienia trafności i rzetelności ocen eksperckich warto przyrzeć się również współczynnikom dotyczącym kryterium IV. A zatem w większości GWO ocena ekspercka jest funkcją wskaźnika bibliometrycznego i siły kadrowej. Jednak na ogół to ten drugi wskaźnik odgrywa większą rolę. Jedynie wśród filologów siła związku z osiągnięciami publikacyjnymi była wyższa niż z potencjałem. Z kolei wśród jednostek ekonomicznych, prawniczych i historycznych ranga osiągnięć ogólnospołecznych była niemalże funkcją potencjału kadrowego. Dodajmy też, że wielkość GWO nie miała tu znaczenia. Ogólnie rzecz biorąc, kryterium IV stanowi zatem mniej lub bardziej dokładne powtórzenie wariancji określonej przez pierwsze dwa parametry, a więc w wielu grupach może zostać uznany za redundantny. Jest to argument na rzecz nie tyle usunięcia tego kryterium, co zmiany jego formuły, by przestała być obciążoną „efektem halo” oceną za „wrażenia artystyczne”. Co ciekawe, przy obecnej postaci parametrów to kryterium III wnosi nowe (tzn. niezależne od innych wskaźników) informacje do ogólnej oceny jednostek, a więc to tu tkwi – jak się wydaje – istotna możliwość podniesienia trafności.

Przedstawione tu dane pozwalają na sformułowanie kilku wniosków ogólnych. Po pierwsze, wkład poszczególnych parametrów do oceny konkretnej JN w znacznej mierze zależy od rozkładu wartości tych parametrów we wszystkich ocenianych wspólnie jednostkach. Po drugie, w niektórych przypadkach, zwłaszcza gdy w grę wchodzi silna konkurencja z innymi jednostkami, znaczenie parametrów II-IV może okazywać się kluczowe. Po trzecie, różnice dotyczące roli, jaką poszczególne parametry pełnią w konkretnych GWO, pokazują, że stosowalność ujednoliconych miar wobec wszystkich JN w NHS jest kwestią problematyczną (dotyczy to zwłaszcza humanistyki). Wydaje się, że przynajmniej wartość wag poszczególnych parametrów powinna być ustalana osobno dla każdej GWO.

Podsumowanie

Kompleksowa ocena JN jest praktyką administracyjną opartą na pomiarze efektów ich pracy, służącą alokacji środków publicznych zgodnie z kierunkami polityki naukowej państwa. Zaproszenie do udziału przedstawicieli środowiska naukowego w postaci utworzenia dodatkowej instytucji (KEJN) stało się krokiem na rzecz trwałego uwzględnienia w tej praktyce głosu samych polskich badaczy. Rozszerzona została „demokratyczna”, środowiskowa kontrola nad przebiegiem przyznawania kategorii, a samo powołanie KEJN można traktować jako wprowadzenie dodatkowego elementu samorządności.

Warto jednocześnie zauważyć, że przyjęta formuła ewaluacji rozproszyła kontrolę nad procesem ewaluacji. Przykładowo, rola przewodniczących zespołów ewaluacji nie była na tyle znacząca, by mogli oni efektywnie panować nad wszystkimi etapami decyzji eksperckich (ich rola została zmarginalizowana zwłaszcza na etapie odwołań). Sytuacja taka stwarza zagrożenie pojawieniem się rozbieżności w ocenach, a więc obniżeniem rzetelności, ale jednocześnie zabezpiecza przed systematycznym wypaczaniem ocen, a więc jest swoistą gwarancją trafności. W efekcie wytworzyło się też kilka odrębnych perspektyw na ewaluację – członkowie KEJN nie zawsze widzą problemy, z którymi stykają się członkowie zespołów dokonujących oceny, i odwrotnie. Inną perspektywę mają też urzędnicy MNiSW, a jeszcze inną podmioty wspierające ewaluację informatycznie. Taka sytuacja wydaje się skłaniać do wymiany uwag i spostrzeżeń, a także bardziej generalnej dyskusji założeń systemu kompleksowej oceny.

