

# Rozdział II.

## Ewaluacja nauki

Jan Kozłowski

---

### 1. Wstęp

Żyjemy w wieku ewaluacji. Ewaluacja staje się integralną częścią rutynowych procedur administracyjnych na wielu poziomach i w wielu rodzajach instytucji. Jest też częścią cyklu polityki, narzędziem zarządzania projektem, budżetem i zasobami ludzkimi, elementem modeli zarządzania, takich jak zarządzanie przez wyniki lub zarządzanie umowami. To także część szerszych tendencji oceny, które obejmują również takie formy, jak audyt, monitoring, kontrola jakości, pomiar dokonań, benchmarking, ocena wpływu, analiza polityki, analiza następstw regulacji i akredytacja<sup>1</sup>.

#### 1.1. Rozkwit ewaluacji

Ewaluację definiuje się w rozmaity sposób, często jako „systematyczną i obiektywną ocenę programu lub polityki, ich założeń, procesu realizacji i rezultatów pod względem stosowności, skuteczności, trwałości, efektywności, a także użyteczności podjętych w ich ramach działań”<sup>2</sup>. Cztery najważniejsze elementy definicji to określenie charakteru ewaluacji jako procesu, cech ewaluacji (systematyczna i obiektywna), przedmiotu ewaluacji (polityki, programy i projekty) oraz połączonych cech tego przedmiotu – efektywność i skuteczność. Stałym elementem innych definicji ewaluacji jest „systematyczny i obiektywny proces”, natomiast pozostałe składniki definicji często określa się szerzej. Przedmiotem ewaluacji

---

<sup>1</sup> G. Emmertsen Lund, *Fifth-Generation Evaluation*, Internet [dostęp: 27.07.2017]; P. Dahler-Larsen, *The Evaluation Society*, Stanford Business Books, Stanford 2001. Fragmenty niniejszego rozdziału zostały opublikowane w artykule: *Ewaluacja instytucji publicznego sektora nauki oparta na pomiarze dokonań – charakter, pochwały, krytyka*, „Nauka Polska. Jej Potrzeby, Organizacja i Rozwój” 2013 oraz w artykule *Grant w opalach*, „Forum Akademickie” 2000, z. 1, a także w raporcie *Ewaluacja instytucji naukowych w Polsce w świetle porównań międzynarodowych i konsultacji*, Departament Strategii Ministerstwa Nauki i Szkolnictwa Wyższego, Warszawa.

<sup>2</sup> K. Ekiert, *Ewaluacja w administracji publicznej. Funkcje, standardy i warunki stosowania*, Rządowe Centrum Studiów Strategicznych, Warszawa 2004.

mogą zatem być „interwencje publiczne” lub, jeszcze szerzej, wszelkie plany, działania, organizacje, systemy (np. innowacji), akty prawne, dyscypliny naukowe, obiekty, takie jak np. program komputerowy, program nauczania, lekarstwo lub rozwiązanie techniczne, a także same ewaluacje (metaewaluacja). Pożądane cechy określa się również szerzej jako takie, które spełniają „wcześniej ustalone kryteria i standardy”, lub też rozwija i doprecyzowuje („jakość”, „stosowność”, „zdatność” itd. – każde z tych pojęć jest przedmiotem osobnych definicji).

W odniesieniu do cech procesu ewaluacji – „systematyczny” oznacza, że ewaluacja jest realizowana w sformalizowany i zaplanowany sposób, z zapisami dotyczącymi zarówno prowadzonego procesu, jak i jego wyniku, a „obiektywny”, że powinna ona być oparta na weryfikowalnych dowodach lub faktach. Z tych dwóch cech częściej podważa się tę drugą – m.in. w koncepcji ewaluacji partycypacyjnej.

Mimo, że ewaluacje sięgają po metody nauk społecznych, często nie są same uznawane za badania naukowe (choć czasami określa się je jako gałąź stosowanych nauk społecznych).

Wpływ ewaluacji zależy od jej charakteru i założeń. Bywa nim podział funduszy, określenie przyszłości ocenianej organizacji, ustalenie jej struktury, przyjęcie programu pracy lub zmiana profilu (na szczeblu organizacji badawczych), względnie bieżąca kontrola, ocena strategii, instrumentów i stylu zarządzania agencji, wprowadzenie zmian do polityki naukowej, planowania i prawodawstwa (na szczeblu rządowym). Ewaluacje mają też silny wpływ na same siebie: są narzędziem uczenia się, służą poprawie metod i stylu zarządzania ewaluacjami.

Ewaluacja, podobnie jak inne mechanizmy autokorekcyjne działań publicznych, takie jak np. foresight, nie ma swojego niezmiennego wzorca z Sevres<sup>3</sup>. Nawet cechy uznane dziś za jej *differencia specifica* mogą się w przyszłości zmienić.

Zróżnicowane są role, jakie odgrywają ewaluatorzy. Dla określenia ich sięga się po takie metafory i analogie, jak np. inżynier społeczny (w procesie podejmowania decyzji), kontroler zgodności z przepisami, doradca i terapeuta, mediator (między różnymi interesami, np. naukowców, agencji finansującej badania oraz użytkowników), akuszerka nowych form itd.

Ewaluacja albo ma charakter odgórny, albo oddolny. Inicjatorami ewaluacji odgórnej są organa rządowe (ministerstwa, agencje technologiczne), organizacje działające w ich imieniu (rady ds. badań) lub też specjalne agencje ds. ewaluacji (np. Comité National d'Évaluation de la Recherche oraz Comité National d'Évaluation des Universités we Francji, Agencia Nacional de Evaluación y Prospectiva [ANEP] w Hiszpanii). Inicjatorami oddolnej ewaluacji są organizacje badawcze, jak uniwersytety czy laboratoria rządowe. Ewaluacje przeprowadza się na mocy ustaw (np. w Stanach Zjednoczonych) lub jako część rutyny, jako *follow-up* kontraktu lub *ad hoc*. Ewaluatorami są zespoły powołane *ad hoc*, krajowi eksperci

---

<sup>3</sup> Sèvres – siedziba Międzynarodowego Biura Wag i Miar, powołanego w 1875 roku przez międzynarodową Konferencję Miar, w którym przechowuje się wzorzec kilograma oraz metra platyny i irydium.

z ocenianej dziedziny (*peers*), eksperci krajowi wspólnie z zagranicznymi lub tylko eksperci zagraniczni, specjalnie do tego celu powołane instytucje, względnie prywatni kontraktorzy (zawodowi ewaluatorzy ze świata nauki oraz konsultingu) czy wreszcie odrębne instytucje publiczne. Wiele ewaluacji przeprowadzają komórki wewnątrz danej organizacji (w Polsce z reguły jest to tylko wąsko pomyślany audyt), oceniając głównie administracyjne, menadżerskie i finansowe aspekty działań. Gdy tylko można, łączy się ewaluację wewnętrzną i zewnętrzną, co pozwala na większy obiektywizm i niezależność ocen.

Istnieje wiele typów ewaluacji, a nawet wiele taksonomii typów (np. wewnętrzna-zewnętrzna; *ex ante*, *ex post*, monitoring, metaewaluacja; ewaluacja struktur, procesów oraz efektów [*outcome*]). Istnieje też wiele elementów organizacyjnych ewaluacji (np. grupa sterująca, zespół ewaluacyjny), technik strukturyzujących problem (SWOT, drzewo celów, mapowanie koncepcyjne wpływów, matryca wpływów krzyżowych), technik gromadzenia informacji (np. ankiety, panele, grupy focusowe, wywiady, studia przypadku, dane naukometryczne), technik analitycznych (np. grupa porównawcza, modele makroekonomiczne, analiza regresyjna, metoda delficka), technik oceny (np. panel ekspertów, benchmarking, analiza wielokryterialna, analiza kosztu i korzyści, analiza kosztu i efektywności). Niemniej szerokie jest zróżnicowanie form organizacyjnych (stałe i zinstytucjonalizowane lub doraźne)<sup>4</sup>.

## 1.2. Geneza ewaluacji

Choć, inaczej niż w XIX wieku, obecnie nie uważamy, że „geneza wyjaśnia istotę zjawiska”, krótka informacja o rozwoju historycznym zarówno pojęcia, jak i zjawiska, ułatwi zrozumienie, czym jest ewaluacja.

Trzy najważniejsze źródła do badania zmian częstotliwości użycia terminu „ewaluacja” to „Ngram Viewer” (Google), WorldCat (OCLC) oraz same Google.

---

<sup>4</sup> E. Kalpazidou Schmidt, *Evaluation and science policy*, [w:] *The use of Evaluation in Europe*, 2002; *The use of R&D evaluations in European science policy*; T. Balogh, *From peer review to portfolio analysis. Evaluation of applied research and development*, msp.; I. Feller, *The academic policy analyst as reporter: the who, what and how of evaluating science and technology programs*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany. red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001, [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhwkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhwkshop.pdf); W. Krull, D. Sensi i D. Sotiriou, eds., *Evaluation of Research and Development. Current Practice and Guidelines. Synthesis report*, Commission of the European Communities Directorate-General Telecommunications, Information Industries and Innovation, Luxembourg 1991; L. Georghiou, *Research evaluation in European national science and technology system*, „Science and Public Policy” 1995, vol. 5, no. 1; E. Kalpazidou Schmidt i K. Siune, *The Use Of R&D Evaluations in European Science Policy*, [www.cfa.au.dk/Staff/eks.htm](http://www.cfa.au.dk/Staff/eks.htm); *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, 1999, vol. 6, European Commission, Luxembourg.

Ngram Viewer pokazuje wartości relatywne: częstotliwość użycia terminu w treści książek (na milion słów w ciągu roku) zeskanowanych przez Google w latach 1900–2008 (rysunek 2.1).

**Rysunek 2.1**  
**Badania zmian częstotliwości użycia terminu „ewaluacja” według Ngram Viewer**

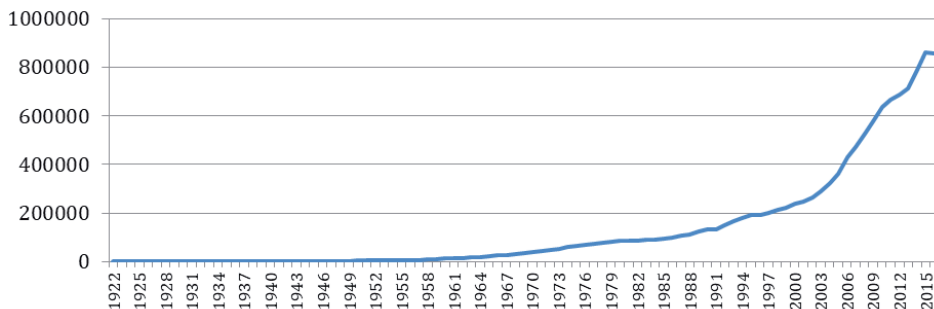


Źródło: opracowanie własne.

WorldCat – wartości bezwzględne (termin „*evaluation*” jako element tytułu lub podtytułu książek i innych materiałów bibliotecznych, takich jak np. artykuły, rozdziały, rozprawy, mikrofilmy itd.) w latach 1922–2016 (rysunek 2.2).

**Rysunek 2.2**  
**Diagram częstotliwości użycia terminu „ewaluacja” według WorldCat**

***Evaluation* jako element tytułu książek i innych materiałów bibliotecznych**



Źródło: opracowanie własne.

Porównanie obu diagramów sugeruje, że „start do wzrostu” w treści książek nastąpił wcześniej niż „start do wzrostu” w tytułach. WorldCat – oprócz dwóch ostatnich lat, 2015 i 2016 – nie notuje stabilizacji i spadku, widocznego w diagramie Ngram Viewer. Możliwe, że spadek częstotliwości użycia poprzedza spadek publikacji poświęconych zjawisku ewaluacji. Jest też możliwe, że dane Ngram Viewer dla książek z ostatnich dekad są mniej pełne niż dla poprzednich dziesięcioleci. Hipotezę tę potwierdza diagram z liczbą materiałów dotyczących ewaluacji w Google, który wykazuje wzrost, ale także późniejszą stabilizację (oznaczającą spadek udziału treści ewaluacyjnych, biorąc pod uwagę nieustanny wzrost platformy).

Wszystkie te wykresy wykazują skok w zainteresowaniu ewaluacją w XX wieku.

Francuskie *evaluation*, stosowane w początkach XVIII wieku w znaczeniu „wyceny” (pieniężnej), z biegiem lat rozszerzyło znaczenie, aby w II połowie tego wieku objąć nim wszelki „akt oceny wartości” (nie tylko pieniężnej). Słowo to w języku angielskim i francuskim przeszło do języka nauki (astronomia, geologia). Fakt, że Oświecenie jest świadkiem narodzin nowego znaczenia ewaluacji, nie dziwi: jednym z jego haseł była „naprawa”, a jedną z idei – branie spraw we własne ręce (zamiast pozostawiania ich królowi, Kościołowi i tradycji). Ulepszając – miasto, gospodarstwo rolne, wojsko – chcemy mieć dowód, że osiągamy założony cel.

Jeśli za istotę ewaluacji uznamy porównywanie – wkładu i efektu, zamiaru i efektu itd. – to porównania te stały się łatwiejsze, bardziej uchwytnie i precyzyjne z chwilą upowszechnienia się kwantyfikacji. Urynkowienie i monetyzacja stworzyły zarówno potrzebę, jak i narzędzia do oceny ilościowej inwestycji. Zmusiła do takich porównań walka rynkowa – zjawisko narastające w miarę przechodzenia od gospodarek autarkicznych i reglamentowanych do wolnorynkowych. Z chwilą gdy takie porównania stały się zarówno możliwe, jak i pożądane, nic nie stawało na przeszkodzie, aby je poszerzać. Choć pojęcia „wkładu” i „produktu” (*input-output*) wchodziły do coraz częstszego użytku z początkiem XX wieku, w dzisiejszym znaczeniu stosowano je od II połowy XVIII wieku<sup>5</sup>. Porównania „wkładu” i „produktu” inwestycji nabierały sensu wraz ze wzrostem ilości danych: porównywania w czasie i w przestrzeni, z uwzględnieniem kolejnych zmiennych. Zanim porównania te nie przyobkleły się w formę klasycznych analiz statystycznych, stosowano je *ad hoc*, niesystematycznie, doraźnie, jednorazowo.

Dwudziestowieczną popularność w różnych językach „ewaluacja” zawdzięcza nie tyle rozszerzonemu znaczeniu (tutaj częściej sięga się po inne słowa, np. w języku angielskim: *assessment*, *appraisal*, *judgement*, *gauging*, *rating*, *estimation*, *ranking*, *weighing up*, *summing up*, *consideration*, *assay*, *analysis* itd.), ile zawężonemu, zgodnemu z definicjami, które rozważyliśmy na początku.

---

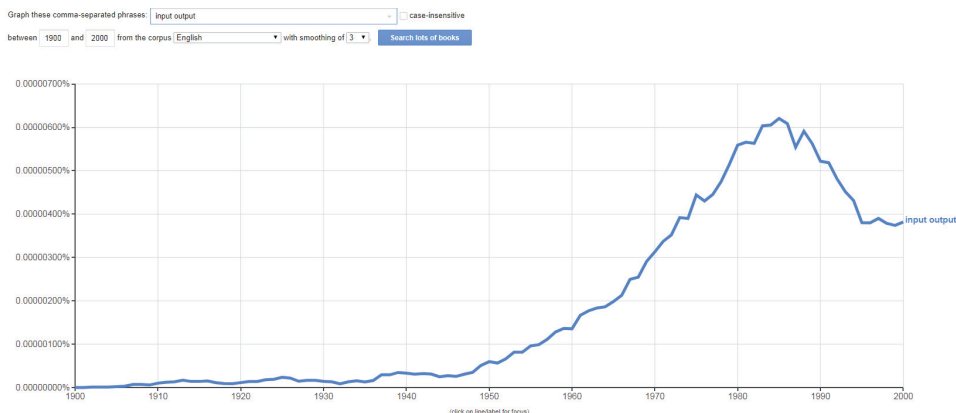
<sup>5</sup> D. Roche, *La France des Lumières*, A. Fayard, Paris 1993. Tłum. ang. A. Goldhammer, *France in the Enlightenment*, Cambridge UP 1997.

Skąd wzięła się popularność ewaluacji jako procedury, narzędzia zarządzania i współ-zarządzania (*governance*), polityki oraz uczenia się? Na to pytanie udziela się najczęściej następujących odpowiedzi:

1. Wykształcenie wystarczająco dużej liczby absolwentów nauk społecznych, którzy byli gotowi wykorzystywać swoją wiedzę praktycznie.
2. Ograniczenia finansowe dotyczące wydatków publicznych.
3. Upowszechnianie Nowego Zarządzania Publicznego<sup>6</sup>.

