

## ESEJ RECENZYJNY

ŁUKASZ REMISIEWICZ 

Uniwersytet Gdański

### „ZOBIEKTYWIZOWAĆ” ARBITRALNOŚĆ

David Beer, *Metric power*, Basingstoke: Palgrave Macmillan, 2016, ss. 223.

Gemma Derrick, *The evaluators' eye: Impact assessment and academic peer review*. Cham: Springer International Publishing, 2018, ss. 230.

### WPROWADZENIE

Współcześnie możemy w Polsce zauważyć daleko idące zmiany dotyczące nie tylko sposobów, ale i podejścia do oceniania dorobku naukowego, zarówno gdy mowa o procedurach awansowych, jak i dystrybucji instytucjonalnych gratyfikacji [Antonowicz 2015; Kulczycki 2017; Kwiek 2015]. Istnieje napięcie między działaniami legislacyjnymi a przekonaniem środowisk naukowych – ideały nauki i oceniania na poziomach politycznym, instytucjonalnym i naukowym pozostają ze sobą w dialogu, ale także oddziałują na siebie; każdy z nich może wywołać reakcję na innym poziomie przekonania i dyskursu.

Jednym z istotnych wątków toczących się debat jest antagonizm pomiędzy obiektywizmem i subiektywizmem oceniania. Zwolennicy podejścia obiektywistycznego opowiadają się za możliwie szerokim rozpowszechnieniem kwantyzacji w ewaluacji dorobku naukowego, która w założeniu ma umożliwić porównywanie go na podstawie kryteriów niezależnych od decyzji lokalnych środowisk. Celem

---

Mgr, Instytut Socjologii, Zakład Teorii Socjologicznej i Metodologii Badań Społecznych;  
e-mail: [lukasz.remisiewicz@ug.edu.pl](mailto:lukasz.remisiewicz@ug.edu.pl); <https://orcid.org/0000-0001-7710-6880>

tego procesu jest zapewnienie jasnych i przejrzystych warunków awansów naukowych, instytucjonalnych oraz przyznawania gratyfikacji, a co za tym idzie pobudzanie również zdrowej rywalizacji o fundusze na badania czy stanowiska naukowe według jasno określonych zasad. Jak pisze Emanuel Kulczycki: „Obecnie główną zasadę organizacji i zarządzania nauką oraz szkolnictwem wyższym można scharakteryzować następująco: wszystko ma być możliwe do zmierzenia i oceniania, a na tej podstawie mają być podejmowane skuteczne i – na ile to możliwe – zobiektywizowane decyzje dotyczące organizacji i finansowania nauki” [Kulczycki 2017: 63–64].

Inni widzą jednak w takim podejściu uosobienie wyrachowanego neoliberalizmu, sprowadzającego naukę do postaci gry, opartej na prostych mechanizmach zdobywania punktów mających przełożenie na kolejne osiągnięcia, lub wymiany tych punktów na inne dobra. Zauważają także szereg negatywnych efektów ubocznych, z których podstawowym jest utrata z pola widzenia rzeczywistego celu nauki, jakim jest poszukiwanie prawdy, lub choćby wyjaśnienie najbardziej adekwatnych. Poszukiwanie to, szczególnie w humanistyce i naukach społecznych, nie jest – argumentują – łatwe do opakowania w popularne dziś ramy projektu, określającego już na samym początku drogi cele, metody i spodziewane rezultaty.

W niniejszym eseju będę omawiał dwie książki, które odwołują się do tego dylematu, choć w sposób dalece nieoczywisty. Przedstawię argumenty na rzecz tezy, że procesy obiektywizacyjne, które już zachodzą w nauce, a postuluje się ich dalsze rozszerzenie, mają zasadniczy wpływ na sam charakter uprawiania tej dziedziny poznania, gdyż nie tyle mierzą stan zastany, ile projektują swój obiekt, przeobrażają go. Tak jak testy standaryzowane nie są jedynie „obiektywnym miernikiem wiedzy uczniów”, ale kulturowym narzędziem odciskającym piętno na wielu poziomach społecznej organizacji [Remisiewicz 2019b], tak i działania wprowadzające pomiar do nauki nie są jedynie „obiektywnym pomiarem osiągnięć naukowych”. Wszystkie działania związane z pomiarem zachowań osób będących świadomymi tego pomiaru rodzą przynajmniej częściowo nieprzewidziane konsekwencje, a ich nawarstwienie zmienia sam system. David Beer, autor pierwszej z omawianych książek, nazywa to „władzą miar” [Beer 2016].

Oczywiście reformatorzy nie ukrywają, że właśnie o zmianę pewnych praktyk im chodzi – polska debata koncentrowała się na przykład na pojęciu umiędzynarodowienia nauki, które przejawiać się miało w znalezieniu sposobu na zwiększenie wkładu polskich naukowców do nauki światowej, w tym radykalny wzrost liczby ich publikacji w zagranicznych czasopiśmie i wydawnictwach. Z tego powodu uznano, że nadanie wyższej rangi czasopiśmom zagranicznym zwiększy zainteresowanie nimi wśród polskich badaczy [Kulczycki 2017].

Jednak odgórnie narzucone kryteria są jedynie instrumentem przeznaczonym do stosowania przez określone osoby lub grupy w ramach konkretnych procesów ocennych i decyzyjnych. Wyróżnić można *peer-reviews*, czyli oceny „koleżeńskie”, w przypadku których zakłada się równość kompetencji pomiędzy ocenianymi i oceniającymi, a także *expert-driven peer reviews*, czyli oceny eksperckie – gdy zakłada się hierarchiczną wyższość osób oceniających (ze względu na ich wiedzę) nad ocenianymi.

Podkreślić należy, że nawet procedury oceniania wykorzystujące docelowo najbardziej obiektywne miary do oceny jakości wymagają na jakimś etapie obecności osoby oceniającej, co oznacza, że w procedurze oceniania zawsze pojawiają się czynniki nazywane przez etnometodologów „prowizorkami” [Garfinkel 2007]. Nawet najdokładniejsze wytyczne oceny, jeśli mają być zastosowane w praktyce, wymagają włączenia czynnika subiektywnego, bo konieczne jest odniesienie ich do innego rodzaju jakości. Na przykład w niektórych grantach wymagane jest, by na podstawie liczby publikacji oraz wskaźnika IF ocenić w skali 1–5 jakość dorobku wnioskodawcy. Okazuje się, że nawet w przypadku tak prostego zadania recenzenci pracujący niezależnie mogą dokonać zróżnicowanej oceny, mając do dyspozycji ten sam zestaw danych. Zawodzi więc mechanizm „translacji” tych danych na język innej kategorii.

Druga omawiana książka jest studium przypadku w sposób pośredni dokumentującym rzeczywiste procesy zachodzące pomiędzy ewaluatorami podczas sesji oceniania projektów w ramach ustalonych wcześniej kryteriów. Publikacja ta jednak nie tylko pokazuje te procesy, ale także – w części teoretycznej – zapewnia możliwość prześledzenia procesów formowania tych kryteriów w ramach polityk naukowych, co pozwala odnieść część wniosków do pierwszej z książek.