Chociaż na potrzeby parametryzacji i kategoryzacji już teraz w proces decyzyjny włączeni zostali liczni przedstawiciele różnych dyscyplin naukowych, to wydaje się, że powinien zostać dokonany kolejny krok ku „decentralizacji” procedur, w ramach którego pozwolono by na większą samorządność w poszczególnych GWO. Komisje KEJN mogłyby powołać zespoły doradcze, które rekomendowałyby konkretne rozwiązania „decentralizacyjne”. Nic nie stoi na przeszkodzie zastosowaniu reguły subsydiarności

również w odniesieniu do jednostek naukowych i ich grup, zwłaszcza że najważniejsze instrumenty służące sterowaniu realizacją polityki publicznej pozostałyby w rękach wyższych instancji (MNiSW, KEJN) – chodzi szczególnie o ustalanie kryteriów, zasady przyznawania kategorii oraz jej finansowe konsekwencje. To tu przecież tkwią i będą tkwiły podstawowe narzędzia do kształtowania preferowanego modelu nauki.

Bardzo ostrożnie należy podejść do kwestii zmian zasadniczych reguł parametryzacji. Jak pisaliśmy na początku tego tekstu, stabilność reguł gry jest niezbędna z punktu widzenia sensowności całego przedsięwzięcia. Ostrożnie trzeba podejść zwłaszcza do propozycji wprowadzenia miar opartych na „surowych” cytowaniach. Zwłaszcza w NHS zasady cytowania rządzą się nieco innymi prawidłami niż w pozostałych GN. W dziedzinach, gdzie nie ma klarownej definicji odkrycia, cytowania są miarą obciążoną wpływem wielu innych czynników.

Wszystkie poczynione tu uwagi okażą się jednak małej wagi, jeśli małe znaczenie będą miały także przyznawane jednostkom kategorie. Poziom finansowania badań naukowych jest w Polsce ciągle bardzo niski, co ilustrują wszystkie zestawienia międzynarodowe. Rekomendowany poziom 3% PKB na badanie i rozwój wydaje się czystą abstrakcją, ale również wydatki publiczne na naukę na tle budżetu państwa nie wyglądają imponująco (w 2012 roku było to 1,6%). W sytuacji dysponowania szczupłymi środkami wypada szczególnie zadbać o ich dobrą alokację. Oparcie się wyłącznie na nauce, będącej koproduktem działalności uczelni wyższych, a także na realizacji grantów badawczych, które nie dają gwarancji stabilnej działalności, z całą pewnością nie doprowadzi do optymalnych efektów działalności naukowej. Potrzebny jest silny strumień, który zagwarantuje stabilność, ale zarazem powinien być to strumień jednoznacznie „oparty” na zasadach merytokratycznych. Obecnie obie części tego postulatu realizowane są w zbyt małym stopniu, zatem rekomendować należy dwa kierunki zmian.

Pierwszy dotyczy podniesienia trafności kompleksowej oceny, zwłaszcza przez eliminację luk istniejących w obecnym systemie ewaluacji, uproszczenie ankiety składanej przez jednostki oraz odpowiedni dobór kryteriów oceny do konkretnych grup dyscyplin naukowych. Jeśli kategoryzacja ma się cieszyć wysokim poziomem środowiskowej legitymizacji, musi być przez wszystkie grupy nauk uznawana za sprawiedliwą – a przede wszystkim odporną na manipulacje. Z tego względu priorytetem powinno być właściwe wykorzystanie *peer review* – w szczególności w odniesieniu do monografii.

Drugi kierunek zmian dotyczy rekonstrukcji strumieni finansowania, w efekcie której doszłoby do silniejszego powiązania środków na naukę z jakością prowadzonych badań. Obecnie finansowe konsekwencje przyznania takiej, a nie innej kategorii są dla większości JN względnie niewielkie, często ważniejszy okazuje się współczynnik kosztowności (którego wyznaczeniem kierują zupełnie inne przesłanki). Ze względu na sam algorytm dystrybucji środków pomiędzy jednostki uczelniane stworzona została