Rysunek 2.3

### Diagram częstotliwości użycia *Input – Output* w języku angielskim według Ngram Viewer



Źródło: opracowanie własne.

Ciekawsze odpowiedzi na to pytanie można jednak znaleźć w pismach filozofujących socjologów, takich jak Anthony Giddens, Ulrich Beck i Scott Lasch, oraz filozofujących badaczy ewaluacji, takich jak Peter Dahler-Larsen, autor książki *The Evaluation Society* (2001). Wszyscy oni uznają ewaluację za zjawisko charakterystyczne dla obecnej fazy rozwoju społecznego, za symptomatyczny składnik „refleksyjnej nowoczesności”. Cechą „refleksyjnej nowoczesności” jest fakt, że znacznie większą rolę niż w okresach poprzednich odgrywa w nich coś, co określają oni jako „refleksyjność”, traktowaną przez nich jako to, co ludzie wiedzą i w co wierzą co do działań własnych i cudzych. Ich zdaniem nowoczesność – skrótowa nazwa nowoczesnego społeczeństwa, w którym świat postrzegany jest jako możliwy do przekształcenia przez ludzkie działania i gdzie funkcjonują złożone instytucje gospodarcze i polityczne – weszła w okres, w którym niezbędna jest znacznie większa niż poprzednio „zwrotność”, monitoring podejmowanych działań. Świat stał się bardziej złożony i dynamiczny, a przez to mniej przejrzysty. Uregulowany, sformalizowany monitoring działań i ich

<sup>6</sup> P. Dahler-Larsen, *The Evaluation Society*, op. cit.

następstw jest w coraz większym stopniu warunkiem przeżycia i powodzenia ludzi oraz organizacji. Ewaluacja odpowiada na te oczekiwania.

„Ewaluacja – pisze Peter Dahler-Larsen – nie byłaby możliwa bez typowo nowoczesnej formy mentalności, którą Berger, Berger i Kellner (1973) nazwali «technologiczną», co oznacza, że niektóre obszary życia podlegające ewaluacji są uważane za abstrakcyjne obiekty ewaluacji, które można kontrolować, porządkować, regulować lub utrzymywać poprzez manipulowanie relacjami między elementami abstrakcyjnymi. (...) Ewaluacja nosi w sobie ideę, że przedmiot oceny – niezależnie od tego jaki – można podzielić na poszczególne elementy, działania, programy, kryteria i rezultaty, aby mógł być zarządzany i reorganizowany. Bez tej szczególnie nowoczesnej kosmologii ocena nie mogła zostać wymyślona. Bez tej kosmologii nie byłoby możliwe, aby współczesna ocena ustaliła się jako jeden ze sposobów, w jaki ludzie uzyskiwali dostęp do organizacji ich własnego społeczeństwa. (...) Nowoczesna ewaluacja jest głęboko zakorzeniona w logice organizacyjnej. Nie możemy myśleć o nowoczesnej ewaluacji poza nowoczesną organizacją”<sup>7</sup>.

Nie znaczy to, że zjawisk podobnych do określanych dziś mianem ewaluacji nie spotykamy w czasach dominowania tradycji. Mówiąc żartobliwie, już w epoce zbieractwa i myślistwa z pewnością dyskutowano, co poszło tak, a co nie tak podczas polowania<sup>8</sup>. W tym sensie „zwrotność” – obserwowanie następstw działań – jest warunkiem rozwoju kultury. Pismo, druk oraz upowszechnianie umiejętności czytania i pisania, umożliwiając „uzewnętrznienie pamięci”, stały się kamieniami milowymi monitoringu<sup>9</sup>. Pismo i jego pochodne (a także pieniądze, druk i zegar) pozwalały na koordynację czasu, przestrzeni, rzeczy i czynności traktowanych jako przedmiot stałej analizy i oceny (od XVII wieku coraz częściej ilościowej), przeprowadzanej z punktu widzenia założonych celów oraz z zamiarem dokonania niezbędnych działań korekcyjnych. Zapewniało ono możliwość porównywania stanu obecnego z docelowym oraz zestrzajanie wspólnych działań na rzecz realizacji planu. Taką funkcję pełniły m.in. podwójna księgowość, kwestionariusze i protokoły potrydenckich wizytacji kanonicznych oraz (oparte na użyciu skali) XVIII-wieczne egzaminy uniwersyteckie. W XIX wieku armia (na początku amerykańska West Point), wielkie przedsiębiorstwa (na początku koleje amerykańskie) oraz rozwijające się instytucje państwowe przekształcały ludzi i rzeczy w obiekt nieustającego pisania, oceniania i rangowania (kwalifikowania według skali) za pomocą danych, pism, planów, harmonogramów, modeli, dyrektyw, przewodników, instrukcji, ewaluacji, rachunków, kosztorysów, analiz oraz takich form prezentacji, jak teksty, mapy, diagramy, tabele, matryce czy

<sup>7</sup> Ibidem.

<sup>8</sup> „Ewaluacja jest stara jak świat. Czy małe grupy myśliwych i zbieraczy nie siedziały wokół ognia, jedząc nogi ugotowanego mięsa jeleni i pytały, czy dziś było dobre polowanie?”. Ibidem.

<sup>9</sup> D. Merlin, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*, Harvard University Press, Harvard 1991.

mapowanie procesów<sup>10</sup>. Jednak – jak podkreśla Peter Dahler-Larsen – to tylko w epoce naszej „refleksyjnej nowoczesności” ewaluacja została urzeczowiona i zinstytucjonalizowana (zestaw metod zwanych „ewaluacyjnymi”, specjaliści zwani „ewaluatorami”, zasoby, środki, prestiż)<sup>11</sup>.

### 1.3. Ewaluacja nauki

Ewaluacja zagarnęła świat nauki w II połowie XX wieku. W sposób obrazowy proces ten opisał John Ziman. W II połowie XX wieku – pisał – gwałtownie rosnące koszty badań oraz fakt, że naukowcy stali się (w lwiej części) pracownikami zatrudnionymi w wielkich hierarchicznych organizacjach (nie tylko uniwersytetach, ale także laboratoriach przemysłowych oraz agencjach rządowych), spowodował zmianę form kierowania badaniami naukowymi i technologicznymi. Budżety państw rozwiniętych gospodarczo nie mogły dłużej wspierać badań prowadzonych przez coraz większe zespoły, pracujące z coraz bardziej kosztowną i skomplikowaną aparaturą. Sfera dawnej samodzielności w doborze tematyki i metod oraz na wespół feudalnych relacji mistrza i ucznia skurczyła się. Nauka stała się szczuplejsza, bardziej zrjonalizowana, zarządzanie instytucją naukową upodobniło się do zarządzania firmą, z indywidualnego przedsięwzięcia przekształciło się w przedsięwzięcie zbiorowe. Proceduralizacja, standaryzacja, oceny efektywności itd. okazały się znakomitym antidotum przeciwko pracom ciągnącym się w nieskończoność, źle zaprojektowanym, niekonkluzywnym, przeciwko dublowaniu aparatury itd. Instytucje naukowe podjęły radykalne restrukturyzacje, tak aby prowadzić badania i lepiej, i efektywniej, i bliżej potrzeb gospodarki i społeczeństwa. W rezultacie wszystkich opisanych zmian ministerstwa oraz agencje rządowe finansujące badania przeszły od prostego podziału funduszy do ich profesjonalnego zarządzania. Od lat 60. w środowiskach naukowych zaczęły krążyć słowa dawniej niesłyszane: „system N+T”, „polityka naukowo-techniczna”, „ustalenie rządowych priorytetów”, „badania strategiczne”, „badania w fazie przedkonkurencyjnej”, „państwowe programy technologiczne”, „wzmocnienie więzi nauki z przemysłem”, „współzawodnictwo o fundusze”, „rozliczanie”, „zarządzanie nauką”, „efektywność BiR”, „ewaluacja badań”, „wskaźniki nauki i techniki”, „input”, „output”, „selektywność”, „masa krytyczna badań”, „centra doskonałości”, „centra multidyscyplinarne”, „sieci badawcze” (łącznie uczonych pracujących w różnych instytucjach), „badania kontraktowe” itp.<sup>12</sup> Żadne z nich nie było pustym hasłem, tylko nazwą praktyk określającą codzienność.

<sup>10</sup> K.W. Hoskin i R.H. Macve, *Accounting And The Examination: A Genealogy Of Disciplinary Power*, „Accounting Organizations and Society” 1986, vol. 11, no. 2; K. Hoskin, R. Macve i J. Stone, *The Historical Genesis of Modern Business and Military Strategy: 1850–1950*, Interdisciplinary Perspectives on Accounting Conference, Manchester, 7–9 July, 1997.

<sup>11</sup> P. Dahler-Larsen, *The Evaluation Society*, op. cit.

<sup>12</sup> J. Ziman, *Prometheus Bound. Science in a dynamic steady state*, Cambridge University Press, Cambridge 1994.



Badania naukowe stały się obiektem polityki oraz zarządzania (politykę naukową od wcześniejszych form mecenatu państwa różni fakt, że jej decyzje to inwestycja publiczna, także w gospodarkę i rozwój społeczny, a nie proste wsparcie konsumpcji dóbr kultury; ponadto, to realizacja celów, zadań i priorytetów rządowych, a nie proste „zaspokajanie potrzeb badawczych” poszczególnych uczonych, instytucji lub dyscyplin).

Ewaluacja nauki okazała się „przemysłem szybkiego wzrostu”. W większości krajów członkowskich OECD wzrasta nacisk na rozliczalność, a także na skuteczność i efektywność badań wspieranych przez rząd, podobnie jak w przypadku wielu innych kategorii wydatków publicznych. Rządy potrzebują takich ewaluacji do różnych celów: optymalizacji alokacji funduszy na badania w czasie rygorów budżetowych, reorientacji wspierania badań, racjonalizacji lub redukcji organizacji badawczych, poprawy produktywności badań itp. Opracowały lub pobudzały działania w dziedzinie ewaluacji badań w celu uzyskania „większej wartości z pieniędzy”<sup>13</sup>.

Nie zaprzeczając roli intuicji, polityka i zarządzanie opiera się na wiedzy. Jak się wydaje, w krajach OECD i Unii występuje 10 różnych, możliwych do wydzielenia za pomocą różnych kryteriów, choć wzajemnie ze sobą powiązanych form (tworzenia i upowszechniania) takiej wiedzy. Są nimi – oprócz ewaluacji – badania nad polityką naukową i innowacyjną; analizy polityk; scoreboardy, wskaźniki efektywności, raporty wskaźnikowe, indeksy, rankingi, tabele ligowe, barometry, monitory, benchmarkingi; serwisy informacyjne, platformy informatyczne, portale, obserwatoria, przeglądy, „perspektywy” (*outlook*); „policy intelligence”, narzędzia „zarządzania polityką naukową i innowacyjną”, takie jak foresight, skanowanie horyzontu, prognozowanie, budowa modeli ekonometrycznych itd.; statystyka nauki i techniki (międzynarodowa i rządowa); debata publiczna; konsultacje publiczne; dokumenty programowe, strategiczne i planistyczne (np. komunikaty Unii Europejskiej, białe i zielone księgi, dokumenty strategiczne) na szczeblu międzynarodowym, rządów i władz regionalnych.

Rozwój tych form wiedzy wspierających decyzje polityczne i menadżerskie to efekt zmian publicznego finansowania badań naukowych. Zmiany te można podsumować jako przejście od względnej autonomii nauki w okresie powojennym, poprzez stadia wzrastającej presji na rozliczanie (finansowe i społeczne) oraz relewantność, aż do dzisiejszej sytuacji, gdy większość badaczy zachęca się do ukierunkowania badań na tematy posiadające (choćby w dalszej perspektywie) znaczenie gospodarcze i społeczne<sup>14</sup>.

Perspektywa akademicka i menadżerska badań naukowych różnią się zasadniczo, a normy i wartości, którymi kierują się badacze, są z reguły przeciwne polityce i zarządzaniu (tabela 2.1).

<sup>13</sup> OECD, *The Evaluation Of Scientific Research. Selected Experiences*, Paryż 1997.

<sup>14</sup> J. Calvert i B. Martin, *Science funding: Europe*, [w:] *International Encyclopaedia of the Social and Behavioural Sciences*.

Tabela 2.1

**Perspektywa akademicka i menadżerska badań naukowych**

	<b>Podejście akademickie</b>	<b>Podejście menadżerskie</b>
<b>Cel nauki</b>	Tworzenie wiedzy jako dobra samego dla siebie.	Tworzenie wiedzy nastawionej (w krótkim oraz dłuższym horyzoncie) na praktyczne zastosowania.
<b>Ocena badań</b>	Kryteria wewnątrznaukowe (wiarygodność, obiektywność, oryginalność).	Kryteria wewnątrznaukowe oraz zewnętrzne (użyteczność, wpływ społeczny i gospodarczy).
<b>Badacz powinien</b>	Być niezależnym i autonomicznym.	Być uczestnikiem zespołu zarządzanego zgodnie z celami społecznymi i organizacyjnymi.
<b>Główne źródło kontroli</b>	Sami uczeni według hierarchii prestiżu.	Profesjonalne zarządzanie instytucji naukowych.
<b>Optymalny rozwój instytucji naukowych</b>	Zagwarantowany dzięki samoorganizacji.	Zagwarantowany dzięki zarządzaniu (na różnych szczeblach).
<b>Obraz natury badań i badaczy</b>	Badania są nieprzewidywalne i wobec tego niemożliwe do kierowania. Perspektywa indywidualistyczna: badacz jako człowiek „samozatrudniony”, wewnętrznie motywowany, artysta.	Badania mają charakter celowy i zamierzony. Perspektywa zatrudnienia wysoko wykwalifikowanego eksperta. Badacz łączy indywidualną pasję z zadaniami zespołowymi, wyznaczanymi przez instytucję.
<b>Główna forma rezultatu badań</b>	Artykuły w czasopismach listy filadelfijskiej.	Usługi techniczne, innowacje techniczne i społeczne, ekspertyzy, konsultacje, kształcenie, patenty, licencje, prototypy, współpraca, książki i artykuły.

Źródło: opracowane za: E. Erno-Kjollhede et al., *Managing university research in the triple helix*, „Science and Public Policy”, February 2001, vol. 28, nr 1.

Jeśli uznać lata 1962–1963 za początki polityki naukowej (pierwsze posiedzenie ministrów ds. nauki krajów OECD, *Podręcznik Frascati* statystyki nauki, artykuł Alvina Weinberga *Criteria for Scientific Choice*), w ciągu kolejnych dekad powstały coraz bardziej liczne i różnicujące się obszary ewaluacji nauki (tabela 2.2).

Tabela 2.2

## Obszary ewaluacji nauki

Nazwa obszaru ewaluacji	Cel i zastosowania	Przykłady metod
<b>Polityka naukowa i innowacyjna oraz system nauki i innowacji</b>	Ogólna ocena systemu i polityki nauki i innowacji – zarówno procedur decyzyjnych ( <i>policy mix</i> , tzw. <i>policy intelligence</i> , priorytety) oraz metod zarządzania ( <i>governance</i> ), jak i produktów oraz naukowych i pozanaukowych efektów polityki N+T – jako wkład do strategii oraz programu reform.	Ekonomika: B+R, nauki, polityki naukowej, szkolnictwa wyższego. <i>Peer-review</i> , <i>review of science (and/or innovation) policy</i> . „Analiza polityki i systemu” („portfolio and systems analysis”). Scoreboardy.
<b>Policy mix polityki naukowej i innowacyjnej</b>	Ocena „portfela” programów i instrumentów dla określenia ich wzajemnego (dodatniego i ujemnego) wpływu na siebie oraz wpływu na realizację założeń polityki.	Wywiady, grupy fokusowe.
<b>Programy i instrumenty polityczne</b>	Ocena: produktywności, pozanaukowego wpływu. Pytania m.in. o dodatkowość.	Metody ilościowe, jakościowe, łączone.
<b>Projekty (realizowane w ramach poszczególnych programów grantowych)</b>	Ocena propozycji ( <i>ex ante</i> ); ocena <i>ex post</i> realizacji projektu.	<i>Peer review</i> , <i>merit review</i> , rozszerzone <i>peer review</i> , bazy <i>Research Information Systems</i> .
<b>Książki i artykuły naukowe, prezentacje na konferencjach</b>	Ocena tekstu (abstraktu).	<i>Peer review</i> .
<b>Aparatura i infrastruktura naukowa</b>	Stosowane do oceny <ul style="list-style-type: none"> <li>• indywidualnych wniosków;</li> <li>• „polityki aparaturowej”.</li> </ul>	Baza aparatury; informacje zbierane na drodze ankiet; <i>peer review</i> .
<b>Ocena wpływu badań</b>	Ocena naukowego i – przede wszystkim – pozanaukowego wpływu badań naukowych.	Metody ilościowe i jakościowe.
<b>Dokonania w dziedzinie badań naukowych i technologicznych</b>	Ocena <i>produktów</i> badań naukowych i technologicznych jako m.in. podstawa formułowania priorytetów badawczych.	Bibliometria; analizy sieciowe; analizy patentowe, z uwzględnieniem tzw. badania wartości patentu; statystyka udziału w programach międzynarodowych.