Celem niniejszego eseju jest nie tylko podsumowanie treści książek i dokonanie ich oceny, ale przede wszystkim przeanalizowanie wniosków autorów w odniesieniu do polskiej rzeczywistości. Nie jest tajemnicą, że przemiany w schematach ewaluacji nauki na różnych poziomach inspirowane są wzorcami zaczerpniętymi z innych światowych systemów.

## TEORIA WŁADZY MIAR

Książka *Metric power* autorstwa Davida Beera ma charakter teoretyczny i traktuje o miarach w sposób możliwie ogólny. Ufundowana jest na paradygmacie analizy zapoczątkowanym przez Michela Foucaulta, w szczególności zaś tych aspektów jego koncepcji, które dotyczą rządomyślności (*governmentality*). Foucault rozumie rządomyślność jako anonimowe relacje władzy, na które składają się

rozmaite społeczne technologie, za pomocą których można (zwłaszcza pośrednio) kierować postępowaniem ludzi. Wśród tematów, które porusza Beer, można wymienić takie, jak historia statystyki pojmowanej jako narzędzie oddziaływania na procesy społeczne, biometria, którą autor łączy z władzą (idąc tropem pojęcia biowładzy Foucaulta), *data mining* czy ruch *Quantified Self*.

Biometria stanowi w najogólniejszym znaczeniu pomiar parametrów biologicznych człowieka. Współcześnie często pisze się o niej w kontekście ustanawiania nowych rodzajów zabezpieczeń (np. skanowanie oka lub linii papilarnych). Jednak dla Foucaulta pomiary ciała zawsze będą uwikłane w konteksty normatywne, wiążące wiedzę z władzą, w konteksty rasy, płci, niepełnosprawności czy wieku.

W tym kontekście warto również wspomnieć o *data mining*, czyli procesie eksplorowania dużych zasobów danych. Współcześnie każdy, kto korzysta z Internetu, wytwarza dane – są one zbierane, agregowane, selekcyjonowane, przetwarzane, łączone, katalogowane przez wiele systemów, o których istnieniu nie każdy wie, a do których architektury dostęp mają bardzo nieliczni. Tymczasem to właśnie działanie tychże algorytmów prowadzi na przykład do wytwarzania różnorodnych zjawisk kulturowych, by wymienić chociażby bańki informacyjne. Powstają one za sprawą selekcyjonowania treści wyświetlanych na portalach społecznościowych w następstwie rozpoznawania preferencji użytkownika.

Równie interesującym przykładem zastosowania systemów numerycznych jest ruch *Quantified Self*, którego uczestnicy za pomocą nowoczesnych technologii (urządzeń, aplikacji) dokonują systematycznych pomiarów swoich aktywności. Taki monitoring ma w założeniu pomóc w podnoszeniu swojej wydajności, zarządzaniu czasem oraz kontrolowaniu parametrów związanych ze zdrowiem.

Beer stawia sobie bardzo ambitne zadanie – znalezienia adekwatnego języka do opisu wpływu społecznego, jaki posiadają systemy numeryczne, których znaczenie we współczesnym świecie wciąż wzrasta na bardzo różnych poziomach i w wielu aspektach życia społecznego.

W pierwszym rozdziale autor wskazuje na historyczne korzenie rządzenia społeczeństwem poprzez gromadzenie danych statystycznych, upatrując początków tych praktyk w XVIII i XIX wieku. W tym kontekście wspomina między innymi o budowaniu infrastruktury archiwów, które pozwalają na tworzenie strategii państwowych służących do „opanowywania” populacji poprzez jej kategoryzowanie i wyznaczanie norm. Kategorie wytwarzane przez uprawnione ciała – osoby, grupy lub instytucje – stają się elementem władzy. Będąc częścią statystycznej struktury, jednostki podlegają procesom kontrolowania „widzialności” i „niewidzialności”, które kształtują się w odniesieniu do kategorii „po-

liczalności” i „niepoliczalności”. To, co niemożliwe do skwantyfikowania, jest uznawane za nieistniejące. Istnieć to znaczy być policzalnym, a mówiąc nieco mniej efektownie – być reprezentowanym w ramach systemów numerycznych. Niezwykle istotne jest tutaj również wprowadzenie koncepcji pomiaru obiektywnego, która staje się dyskursywnym narzędziem legitymizującym proces wytwarzania i kategoryzowania danych.

W drugim rozdziale *Metric power* Beer skupia się na pojęciu cyrkulacji miar, a jego celem jest wskazanie mechanizmów, w ramach których powstaje wartość. Koncentruje się więc na pojęciach opowieści i narracji, które w pewnym sensie są wytwarzane dzięki danym, ponieważ dane i ich akumulacja stanowią podstawy narracji, choć dużo ważniejsze są procesy: sortowania, filtrowania i ukierunkowania strumieni danych, ponieważ to właśnie one stanowią główny oręż władzy.

W rozdziale dotyczącym możliwości Beer koncentruje się na pokazaniu, w jaki sposób miary wyznaczają granice działania, co jest istotne w szczególności z perspektywy badań nierówności. Zwraca przy tym uwagę na ważną kwestię – otoczenie miar nimbem obiektywności zamyka usta krytykom i redukuje możliwe obiekcje. Tymczasem często trudno zauważyć oraz skorygować, gdy inni w sposób niewłaściwy posługują się miarami, ponieważ nawet najistotniejsze problemy z zakresu porównywania obiektów różnych typów miar są deprecjonowane. Ten problem nie jest obcy nauce polskiej, na co w swojej rozprawie doktorskiej zwracała uwagę Ewa Rozkosz [2018]. Jej badania dotyczące recenzji w postępowaniach habilitacyjnych pokazywały rozmaite błędy, jakie popełniali recenzenci, posługując się standaryzowanymi miarami opisującymi dorobek habilitantów, co przynajmniej w pewnych przypadkach mogło mieć wpływ na wynik recenzji. W omawianym rozdziale Beer pisze także o współczesnych tendencjach do samodoskonalenia się dzięki wykorzystaniu różnego typu mechanizmów kwantyfikujących własną aktywność i usprawniających własny rozwój. Na przykład naukowcy mogą wykorzystać narzędzia, które mają ich zmotywować do cięższej, bardziej wyteżonej pracy. Wśród nich są między innymi programy pisarskie posiadające funkcjonalności umożliwiające ustawienie sobie „wyzwań” do wykonania – na przykład osiągnięcie dziennego limitu słów. Z drugiej strony istnieją także programy, które mają w sposób pośredni pomóc przyspieszyć pracę naukową – na przykład sprawić, żeby miała ona bardziej płynny charakter. Można tutaj wymienić między innymi menadżery cytowań czy programy wspomagające proces pisania. Analiza praktyk wykorzystania tego typu narzędzi zahacza o tematykę złożonych zależności związanych z miarami produktywności z jednej strony i organizacją własnego czasu w sposób maksymalnie efektywny z drugiej.