sytuacja, w której naczelnym zadaniem uniwersytetu jest dydaktyka, a dopiero dodatkowym badania naukowe. Prawodawcy muszą mieć na względzie, że w warunkach systemu wolnorynkowego to sposób finansowania definiuje cele instytucji (kierownicy jednostek dbają w pierwszym rzędzie o płynność finansową), a pośrednio i pojedynczych aktorów – pracowników naukowych i naukowo-dydaktycznych. „Jeśli (...) alokacja środków na dydaktykę będzie znacząco wyższa od alokacji na badania, uczelnia będzie miała bodźce do uczenia większej liczby studentów, co może doprowadzić do obniżenia jakości prac naukowych” (Bukowska, Łopaciuk-Gonczaryk 2013: 77). Wydaje się, że uniwersytet powinien być niezależnym stowarzyszeniem badaczy, którzy przy okazji swojej pracy naukowej wychowują kolejne pokolenia polskiej inteligencji. Obecnie uniwersytet pełni raczej funkcję wyższej szkoły zawodowej, która przy okazji realizacji zadań dydaktycznych nakazuje swoim pracownikom prowadzić badania. Motywacją pracowników naukowo-dydaktycznych jest obecnie nie tyle status finansowy jednostki, co potrzeba zrobienia kolejnego stopnia naukowego. Ich bazowe wynagrodzenie pochodzi z dotacji na kształcenie studentów, a środki na badania statutowe stanowią jedynie „podręczny” budżet kierowników katedr. W takich warunkach możliwość prowadzenia sensownych programów badawczych, z niezbędną do tego kadrami, zależy bezpośrednio od liczby przyjętych studentów. Zadania dydaktyczne mają ewidentny priorytet w stosunku do zadań naukowych. W efekcie środki ulegają rozproszeniu i powstaje gęsta sieć placówek wiążących prowadzenie „dochodowych” kierunków z robionymi często *ad hoc* badaniami.

Powyższy problem można rozwiązać poprzez wzmocnienie strumienia skierowanego na finansowanie badań statutowych (i utrzymanie potencjału) kosztem strumienia na szkolnictwo wyższe, a także ściślejsze uzależnienie tego pierwszego od wyników parametryzacji. Jakość prowadzonych przez jednostki badań powinna decydować o większej części nakładów państwa na naukę, niż o jednej z każdych sześciu złotych, jak to się dzieje obecnie. Merytokratyczny mechanizm musi zostać oparty również na silniejszej gradacji finansowych skutków przyznanej kategorii. Z punktu widzenia rozwoju nauki polskiej właściwsze, bo motywujące do bardziej wyężonych starań, byłoby na przykład wprowadzenie przynajmniej pięciu kategorii i powiązanie ich z większą rozpiętością przelicznika finansowego (nie od 40% do 150%, ale np. od 30% do 200%, przy bardziej równomiernym rozkładzie jednostek między kategorie). Dopiero spełnienie wspomnianych warunków sprawi, że o kondycji finansowej jednostki naukowej decydowała będzie waga i sprawność prowadzonych programów badawczych, a nie skuteczność naboru na studia. Będzie to tym samym istotny krok na drodze ku reinstytucjonalizacji polskiej nauki (zob. Kwiek 2012). Ta reinstytucjonalizacja wymagać też będzie od osób odpowiedzialnych za następną ewaluację przygotowania, jak to ujmuje Antoni Sułek (2011), „preambuły” do kolejnego ewaluacyjnego Rozporządzenia MNiSW, w której jasno wyłożone zostaną zasadnicze cele ewaluacji i model nauki leżący u jej podstaw.

Przywoływane dokumenty

- Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 5 listopada 2010 r. w sprawie kryteriów i trybu przyznawania oraz rozliczania środków finansowych na naukę na finansowanie działalności statutowej (Dz. U. z 2010 r., Nr 218, Poz. 1438).
- Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 8 lutego 2012 r. w sprawie sposobu podziału dotacji z budżetu państwa dla uczelni publicznych i niepublicznych (Dz. U. z 2012 r., Nr 0, Poz. 202).
- Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 13 lipca 2012 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym (Dz. U. z 2012 r., Nr 0, Poz. 877).
- Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 lutego 2013 r. zmieniające rozporządzenie w sprawie sposobu podziału dotacji z budżetu państwa dla uczelni publicznych i niepublicznych (Dz. U. z 2013 r., Nr 0, Poz. 273).
- Strategia rozwoju nauki do 2015 roku, projekt dokumentu strategicznego, MNiSW, Warszawa, kwiecień 2008 r.
- Ustawa z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (Dz. U. z 2005 r., Nr 164, Poz. 1365).
- Ustawa z dnia 30 kwietnia 2010 r. o zasadach finansowania nauki (Dz. U. z 2010 r., Nr 96, Poz. 615).