Tabela 2.2 – cd.

Nazwa obszaru ewaluacji	Cel i zastosowania	Przykłady metod
<b>Dyscypliny naukowe technologiczne</b>	Stosowane jako wkład do strategii i priorytetów N+T oraz decyzji odnośnie do rozwoju dyscyplin i pól.	<i>Peer review</i> (często oparty na udziale ekspertów zagranicznych), naukometria.
<b>Udział kraju w Programach Ramowych oraz innych programach EU</b>	Podstawa ulepszenia działań MNiSW, KPK oraz Komisji Europejskiej.	Różne metody rozciągające się od statystyki po ocenę wpływu.
<b>Resorty ds. nauki oraz agencje ds. badań naukowych i/lub technologicznych (zarządzanie oraz systemy grantowe)</b>	Ocena: <ul style="list-style-type: none"> <li>• produktywności;</li> <li>• pozanaukowego wpływu.</li> </ul>	Zróżnicowane metody.
<b>Instytucje naukowe / dyscypliny/ zespoły (finansowanie instytucjonalne oparte na dokonaniach)</b>	Podstawa alokacji środków finansowych dla uniwersytetów (oraz, w Polsce, także pozauczelnianych instytutów naukowych) oparta na dokonaniach ( <i>performace funding</i> ): <ul style="list-style-type: none"> <li>• stały system informacji o badaniach;</li> <li>• cykliczne zbieranie informacji;</li> <li>• ewaluacja <i>ad hoc</i> pewnego typu jednostek dziedzinowych.</li> </ul>	Bazy danych o badaczach i instytucjach naukowych, oparte na danych pochodzących z baz bibliometrycznych i patentowych oraz na informacjach ankietowych (tzw. <i>Research Information System</i> )
<b>Instytucje naukowe – ewaluacje indywidualne</b>	Podstawa strategii instytucji oraz decyzji dotyczących poszczególnych badaczy i zespołów badawczych.	Metody bibliometryczne, wywiady.
<b>Uczelnie (ocena dwóch misji – badań naukowych i nauczania, oraz tzw. trzeciej misji)</b>	Jednorazowa ewaluacja jako podstawa przewidywanej reformy sektora.	Bazy danych (informacja zarządcza). Zróżnicowane metody.

Nazwa obszaru ewaluacji	Cel i zastosowania	Przykłady metod
<b>Pozauczelniane instytuty badawcze; ocena badań naukowych oraz tzw. innych form działalności N+T, jak pomiar, standaryzacja itd.</b>	Jednorazowa ewaluacja pozauczelnianych instytutów badawczych jako podstawa przewidywanej reformy sektora.	Bazy danych (informacja zarządcza). Zróżnicowane metody, studia przypadku (statystyka dokonań, wywiady, kwestionariusz), <i>peer review</i> , analiza kosztów i korzyści.
<b>Narzędzia <i>policy intelligence</i>, takie jak ewaluacja, foresight, mapowanie technologii itd.</b>	Stosowane do oceny narzędzi informacyjno-analitycznych systemu wspierających politykę N+T.	<i>Peer review</i> .

Źródło: opracowanie własne.

Te piętnaście obszarów ewaluacji oddaje całe zróżnicowanie ewaluacji, obszary te różnią się co do:

1. Przedmiotu:
  - a) organizacje (resorty, agencje rządowe, uczelnie, instytuty pozauczelniane);
  - b) system nauki i innowacji;
  - c) polityka naukowa, programy i instrumenty;
  - d) aktywność badawcza (dyscypliny, udział programach);
  - e) wnioski (propozycje projektów badawczych, aparaturowych i infrastrukturalnych, teksty do druku);
  - f) *policy intelligence* (włącznie z samą ewaluacją).
2. Genezy i ewolucji.
3. Celu i zastosowań.
4. Podstaw prawnych.
5. Zakresu geograficznego i instytucjonalnego.
6. Metod.
7. Zleceniodawców.
8. Ewaluatorów.
9. Organizacji procesu.
10. Podstaw informacji.
11. Standardów.
12. Kosztów.
13. Problemów i dylematów.
14. Następstw i oddziaływania.

Spróbujmy przybliżyć każdy obszar, zwracając uwagę na jego najciekawsze elementy.

## 2. Obszary ewaluacji nauki

### 2.1. Ewaluacja krajowego systemu badań i innowacji (lub systemu badań i innowacji oraz polityki naukowej i innowacyjnej)

Ewaluacja ta, przeprowadzana przez zagranicznych ekspertów z ramienia OECD, Unii Europejskiej lub firm konsultingowych w porozumieniu i na zaproszenie ministerstwa ds. nauki (i/lub innowacji), ma na celu dokonanie zewnętrznej bezstronnej oceny przez wybitnych naukowców lub administratorów nauki niewiązanych w krajową walkę interesów. Raporty noszą zwykle w tytule słowa i zwroty: *reviews*, *peer review*, *policy mix*. Ewaluacja zewnętrznych ekspertów ma charakter „wizyty terenowej” składającej się z wielu spotkań i rozmów z miejscowym środowiskiem naukowym i/lub innowacyjnym. Często wizytację poprzedza przygotowanie przez zamawiającego raportu naświetlającego stan i problemy będące przedmiotem ewaluacji. Raport wstępny ułatwia zaplanowanie programu wizyty. Zwykle ministerstwo zamawia raport, gdy zamierza dokonać istotnych zmian w systemie badań i/lub innowacji. Raport końcowy ułatwia ich przeprowadzenie.

Najbardziej znane są serie:

- OECD;
- EU (CREST, ERAC, UNU-MERIT, Trendchart, Erawatch, Policy Support Facility).

**OECD.** Raporty zostały zapoczątkowane w roku 1963 przez serię *OECD Country reports on the organisation of scientific research*, poświęconą krajom członkowskim organizacji. Od lat 70. zastąpiła ją seria *OECD Reviews of National Science Policy*. W przeciwieństwie do pierwszej serii, opis systemu i polityki (przygotowany przez przedstawicieli ocenianego kraju) łączyła ona z raportem egzaminatorów, przedstawiającym wnioski i zalecenia. Od połowy lat 80. seria zmieniła nazwę na *OECD Reviews of national science and technology policy*. W I połowie lat 90. nazwę uzupełniono o jeszcze jeden człon: *OECD Science, technology and innovation policies* (choć nadal ukazywały się raporty z poprzednim tytułem, np. *OECD Reviews of national science and technology policy: Poland*, 1996). Łącznie w latach 1963–1996 ogłoszono drukiem 38 przeglądów. Po dziesięcioletniej przerwie od połowy I dekady XXI wieku raporty noszą z reguły tytuł *OECD Reviews of innovation policy* (najczęściej) lub *Peer review of the policy mix for innovation*, względnie *Policy mix for innovation* (np. *OECD Peer review of the policy mix for innovation in Poland* 2006), przy czym obejmują zarówno problematykę B+R, jak i innowacji: główną osią koncepcyjną przeglądów jest krajowy system innowacji. Obecnie OECD myśli o rozszerzeniu formuły i jednocześnie uwzględnieniu w niektórych raportach problematyki nauki, innowacji i szkolnictwa wyższego.

Celem ostatniej edycji raportów OECD jest całościowa analiza krajowego systemu innowacji, z naciskiem na rolę polityki publicznej. Podstawowe elementy przeglądu to:

- innowacje i wyniki gospodarcze;
- warunki ramowe dla innowacji (stabilność makroekonomiczna, ramy prawne, konkurencja itp.);
- międzynarodowa analiza porównawcza innowacyjności;
- wspieranie działań w zakresie badań i rozwoju oraz innowacji;
- rola uniwersytetów i publicznych organizacji badawczych – powiązania uniwersytetów i gospodarki;
- umiędzynarodowienie B+R;
- zasoby ludzkie dla nauki i techniki;
- infrastruktury wiedzy;
- zarządzanie w systemie innowacji, *policy mix*, ewaluacje.

Przegląd jest zaprojektowany tak, aby przyczynić się do:

- rozwoju świadomości innowacji;
- rozwoju agendy polityki innowacyjnej;
- poprawy integracji polityki N+T+I w polityce gospodarczej;
- pobudzenia dialogu między głównymi zainteresowanymi stronami;
- rozpoznania barier ograniczających poprawę stanu innowacji;
- pomocy w lepszej koordynacji polityki;
- poprawy rozwiązań instytucjonalnych oraz mechanizmów zarządzania;
- optymalizacji *policy mix* oraz projektowania i realizacji poszczególnych instrumentów polityki.

Standardowa procedura przeglądu obejmuje następujące fazy:

- porozumienie w sprawie zakresu zadań między OECD i Ministerstwem;
- raport wstępny (zazwyczaj przygotowany przez ekspertów krajowych);
- misja zespołu ekspertów OECD, wywiady z głównymi zainteresowanymi stronami w krajowym systemie innowacji;
- zespół OECD przygotowuje projekt sprawozdania, w tym ogólną ocenę i zalecenia;
- konsultacje z krajowymi ekspertami;
- przesłanie raportu do Komitetu OECD ds. Polityki Naukowej i Technologicznej (CSTP) Grupy Roboczej ds. Polityki Technologicznej i Innowacyjnej (TIP);
- publikacja raportu końcowego pozostawiona w gestii Sekretarza Generalnego OECD;
- wnioski i zalecenia przeglądu prezentowane na jednej lub więcej konferencji w ocenianym kraju.

Zazwyczaj całość powstaje w ciągu 12 miesięcy.

W ostatnich latach OECD publikuje także raporty ewaluacyjne dotyczące polityki regionalnej.

Pomocą w ewaluacji systemów i polityk B+R i innowacji są też wydawnictwa OECD *Science, Technology and Industry Outlook* oraz OECD *Science, Technology and Industry Scoreboard*, ukazujące się od II połowy lat 90.

**CREST/ERAC.** Inne serie raportów opartych na ocenie zagranicznych raportów wprowadził CREST. Komitet CREST (ang. *Scientific and Technical Research Committee*) – Komitet Badań Naukowo-Technicznych, organ doradczy wspomagający zarówno Radę Unii Europejskiej, jak też Komisję Europejską we wszystkich sprawach dotyczących polityki UE w dziedzinie badań i rozwoju technicznego. Komitet działał w latach 1974–2000. Do jego zadań należało m.in. dokonywanie ocen krajowych systemów badań i technologii w ramach Otwartej Metody Koordynacji oraz 3% Action Plan. CREST przeprowadził ewaluacje w ramach czterech cykli: I (2004), II (2006), III (2007), IV (2008).

Raporty i niektóre zadania CREST zostały podjęte przez ERAC (Komisję Europejskiej Przestrzeni Badawczej), powołaną w 2010 roku. ERAC jest, podobnie jak CREST, organem doradczym wspierającym politykę Rady Unii Europejskiej i Komisji Europejskiej w dziedzinie badań, innowacji i rozwoju technologicznego. Główną misją ERAC jest zapewnienie wkładu strategicznego w sprawach badań i innowacji istotnych dla rozwoju ERA. ERAC publikuje raporty ewaluacyjne od 2010 roku. Obecnie raporty unijne nazywane *peer review* sporządzane są w ramach Policy Support Facility<sup>15</sup>. *Peer Review* krajowych systemów badań i rozwoju to pogłębiona ocena systemu badań i innowacji danego kraju, przeprowadzana przez zespół ekspertów z różnych dziedzin polityki, która prowadzi do zaleceń operacyjnych dla władz krajowych w zakresie reform niezbędnych do wzmocnienia ich systemów B+R. *Peer Review* może przybrać formę ogólnej oceny silnych i słabych stron w dziedzinie B+R lub też może się koncentrować na konkretnych elementach systemów badań i innowacji (np. reformie uniwersytetów, systemie transferu wiedzy itp.), zgodnie z uzgodnieniami z krajem poddanym przeglądowi. Przegląd może być poprzedzony „przełogdem wstępnym” (etap przygotowawczy dla określenia obszarów fokusowych), a następnie „przełogdem po *Peer Review*” (jako kontynuacja kroków w celu dostarczenia konkretnych porad na temat sposobów wdrażania zaleceń)<sup>16</sup>.

**EU/UNU-MERIT.** W latach 2006–2009 na zlecenie Komisji Unii Europejskiej instytut UNU-MERIT w Maastrich realizował projekt badawczo-ewaluacyjny mający na celu zarówno pogłębienie rozumienia problematyki *policy mix* w dziedzinie B+R, jak i przeprowadzenie oceny „mieszanek polityki” naukowej w niektórych krajach członkowskich. W ramach projektu opublikowano wiele

---

<sup>15</sup> Instrument Wsparcia Polityki, instytucja analityczno-doradcza, stara się, na żądanie państw członkowskich EU i krajów stowarzyszonych, zapewniać najlepsze wzorce, niezależne ekspertyzy oraz zalecenia w dziedzinie polityki naukowej i innowacyjnej.

<sup>16</sup> <https://rio.jrc.ec.europa.eu/en/policy-support-facility/peer-reviews>. Coroczne raporty dotyczące polityki naukowej i innowacyjnej publikowano w ramach serwisów informacyjnych Komisji UE Inno-Policy Trendchart i Erawatch (połączonych w roku 2011). Obecnie corocznie ukazują się RIO (*Research and Innovation Observatory*) *Country Reports*.



raportów metodologicznych, problemowych i ewaluacyjnych, m.in. *UNU-Merit Policy Mixes for R&D in Europe 2009* oraz *UNU-Merit Monitoring and analysis of policies conducive to higher levels of R&D. Country Review Poland 2007*.

**Inne ewaluacje.** Ministerstwa ds. nauki i innowacji zlecają nieraz sporządzenie raportów niezależne od Unii Europejskiej i organizacji międzynarodowych – przez konsorcja firm konsultingowych i instytutów naukowych, np. *Evaluation of Government Funding in RTDI from a Systems Perspective in Austria (2009)* lub *Evaluation of the Finnish National Innovation System – Full Report (2009)*. Podstawowe metody to analiza dokumentów, analiza statystyk oraz rozmowy z kluczowymi interesariuszami<sup>17</sup>.

## 2.2. Policy mix w dziedzinie B+R i innowacji

Jak dotąd polityka rozwija się bardziej wskutek kolejnych bieżących „przyrostów” (dodawania nowych i usuwania dawniejszych instrumentów) niż wskutek świadomych i przemyślanych działań w celu konstruowania *policy mix*. Ta sytuacja powoli się zmienia, w miarę jak koncepcja ta staje się coraz bardziej znana politykom i szerokim kręgom zainteresowanych<sup>18</sup>.

Mieszanka polityki (ang. *policy mix*) w dziedzinie B+R i innowacji to zestaw instrumentów polityki, które oddziałując na siebie, wpływają na ilość i jakość inwestycji B+R w sektorze publicznym i prywatnym<sup>19</sup>. Celem koncepcji *policy mix* jest pomoc w uniknięciu stosowania różnych instrumentów polityki oddzielnie lub w sposób chaotyczny i nieskoordynowany<sup>20</sup>, co ma zapewnić polityce spójność (ang. *policy coherence*).

W ostatnich latach przedmiotem wielu analiz i dyskusji stał się problem właściwego doboru instrumentów, czyli stworzenia najlepszego tzw. *policy mix*. Pojęcie mieszanki polityki stało się bardzo popularne w OECD i w Unii Europejskiej. Stosowane sporadycznie w II połowie XX wieku, weszło do języka polityki w II połowie lat 90., a stało się tematem wielu książek, artykułów i raportów od połowy I dekady XXI wieku (WordCat).