Miary, które wytwarzają systemy, mogą służyć jako pewne narzędzia klasyfikacyjne. Niektórzy uważają, że powinny być one podobne do tabeli ligowej, bowiem to tabela ujawnia najlepszych i najgorszych, a także – co być może najważniejsze – tych, którzy są po środku. A według Beera to właśnie osoby ze środkowych miejsc w tabeli stanowią główny punkt odniesienia, ponieważ wyznaczają normy. David Beer idzie tutaj tropem Foucaulta, który również uważał, że podstawą dla miary jest rozkład normalny, tworzący swoisty balans. Dopiero znając normę, jesteśmy w stanie ustalić, czego jest zbyt mało, a czego za dużo, kto jest lepszy, a kto gorszy. Miary pozwalają więc w pewien sposób określić parametry życia społecznego.

Struktury rankingów wytwarzają niepewność, ale niepewność ta jest konieczna w systemie neoliberalnym, ponieważ napędza ona rywalizację. W tym znaczeniu nie można jej traktować jako jedynie negatywny efekt uboczny, ale raczej jako element inherentnie wpisany w cały system, a nawet jego siłę napędową. W tym kontekście trudno mówić o tym, że koincydencja czasowa wprowadzenia systemu oceny pracy naukowców opartego na miarach i problemu prekaryzacji pracy młodych naukowców [Rykiel 2014] jest zupełnie przypadkowa.

Powołując się na Wendy Espeland i Michaela Saudera [2007], a także Wendy Espeland i Mitchella Stevensa [2008], Beer wskazuje, że miary są rodzajem aktów komunikacji, z przypisanymi im znaczeniami i funkcjami. Prowokują one jednostki do emocjonalnego zaangażowania w praktyki mierzenia i bycia mierzonym. W przywołanych pozycjach wprowadzona została koncepcja ucieleśnienia, zgodnie z którą ciało, w tym przede wszystkim procesom neuronalnym, przypisuje się rolę pasa transmisyjnego w procesie dopasowywania się do miar.

System miar może budzić różne wątpliwości, dlatego jego pełne zinternalizowanie wymaga odpowiedniego treningu, by nauczyć się funkcjonować według miar, rozumieć te miary, a także swoje miejsce w ramach systemu oraz poznać schematy uzyskiwania poprawnych efektów, które zwiększają szansę na zdobywanie wyższych miejsc w rankingach, będących efektem wprowadzenia tego systemu.

Miary nie tylko zmuszają do pewnych działań, ale także zakładają, że będziemy próbowali przewidzieć to, w jaki sposób systemy te zadziałają, biorąc pod uwagę wykonywane przez nas czynności. Dopiero tego rodzaju antycypacja prowokuje nas do pewnych działań [Sauder, Espeland 2009]. Tym samym niektórzy naukowcy próbują ujmować swoją własną pracę w poetyce gry, rozważając, która strategia pomoże im wydatnie wzmocnić swoje szanse w tym wyścigu. W związku z tym dostosowują na przykład wzorce publikowania do schematu, który zapewnia im, według ich rozeznania, jak najlepsze efekty [Kulczycki 2017].

Z zakończeniu książki Beer podkreślił, że mamy do czynienia z teorią jeszcze nieukończoną, wskazując jednocześnie na nowe, aktualne wątki, które powinny stać się obiektem zainteresowania socjologów. Skupił się na tym, jak miary wpływają na społeczne życie codzienne oraz na życie psychiczne jednostek. Chodzi tutaj o reaktywność jednostek na proces pomiaru, o to w jaki sposób zmieniają one swoje zachowania oraz stan emocjonalny w przypadku bycia obiektem pomiaru. Autor wymienia tu przede wszystkim stany emocjonalne związane z lękiem oraz poczucie niepewności. To właśnie stany emocjonalne, a także, w dalszej kolejności, wywoływane przez nie myśli i odczucia są podstawą tworzącą wzorce reaktywności.

Dobre podsumowanie poglądów autora może stanowić cytata z wprowadzenia książki: „Jesteśmy tworzeni i przetwarzani przez miary, żyjemy poprzez nie, z nimi i wewnątrz nich. Miary ułatwiają formułowanie i przeformułowywanie sądów na nasz temat; sądy które wydajemy o innych, a także ich konsekwencje są odczuwane i doświadczane w ten sposób w naszym życiu” [Beer 2016: 3].

Książka Beera stanowi niezwykle istotny wkład w rozwój badań nad systemami numerycznymi, zawiera także podsumowanie koncepcji innych autorów, których prace były dla niego inspiracją. Aktualność publikacji nie budzi dziś żadnych wątpliwości, a wręcz wydaje się, że będzie jeszcze rosła. Tym samym fakt, że mamy do czynienia z dziełem otwartym daje nadzieję na to, że sam autor pozostanie „reaktywny”, co być może pozwoli mu w przyszłości, kiedy zajdą kolejne zmiany, na podjęcie nowych wątków i uzupełnianie teorii o nowe aspekty lub zastosowania.

## WŁADZA NAD MIARAMI – PERCEPCJA MIAR

Doceniając omawianą pracę, warto jednakże zwrócić uwagę na jej pewne niedoskonałości. Wadą książki jest moim zdaniem to, że przedstawia ona reaktywność na miary w poetyce bezwładności, to znaczy nie pokazuje żadnych mechanizmów, które świadczyłyby o subwersywności czy możliwych konfliktach w związku z aktywnym przeciwstawianiem się pomiarom. Tymczasem właśnie w Polsce istnieje wiele środowisk, które próbują negocjować wprowadzanie miar i sposobów ewaluacji. W momencie gdy piszę ten tekst, nauka polska jest po okresie burzliwych debat dotyczących ewaluacji nauki [Antonowicz 2015; Dziedziczak-Foltyn 2017; Kwiek 2015; Ostrowicka i in. 2019; Stankiewicz 2018], które jednakowoż nie doprowadziły do środowiskowego konsensusu. Istnieją wszakże naukowcy, którzy dążą do wyzwolenia się spod reżimu wprowadzonych miar; tacy, którzy próbują różnymi środkami realizować własne cele, a nawet



niejako „oszukać” system, wiedząc, w jaki sposób on działa. Oczywiście można powiedzieć, że jest to również rodzaj reaktywności, ponieważ działania te mimo wszystko pozostają w związku z systemem ewaluacji, ale należy pamiętać o tym, że jest to reaktywność innego rodzaju niż ta, o której pisze David Beer. Odwołuje się bowiem do środowiskowych reguł interakcyjnych, które pozostają z miarami w pewnych zależnościach, a nie do samych miar.