Przywoływana literatura

- Abramo G., D'Angelo C.A. (2009) *A Decision Support System for Public Research Organizations Participating in National Research Assessment Exercises*, Journal of the American Society for Information Science and Technology 60(10): 2095-2106.
- Aksnes D.W., Taxt R.E. (2004) *Peer reviews and bibliometric indicators: a comparative study at a Norwegian university*, Research Evaluation 13(1): 33-41.
- Allik J. (2013) *Factors affecting bibliometric indicators of scientific quality*, Trames 17 (3): 199-214.
- Antonowicz D., Brzeziński J.M. (2013) *Doświadczenia parametryzacji jednostek naukowych z obszaru nauk humanistycznych i społecznych 2013 – z myślą o parametryzacji 2017*, Nauka 4: 51-85.
- Brzeziński J.M. (2013) *Ocena parametryczna – co dobre, co zmienić*, Forum Akademickie 11/13: 30-33.
- Bukowska G., Łopaciuk-Gonczaryk B. (2013) *Determinanty sukcesów publikacyjnych naukowców*, Nauka 3: 59-86.
- Cieśliński J. (2014), *Pracownicy naukowci uczelni wyższych od lat prowadzą badania naukowe... za darmo*, Gazeta Wyborcza (wydanie internetowe z dnia 16.01.2014).
- Cole S., Cole J.R., Simon G. (1981) *Chance and Consensus in Peer Review*, Science, 214 (4523): 881-886.
- Kwiek M. (2012) *Changing higher education policies: From the deinstitutionalization to the re-institutionalization of the research mission in Polish universities*, Science and Public Policy 39: 641-654.
- Lewicka M. (2013) *Parametryzacja 2013 – co, jak i jakie wnioski na przyszłość?* (pełna, niepublikowana wersja tekstu, którego część ukazała się jako *Czy oceny w nauce są konieczne* w czasopiśmie uczelnianym Uniwersytet Warszawski. Pismo uczelni 6(66), grudzień 2013).

- Moed H.F. (2007) *The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review*, Science and Public Policy 34 (8): 575-583.
- Preston C.C., Colman A.M. (2000) *Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences*, Acta Psychologica 104: 1-15.
- Sulek A. (2011) *Kilka uwag do (roboczego) Projektu szczegółowych parametrów i kryteriów oceny jednostek naukowych etc.*, maszynopis.
- Tourangeau R., Rips L.J., Rasinski K. (2000) *The Psychology of Survey Response*, Cambridge: Cambridge University Press.
- Van den Besselaar P., Leydesdorff L. (2007) *Past performance as predictor of successful grant applications*, SciSA rapport 0704, Haga: Rathenau Instituut.
- Życzkowski K., Zakrzewski J. (2012) *Pełzający potwór bibliometrii*, PAUza Akademicka, nr 184: 2-3.

**Parameterization and categorization of scientific units in 2013 (in Poland)
as an evaluative practice and institutional process:
The case of humanities and social sciences**

The article aims to review and discuss the 2013-14 evaluation of scientific units in Poland in two perspectives. As an evaluative practice the currently concluded parameterization has had its advantages and drawbacks. Since there are reasons to believe that this practice will become a permanent element of scientific policy, all the deficiencies and especially some crucial loopholes should be eliminated. We try to indicate the latter and give some guidelines on how this could be done. At the same time the evaluation of scientific units can be seen as a part of a broader scientific policy of Polish state, and apart from the problems of reliability and validity, the issues of legitimacy, autonomy and self-direction are at play. We claim that some more general considerations concerning public funding of science are needed if the further reinstitutionalization of Polish science (see Kwiek 2012) is to be achieved.

Key words: parametrization, evaluation, reform of science in Poland, funding of science, Committee for Evaluation of Scientific Units