Problematykę *policy mix* podejmuje tzw. trzecia generacja teorii wyboru instrumentu polityki: pierwsza generacja zajmowała się analizą oddzielnych instrumentów, druga generacja badaniami porównawczymi nad instrumentami,

<sup>17</sup> Technopolis, *Evaluation of the design and implementation of Estonian RTDI policy: implications for policy planning*, Final Report, December 2005.

<sup>18</sup> *Designing Policy Mixes: Enhancing Innovation System Performance and R&D Investment Levels*, March A study funded by the European Commission – DG Research 2009.

<sup>19</sup> *Policy Mixes for R&D in Europe*, UNU-MERIT, A study commissioned by the European Commission – Directorate-General for Research, April 2009.

<sup>20</sup> *Raising EU R&D Intensity. Report to the European Commission by an Independent Expert Group Improving the Effectiveness of the Mix of Public Support Mechanisms for Private Sector Research and Development*, European Commission 2003.

a trzecia – zagadnieniami projektowania i doboru optymalnej mieszanki polityki w złożonych sytuacjach decyzyjnych i wykonawczych<sup>21</sup>.

Koncepcja mieszanki polityki zwraca uwagę, że instrumenty polityki nigdy nie działają bez interakcji z innymi instrumentami, najczęściej (dobrze dobrane) wspomagają się wzajemnie lub też (niezharmonizowane ze sobą) obniżają wzajemnie swoją skuteczność albo też niwelują wzajemnie swoje efekty. Poszczególne instrumenty osiągają nieraz zakładany cel cząstkowy (np. zwiększenie wydatków biznesu na B+R), ale cel ten często ma szerszy sens społeczny tylko pod warunkiem, że jest sprzęgnięty z wieloma innymi celami, wspieranymi przez inne instrumenty<sup>22</sup>.

Nowatorstwo koncepcji *policy mix* polega na założeniu, że nie instrumenty stosowane w izolacji, tylko połączenie oddziałujących na siebie instrumentów politycznych wpływa na B+R i innowacje. Drugim kluczowym pomysłem jest założenie, że na B+R i innowacje wpływ mają nie tylko instrumenty z tego obszaru polityki (np. dotacje rządowe lub zachęty podatkowe itp.), ale również polityki z innych dziedzin, takie jak np. przepisy ochrony środowiska<sup>23</sup>.

Z reguły każdy instrument ma sens pod warunkiem istnienia innych instrumentów bądź też mechanizmów rynkowych i społecznych. Na przykład finansowany przez rząd „fundusz zasiewny” na wstępne fazy innowacji ma sens pod warunkiem istnienia źródeł finansowania kolejnych faz (np. kapitału ryzyka). W systemie finansowania badań naukowych wielkie projekty i programy badawcze, granty tradycyjne oraz granty specjalne, służące np. finansowaniu propozycji nadsyłanych przez młodych badaczy oraz idei nieortodoksyjnych i oryginalnych, uzupełniają się wzajemnie; granty pełnią funkcję „laboratorium nowych form” w stosunku do wielkich programów. Instrumenty finansowania badań przemysłowych prowadzonych w uczelniach mają sens pod warunkiem równoległego istnienia mechanizmów transferu technologii, mobilności naukowców i inżynierów pomiędzy nauką a gospodarką oraz partnerstwa publiczno-prywatnego. Korzyści ze wspierania prac B+R w przedsiębiorstwach wzrastają, gdy towarzyszy temu pomoc w zakresie zarządzania, podnoszenia kompetencji oraz wprowadzenia na rynek nowych produktów i usług. Myśląc o pojedynczym instrumencie nastawionym na osiągnięcie pewnego celu, politycy powinni brać pod uwagę szersze spektrum już istniejących lub możliwych do wprowadzenia narzędzi.

Mieszanki polityki można opisywać według dziedzin polityki, których dotyczy (*policy mix* może nie ograniczać się do jednej domeny), według agend

---

<sup>21</sup> M. Howlett i J. Rayner, *Design Principles for Policy Mixes: Cohesion and Coherence in 'New Governance Arrangements'*, „Policy and Society”, 01/2007, 26(4), s. 1–18.

<sup>22</sup> *Benchmarking Of National Policies Public And Private Investments In R&D*. Final Report Expert Group, OECD June 2002.

<sup>23</sup> *Policy-Mix Methodological Report* 2006.

polityki i uzasadnień polityki, według grup i procesów docelowych oraz według typu zastosowanych instrumentów<sup>24</sup>.

W polityce naukowej i innowacyjnej OECD dzieli *policy mix* instrumentów na mieszanki instrumentów wspierających: (a) podstawy nauki i technologii (ang. *science and technology base*), (b) B+R biznesu i innowacji, (c) powiązania nauki i gospodarki, (d) zasoby ludzkie dla N+T, (e) system zarządzania badaniami i innowacjami<sup>25</sup>.

Niezależnie od pojęcia mieszanki polityki proponuje się także pojęcie „mini-mieszanki” oznaczającej zestaw instrumentów (takich jak np. programy mobilności, bodźce fiskalne, granty, nowe regulacje) ukierunkowanych na osiągnięcie jednego konkretnego celu, takiego np. jak podnoszenie badań i poziomu inwestycji w określonych obszarach technologii czy wspieranie konkretnych grup docelowych (np. nowych firm opartych na technologii)<sup>26</sup>. Ewaluacja *policy mix* może się opierać na ustaleniach ewaluacji „minimieszanek”.

Pojęcie *policy mix* odnosi się także do łączenia ze sobą rozmaitych narzędzi analizy politycznej, zwłaszcza takich, które ze względu na szeroki udział grup nierządowych mają równocześnie charakter instrumentów politycznych. Między innymi dla wspomnienia procesu podejmowania decyzji zaleca się stosowanie kombinacji ewaluacji polityk, foresightu i wartościowania technologii (efekt synergii, wzajemnego wspomagania)<sup>27</sup>.

Zalecenia. Podstawowe wymiary mieszanki polityki to:

- głębokość – ranga instrumentów;
- gęstość – liczba instrumentów;
- luka pomiędzy instrumentami stosowanymi w danym kraju a stosowanymi na świecie (np. brak zachęt podatkowych na B+R lub innowacyjnych zamówień publicznych);
- równowaga między różnymi rodzajami instrumentów, np. instrumentami bezpośredniego wsparcia, jak dotacje, oraz instrumentami pośrednimi, jak zachęty podatkowe na B+R; lub pomiędzy konkurencyjnym finansowaniem projektów oraz finansowaniem instytucjonalnym<sup>28</sup>.

---

<sup>24</sup> P. Cunningham, J. Edler, K. Flanagan i Ph. Larédo, *Innovation policy mix and instrument interaction: a review*, NESTA, November 2013; OECD *Public Funding Of R&D: Emerging Policy Issues*, 13–14 March 2001.

<sup>25</sup> OECD, *Policy Mix for Innovation In Poland. Key Issues and Recommendations*, 2007.

<sup>26</sup> *Designing Policy Mixes: Enhancing Innovation System Performance and R&D Investment Levels*, March A study funded by the European Commission – DG Research 2009.

<sup>27</sup> S. Kuhlmann, *Evaluation as a source of strategic intelligence*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001, [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhworkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhworkshop.pdf).

<sup>28</sup> *Designing Policy Mixes: Enhancing Innovation System Performance and R&D Investment Levels*, March A study funded by the European Commission – DG Research 2009.

W latach 2006–2009 na zlecenie Komisji Unii Europejskiej instytut UNU-MERIT w Maastrich realizował projekt badawczo-ewaluacyjny mający na celu zarówno pogłębienie rozumienia problematyki *policy mix* w dziedzinie B+R, jak i przeprowadzenie oceny mieszanek polityki naukowej w niektórych krajach członkowskich. W ramach projektu opublikowano wiele raportów metodologicznych, problemowych i ewaluacyjnych, m.in. *UNU-Merit Policy Mixes for R&D in Europe 2009* oraz *UNU-Merit Monitoring and analysis of policies conducive to higher levels of R&D. Country Review Poland 2007*.

**Ewaluacja policy mix. Pierwszy krok.** Ocena dynamiki i wpływu pakietu programów i instrumentów wymaga, jako pierwszego kroku, odpowiedzi na pytanie, czy stosowany zestaw programów i instrumentów jest prostą sumą różnych narzędzi, czy też przemyślanym zestawem.

Na podstawie porównania poszczególnych instrumentów należy zbadać, jakie istnieją relacje pomiędzy instrumentami oraz pomiędzy poszczególnymi grupami instrumentów (np. instrumentami tematycznymi i ogólnymi; instrumentami wzmacniającymi podstawy nauki, B+R biznesu i innowacji, powiązania nauki z przemysłem, zasoby ludzkie itd.) – czy są one wobec siebie komplementarne, czy wzmacniają się wzajemnie, czy dublują się lub kolidują ze sobą; jaki wpływ na funkcjonowanie instrumentów polityki naukowej mają instrumenty innych polityk – ekonomicznej (finansowej, makroekonomicznej, podatkowej), konkurencyjności, handlu, innowacyjnej, przedsiębiorczości, edukacyjnej, zatrudnienia, rozwoju regionalnego, praw własności intelektualnej, ochrony konsumenta, obronnej, zdrowia, środowiska; jakie istnieją relacje pomiędzy instrumentami na różnych poziomach polityki – krajowym, unijnym i regionalnym? Czy ogólna równowaga pomiędzy instrumentami jest właściwa, czy też istnieje nadmiar pewnego typu instrumentów lub mają one zbyt wielką skalę? Czy poszczególne instrumenty są dobrze zaprojektowane dla rozwiązania problemu, który mają adresować? Czy struktura zarządzania instrumentami jest właściwa?

Ocena powinna też brać pod uwagę m.in. zasady ewaluacji programów i instrumentów, informacje na temat postrzegania ich przez naukowców oraz różnice między dyscyplinami badawczymi.

Jako punkt wyjścia ewaluacji byłoby użyteczne mieć dostęp nie tylko do danych zagregowanych dotyczących instrumentów finansowania, ale przede wszystkim danych w rozbiciu na dyscypliny i sektory instytucjonalne (szkoły wyższe, pozauczelniane instytucje naukowe, biznes, organizacje *non profit*), pochodzenie (publiczne, prywatne), typ (konkurencyjny, niekonkurencyjna) i cel (ogólny, ukierunkowany na rozwiązywanie konkretnych problemów).

**Ewaluacja policy mix. Drugi krok.** Drugi etap polega na analizie indywidualnych i łącznych skutków różnych instrumentów w takich wymiarach publicznego sektora nauki, jak struktury, powiązania z przemysłem, umiędzynarodowienie, wyniki i dostęp do danych badawczych. Oznaczałoby to bardziej całościowe oceny wpływu programów i instrumentów na publiczny system nauki, wykraczający poza ocenę wpływu poszczególnych instrumentów i unikający ryzyka

popadania w mechaniczne metody ilościowe (np. proste liczenie umów, patentów i publikacji).

Na podstawie porównania poszczególnych instrumentów należy zbadać, czy pakiet instrumentów właściwie odpowiada na najważniejsze problemy systemu nauki i innowacji w perspektywie długofalowych celów rządowych; czy ogólny *policy mix* jest spójny, adekwatny, skuteczny i efektywny, czy też czy istnieją ważne problemy, których istniejące instrumenty nie podejmują<sup>29</sup>.

### 2.3. Programy i instrumenty polityczne

Pojęcie „instrumentu politycznego” szybko i stosunkowo niedawno awansowało do roli jednego z podstawowych pojęć polityki publicznej. Statystyka użycia pojęcia „instrument polityczny” w tytułach książek (w języku angielskim: *policy instruments*, *policy measures*, *policy tools*) pokazuje, że pojęcie to zaczęło sięgać sporadycznie od lat 60., coraz częściej od końca lat 70., a lawinowo od początku lat 90. W dokumentach OECD pojęcie *innovation policy instruments* pojawia się od roku 1992. W roku 1997 Komitet Zarządzania Publicznego OECD opublikował raport poświęcony instrumentom politycznym jako takim (*Choice of policy instruments*, PUMA/SBO(97)9).

W odniesieniu do polityk naukowej, technologicznej i innowacyjnej, opracowanie *An introduction to Policy Analysis in Science and Technology*, UNESCO 1979, kodyfikujące standardy myślenia o polityce naukowej i technologicznej, nie uwzględniło jeszcze w ogóle tego pojęcia. Natomiast w dzisiejszych systemach monitoringu Unii Europejskiej (*INNO-Policy TrendChart*, *Erawatch* oraz przede wszystkim *Policy Mix Portal*) oraz OECD (*Science, Technology and Innovation Outlook*) „instrument polityczny” stał się już osią organizującą przekazywaną wiedzę o polityce naukowej, technologicznej i innowacyjnej.

Instrumenty polityczne definiuje się jako „narzędzia realizacji polityk przez instytucje publiczne, dzięki którym osiąga się założone cele, polegające na mobilizacji zasobów (finansowych, ludzkich, organizacyjnych) za pomocą programów i projektów, informacji (np. tzw. mapy drogowe, informacja i promocja, koordynacja), oraz prawodawstwo (ustawy, rozporządzenia). Instrumenty polityczne mogą przybierać formę dotacji; *quasi*-dotacji (vouchery); usług, takich jak bezpieczeństwo publiczne; przymusu (np. prawa ruchu drogowego lub prawodawstwo dotyczące środowiska); informacji (np. prognoza pogody); perswazji (ogłoszenia publiczne)”<sup>30</sup>.

Najczęściej instrumenty polityczne są częścią szerszej całości „interwencji politycznej” („inicjatywy politycznej”, „politycznego działania”). Interwencję polityczną określa się jako uprawniony akt organu rządowego (publicznego), mający na celu modyfikację instytucji społecznych, gospodarczych i politycznych

<sup>29</sup> OECD IPP, Innovation Policy Platform, *Issue Brief: Public Sector Research Funding*.

<sup>30</sup> OECD *Choice of policy instruments*, 1997.

państwa, regionu lub sektora gospodarki, podjęty w zgodzie z ogólnymi celami i priorytetami rządowymi. Interwencja polityczna to najbardziej ogólny termin oznaczający działanie polityczne. Po nim – na niższych szczeblach abstrakcji – następują polityki (horyzontalne lub sektorowe strategie polityczne), programy oraz instrumenty polityczne i polityczne decyzje.

Termin „instrument polityki” stosuje się do interwencji szczebla rządowego. Instrumentem polityki są także zarówno rządowe agencje, jak i prowadzone przez nie programy (w Narodowym Centrum Nauki nazywane „konkursami”).

Instrumenty polityki są zróżnicowane. Można je podzielić w różny sposób, np.:

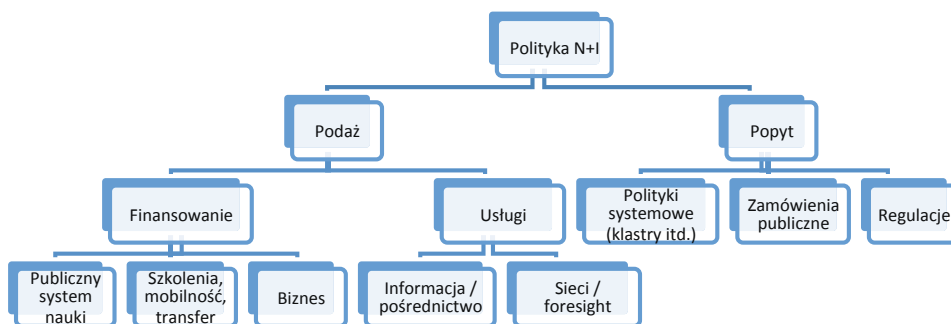
- według dominujących mechanizmów zastosowania, np. zachęty finansowe, regulacje i informacje (czasami określane żartobliwie jako „marchew, kij i kazanie”);
- według grup docelowych celów, np. firm, uczelni itp.;
- według celów, których mają one dotyczyć.

Z reguły poszczególne kraje mają pełny zestaw instrumentów polityki naukowej i innowacyjnej, ale sposoby ich opracowania i wdrażania mogą być bardzo różne<sup>31</sup>.

Spośród kilkudziesięciu klasyfikacji polityk nauki i innowacji – w ramach których tworzy się instrumenty – najbardziej przekonująca jest zaprezentowana poniżej (rysunek 2.4).