Jedyny jak dotąd polski recenzent książki Beera wskazywał na nieobecność w niej wątków ekonomicznych, a w szczególności brak odniesienia do problemu kapitalizmu, którego mechanizmy niejako wymuszają procesy standaryzacji (np. produkcji), a tym samym stanowią siłę napędową wytwarzania kolejnych systemów miar [Krzeski 2018]. Neoliberalizm jako taki pojawia się w książce często, ale raczej – jak twierdzi recenzent – na poziomie narracji czy dyskursu, gdyż nie przedstawiono badań bardziej bezpośrednich efektów jego oddziaływania w świecie [Krzeski 2018]. Wydaje mi się, że podobnie rzecz się ma z obecnością szeroko pojętych polityk, co w szczególności widać, gdy rozważa się funkcjonowanie miar w nauce. Wprowadzenie miar jako obowiązujących jest każdorazowo decyzją polityczną, a ich oddziaływanie nie jest wcale bezpośrednie, ale zależy od tego, jakie praktyki zostaną przyjęte na różnych etapach ich stosowania. Tym samym naukowcy oczywiście w pewnym sensie są „reaktywni” i istnieją sposoby, by uchwycić zmiany praktyk publikacyjnych po wprowadzeniu określonych miar. Ale reaktywność ta nie jest, jak się wydaje, związana z samą tylko świadomością bycia mierzonym, ale raczej ze świadomością, jak miary te będą wykorzystywane w procesach rzeczywiście decydujących o kluczowych aspektach kariery naukowej. Obiektem zainteresowania nie jest więc sama miara, ale praktyki środowiskowe dotyczące jej wykorzystania. Te zaś mogą być bardzo zróżnicowane, co ma swoją przyczynę w tradycjach, przyzwyczajeniach i nieformalnych rozwiązaniach przyjętych wcześniej, a zmienianych dopiero w następstwie czynników zewnętrznych. W książce Beera brakuje mi tej dbałości o kontekst, szczegól i powiązania reaktywności nie tylko z makrohistorycznymi zmianami, ale także tymi lokalnymi uwarunkowaniami, które mają największy i najbardziej bezpośredni wpływ na konkretne zachowania.

## O POTRZEBIE PERSPEKTYWY INTERAKCYJNEJ

Beer pisze pod koniec, że należy zwrócić uwagę na „praktyki pomiaru praktyk”, a także odwołuje się do koncepcji zakorzenienia według schematu mózg–ciało–świat. Wydaje się, że w przypadku nauki ten trop jest najbardziej zasadny. To właśnie praktyki pomiaru, czyli, biorąc to zagadnienie dosłownie: wszystko to,



co środowiska decyzyjne robią z miarami, jak je przekształcają, reinterpretują, teoretyzują, a także to, czy dzieje się to według reguł, które czynią te procesy przewidywalnymi, jest w głównej mierze czynnikiem, na który reagują zainteresowani. Reguły, które nie mają zastosowania i miary, które nie są użytkowane, nie stanowią obiektu zainteresowania. Pomiar jest tylko jednym z etapów, równie istotne są praktyki przeobrażenia w miary jakościowe, ale także struktura zysków i strat, które można ponieść, na podstawie wyników pomiaru. O strukturze korzyści i strat decydują zwykle polityki naukowe wyznaczające horyzont instytucjonalnych gratyfikacji oraz zakres odpowiedzialności, natomiast środowisko naukowe, w ramach różnorodnych mechanizmów interakcyjnych, decyduje o transformowaniu wyników pomiaru w decyzje.

Na przykład w ogólnym sensie można zgodzić się z Beerem, że założenia neoliberalizmu wspierają instytucjonalne procesy wytwarzania i stosowania miar, a przykład polskiej nauki dobrze to pokazuje. Wraz z przemianami neoliberalnymi zmieniły się kryteria ocen, które uzasadniane były w retoryce desubiektywizacji [Antonowicz 2015]. Niemniej użycie socjologicznej lupy, która ukazywałaby interakcyjny charakter abstrakcyjnych, opisywanych zjawisk, uzmysłowiłoby, jak wiele aspektów oceny musi być rozwiązywanych w sposób lokalny.

Rozważmy przykład postępowania habilitacyjnego. Habilitant składa wniosek o wszczęcie postępowania, dołącza do niego pakiet informacji o swoim dorobku naukowym. Już same te informacje dają możliwość pomiaru, jak wiele jest tych publikacji (a biorąc pod uwagę okres, jaki upłynął od nadania stopnia doktora, można zbadać także produktywność), jakiego rodzaju publikacje dominują (krajowe czy zagraniczne? za ile punktów?). Oprócz tego habilitanci zobowiązani są także dostarczyć informacje dotyczące ich wskaźnika Impact Factor oraz liczby cytowań. Wniosek jest oceniany przez trzech recenzentów, którzy mają rozstrzygnąć, czy dorobek naukowy stanowi znaczny wkład do dyscypliny naukowej, w której prowadzone jest postępowanie. Muszą wziąć oni pod uwagę zarówno jakość dorobku, jak i jego aspekty ilościowe. A zatem same miary nie mają tutaj znaczenia – dla decyzji recenzenta kluczowa jest interpretacja miar, waga, jaką nada im przy ocenie całościowej oraz transformacja tych danych w końcowy wynik, którym jest poparcie lub odmowa poparcia wniosku o nadanie stopnia. Jak pokazuje doświadczenie, recenzenci nie zawsze są w tym ze sobą zgodni [Remisiewicz 2019a; Rozkosz 2018]. Kolejnym etapem jest posiedzenie komisji habilitacyjnej, która na podstawie głosów recenzentów oraz innych członków i w następstwie wewnętrznej dyskusji, podczas której członkowie komisji mogą wpływać na siebie bezpośrednio, podejmuje uchwałę o wydaniu opinii w sprawie nadania lub odmowie nadania stopnia. Ostateczna decyzja podejmowana jest na-

tomiast na posiedzeniu kolektywnego organu przeprowadzającego postępowanie. Zwróćmy więc uwagę, jak wiele ogniw występuje w łańcuchu prowadzącym od pomiaru osiągnięć naukowych do ich transformacji w korzyść instytucjonalną.

Moglibyśmy zapytać oczywiście, na ile miary warunkowały decyzje podejmowane na kolejnych etapach procesu opiniowania – poszczególnych recenzentów, komisji oraz rad. O ile w przypadku recenzentów jesteśmy w stanie prześledzić tok rozumowania (choć należy wziąć pod uwagę, że tekst stanowi jedynie rekonstrukcję procesu rozumowania i nie jest z procesami myślowymi tożsamy<sup>1</sup>), o tyle dużo trudniej o to w przypadku niejawnych obrad ciał instytucjonalnych. Pewien wgląd w procesy kolektywnego podejmowania decyzji daje druga książka, która zostanie przedstawiona poniżej.

### KONSTRUUJĄC „OKO EWALUACYJNE”

Druga praca, *The Evaluators' eye: Impact assessment and academic peer review*, jest ukierunkowana na procesy interakcyjne i koncentruje się raczej na konstruowaniu znaczeń, interpretacji doświadczeń, a także na formowaniu kultur ewaluacyjnych [Lamont 2009], które stanowią dla członków grupy oceniającej punkt odniesienia. Sama autorka nie przedstawia własnych preferencji teoretycznych tak zdecydowanie, jak czyni to Beer.

Przedmiotem książki jest ewaluacja REF (*Research Excellence Framework*), która stanowi mechanizm rozporządzania środkami przyznawanymi instytucjom naukowym w Wielkiej Brytanii<sup>2</sup>. Podstawowym założeniem jest tu transparentność publicznej inwestycji w naukę i ukazywanie płynących z niej korzyści, ale także kształtowanie reputacji poszczególnych instytucji.

Obiektem oceny był raport przygotowywany przez daną jednostkę, zawierający szereg informacji na temat zatrudnionych pracowników naukowych, ich

---

<sup>1</sup> Zewnętrzny obserwator nie wie nawet tego na ile, w przypadku konkretnej recenzji, argumentacja przygotowuje grunt pod wydanie decyzji, a na ile jest ona *ex post* dopisana w taki sposób, aby uzasadniała decyzję powziętą na wcześniejszym etapie. Nieliczne prace dotyczące recenzji w postępowaniach awansowych pokazują pracę dyskursywną, jaką wykonują niekiedy recenzenci, aby zminimalizować znaczenie niektórych zarzutów [por. Hyland 2007; Woroniecka 2015].