Rysunek 2.4

#### Klasyfikacja polityk nauki i innowacji



Źródło: L. Georghiou, E. Amanatidou, H. Belitz, L. Cruz, J. Edler, C. Edquist, O. Granstrand, J. Guinet, E. Leprince, L. Orsenigo, J. Rigby, J. Romanainen, M. Stampfer i J. van den Biesen, *Improving the Effectiveness of Direct Public Support Measures to Stimulate Private Investment in Research*, Report of the ETAN Working Group on Direct Measures for Directorate General Research, European Commission, 2002. Cyt. za: L. Georghiou, *Evaluating Foresight and Lessons for Its Future Impact*, The Second International Conference On Technology Foresight – Tokyo, 27–28 Feb. 2003, <http://www.nistep.go.jp/IC/ic030227/pdf/p6-1.pdf>.

<sup>31</sup> IPP, Innovation Policy Platform, *Policy goals and means*, 2014–01–13.

Pomocą dla badacza i praktyka ewaluacji jest europejskie Repozytorium Ewaluacji Polityki Naukowej i Polityki Innowacyjnej (SIPER). SIPER powstał dzięki Manchester Institute of Innovation Research (MIOIR), Uniwersytet w Manchesterze, w ramach projektu RISIS Komisji Europejskiej (<http://risis.eu/>). Baza danych SIPER jest w trakcie ciągłej aktualizacji.

Główne kategorie SIPER informują jednocześnie o kategoriach, w jakich opisuje się ewaluacje polityk naukowej i innowacyjnej:

#### 1. Charakterystyka instrumentu

- a) zakres geograficzny;
- b) grupy docelowe (jednostki, uczelnie, instytuty pozauczelniane, instytucje transferowe, firmy, sektory);
- c) tryb (bezpośrednie wsparcie finansowe: dotacje, pożyczki, gwarancje, umowy, stypendia, instytucjonalne dotacje blokowe; pośrednie wsparcie finansowe: zachęty podatkowe i fiskalne (np. kredyty na B+R); wsparcie infrastruktury; wsparcie niefinansowe (np. szkolenie, koordynacja i doradztwo, dostarczanie informacji); nagrody;
- d) cel (np. wzmocnienie edukacji; ułatwienie mobilności personelu (w tym poprawa kariery zawodowej; umiędzynarodowienie; wspieranie akceptacji społecznej; wzmocnianie doskonałości badawczej i zarządzania; poprawa zdolności absorpcyjnych; wspieranie współpracy i komercjalizacji badań; mobilizowanie dodatkowego (niepublicznego) finansowania innowacji (np. aniołów biznesu, kapitału ryzyka itp.); dyfuzja innowacji; wspieranie udziału i rozwoju kobiet i grup mniejszościowych w dziedzinie badań i innowacji).

#### 2. Charakterystyka ewaluacji

- a) zakres geograficzny;
- b) rok publikacji;
- c) ewaluator (wewnętrzny, zewnętrzny);
- f) „timing” (*ex ante*, monitoring, interim, *ex post*);
- g) cel (posumowujący, kształcący, inny);
- h) odniesienie do logiki programu (w pełni włączona, nie w pełni, w ogóle nie włączona).

#### 3. Kryteria

- a) odpowiedniość:
  - uzasadnienie,
  - cel,
  - projekt.
- b) spójność – komplementarność z innymi instrumentami;
- c) osiągnięcie celów / skuteczność;
- d) produkty i ich jakość;
- e) produkty i wpływ:
  - regionalny,
  - krajowy,

- międzynarodowy,
  - naukowy i technologiczny,
  - gospodarczy,
  - społeczny,
  - edukacyjny,
  - niezamierzony.
- f) zwrot z inwestycji / „wartość za pieniądze”;
  - g) efektywność wdrożenia programu;
  - h) dodatkowość:
    - wkładu,
    - produktu,
    - behawioralna.
4. Charakterystyka ewaluacji
- a) eksperymentalna;
  - b) *quasi*-eksperymentalna;
  - c) nieeksperymentalna.
5. Metody zbierania danych
- a) istniejące bazy danych;
  - b) ankiety;
  - c) wywiady;
  - d) grupy fokusowe;
  - e) *peer review*;
  - f) dane dotyczące własności intelektualnej (patenty itd.);
  - g) altmetryka;
  - h) dane CV;
  - i) wizyty studyjne.
6. Metody analizy danych
- a) studia przypadku;
  - b) analiza sieci;
  - c) analiza ekonometryczna;
  - d) statystyka opisowa;
  - e) analiza kosztów i korzyści, wkładów i efektów, analiza zwrotu z inwestycji;
  - f) analiza danych własności intelektualnej (np. analiza cytowań);
  - g) analiza danych altmetrycznych;
  - h) ilościowa i/lub jakościowa analiza tekstu<sup>32</sup>.

Wraz z „agencyfikacją”, czyli przekazywaniem realizacji programów i instrumentów z ministerstw do agencji rządowych (spóźniona o 20 lat w porównaniu z krajami zachodnimi fala agencyfikacji dotarła do Polski w ostatnim ćwierćwieczu), odpowiedzialność za ewaluacje przeszła głównie do rąk agencji rządowych i prywatnych fundacji, podczas gdy ministerstwa zostawiły sobie funkcje

---

<sup>32</sup> <http://si-per.eu/Search/MainSearch>.



strategiczne. Granice pomiędzy tymi dwoma typami organizacji – agencjami i fundacjami – bywają zatarte, gdyż fundacje korzystają często z pieniędzy rządowych lub unijnych, a ponadto rządy, projektując instrumenty polityczne, albo uwzględniają ich istnienie i działalność, albo także stwarzają warunki do ich rozwoju, np. wprowadzają specjalne zachęty podatkowe.

Agencje i fundacje prowadzą z reguły trzy rodzaje ewaluacji: monitoring wewnętrzny oraz ewaluacje wewnętrzne i zewnętrzne. „Monitoring służy bieżącej analizie prowadzonej działalności (stanowiąc swoisty system ‘wczesnego ostrzeżenia’ o zaistniałych trudnościach), włączeniu koordynatorów w ocenę prowadzonych przez siebie programów, oraz przeprowadzanej okresowo ewaluacji”<sup>33</sup>. Celem ewaluacji wewnętrznej jest „analiza zasadności i jakości prowadzonych działań, identyfikacja ‘punktów krytycznych’ oraz zalecenia na przyszłość”. Ewaluacja wewnętrzna obejmuje analizę procedur i dokumentów programowych, badania ankietowe wśród beneficjentów i innych osób zaangażowanych w program (np. recenzentów), wywiady z beneficjentami oraz innymi interesariuszami, analizę danych statystycznych oraz analizę kontekstu programów (porównanie ich z ofertą innych instytucji). W końcu, celem ewaluacji zewnętrznej jest też zarówno ocena poszczególnych programów, jak i ocena całokształtu działań organizacji<sup>34</sup>. Ten ostatni typ ewaluacji zostanie omówiony oddzielnie.

#### **2.4. Projekty (realizowane w ramach poszczególnych programów agencji grantowych)**

Agencje uznaje się za zjawisko charakterystyczne dla przemian administracji publicznych (nowe zarządzanie publiczne). Podkreśla się ich zróżnicowanie oraz płynność przejścia do innych form organizacyjnych. Zwraca się uwagę na ich usytuowanie pomiędzy instytucjami rządowymi a organizacjami realizującymi zadania publiczne bądź też usługobiorcami<sup>35</sup>.

Cechy charakterystyczne agencji to:

- pełnienie funkcji publicznych, w szczególności:
  - finansowanie badań i działalności innowacyjnej;
  - doradztwo dla rządu;
  - rola mediacyjna, łącznikowa i buforowa pomiędzy rządem a środowiskiem naukowym (lub rządem, środowiskiem naukowym i innymi interesariuszami);

---

<sup>33</sup> Fundacja Nauki Polskiej, [http://www.fnpp.org.pl/o\\_fundacji/ewaluacja-programow-fundacji/](http://www.fnpp.org.pl/o_fundacji/ewaluacja-programow-fundacji/).

<sup>34</sup> Ibidem.

<sup>35</sup> W sferze nauki i innowacji agencje zaliczają się do szerszej grupy „organizacji pomostowych” lub „pośredniczących”, do których należą także wielkie organizacje badawcze (np. CNRS), rady konsultacyjne, konferencje rektorów szkół wyższych, programy badawcze i technologiczne (organizacje zadaniowe typu *ad hoc*), centra doskonałości (finansowane na zasadach konkursowych, np. austriacki Kplus Competence Centrem Programme), sieci i konsorcja badawcze (finansowane na zasadach konkursowych) itd.

- ograniczony, choć zróżnicowany zakres niezależności, pozostawianie „w zasięgu ręki” rządu reprezentowanego przez odnośne ministerstwo;
- finansowanie z budżetu państwa;
- status służby publicznej zatrudnionych.

Tak zdefiniowane agencje można podzielić według następujących kryteriów:

- a) typ badań (podstawowe, stosowane, prace rozwojowe);
- b) cel badań – badania dla rozwoju wiedzy lub badania dla rozwoju innowacji (wraz z innymi formami działalności naukowo-technicznej);
- c) uniwersalność dziedzinowa vs dyscyplinowa lub sektorowa specjalizacja;
- d) uniwersalność zadań vs realizacja specjalnych zadań.

Stosując te kryteria, otrzymujemy sześć grup, z których trzecia – agencje technologiczne – odpowiada kategorii wskazanej w zapytaniu (tabela 2.3).

**Tabela 2.3**

**Typy agencji i rodzaje finansowanej działalności**

Rodzaj finansowanej działalności	Typ agencji
Badania podstawowe i/lub strategiczne	Agencje całościowe, rady ds. badań
Badania podstawowe i/lub strategiczne	Agencje dyscyplinowe, rady ds. badań (np. brytyjskie)
Badania stosowane i prace rozwojowe	Agencje technologiczne
B+R w określonych sektorach gospodarki	Agencje sektorowe
B+R na rzecz realizacji określonych zadań	Agencje specjalne zadaniowe
Projekty innowacyjne w przedsiębiorstwach	Agencje innowacyjne

Źródło: opracowanie własne.

Typ agencji przesądza o wielu rozwiązaniach organizacyjnych. Na przykład agencje ukierunkowane na wspieranie badań podstawowych częściej w ocenie propozycji biorą pod uwagę tylko kryteria doskonałości badawczej, rzadziej niż inne typy agencji stosują priorytety badawcze (finansując, w ramach jednej lub kilku agencji, cały obszar badań naukowych), częściej nastawione są na kształcenie – poprzez dotacje – badaczy (a w mniejszym stopniu na to, by osiągnęli oni założone rezultaty).

Agencje z reguły stosują metodę *peer review*. *Peer review* to osąd wartości naukowej (badaczy, publikacji, instytucji, dyscyplin) dokonywany przez samych badaczy pracujących w tej samej (lub zbliżonej) dyscyplinie. *Peer review* oparte jest na założeniu, że ocena pewnych aspektów nauki – przede wszystkim jej jakości – powinna zostać pozostawiona ekspertom posiadającym dostateczną wiedzę o rozwoju ocenianego pola badań i uprawiającym to pole. W zasadzie oceny dokonuje się anonimowo. Badacze mogą występować w roli wykonawców lub w roli oceniających; nigdy jednak w obu rolach jednocześnie. Ocenia-

jący nie powinni mieć osobistego interesu w wyniku oceny. Najczęściej *peer review* polega na ocenie dokonanej przez agencje grantowe, czasopisma naukowe i wydawców, organizatorów konferencji oraz przez instytucje naukowe (Kogo zatrudnić? Kto zasłużył na awans? itp.).

Termin *peer review* (dosł. ocena dokonana przez równych) jest nieco mylący; oceniać powinni najlepsi, a nie przeciętni. Zazwyczaj głównym kryterium oceny jest wkład badań dla nauki oraz zgodność z misją instytucji. *Peer review* ma być mechanizmem kontroli jakości w nauce, metodą odróżniania „ziaren od plew”, oddzielania idei istotnych i oryginalnych od wtórnych i błahych. *Peer review* spełnia jednocześnie dwie przeciwstawne funkcje. Jako uznany przez większość środowiska mechanizm wzmacnia autonomię nauki. Dla podatnika stanowi gwarancję, że publiczne środki na rozwój badacze dzielą między sobą w sposób odpowiedzialny<sup>36</sup>. Istnieje kilka typologii *peer review*. Według jednej z nich dzieli się ono na bezpośrednie (ocena przez ekspertów z tej samej dyscypliny oparta na kryteriach naukowych) oraz zmodyfikowane, rozszerzone (uzupełnienie kryteriów naukowych o kryteria pozanaukowe, wpływu na rozwój technologiczny, gospodarczy, społeczny).

W Polsce terminy *peer review* i *peers* (oceniający) są mało znane. Mówi się o „systemie grantowym”; unika określeń obejmujących cały system ocen.

Zazwyczaj agencje stosują dwu- lub trójfazowy proces oceny wniosków, np.

- faza preselekcji – recenzenci oceniają na podstawie skróconych wniosków, odrzucają wnioski, które są wyraźnie słabsze, przekraczają możliwości finansowe agencji lub też nie spełniają pewnych formalnych wymogów;
- faza selekcji – recenzenci oceniają według skali pakiet kilku, kilkunastu wniosków, biorąc pod uwagę ustalone wcześniej kryteria, takie jak np. dorobek aplikanta (zespołu badawczego), założenia metodologiczne i opis planu zarządzania, a także wartości instytucji (jednej lub kilku), w których mają zostać przeprowadzone badania. Recenzenci czytają większą liczbę wniosków grantowych, przez co mają lepszą skalę porównawczą;
- panel ekspercki, na który mogą być zapraszani wnioskodawcy (którzy wcześniej otrzymują anonimowe recenzje).

Często proces oceny łączy „recenzje na odległość” z „panelami eksperckimi”, w których recenzenci wspólnie obradują (nieraz także spotykając się z wnioskodawcami). Z reguły recenzenci składają oświadczenia o braku konfliktu interesów, a agencje zapewniają wnioskodawcom możliwość apelacji (w zasadzie w sprawach formalnych, a nie od ocen merytorycznych). Wielkie i kosztowne projekty z reguły poddawane są ocenie śródkresowej. Na ogół agencje utrzymują bazy danych z opisami publikacji (lub innego typu efektów) powstałych dzięki dotacji. Informacje o grantach, z których sfinansowano badania, podają międzynarodowe bazy danych, takie jak np. Web of Science. Zasadą jest zapraszanie recenzentów z zagranicy. Regułą w agencjach są wewnętrzne komórki

<sup>36</sup> OECD IPP Innovation Policy Platform *Issue Brief Peer Review*, 2014.

ewaluacyjne, które oceniają (m.in. na podstawie ankiet do uczestników oraz osiągniętych rezultatów) sukces programu. Sprawozdanie z ewaluacji przesyłane jest do zarządu, który decyduje, czy w jego świetle przedłużyć, zawiesić lub ulepszyć program. Agencje oceniają także samych recenzentów (czy są terminowi, wnikliwi i bezstronni).

Aby zaradzić wbudowanym w *peer review* skrzywieniom, agencje wprowadzają modyfikacje promujące kategorie badaczy krzywdzonych wcześniej przez ich decyzje: młodych, zainteresowanych badaniami interdyscyplinarnymi, zdecydowanych na podjęcie ryzyka badań pionierskich, planujących przeprowadzenie badań na rzecz przemysłu. I tak np. zgłaszane są pomysły tzw. „dotacji zasiewnych” dla badaczy zdecydowanych na zmianę zainteresowań badawczych i wkroczenie na nowy obszar problemowy oraz dotacji połączonych z ulgami podatkowymi na badania przemysłowe<sup>37</sup>.

Aby zmniejszyć skrzywienia tradycyjnego *peer review*, w wielu krajach zmodyfikowano jego procedury.

Dla przeciwdziałania dyscyplinowemu separatyzmowi amerykańska Narodowa Fundacja Nauki już w latach 60. za radą Alvina Weinberga włączyła do paneli dyscyplinowych przedstawicieli innych, sąsiednich dyscyplin. Zadaniem ich jest ocena, czy przedkładane propozycje mają szansę wywrzeć na naukę wpływ wykraczający poza dyscyplinowe opłotki. Aby osłabić syndrom „wieży z kości słoniowej”, do paneli rozstrzygających o dotacjach na rozwój badań stosowanych włącza się nieraz przedstawiciele użytkowników (np. inżynierów zatrudnionych w przedsiębiorstwach), a w kryteriach oceny wniosków zwiększa znaczenie informacji o planach wykorzystania wyników badań.