<sup>2</sup> Zgodnie z Rozporządzeniem Ministra Nauki i Szkolnictwa Wyższego z dnia 22 lutego 2019 r. w sprawie ewaluacji jakości działalności naukowej instytucje naukowe będą oceniane również według kryterium wpływu społecznego, czy mówiąc inaczej: pozanaukowego znaczenia prac naukowych, co stanowi jawny przykład inspiracji systemem brytyjskim, który opisuje autorka [MNiSW 2016]. Kontekst i szczegółowy opis koncepcji wpływu społecznego czytelnik znajdzie w pracy Marty N. Wróblewskiej [2017].

publikacji, danych dotyczących „wpływu społecznego” czy też przyznanych nagród. Ewaluatorami byli naukowcy-specjaliści w danym obszarze oraz przedstawiciele instytucji i przedsiębiorstw wykorzystujących wyniki badań. Na ich sumaryczną ocenę składały się trzy elementy – ocena jakości publikacji pod względem ich oryginalności, znaczenia i rygoru (65%), wpływ społeczny w rozumieniu „zasięgu i znaczenia” badań (20%), a także ocena jakości środowiska naukowego pod względem jego „witalności i trwałości” (15%), w której uwzględnia się także takie elementy jak kształcenie doktorantów czy podejście do zagadnienia równości.

Derrick skupia się na analizie procesów grupowych w panelowej ocenie koleżeńskiej. Recenzowanie grupowe stosowane jest także w polskiej nauce. Tego rodzaju ocena znacząco różni się oczywiście od indywidualnej, niezależnej recenzji. Zakłada się, że praca grupowa pozwoli osiągnąć lepsze efekty, gdy chodzi o ocenę większej liczby projektów w krótszym czasie – oceniający bowiem mogą nawzajem weryfikować swoje sądy oraz wskazywać na aspekty pominięte przez innych. Wymaga to jednak ponadstandardowych wysiłków związanych z kalibracją kryteriów, którymi posługiwać ma się grupa, łagodzenia ewentualnych konfliktów, a także ścisłej współpracy w osiągnięciu powziętego wcześniej celu. Autorka próbowała zbadać, w jaki sposób procesy te zachodzą w rzeczywistości i jaki wpływ mają na efekty pracy zespołów oceniających. Jej praca ma także na celu wskazanie szeregu rekomendacji, które mają upłynnić i usprawnić proces prowadzenia oceny grupowej. Niestety, jak wskazuje autorka, badań o podobnym charakterze jest bardzo niewiele, a mechanizm i niejawnosc samego procesu znacząco utrudnia to zadanie.

Badanie Derrick miało charakter jakościowy. Przeprowadzono dwie serie wywiadów z ewaluatorami. W pierwszej serii, wykonanej przed sesją ewaluacyjną, naukowcy dzielili się swoimi intuicjami dotyczącymi pojęcia wpływu. Druga seria została wykonana po sesji, co umożliwiło uchwycenie, jakie zmiany w konceptualizacji tego pojęcia zaszły za sprawą uczestniczenia w panelu w charakterze oceniającego. Niestety autorka nie zdołała wykonać badania etnograficznego, które uchwyciłoby „na żywo” procesy negocjowania znaczeń, a to prawdopodobnie byłoby najbardziej interesujące. Niemniej za pomocą rozmaitych porównań i dzięki bardzo rygorystycznemu podejściu analitycznemu udało się z tego materiału wyłonić konstrukcję, która może zastępować etnograficzny sposób prowadzenia badania pod względem zapewnianego potencjału informacyjnego.

## UŻYTKOWANIE MIAR – PRAKTYKI OCENIANIA

Wyniki badania są wielowątkowe, streszczę więc tylko te, które uważam za najistotniejsze dla rozumienia procesów ocennych w warunkach interakcji pomiędzy ewaluatorami. Autorka wskazuje na fakt, że pojęcie wpływu społecznego jest w trakcie oceniania operacjonalizowane. Grupa musi skonstruować własne wskaźniki, sposoby rozumienia tego pojęcia, w taki sposób, aby konkretnym projektem przypisać określone wartości. Praca ta ma więc charakter kreatywny – zewnętrzne kryteria są poddawane „obróbce”, w ramach której grupa oceniających ustala adekwatne sposoby rozumienia określonych pojęć. Proces oceny jest więc sterowany z zewnątrz tylko w niewielkim stopniu, biurokraci wytwarzający kryteria nie mają wpływu na to, jakie znaczenia określonym pojęciom nada grupa w konkretnej sytuacji. Warto też pamiętać, że choć punkt odniesienia jest ten sam, w każdym panelu pojęcia te mogą być konstruowane w inny sposób. Zapewnienie więc, że ocena jest dla każdego równa, jest całkowicie iluzoryczne. Zmiana składu grupy mogłaby zmienić kryteria, a tym samym także ostateczne decyzje.

Gdy w grę wchodzi ocenianie na skali, oceniający mogą mieć różne podejście do „punktu wyjścia”. Można rozpocząć od oceny najniższej i stopniowo poszukiwać elementów, które będą ją podwyższały, można także rozpocząć od oceny najwyższej, szukając błędów, które będą ją obniżały. Można także próbować rozpocząć od jakiegoś punktu archimedesowego, a lektura będzie przeważać szalę to w jedną, to w drugą stronę. Wnioskując z opinii recenzentów, w przypadku ewaluacji REF2014 dominowało podejście drugie, co spowodowało, że zdaniem niektórych oceny były zawyżone. Z drugiej jednak strony wśród oceniających może zaistnieć podobny konflikt dotyczący tego, czy oceny powinny układać się podobnie do rozkładu normalnego, czy też należy przyjąć, że nawet duża skośność rozkładu nie stanowi problemu. Na przykład, czy jeżeli wybitny projekt otrzyma najwyższą ocenę, to czy powinna ona być przyznawana jedynie projektom równie wybitnym?

Istnieją także różne podejścia do samego procesu nadawania ocen. Jeden z ekspertów porównuje to raczej do „sygnalizowania jakości” niż rzeczywistej oceny wartości zgłoszenia. Tak jak niekiedy recenzenci muszą transformować wskaźniki ilościowe na język jakościowego opisu lub podejmować na ich podstawie decyzje wyrażone w ramach innych niż ilościowe kategorii, tak w opisywanym przez Gemmę Derrick przypadku to raczej kategorie jakościowe (oparte na „znaczeniu” i „zasięgu” badań) miały być przekładane na język oceny ilościowej. Trudno więc dziwić się temu, że taka ocena miała charakter „umowny” i wymagała pewnego porozumienia dotyczącego tego procesu. Warto zacytować

jedną z respondentek, która wskazuje na skrajny mechanizm przyznawania ocen: „jeśli nie wiesz, co z tym zrobić, przyznaj trzy punkty; jeśli uważasz, że jest to fantastyczne, daj cztery” [Derrick 2018: 191].