Często reformy polegają nie tylko na modyfikacji procedur, ale także na wprowadzeniu nowych form grantów. Celem wzmocnienia edukacyjnego oddziaływania badań agencje grantowe wprowadziły dotacje wspierające powiązania badań z dydaktyką<sup>38</sup>. Aby związać badania z rządowymi priorytetami, holen-

---

<sup>37</sup> Np. program Small Grants for Exploratory Research (SGERR) amerykańskiej National Science Foundation (NSF) finansuje badania o charakterze rozpoznawczym i wysokim stopniu ryzyka. Warunkiem uzyskania dotacji jest przedstawienie krótkiego (dwie do pięciu stron) wniosku na badania nad nowymi ideami, nowymi obszarami problemowymi lub nad zastosowaniem nowych metod. Stypendia nie przekraczają 100 tys. dolarów i są z reguły niższe od przeciętnego grantu badawczego. Zazwyczaj realizacja projektu trwa rok, czasem dwa lata. Odejściem od standardowej formuły *peer review* są także elitarne stypendia wprowadzane przez brytyjską Radę ds. Badań (ok. 2–3 w ciągu roku). Mają one na celu „wykupienie” wybitnych nauczycieli akademickich od obowiązków dydaktycznych (przez oznaczony okres uniwersytet nie wypłaca stypendyście pensji, udziela mu natomiast „*sabbatical year(s)*”; stypendysta pobiera od MRC wynagrodzenie równe wypłacanemu wcześniej przez uniwersytet oraz grant na realizację projektu badawczego). Zgłoszenia na stypendium składane są jednocześnie z propozycją projektu badawczego. Warunkiem uzyskania stypendium jest wysoka ocena zarówno dorobku badacza, jak i planu badań.

<sup>38</sup> Np. holenderska NWO program Stimulans (na rozwój studiów doktoranckich), Research Corporation, prywatna fundacja w Tuscon (Arizona) i Cotrell Scholars Awards (na badania z udziałem doktorantów).

derska NWO udziela na zasadzie konkursu granty na podjęcie badań nad wskazanymi tematami lub na rozwiązywanie określonych problemów.

**Warunki skuteczności.** *Peer review* działa dobrze, gdy istnieje równowaga pomiędzy liczbą propozycji nowatorskich a pulą funduszy. Gdy propozycji nowatorskich jest zbyt mało, marnują się pieniądze; gdy jest ich zbyt wiele, marnują się talenty i szanse badań.

Zdaniem M. Gibbonsa *peer review* nadaje się szczególnie do oceny dojrzałych pól badań, legitymujących się własną społecznością międzynarodową, własnymi czasopismami, konferencjami i instytucjami; pól badań stosunkowo oddalonych od praktyki; badań prowadzonych w krajach cieszących się choćby minimalnym wzrostem gospodarczym. Odwrotnie, jest mniej użyteczny, gdy środki na badania się kurczą, badania są ukierunkowane na problemy o charakterze praktycznym, gdy potrzeba badań interdyscyplinarnych i gdy oceniający są zarazem wykonawcami (jak to się dzieje w wypadku wielkich programów naukowych). *Peer review* sprawdza się zatem tylko pod pewnymi warunkami.

**Zarzuty przeciwko *peer review*.** Jeden z badaczy *peer review* ocenił system z przekąsem: „Gdyby ktoś chciał opublikować badania oparte na metodzie o wiarygodności podobnej do *peer review*, bez dwóch zdań spotkałby się z odmową druku”. Od swoich narodzin *peer review* jest stawiane pod pręgierzem krytyki. Krytyka dotyczy zasad, praktyki oraz nieuniknionych i ubocznych, ujemnych następstw tego systemu. Kiedy naukowcy mają opiniować innych naukowców, trudno uniknąć konfliktu interesów. Mogą oni ulec pokusie, aby utrudnić publikację wyników sprzecznych z wynikami własnej pracy, lub opóźnić publikację badania podobnego do ich własnego, gdy w środowisku naukowym istnieje ciśnienia „publikuj pierwszy”. To budzi istotne wątpliwości dotyczące odpowiedzialności i jawności *peer review*.

Inny ważny problem z *peer review* dotyczy trudności uznawania heterodoksyjnych lub przełomowych badań. Peers – najczęściej starsi naukowcy – mogą mieć trudności z akceptacją nowych paradygmatów. Najczęściej *peer review* zarzuca się, że tłumi w nauce nowatorstwo, czyli jej najbardziej pożądaną cechę. Wielkie odkrycia rodzą się dziś częściej na styku dyscyplin niż w ich obrębie. *Peer review* z reguły premiuje badania nad okrzepłymi problemami zgłaszanyymi przez ustabilizowane dyscypliny. Rzadko dostrzega problemy wyłaniające się na styku różnych pól badań. Badaczom wkraczającym na zupełnie nowe obszary problemowe brak w agencjach grantowych siły przebiccia. *Peer review* wzmacnia ustalone dyscypliny i dominujące paradygmaty badawcze. Utrwala *status quo*. Uśrednia; eliminuje najgorszych, ale też i najlepszych. Projekty rzeczywiście pionierskie nie mogą być oceniane przez *peers*, gdyż takich – na nowym obszarze badań – jeszcze nie ma. Szansę uzyskania grantu na badania ryzykowne i nieortodoksyjne mają z reguły uczeni zatrudnieni w prestiżowych placówkach lub legitymujący się imponującym dorobkiem. Często jednak, gdy osiągną oni szczybel w karierze pozwalający na zgłaszanie rewolucyjnych tez, nie mają w sobie wewnętrznego żaru, by burzyć zastany świat.

Zwraca się uwagę, że gdyby dzisiaj Einstein ubiegał się o grant na opracowanie teorii względności, jego wniosek przepadłby z kretesem. W chwili składania wniosku Einstein nie miałby żadnego poważnego dorobku. Na wnioskowany temat nic w ogóle nie napisał. Problematyki nie studiował na uniwersytecie. Nie pracował naukowo i nie miał żadnej afiliacji uniwersyteckiej. Był jedynie niskim rangą urzędnikiem patentowym. Nie przeprowadził badań pilotażowych, nie rozporządzał wstępnymi danymi. Nie znał nikogo, kogo mógłby zaproponować na kierownika tematu. Zaproponowałby zatem samego siebie! Dziedzina jego badań znajdowała się poza obszarem zainteresowań wydziałów szkół wyższych. Każdy poważny profesor powitałby problem powiązań masy i energii z uśmiechem na twarzy. Einstein nie mógłby też przedstawić żadnych sensownych zastosowań praktycznych uzyskanych wyników.

Trudno ustalić, co to znaczy „dobra nauka” czy „oryginalność” (czy chodzi o oryginalność metod, danych, podejść, ustaleń, koncepcji, tematów, obszarów...).

*Peer review* często źle radzi sobie z oceną badań inter- i multidyscyplinarnych (wyjściem jest tworzenie osobnej kategorii oceny).

Podobnie, *peer review* ma nieraz problemy z oceną badań ukierunkowanych na rozwiązywanie praktycznych problemów, nie sprawdza się w ocenie badań stosowanych, ukierunkowanych na pomoc w realizacji zadań zewnętrznych w stosunku do nauki (tutaj wyjściem jest stosowanie podwójnych ocen w ramach rozszerzonego *peer review*: oceny dotyczącej jakości badań oraz oceny dotyczącej użyteczności społecznej i gospodarczej).

Z reguły prowadzi do odrzucenia negatywnych ustaleń, a te są równie istotne dla rozwoju nauki, co wyniki pozytywne.

Nie zapewnia powiązania oceny *ex ante* z oceną *ex post* (w jaki sposób dzisiejszy wnioskodawca wykorzystał przyznane dawniej fundusze, jakie uzyskał wyniki?).

Jest podatny na różnego typu „skrzywienia”: dyskryminuje badaczy o niższej pozycji naukowej, pracujących w mniej prestiżowych instytucjach, reprezentujących podejścia dalekie od „głównego nurtu”.

Jest kosztowny: koszty pośrednie (czas wnioskodawców i recenzentów) i bezpośrednie (koszty administracyjne).

Cykl pozyskiwania grantu jest niezgodny z cyklem twórczości naukowej. Perspektywa czekania dziewięciu czy dwunastu miesięcy na dotację potrafi zdusić zapał badacza, który doznał objawienia zarysu nowej naukowej prawdy.

W sytuacji, gdy dorobek jest zazwyczaj istotnym czynnikiem oceny, młodzi badacze stoją często na straconych pozycjach. *Peer review* działa zgodnie z opisanym przez socjologa Roberta Mertona „efektem Mateusza”: tym, co mają, dane jest więcej. System dyskryminuje badaczy o szczuplejszym dorobku naukowym, pracujących w mniej prestiżowych instytucjach.

Częstym zarzutem stawianym *peer review* jest jego podatność na kumoterstwo. Oceniający tworzą nieraz tzw. „old boy's network”, czyli grupę badaczy

stale popierających wzajemnie swoje wnioski. Kliki lubią się utrwaląć. Zarzut kumoterstwa trudno udowodnić.

*Peer review* zakłada obiektywizm i uczciwość ocen. Nie zawsze jednak recenzenci spełniają pokładane w nich nadzieje. Choć rzadko, zdarza się, że przechwytyują oni idee odrzuconych przez siebie propozycji. Rozproszenie odpowiedzialności za decyzje na kilku oceniających pozwala recenzentom ukryć stronniczość. System oceny propozycji badawczych pozwala na manipulacje (zanizanie punktów konkurencyjnych wniosków).

Recenzenci zgadzają się zazwyczaj w ocenie wyjątkowo wybitnych i wyjątkowo marnych wniosków. Różnią się jednak często w ocenie pozostałych propozycji (czyli ich zdecydowanej większości). Różnice ocen wpływają z różnic wykształcenia, perspektyw badawczych, kalibru naukowego oceniających oraz ich doświadczenia w roli recenzenta. Słabsi recenzenci dyskwalifikują nieraz projekty na podstawie uchybień nieistotnych (jak np. brak wypełnienia całej powierzchni formularza wnioskowego) lub nieprawdziwych (jak np. brak pewnego koniecznego elementu, który recenzent przeoczył).

Często zdarza się, że wniosek posiadający bezdyskusyjne walory zostaje oddalony z bardzo wątpliwego powodu. Propozycja badań nad niezwykle ważnym problemem, z zastosowaniem nowatorskich metod i siłami zespołu światowej klasy, może zostać odrzucona, gdy recenzenci dojdą do przekonania, że jej właściwym adresatem powinien być przemysł. Recenzenci zgodni zupełnie w ocenie pewnego wniosku mogą swoją ocenę wyrażać przyznaniem różnej liczby punktów. Wnioski łatwo odrzucać z uzasadnieniem, że są zbyt nachylone praktycznie, lub przeciwnie, że nie dają podstaw do oczekiwania przyszłych zastosowań; kategorie „ładunku teoretycznego” i „użyteczności” są bardziej intuicyjne niż uchwytne empirycznie.

Podczas gdy badacze spędzają nad przygotowaniem wniosku od trzech do sześciu miesięcy, dyskusja nad nim rzadko przekracza 10 do 20 minut. Proporcje takie częściej wynikają z przepracowania niż lenistwa oceniających.

Nie sposób ustalić optymalnej metody wyboru recenzentów. Nierzadko mniej wybitni oceniają wybitniejszych. Często pole badań jest zbyt małe, by mechanizm *peer review* mógł być stosowany.

Na podstawie propozycji niepodobna przewidzieć wyniku badań. Wnioskodawcy mają skłonność do pisania optymistycznych scenariuszy i składania obietnic, których nie są w stanie spełnić. Brak rzetelnej oceny raportów końcowych zachęca do czynienia obietnic bez pokrycia.

*Peer review* zarzuca się też niedostateczny *feedback* wnioskodawca – recenzenci (dlaczego mój wniosek został odrzucony?) oraz brak mechanizmu apelacyjnego (dlaczego nie mam prawa odwołać się od krzywdzącej decyzji?). Zwraca się też uwagę, że recenzenci rzadko wiążą ocenę *ex ante* z oceną *ex post* (w jaki sposób dzisiejszy wnioskodawca wykorzystał przyznane dawniej fundusze, jakie uzyskał wyniki?).

*Peer review* kosztuje; wysokie są jego koszty pośrednie (czas wnioskodawców i recenzentów) i bezpośrednie (koszty administracyjne). *Peer review* stawia się też zarzuty ogólne. Mówi się, że wzmacnia syndrom zamknięcia uczonych w wieży z kości słoniowej, zwiększa dyscyplinowy separatyzm oraz nie daje podstaw do podziału funduszy pomiędzy dyscyplinami (a nie wewnątrz dyscyplin).

Zarzuty przeciwko *peer review* można podzielić na takie, które dotyczą jego (niemożliwej do zmiany) istoty, oraz takie, które odnoszą się do jego (możliwych do modyfikacji) cech zewnętrznych. Niektóre zarzuty trafiają w system jako taki, inne – w jego użyteczność do oceny badań w danym kraju, w danej dyscyplinie lub w danym okresie (dobry dawniej, przeżytek dziś).

Anglicy, słysząc zarzut, mają w zanadrzu pytanie: A co proponujesz w zamian? Stosownie do różnic w charakterze obiekcji, również i propozycje zmian *peer review* rozciągają się od kodeksów „dobrego recenzenta i panelisty”, poprzez sugestie ulepszeń tego systemu, aż po propozycje wprowadzenia nowego.

Z jednym zgadzają się wszyscy: funkcjonowanie systemu należy stale kontrolować i monitorować. Co kilka lat każda agencja grantowa powinna zlecić zbadanie działania *peer review* niezależnym ekspertom. Trudno znaleźć poważną agencję za granicą, która by zaniechała tego obowiązku. W Stanach Zjednoczonych sporządzono raporty o *peer review* m.in. National Science Foundation (NSF) oraz National Institute of Health. W Wielkiej Brytanii raporty o działalności rad ds. badań przygotowały m.in. Royal Society oraz Committee on Science and Public Policy.

**Propozycje usprawnienia.** Sformułowano wiele propozycji usprawnienia *peer review*. Zaleca się, aby system był bardziej przejrzysty. Agencje grantowe powinny informować o zasadach doboru recenzentów i panelistów; oceniani – mieć prawo do repliki nie po, lecz przed wydaniem decyzji; wnioskodawcy – otrzymywać niezadawkowe informacje o powodach odrzucenia ich propozycji. Powinno się dopuszczać możliwość składania apelacji. Należy odwrócić kwestię anonimowości: wnioski powinny być anonimowe dla recenzentów i panelistów (z usunięciem możliwie wszystkich oznak pozwalających na identyfikację autora), natomiast nazwiska oceniających powinny być ujawniane wnioskodawcom. *Peers* powinni brać pełną odpowiedzialność za swój werdykt, a nie ukrywać się za zasadą anonimowości.

Proces *peer review* należy usprawnić. Formularze wnioskowe powinny być krótsze, a terminy składania propozycji zniesione (panele powinny pracować stale, a nie sesyjnie). Okres realizacji projektów badań podstawowych należałoby przedłużyć. Bazy danych o recenzentach trzeba częściej weryfikować. Co pewien czas powinno się badać skuteczność zasad punktacji i wprowadzać potrzebne zmiany. Powinno się utrzymywać bazę o efektach badań i sięgać po jej dane dla oceny wniosków.

Niektóre z propozycji modyfikacji *peer review* polepszają pewne parametry systemu kosztem innych (np. *feedback* i możliwość apelacji od decyzji zwiększają i tak wielkie obciążenie administracyjne). Inne niosą za sobą nowe możliwe



pułapki (odwrócenie kwestii anonimowości). Zazwyczaj nie sposób poprawić pewnej cechy systemu, nie pogarszając przy tym innej.

**Propozycje alternatywne.** Od czasu do czasu pojawiają się w poważnych pismach radykalne propozycje odejścia od *peer review*. Zamiast finansowania projektu proponuje się powrót do finansowania instytucji, najlepiej według pewnego ustalonego algorytmu.

Inne propozycje finansowania wykonawców dotyczą nie instytucji, tylko badaczy lub zespołów badawczych. Rosemary Yalow zaproponowała finansowanie badań biomedycznych poprzez trzyletnie dotacje, odnawiane pod warunkiem osiągnięcia sukcesu naukowego. Odrębny mechanizm regulowałby wchodzenie do systemu nowych badaczy. Podobna propozycja zaleca wynagradzanie już dokonanych wyników badań, a nie ich projektów. Ma ona za sobą stwierdzony empirycznie fakt, że wnioskodawcy, którzy nie dostali grantu, najczęściej i tak wykonują zaplanowane badania pomimo niekorzystnego wyroku agencji. Zgłoszono też propozycje finansowania zespołów badawczych na podstawie przeprowadzanej co kilka lat międzynarodowej ewaluacji. Richard Muller (USA) przedstawił pomysł stałej ankietyzacji badaczy mającej na celu ustalenie listy 1000 najwybitniejszych uczonych amerykańskich. Wybrańcy otrzymywaliby po milionie dolarów, który mogliby wykorzystać według własnego wyboru. Zgłoszono też ideę wybierania propozycji badawczych na zasadzie loterii. Eksperti odrzucaliby tylko te wnioski, które nie spełniałyby pewnych warunków brzegowych. O wyborze zwycięskich wniosków decydowałby ślepy los. Odpadłyby wysokie koszty administracyjne.