Podczas oceniania ewaluatorzy muszą odwoływać się do pojęć, którymi sami posługują się na co dzień, ale często definicje tych pojęć nie pokrywają się z ich rozumieniem przewidzianym w oficjalnych dokumentach będących podstawą ewaluacji. Posługiwanie się przez każdego inną definicją uniemożliwiłoby osiągnięcie porozumienia, dlatego pierwszym krokiem podczas ocen jest wspólne ustalenie zakresu pojęciowego. Tym niemniej w praktyce dochodzi często do niekontrolowanych „powrotów” do własnych definicji. Tego typu procesy mogą też doprowadzić do powstania nowej – grupowej definicji, mającej mniej teoretyczny, a bardziej operacyjny charakter.

Bardzo istotnym wątkiem jest to, kto zabiera głos w trakcie debat. Wkład poszczególnych osób nie jest równy i zależy od wcześniejszych osiągnięć danej osoby, ale także od statusu, jaki ma ona wśród innych członków grupy, a także od jej własnych predyspozycji oraz umiejętności dotyczących dominowania w dyskusji. Autorka zwraca uwagę na trudności, jakie pojawiają się w relacjach pomiędzy praktykami a przedstawicielami środowiska akademickiego. Akademicka kultura dyskusji, w tym reguły zabierania głosu, minimalizowała wpływ praktyków na przebieg rozmowy. W związku z tym można nawet powiedzieć, że w szczególnych wypadkach praktycy stawali się outsiderami.

Dostęp do głosu może wydatnie wpływać na kształtowanie się stosunku całej grupy do określonego tematu. Gdy istnieje możliwość klarownego przedstawienia własnych opinii przez każdego z członków grupy, ustosunkowania się do tych opinii przez innych członków, a także gdy na podejmowanie i zarzucanie kolejnych tematów ma wpływ każdy członek, zapewniona jest równość. W konwersacjach często jednak wyłaniają się osoby dominujące, bardziej pewne siebie – Randall Collins [2011] powiedziałby o nich, że posiadają wyższą energię emocjonalną. Ich dominacja staje się barierą dla innych w zakresie zabierania głosu, ale także wpływa na kierunek dyskusji, stawiania i wartościowania pojawiających się kwestii. Autorka słusznie zauważa, że warto ten problem zbadać szerzej także w odniesieniu do zmiennych, które nie powinny mieć znaczenia przy ocenie – na przykład płci.

W pracy grupowej niezwykle istotne są zagadnienia dotyczące konsensusu. Pojęcie to jednak może wywieść analityków interakcji na manowce, bowiem sugeruje, że ewaluatorzy przyjmują niejako wspólny, całościowy pogląd, razem stanowią „jedno ciało oceniające”. Ten ontologiczny wątek wykorzystują – jak się wydaje – biurokratyczne narracje. Wynik oceny jest „stabilizowany”

i „obiektywizowany *post factum*” jako ocena „ponadjednostkowego ciała” lub, zgodnie z tytułem, „ewaluacyjnego” oka, co może sprawiać wrażenie, że posiada ona większe znaczenie, niż gdyby została dokonana przez jedną osobę. Pomiędzy przy tym fakt, że ocena ta była wynegocjowana i stanowi efekt bardzo niedoskonałych ludzkich procesów poznawczych. Zasadność wprowadzenia oceny grupowej uzasadniana jest tym, że w procesie interakcji poszczególne perspektywy mogą zostać porównane, a poszczególni recenzenci mogą wzajemnie się kontrolować w duchu zachowania jak największej obiektywności. Książka pokazuje, że owo idealistyczne spojrzenie nie zawsze jest prawdziwe.

Jak już wspomniałem, oceniający muszą osiągnąć w grupie pewien konsensus. Można zaobserwować, że wraz z upływem czasu ocenianie staje się coraz bardziej płynne, że mechanizmy prowadzenia interakcji niejako docierają się: proces przebiega coraz sprawniej, oceniający mają coraz mniej wątpliwości, a w ich opiniach pojawia się coraz mniej rozbieżności. Autorka słusznie podaje w wątpliwość założenie, że takie interakcyjne dopasowanie ma jedynie zalety. Jego skutkiem ubocznym może być rozluźnienie kontroli, wpływające na jakość oceny, ale także wycofywanie się niektórych osób z wpływania na przebieg dyskusji, jeżeli wcześniejsze próby zawodziły. Ich milcząca zgoda wpływa więc negatywnie na rzetelność procesu.

Oprócz sytuacji, w których jednostki znajdują się w niesprzyjających warunkach interakcyjnych, przykładem negatywnego zjawiska podczas pracy w grupie może być też „próżniactwo społeczne”. Możemy o nim mówić wówczas, gdy efektywność pracy jednostki w grupie jest mniejsza niż w pojedynkę. Zjawisko to autorka opisuje na przykładzie osób, które posiadają mniejszą wiedzę ekspercką, czują się mniej kompetentne i przez to oddają pole innym.

Myślę, że jednym z ważniejszych wniosków płynących z opisywanej pracy badawczej jest ten dotyczący pragmatycznego podejścia ewaluatorów do oceniania w warunkach, gdy czas jest ograniczony. Z punktu widzenia oceniających, którzy muszą pochylić się nad blisko dwustoma wnioskami, niemożliwe jest szczegółowe przesłanie ich wszystkich. Chcąc czy nie chcąc oceniający posługują się więc pewnymi heurystykami, uproszczonymi sposobami wyciągania wniosków, które zapewniają im bardziej efektywną pracę, co jednak może wpływać na rzetelność samego procesu. Widać to szczególnie pod koniec procesu oceniania, gdy zmęczeni oceniający gotowi są iść na większe ustępstwa.

W związku z tym należy podkreślić, że nie istnieje obiektywna, idealna wartość danego projektu, która wyłania się jako efekt procesów oceniania. Przeciwnie, wartościowanie go ma zawsze charakter interakcyjny, zachodzi w sposób złożony i podlega dynamice dyskusji. Podczas panelu ranga określonego projektu



nadawana jest w procesie negocjacji i bieżących ustaleń, a zatem pomimo że za negocjacjami tymi stoją określone obiektywistyczne kryteria, sam akt uznania wartości zawsze ma charakter społeczny.

## PORÓWNANIE PERSPEKTYW

Przedstawione książki różnią się na wielu poziomach. Pierwsza z nich przedstawia ramę analityczno-teoretyczną, która wykorzystuje pewne uogólnienia dokonane na podstawie analizy wielu konkretnych przypadków. Druga dotyczy jednego, gruntownie przeanalizowanego przypadku działania systemu ewaluacyjnego, który brał pod uwagę kryterium wpływu. W pierwszej książce autor przyjmuje perspektywę socjologii krytycznej, widząc w miarach narzędzie władzy i poniekąd opresji, którym jednak posługują się nie tylko grupy uprzywilejowane wobec grup podporządkowanych, ale raczej opresji, która działa w znacznie bardziej rozproszony sposób. Niektóre jej formy, biorąc pod uwagę bezpośrednie działania, które ją umożliwiają, są aktywizowane przez same jednostki rozpoczynające pomiar własnych osiągnięć na różnych polach. Rzecz jasna, miary mogą także być narzucane, jak zdarza się w przypadku systemów politycznych adaptujących różnorodne mechanizmy naukometryczne w celu dystrybucji korzyści. Perspektywa przyjęta w drugiej książce jest interakcyjna – miary są widziane raczej jako dyskursywne narzędzie, pewnego rodzaju punkt odniesienia w dyskusjach mających na celu doprowadzenie do pewnego konsensusu dotyczącego oceny poszczególnych projektów. Tym niemniej wydaje się zasadne, by potraktować perspektywę pierwszą jako nadrzędną, natomiast drugą jako jej pochodną, interakcyjną stronę, która może pojawić się dopiero wówczas, gdy pewne miary istnieją, są zapisane jako kryteria konieczne do uwzględnienia w ocenach, bo dopiero wówczas można podjąć aktywności, w ramach których decyduje się o zasadności decyzji dotyczących uznania wartości danego projektu.