Jak dotąd każda zgłaszana alternatywa *peer review* natychmiast wzbudzała uzasadnione obiekcje. Podkreślano, że rozwiązując pewne problemy, propozycje alternatywne stwarzałyby nowe, często potencjalnie groźniejsze. Na przykład system finansowania według algorytmu skłaniałby do publikowania „byle szybko, byle dużo” oraz do obniżania wymagań związanych z uzyskiwaniem stopnia doktora. Propozycja Mullera, podobnie jak tradycyjne *peer review*, utrwałaby *status quo* (nowe kierunki i dyscypliny nie miałyby szans przebicia).

**Otwarte *peer review*.** Epoka cyfryzacji i Internetu zgłasza nowe propozycje ulepszeń *peer review*. Dokument OECD stwierdza: „istnieją obawy, że system *peer review* znajduje się pod presją i traci zaufanie wśród użytkowników, gdyż zależy on od subiektywnych decyzji kolegów i jest coraz bardziej czasochłonny i kosztowny. Nie jest przesadą stwierdzenie, że *peer review* jest obecnie w obliczu najsilniejszych wyzwań od kilku dekad”<sup>39</sup>.

Podkreśla się, że *peer review* sprawdza się tylko pod pewnymi warunkami. Wiele wskazuje na to, że warunki te zmieniają się. „Wskaźniki sukcesu” wniosków grantowych stale maleją. Aby zaradzić spadkowi szans, badacze piszą i przesyłają coraz więcej propozycji, tracąc w ten sposób czas, jaki mogliby poświęcić badaniom. Już w roku 1961 Leo Szilard obliczył, że gdy „wskaźnik

<sup>39</sup> OECD *Peer Review: Its Uses, Demands, And Issues*, 2007.

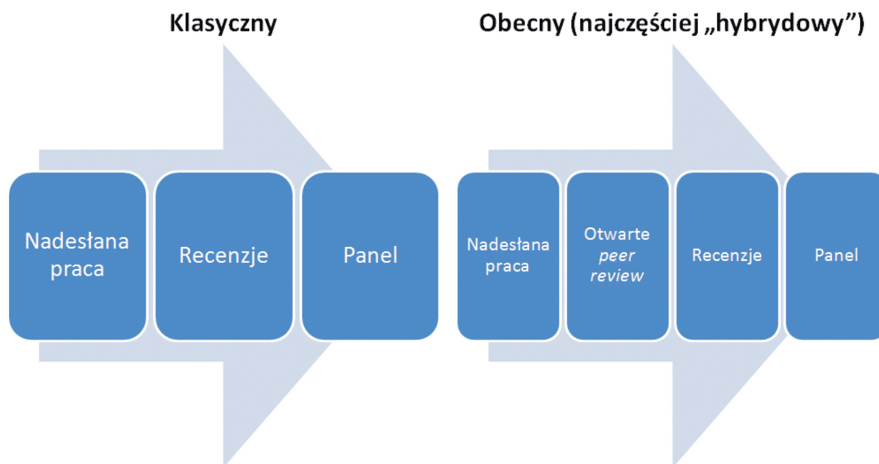
sukcesu” spadnie do 10%, 90% badaczy poświęci cały swój czas pisaniu propozycji badawczych (zaniedbując przy tym badania, dydaktykę i publikacje). W pewnych dziedzinach bliscy jesteście osiągnięciu „punktu Szilarda”. Jedną z form odpowiedzi jest otwarte *peer review*, którego pionierem są czasopisma i repozytoria naukowe<sup>40</sup>.

---

<sup>40</sup> Podrozdział oparty na następujących publikacjach: A. Mulligan, *Is peer review in crisis?*, „Oral Oncology” 2005, 41; A.L. Porter i F.A. Rossini, *Peer Review of Interdisciplinary Research Proposals*, „Science, Technology, & Human Values”, Summer 1985, vol. 10, no. 3; A. Scott, *Peer review and the relevance of science*, „Futures” 39, 2007; A. McQueen i A. Beard, *A Creative Approach to Assessing Proposals*, University of Malta, The Edward De Bono Institute For The Design And Development Of Thinking, *The Fifth International Conference On Creative Thinking: Designing Future Possibilities* 21–22nd June 2004; B.S. Noveck, *Peer to Patent. Collective Intelligence, Open Review and Patent Reform*, 2006; C. Berkenkotter, *The Power and the Perils of Peer Review*, „Rhetoric Review”, Spring 1995, vol. 13, no. 2; Ch. Smit, *Peer Review: Time for a Change?*, „BioScience”, September 2006, vol. 56, no. 9; D.E. Chubin i E.J. Hackett, *Peerless Science. Peer review and US Science Policy*, 1990; D.F.J. Campbell, *The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria*, [w:] *Learning from Science and Technology Policy Evaluation*, red. P. Shapira, S. Kuhlmann, Cheltenham, UK, Northampton, MA, USA: Edward Elgar, 2003; D.H. Guston, *The expanding role of peer review processes in the United States*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001, [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhworkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhworkshop.pdf); ESF, *Evaluation in National Research Funding Agencies, approaches, experiences and case studies*, A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes, September 2009; B. Bozeman i J. Melkers, red., *Evaluating R&D Impacts: Methods and Practice*, Springer, Boston 1993; F.Q. Wood, *The Peer Review Process*, Canberra Australian Govt. Pub. Service, 1997; G. Ellison, *Is Peer Review In Decline?* NBER Working Paper Series 2007; H. Diane i S. Krzys, *Peer Review in Academic Promotion and Publishing: Its Meaning, Locus, and Future*, Center for Studies in Higher Education, Berkeley 2011; H.W. Marsh, U.W. Jayasinghe i N.W. Bond, *Improving the Peer-Review Process for Grant Applications. Reliability, Validity, Bias, and Generalizability*, „American Psychologist”, April 2008, vol. 63, no. 3; House of Commons. Science and Technology Committee, *Peer review in scientific publications*, Eighth Report of Session 2010–12, 2011; J.S. Dietz, *Building a social capital model of research development: the case of the Experimental Program to Stimulate Competitive Research*, „Science and Public Policy” 2000, vol. 27, no. 2; M. Nedeava, *Peer Review: omnipresence vs. omnipotence*, „Evaluation of Science and Innovation Policies”, 17–21 May 2010, Manchester; M.E Falagas, *Peer review in open access scientific journals*, „Open Medicine” 2007, vol. 1, no. 1; M. Mandviwalla, R. Patnayakuni i D. Schuff, *Improving the peer review process with information technology*, „Decision Support Systems” 2008, 46; P. Nightingale i A. Scott, *Peer review and the relevance gap: ten suggestions for policy-makers*, „Science and Public Policy”, October 2007, 34(8); *Peer Review: the challenges for the humanities and social sciences. A British Academy Report*, The British Academy 2007; P. Gruss, *Peer review at the Max Planck Society*, [w:] *Science between evaluation and innovation: a conference on peer review*, DFG 2002; F.Q. Wood i V.L. Meek, ed., *Research Grants. Management and Funding*, Australian Academy of Science Canberra 1993; R.E. Malone, *Should peer review be an open process?*, „Journal Of Emergency Nursing” 1999, 25; *Science between evaluation and innovation: a conference on peer review*, DFG 2002; S. Hornbostel i M. Olbrecht, *Peer Review In Der DFG: Die Fachkollegiaten*, iFQ-Working Paper No. 2, November 2007.

Rysunek 2.5

### Klasyczny i obecny *peer review*



Źródło: opracowanie własne.

## 2.5. Książki i artykuły naukowe, prezentacje na konferencjach

Książki i artykuły naukowe, podobnie jak wnioski grantowe, nadal oceniane są metodą klasycznego *peer review*.

Zasady klasycznego *peer review*:

- autor (autorzy) proponuje temat i metodę oraz nadaje ostateczny kształt – propozycji projektu lub artykułu;
- autorzy są zidentyfikowani, a ich liczba jest zamknięta;
- pole badań jest uformowane i uznane;
- oceniający są anonimowi wobec ocenianych;
- oceny dokonuje wąska i zamknięta grupa badaczy o wysokiej reputacji;
- ocena ma charakter *ex ante*;
- najbardziej popularne formy: pisemne recenzje przesyłane pocztą oraz panel. Klasyczne *peer review* ulega jednak stałej erozji.

Jakie są jej powody?

- ICT i jego konsekwencje (np. rozwój modeli otwartych – *Open Source*, *Open Knowledge Initiative*, *OpenLaw*, *Open Archive Initiative*, *e-printArchives* itp.);
- narodziny nowych form życia społecznego i kultury (np. *open source communities*, portale społecznościowe);
- narastanie „społeczeństwa postnaukowego” (innowacje oparte nie tyle na rozwoju specjalistycznej wiedzy naukowej i technicznej, ile na łączeniu wiedzy naukowej, inżynierskiej, artystycznej, humanistycznej, społecznej, menadżerskiej);

- nowe idee (Partycypacja, Otwartość, Kolektywna/Rozproszona Inteligencja, „Mądrość Tłumu”, *Feed-Back*, Refleksyjność, Twórczość i Innowacje);
- wzrost badań wielkiej skali prowadzonych przez wielkie międzynarodowe zespoły badawcze, np. nad genomem, w których forma „komunikatu z frontu badań” jest bardziej właściwa niż tradycyjny artykuł prezentujący ostateczne wnioski;
- nowe formy instytucjonalne badań (np. wirtualne instytuty i obserwatoria, m.in. *Astrophysical Virtual Observatory*; telepraca; zespoły rozproszone; zadaniowe klastry badawcze; badania prowadzone równoległe przez różne zespoły badawcze, łączące współpracę ze współzawodnictwem; „badania partycypacyjne” z udziałem przyszłych użytkowników aktywnych na wszystkich etapach badań; wielkie sieci powiązanych ze sobą laboratoriów czy obserwatoriów);
- nowe formy metodologii badań (np. symulacja, wielkie bazy danych);
- nowe formy komunikacji, archiwizacji i prezentacji badań (np. e-archiwa, wiki, blogi, *hypertext*, łączenie tekstu z wideo, dynamiczne diagramy – prezentacja dynamiki zmian złożonych układów);
- wzrost zakresu badań inter- i multidyscyplinarnych;
- nowe modele innowacji („otwarta innowacja”, innowacje zorientowane na użytkownika);
- wzrost nacisku opinii publicznej na efekty zewnętrzne i aplikacje badań oraz na udział w ustalaniu priorytetów badawczych;
- rozszerzenie pojęcia efektu badań naukowych (nie tylko nowe ustalenia, ale także rozwój kompetencji, sieci naukowych, mobilności, aparatury itd.).

W rezultacie:

- zmieniają się metody i formy instytucjonalne prowadzenia badań naukowych, rozumienie ich celu i orientacji, formy komunikacji naukowej i archiwizacji ustaleń itd.;
- rozwijają się nowe reguły gry w nauce, a warunki, w których zarówno zasady klasycznego *peer review* (anonimowość, specjalizacja, sprecyzowany temat, określeni autorzy itd.), jak i jego zastosowania (propozycja badawcza, artykuł zawierający ostateczne ustalenia itd.) najlepiej sprawdzały się – powoli się kurczą.

*Peer review* staje się przedmiotem skrajnie krytycznych opinii: „Wydaje się, że nie ma badań zbyt rozdrobnionych, hipotez zbyt banalnych, literatury zbyt stroniczej lub zbyt egoistycznej, projektu zbyt wypaczonego, metodologii zbyt partackiej, prezentacji wyników zbyt niedokładnej, niejasnej i sprzecznej, analizy zbyt samopotwierdzającej, argumentu zbyt okrągłego, zbyt błędnego lub zbyt nieuzasadnionego, gramatyki i składni zbyt obraźliwej, aby nie ukazały się w końcu w druku” (Drummond Rennie, editor of *Journal of the American Medical Association*)<sup>41</sup>. „Oceniamy *peer review* publicznie jako *quasi-sakralną*

<sup>41</sup> R. Drummond, *Misconduct and journal peer review*, [w:] *Peer Review In Health Sciences*, red. F. Godlee i T. Jefferson, 2nd ed. BMJ Books, London 2003.

procedurę, która pomaga uczynić naukę naszym najbardziej obiektywnym wyrazicielem prawdy. Ale wiemy, że system *peer review* jest tendencyjny, niesprawiedliwy, niewiarygodny, niekompletny, często obraźliwy, zazwyczaj ignorancki, czasem głupi i często zły” (Richard Horton, redaktor brytyjskiego czasopisma medycznego „The Lancet”)42. „[Michael] Nielsen podaje liczne przypadki, kiedy recenzenci najlepszych tytułów naukowych nie byli w stanie ujawnić zawnazs słynnych później skandali naukowych. Tak było w przypadku fizyka Jana Hendrika Schoena, który w 2000 i 2001 r. opublikował wiele artykułów o przełomowych odkryciach w dziedzinie nadprzewodnictwa w związkach organicznych. Schoen sfabrykował wyniki, czego nie wykryli ani recenzenci «Science», ani «Nature». Recenzenci z kolei doskonale potrafią blokować postęp nauki, o czym przekonało się wielu noblistów, których prace zostały odrzucone. I tak George Zweig miał problemy z opublikowaniem artykułu ogłaszającego odkrycie kwarków, Krebs nie mógł przebić się z artykułem o cyklu kwasu cytrynowego. Dość długa lista takich niepowodzeń sugeruje, że system *peer review* nie dość, że nie spełnia podstawowego warunku – nie chroni przed oszustwami, to jeszcze na dodatek wstrzymuje rozwój nauki przez tępienie prac najbardziej innowacyjnych”43.

Nastąpił okres przejściowy – erozji klasycznego *peer review* i kształtowania nowych oczekiwań od badaczy oraz nowych zasad ocen ich pracy.

Najwcześniejsze propozycje otwartego *peer review* to m.in.:

- 1959 – *Open peer commentary*, „Current Anthropology” – komentowanie tytułowego artykułu przez kilkunastu specjalistów wraz z odpowiedzią autora. Naśladowcy;
- lata 70. – „Behavioral and Brain Sciences” inicjuje publikowanie kontrowersyjnych artykułów wraz z opiniami recenzentów;
- 1991 – internetowe archiwum arXiv.org Paula Ginsparga z możliwością komentarzy czytelników. 470 000 prac. Naśladowcy;
- II połowa lat 90. – początki zamieszczania w Internecie wersji roboczych prac przed przesłaniem ich do redakcji czasopism papierowych; rozwój czasopism elektronicznych;
- 1997 – „Electronic Transactions on Artificial Intelligence” (ETAI) – jedno z pierwszych czasopism elektronicznych stosujących otwarte *peer review* (dwustopniowa ocena – *reviewing* i *refereeing*);
- 2001 – *peer review* otwarte dla publiczności – „Atmospheric Chemistry and Physics”. Dwa stadia oceny: wstępna redakcyjna – umieszczenie w Internecie; opinie i komentarze czytelników w połączeniu z formalnym *peer review* i odpowiedzią autorów. Na końcu decyzja co do publikacji;

<sup>42</sup> Cyt. za: E.E. van der Wall, *Peer review under review: room for improvement?*, „Neth Heart” J. 2009 May, 17(5).

<sup>43</sup> E. Bendyk, *Makroskop*, <https://edwinbendyk.wordpress.com/tag/recenzje/>; M. Nielsen, *Three myths about scientific peer review*, Blog January 8, 2009, <http://michaelnielsen.org/blog/three-myths-about-scientific-peer-review/>.