Wydaje się, że w Polsce brakuje pogłębionych studiów, które ujmowałyby grupowe procesy ocenne w sposób interakcyjny. Ograniczenia wynikające z przyjęcia takiego podejścia są dość oczywiste: aby podjąć się takiego projektu, należałoby pokonać mnóstwo intencjonalnych barier, które mają za zadanie nic innego, jak właśnie zasłonić proces oceniania. Jak pokazywałem w innym miejscu, sam projekt fizycznej przestrzeni, w jakiej odbywa się ocenianie, zakłada, że to, co się w niej dzieje, musi pozostać niejawne [Remisiewicz 2016]. Nadto także biurokratyczna specyfika oceniania, która zakłada udostępnienie publiczności jedynie wyników i – w niektórych przypadkach – sprawozdania lub protokołu z dyskusji, przedstawiającego oficjalną wersję wydarzeń, autoryzowaną przez



jej członków, pokazuje dobitnie, że znaczenie rzeczywistego procesu zostało zredukowane do minimum. Dowody takiego myślenia można znaleźć choćby w ujawnionym przez dziennikarzy nagraniu spotkania szkoleniowego dla dziekanów, na którym znany prawnik radzi, by niszczyć nagrania z posiedzeń komisji habilitacyjnych [Dziubka 2019]. W tym przypadku *explicite* sformułowany cel był czysto użyteczny – chodziło o zredukowanie ryzyka podważenia wyników oceny. Derrick wskazuje, że mechanizm oceny eksperckiej lub koleżeńkiej staje się politycznym instrumentem, a powołanie się na niego „uświęca” system [Derrick 2018: 26]. Demaskowanie tego systemu, wskazywanie na czysto ludzkie, dynamiczne procesy, w jakie uwikłane są oceny, niektórzy mogliby więc uważać za podważające tak starannie konstruowane mity o obiektywizmie.

Na pytanie o to, dlaczego alokować publiczne fundusze w taki, a nie inny sposób, można w ramach obiektywistycznego podejścia udzielić jednej jedynej odpowiedzi: ponieważ taki sposób alokacji eksperci uznali za najlepszy. Któż bowiem może oceniać ekspertów? Być może jedynie eksperci wyższego rzędu. Prowadzi to oczywiście do dalszych pytań o zaufanie w stosunku do określonych osób, które zyskały miano lub status eksperta. Ekspertkość wiąże się z przypisywanymi dyskursywnie cechami, takimi jak wiedza przedmiotowa, obiektywizm, umiar, brak uprzedzeń. Ponadto także przestrzenna izolacja oceniających wiąże się z kontrolą dostępu do informacji, który ma nie tylko charakter użyteczny, ale jest także społeczną oznaką prestiżu.

W odniesieniu do książki Beera wnioski Derrick sugerowałyby, że nie należy przyjmować, że to miary (kryteria obiektywistyczne) same w sobie „posiadają” władzę. Władzę posiada ten, kto ma legitymację do bycia ekspertem od ich stosowania, ponieważ to on wpływa na rzeczywiste kształtowanie i przeobrażanie rzeczywistości poprzez systemy numeryczne. To wiedza o sposobach wykorzystywania miar staje się najważniejsza i to właśnie ona staje się podstawą określonych typów reaktywności. Przynajmniej dopóty, dopóki cały system nie stanie się bardziej zautomatyzowany, a kolejne decyzje będą podejmowane bez względu na bardziej subiektywne kryteria.

Transparentność przyznawania ocen jest niezwykle ważna zarówno z punktu widzenia wewnątrzśrodowiskowej legitymizacji procesu oceniania, jak i szerszej odpowiedzialności społecznej. Szeroka publiczność, która jest także jednym z odbiorców, a przede wszystkim fundatorów badań, ma prawo do informacji na temat podstaw alokacji funduszy publicznych. Poszczególni naukowcy natomiast mają prawo znać kryteria ocen ich wniosków czy projektów. W szczególności, jak się wydaje, ważne jest to dla osób, którym się nie powiodło – mają prawo wiedzieć, z jakich konkretnie powodów nie otrzymały grantu lub awansu, co po-

winni zrobić lepiej lub inaczej. Niestety niekiedy pokutuje opinia, że otrzymanie grantu przypomina raczej rzut monetą, a rozbieżności w ocenach lub niejasność uzasadnień tylko ten wniosek pogłębiają (w badaniu Lambrosa Roumbanisa [2017] ewaluatorzy wprost porównują przyznawanie oceny do loterii).

Współczesne systemy oceny są stworzone na zasadzie czarnej skrzynki – metafora ta zresztą pojawia się zarówno w odniesieniu do ewaluacji nauki, jak i na przykład w stosunku do opisu procesów egzaminowania. Mianowicie istnieją pewne jawne dane na wejściu oraz końcowy rezultat oceny – jak w przypadku wyników konkursu grantowego, jednak niewiele jest informacji o tym, co w rzeczywistości zadecydowało o podjęciu konkretnej decyzji. Tym samym praca Gemmy Derrick jest niezwykle istotna także ze względu na to, że odsłania charakter (co może przyczynić się do ewentualnej naprawy) procesów ocennych. Wydaje się, że pozbawienie recenzji i ocen grupowych nimbu tajemniczości może pomóc także w pozbyciu się wrażenia, że w tego rodzaju procesach dominują arbitralne rozstrzygnięcia. Autorce ponad rok zajęło uzyskanie dostępu do paneli. Świadczy to o tym, że w zakresie transparentności ocen w nauce jeszcze wiele jest do zrobienia. Jak pisze autorka, wszystko, co jest widoczne, to końcowy rezultat oceny. W czarnej skrzynce oceniania kryć się jednak mogą „cały cynizm, wątpliwości, niepoprawność polityczna, nepotyzm i ogólne uprzedzenia” [Derrick 2018: 24], które mogą oddziaływać na końcowy wynik. Istnieją jedynie nieliczne badania, które biorą pod uwagę rzeczywiste interakcje pomiędzy recenzentami czy ewaluatorami [Langfeldt 2001; Pier i in. 2017; Roumbanis 2017]. Inne opierają się na wywiadach rekonstruujących sposoby myślenia i procedury oceniania, tak jak zrobiła to autorka omawianej pracy [Lamont 2009; Langfeldt 2004]. Warto coraz lepiej poznawać mechanizmy funkcjonowania czarnej skrzynki oceniania. Można mieć nadzieję, że podobnych prac będzie powstawało więcej.

Prace Beera i Derrick uważam za komplementarne. *Metric power* proponuje teoretyczny język opisu procesów będących wynikiem uwikłania w miary. Ogólność tej analizy należy postrzegać w tym wypadku jako zaletę – nie mamy tutaj bowiem do czynienia ze sztywną ramą analityczną, ale raczej otwartym projektem, który może być uzupełniany i korygowany. Wydaje się jednak, że intuicje stanowiące podstawę pracy, w tym w szczególności krytyka neoliberalizmu w nauce, z konstytutywną dlań logiką konkurencji, są głęboko zakorzenione w polskim piśmiennictwie. Tezy Beera bez szczególnych problemów mogą pojawić się w dyskusjach poświęconych kolejnym reformom, będącym wielo-  
poziomym efektem i jednocześnie siłą napędową dla środowiska uczonych. Z kolei *The Evaluator's eye* jest zapisem pracochłonnego badania empirycznego, które uwidocznilo mikroskalę procesów ocennych w nauce, ujawniło mozaikę

kontekstów i lokalnych uwarunkowań, z jakimi muszą mierzyć się uwikłani w miary recenzenci. Ta perspektywa badania w moim przekonaniu nie zajmuje jeszcze w Polsce miejsca, na jakie zasługuje. Wydaje się, że wzmożona uwaga skierowana na same procesy oceniania mogłaby zaowocować interesującymi wnioskami wykraczającymi poza czysto hipotetyczne, lub w najlepszym razie, retrospektywne opowieści o ocenianiu w nauce.

## BIBLIOGRAFIA

- Antonowicz Dominik.** 2015. *Między siłą globalnych procesów a lokalną tradycją. Polskie szkolnictwo wyższe w dobie przemian.* Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.
- Beer David.** 2016. *Metric power.* Basingstoke: Palgrave Macmillan.
- Collins Randall.** 2011. *Łańcuchy rytuałów interakcyjnych.* Kraków: Zakład Wydawniczy Nomos.
- Derrick Gemma.** 2018. *The Evaluators' eye. Impact assessment and academic peer review.* Cham: Springer International Publishing.
- Dziedziczak-Foltyn Agnieszka.** 2017. *Reforma szkolnictwa wyższego w Polsce w debacie publicznej. Bilans dyskusji o uniwersytetach.* Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Dziubka Kamil.** 5 lipca 2019. Mówił o Andrzeju Dudzie, że „zglupiał”, radził „niszczyć nagrania”. Po materiale Onetu ważny sędzia będzie się tłumaczył. Onet. <https://wiadomosci.onet.pl/tylko-w-onecie/mowil-o-andrzeju-dudzie-ze-zglupial-radzil-niszczyc-nagrania-po-materiale-onetu-wazny/kqffnsw>.
- Espeland Wendy N., Michael Sauder.** 2007. “Rankings and reactivity”. *American Journal of Sociology* 113(1): 1–40.
- Espeland Wendy N., Mitchell L. Stevens.** 2008. “A sociology of quantification”. *European Journal of Sociology* 49(3): 401–436.
- Garfinkel Harold.** 2007. *Studia z etnometodologii.* Warszawa: Wydawnictwo Naukowe PWN.
- Hyland Ken.** 2007. *Disciplinary discourses. Social interactions in academic writing.* Ann Arbor, Michigan: University of Michigan Press.
- Krzęski Jakub.** 2018. Władza miary a materialne procesy produkcji. (preprint) [https://www.researchgate.net/publication/325269885\\_Wladza\\_miary\\_a\\_materialne\\_procesy\\_produkcji](https://www.researchgate.net/publication/325269885_Wladza_miary_a_materialne_procesy_produkcji) [dostęp: 20.12.2020].
- Kulczycki Emanuel.** 2017. „Punktoza jako strategia w grze parametrycznej w Polsce”. *Nauka i Szkolnictwo Wyższe* 49(1): 63–78.
- Kwiek Marek.** 2015. *Uniwersytet w dobie przemian. Instytucje i kadra akademicka w warunkach rosnącej konkurencji.* Wydanie pierwsze. Warszawa: Wydawnictwo Naukowe PWN.
- Lamont Michèle.** 2009. *How professors think. Inside the curious world of academic judgment.* Cambridge, Mass.: Harvard University Press.
- Langfeldt Liv.** 2001. “The decision-making constraints and processes of grant peer review, and their effects on the review outcome”. *Social Studies of Science* 31(6): 820–841.
- Langfeldt Liv.** 2004. “Expert panels evaluating research”. *Research Evaluation* 13(1): 51–62.
- MNiSW.** 2016. *Biała Księga Innowacji.* <http://bit.ly/2peCjFy> [dostęp: 05.04.2020].

- Ostrowicka Helena, Aleksandra Rzycka, Tomasz Falkowski, Justyna Spsychalska-Stasiak, Dominik Chomik, Łukasz Stankiewicz.** 2019. *Dyskursywny obraz reformy szkolnictwa wyższego w Polsce 2011–2014*. Warszawa: Wydawnictwo Naukowe PWN.
- Pier Elizabeth L., Joshua Raclaw, Anna Kaatz, Markus Brauer, Molly Carnes, Mitchell J. Nathan, Cecilia E. Ford.** 2017. “‘Your comments are meaner than your score’: Score calibration talk influences intra- and inter-panel variability during scientific grant peer review”. *Research Evaluation* 26(1): 1–14.
- Remisiewicz Łukasz.** 2016. *Egzamin w perspektywie socjologicznej*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- Remisiewicz Łukasz.** 2019a. „Niespójność recenzji a wynik postępowania habilitacyjnego. Próba analizy w dyscyplinie socjologia”. *Nauka* 4: 109–127.
- Remisiewicz Łukasz.** 2019b. „Testyzowanie edukacji i jego efekty”. *Studia Socjologiczne* 233(2): 205–233.
- Roumbanis Lambros.** 2017. “Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups”. *Social Studies of Science* 47(1): 95–116.
- Rozkosz Ewa A.** 2018. *Ewaluacja osiągnięć naukowych w postępowaniach habilitacyjnych*. Niepublikowana rozprawa doktorska. Toruń: Uniwersytet Mikołaja Kopernika w Toruniu.
- Rykiel Zbigniew.** 2014. „Szaleństwo czy metoda?”. *Przestrzeń Społeczna (Social Space)* 8(2): 7–60.
- Sauder Michael, Wendy N. Espeland.** 2009. „The discipline of rankings: Tight coupling and organizational change”. *American Sociological Review* 74(1): 63–82.
- Stankiewicz Łukasz.** 2018. *Wizje uniwersytetu w polskiej debacie publicznej 2007–2010*. Wydanie I. Kraków: Impuls.
- Woroniecka Grażyna.** 2015. „Recenzje naukowe jako materiał badawczy”. *Przegląd Socjologiczny* 64(4): 9–25.
- Wróblewska Marta N.** 2017. „Ewaluacja «wpływu społecznego» nauki. Przykład REF 2014 a kontekst polski”. *Nauka i Szkolnictwo Wyższe* 49(1): 79–104.