- ok. 2004 – Signaling Gateway, brama sygnalizacji, baza relacyjna o otwartym dostępie mająca na celu zbieranie informacji o molekułach komórek <http://www.signaling-gateway.org/>, zastosowanie zasad otwartego *peer review* do baz danych.  
Przełomowy okazał się rok 2006:
- 2006 – „Nature” – eksperyment z otwartym *peer review* niektórych wybranych artykułów;
- 2006 – czasopisma elektroniczne / „otwartego dostępu” eksperymentują z otwartym *peer review* (np. „Phyllica”, „Naboj”, „Biology Direct”, „BMC Medicine”, „PLoS Medicine”, „British Medical Journal”, „Atmospheric Chemistry and Physics”, „Journal of Economic Perspectives”, „American Psychologist”);
- 2006 – debata „Nature” nad przyszłością otwartego *peer review*. Reakcja: mieszanina akceptacji z krytycyzmem. „Science” – podobne próby.
- 2006 – B.S. Noveck, *Peer to Patent: Collective Intelligence, Open Review, and Patent Reform*, „Harvard Journal of Law & Technology”; 2007 – *The Peer to Patent Project – Community Patent Review* połączenie eksperckiej z otwartą oceną wniosków patentowych;
- 2008 – „społeczny eksperyment sieciowy” „Scientific American”.

Wady otwartego *peer review* to możliwość eksplozji internetowej nienawistnej głupoty („hejtowanie”). Zalety otwartego *peer review* to: możliwość ulepszenia tekstu, większy poziom rzetelności i odpowiedzialności opiniujących (skoro występują „z otwartą przyłbicą”, pod własnym nazwiskiem), większa łatwość demaskowania negatywnych recenzji dyktowanych względami niemerytorycznymi, naruszanie monopoli naukowego establishmentu skutkujące odblokowaniem energii środowiska naukowego.

## 2.6. Aparatura i infrastruktura naukowa

Aparatura badawcza (instrument naukowy) to każdy rodzaj sprzętu, aparatury lub urządzenia zaprojektowany tak, aby na podstawie sprawdzonej zasady fizycznej ułatwiał lub umożliwiał pozyskiwanie powtarzalnych, weryfikowalnych danych, zwykle składających się z pomiarów ilości, właściwości zjawisk i materiałów, sił itp. Pojęcie dużej infrastruktury badawczej obejmuje wielką aparaturę naukową lub zestawy instrumentów naukowych, zbiory i inne zasoby oparte na wiedzy, pliki i bazy danych, systemy komputerowe i oprogramowania, sieci łączności, jak też inną infrastrukturę o unikalnym charakterze niezbędną do prowadzenia badań naukowych. Infrastruktura badawcza może być krajowa lub międzynarodowa, umieszczona w jednej placówce lub też rozproszona albo wirtualna<sup>44</sup>. Podkreśla się, że pojęcie infrastruktury badawczej stało się bardziej zróżnicowane. Dawniej mianem tym określano tylko wielkie obiekty, takie jak akceleratory lub

<sup>44</sup> FCT, *Evaluation Guidelines*, 2013.

statki badawcze, dziś pojęcie to obejmuje nie tylko rozproszoną infrastrukturę badawczą, ale także zbiory i bazy danych, e-infrastrukturę oraz infrastrukturę dla badań społecznych<sup>45</sup>.

Do form finansowania aparatury należą:

- dotacje na studia projektowe i planowanie budowy aparatury/infrastruktury;
- dotacje na inwestycje w urzędnictwo lub bazy danych dla budowy krajowej lub międzynarodowej aparatury/infrastruktury;
- dotacje na pokrycie kosztów operacyjnych utrzymania w dłuższej perspektywie aparatury/infrastruktury;
- dotacje na zaprzestanie pracy infrastruktury badawczej<sup>46</sup>.

Najczęstsze formy ewaluacji aparatury i infrastruktury badawczej obejmują: ewaluację *ex ante* wniosków o sfinansowanie zakupu aparatury/infrastruktury oraz ewaluację *ex post* programów jej finansowania.

**Ewaluacja *ex ante* wniosków aparaturowych.** W Australii Australian Science and Technology Council (ASTEC), oceniając wnioski dotyczące wielkiej aparatury, brał pod uwagę odpowiedzi na następujące pytania:

- Czy propozycja odnosi się do dziedziny badań posiadającej duże znaczenie jednocześnie dla kraju oraz nauki światowej?
- Jakie kluczowe problemy naukowe można by starać się rozwiązać dzięki proponowanej aparaturze?
- Dlaczego problemy te mają zasadnicze znaczenie dla australijskiej nauki i techniki?
- Czy rozpatrywane urządzenie jest zgodne z bieżącymi priorytetami naukowymi i technologicznymi Australii?
- Czy będzie ono służyć znakomitym uczonym i inżynierom?
- Czy można oczekiwać, że będzie ono pobudzać badania interdyscyplinarne?
- Czy otworzy nowe możliwości kształcenia doktorantów i doktorów?
- Czy zwiększy ono atrakcyjność karier naukowej i inżynierskiej?
- Czy pobudzi międzynarodową współpracę naukową, ściągając do Australii czołowych uczonych zagranicznych?
- Czy rozważane urządzenie powinna zbudować sama tylko Australia, czy też Australia wspólnie z innymi krajami?
- Jeśli zbudowane ono zostanie w Australii, czy stanie się magnesem ściągającym fundusze zagraniczne?
- Czy budowa rozważanego urządzenia stanie się bodźcem rozwoju przemysłu australijskiego? Czy będzie ono świadczyć usługi dla firm?
- Czy będzie ono wzmacniać więzi między uniwersytetami oraz państwowymi i prywatnymi laboratoriami?

<sup>45</sup> Wissenschaftsrat, *Evaluation of large infrastructure projects*, 2013.

<sup>46</sup> FIRI, *Evaluation criteria for research infrastructure*, cyt. za: *Applied Multivariate Research. Design and Interpretation*, Sage 2006.

- Czy prowadzone dzięki niemu badania naukowe przyczynią się do zakładania nowych firm?
- Czy urządzenie to jest zgodne z narodowymi priorytetami w zakresie rozwoju wiedzy, wzrostu gospodarczego, zdrowia, jakości życia, bezpieczeństwa narodowego?
- Czy będzie ono służyć ochronie środowiska lub lepszemu zarządzaniu zasobami naturalnymi?
- Czy można oczekiwać, że zwiększy ono zrozumienie i uznanie opinii publicznej dla osiągnięć australijskiej nauki i techniki?
- Czy urządzenie to poprawi obraz Australii jako kraju zaawansowanego technologicznie?
- Czy z chwilą oddania do użytku rozważanego urządzenia pozycja Australii w negocjacjach międzynarodowych wzmocni się?

W Finlandii Akademia Fińska przygotowała zestaw ogólnych kryteriów oceny infrastruktury badawczych. Infrastruktura badawcza musi m.in.:

- zapewniać podstawę dla światowej klasy badań naukowych;
- wzmacniać międzynarodowe oddziaływanie fińskiej nauki;
- posiadać długoterminowy plan finansowania, utrzymania i użytkowania;
- być używana przez kilka grup badawczych dla badań o wysokiej jakości;
- być otwarta i łatwo dostępna dla naukowców, przemysłu i innych podmiotów;
- posiadać plan ochrony gromadzonych danych i / lub materiałów;
- wprowadzić nowoczesną technologię (jeśli to możliwe)<sup>47</sup>.

Ewaluację projektów infrastruktury badawczej przeprowadza się w pięciu różnych wymiarach. Każdy projekt infrastruktury badawczej jest oceniany indywidualnie w każdym oddzielnym wymiarze, a także w stosunku do innych projektów w innych dziedzinach nauki. Wspomniane wymiary to:

1. Jakość i potencjał naukowy.
2. Otwarty dostęp i wykorzystanie.
3. Znaczenie dla strategii instytucji przyjmującej.
4. Krajowe i międzynarodowe znaczenie.
5. Wykonalność.

**Ewaluacja *ex post* programów.** Przekłady ewaluacji z ostatnich lat:

- W Wielkiej Brytanii Rada ds. Badań Naukowych w Dziedzinie Biotechnologii i Nauk Biologicznych w ewaluacji *ex post* korzystała z wyników ankiety adresowanej do wszystkich beneficjentów programu aparaturowego. Pytania dotyczyły m.in.:
  - wpływu aparatury na rozszerzenie tematyki badań, publikacje lub inne osiągnięcia;
  - obecnego stanu i stopnia wykorzystywania aparatury;

<sup>47</sup> Academy of Finland, *National criteria for research infrastructures*, 3.4.2017, [http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/firi/firi\\_national\\_criteria\\_ri\\_2017.pdf](http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/firi/firi_national_criteria_ri_2017.pdf).



- innych źródeł finansowania aparatury<sup>48</sup>.
- Brytyjska Ekonomiczna i Społeczna Rada Badawcza zleciła studium na temat wpływu gospodarczego programu usług w zakresie danych ekonomicznych i społecznych (*Economic and Social Data Service*). Analiza ekonomiczna korzystała z wielu metod ekonometrycznych<sup>49</sup>.
- Niemiecka Rada Nauki i Humanistyki zleciła ocenę projektów infrastrukturalnych dużej skali. Ewaluacja odbyła się w dwóch kolejnych etapach: dokonano indywidualnej oceny każdego projektu oraz ogólnej oceny porównawczej. Obie oceny oparto na czterech wymiarach: „naukowy potencjał”, „wykorzystanie”, „wykonalność” oraz „przydatność dla Niemiec jako miejsca nauki i badań naukowych”. W ewaluacji brali udział zagraniczni eksperci.
- Dyrekcje Generalne ds. Badań Naukowych (DG RTD) oraz ds. Społeczeństwa Informacyjnego i Mediów (DG IN FSO) zleciły przeprowadzenie ewaluacji wpływu wsparcia UE dla infrastruktur badawczych w ramach VI Programu Ramowego (2002–2006)<sup>50</sup>.

## 2.7. Ocena wpływu badań

Ocena wpływu badań (ang. *Research Impact Assessment*) ma na celu pomiar efektów realizacji wcześniej postawionych celów polityki naukowej, jej programów i instrumentów (ocena *ex post*), lub też ustalenie kto, gdzie i ile powinien inwestować w badania naukowe, a także przewidywanie, co społeczeństwo dostanie w zamian (ocena *ex ante*). Celem oceny wpływu jest określenie zarówno efektów ekonomicznych inwestycji publicznych w badania i rozwój, jak również skutków społecznych (np. poprawy stanu zdrowia społeczeństwa) i środowiskowych<sup>51</sup>.

Ocena wpływu może mieć dwie formy:

- ocena wpływu *ex ante* polega na analizie prospektywnej, jaki może być wpływ interwencji;
- ocena wpływu *ex post* ma na celu zrozumienie, w jakim stopniu i w jaki sposób interwencja polityki rozwiązuje problem, którym zamierzano się zająć<sup>52</sup>.

Ocena *ex ante* i ocena *ex post* powinny być ze sobą zharmonizowane w ramach cyklu polityki; ocena *ex ante* powinna być punktem wyjścia dla oceny *ex post*<sup>53</sup>.

<sup>48</sup> BBSRC, *Evaluation of BBSRC's Research Equipment Initiative appendices* 2012.

<sup>49</sup> ESRC Economic and Social Research Council, *Economic Impact Evaluation of the Economic and Social Data Service* 2012.

<sup>50</sup> EU, *Community Support for Research Infrastructures in the Sixth Framework Programme Evaluation of pertinence and impact*, 2009.

<sup>51</sup> M. Cervantes, *CSTP work on Evaluation and Impact Assessment: Key Findings*, Presentation at the 1stRIHR Meeting 18–19 May 2009.

<sup>52</sup> OECD, *Assessing The Impact Of State Interventions In Research – Techniques, Issues And Solutions* DSTI/STP(2014)24.

<sup>53</sup> *Optimizing the research and innovation policy mix: The practice and challenges of impact assessment in Europe*, 2011.

Głównym celem analizy wpływu jest wspieranie rozwoju polityki. Wybór technik stosowanych zależy zatem od potrzeb politycznych, a nie od realizacji jakichś „najlepszych praktyk metodologicznych”. Dobrą praktyką w zakresie oceny jest wykorzystywanie wielu metod. Pojedyncze podejścia wydają się zawodne; wykorzystywanie wielu metod to dobry sposób na zwiększenie zaufania do wiarygodności wyników<sup>54</sup>.

Relacje pomiędzy oceną wpływu badań a ewaluacją polityki naukowej i innowacyjnej, jej programów i instrumentów są różnie ujmowane; dla niektórych ocena wpływu jest częścią ewaluacji, inni traktują oba terminy zamiennie. Nie zawsze da się jednoznacznie odróżnić ocenę wpływu badań od innych ewaluacji<sup>55</sup>. Głównym celem ewaluacji jest wspieranie i wzmacnianie jakości badań. Oceny wpływu mają szerszy cel: uzyskanie gruntownej wiedzy o związku między działaniami i efektami<sup>56</sup>. Według jednego z ujęć ocena wpływu badań rozwinęła się ze wcześniejszej ewaluacji *ex post*, wzbogacając ją o interaktywny proces uczenia się ukierunkowany na budowanie długofalowych strategii, ustalanie priorytetów i projektowanie programów badań<sup>57</sup>. Podkreśla się też, że ocena wpływu koncentruje się na skutkach interwencji, podczas gdy ewaluacja może obejmować szerszy zakres zagadnień, takich jak stosowność projektu interwencyjnego, koszty i efektywności interwencji czy też jego niezamierzone następstwa, a także kwestię, jak wykorzystać doświadczenia z interwencji dla poprawy przyszłych działań<sup>58</sup>.

Ocenę wpływu przeprowadza się na poziomie krajów, ale także programów i instrumentów, agencji finansujących B+R<sup>59</sup>, dziedzin nauki (np. nauk medycznych i rolniczych) oraz uczelni. Metody oceny wpływu stosuje się m.in. w trakcie oceny skutków regulacji (ang. *Regulatory Impact Assessment*), gdzie służą ulepszeniu instrumentów polityki i decyzji budżetowych<sup>60</sup>, oraz w ocenie *policy mix*, portfela programów i instrumentów rządowych<sup>61</sup>.

Wpływ publicznego B+R na społeczeństwo i gospodarkę stał się przedmiotem wzmożonego zainteresowania polityków od połowy lat 90.; od tego czasu rządy coraz usilniej starają się ustalić, ile powinny inwestować w B+R, w co inwestować i co społeczeństwo otrzymuje w zamian. Analiza wpływu pomaga w określeniu

---

<sup>54</sup> OECD *Workshop On Assessing The Impact Of Science, Technology And Innovation Policy Instruments With Common Objectives* DSTI/STP/AH/A(2014)2.

<sup>55</sup> M. Aymerich et al., *Measuring the payback of research activities: A feasible ex-post evaluation methodology in epidemiology and public health*, „Social Science & Medicine” 2012, 75, s. 505–510.

<sup>56</sup> ESF, *The Challenges of Impact Assessment*, 2012.

<sup>57</sup> OECD *Assessing The Socio-Economic Impacts Of Public Investment In R&D*, 2007.

<sup>58</sup> OECD, *Assessing The Impact Of State Interventions In Research – Techniques, Issues And Solutions* DSTI/STP(2014)24.

<sup>59</sup> MRC Medical Research Council 2011, *Measuring the link between research and economic impact*.

<sup>60</sup> OECD *Improving Policy Instruments Through Impact Assessment*, Sigma Paper: No. 31, CCNM/SIGMA/PUMA(2001)1.

<sup>61</sup> *Optimizing the research and innovation policy mix: The practice and challenges of impact assessment in Europe*, 2011.

zarówno efektów ekonomicznych inwestycji publicznych w B+R, takich jak wkład we wzrost gospodarczy, jak i skutków społecznych, takich jak poprawa stanu zdrowia społeczeństwa. Co więcej, w coraz większym stopniu decydenci chcą, aby inwestycje publiczne w B+R odpowiadały na globalne wyzwania, takie jak energia, bezpieczeństwo i zmiany klimatu<sup>62</sup>. Temu zapotrzebowaniu odpowiada rozwój ewaluacji od pomiaru wkładu do pomiaru „produktu”, „efektu” i „wpływu” inwestycji publicznych, co wynika zarówno z przenikania idei tzw. nowego zarządzania publicznego, jak i przemian praktyk w działalności badawczej, innowacyjnej i gospodarczej (wzrost rynkowego ukierunkowania badań i nasilenie powiązań nauki z gospodarką oraz sferą publiczną i społeczną, m.in. wskutek wpływu ICT).

Szczególna potrzeba, a zarazem trudność w przeprowadzaniu oceny wpływu bierze się z faktu, że w wielu krajach wzrasta liczba instrumentów polityki naukowej i innowacyjnej, sięgając często ponad sto. Często rządy korzystają z kilku instrumentów, aby osiągnąć jeden cel, lub tworzą jeden instrument dla realizacji wielu celów, takich jak promowanie doskonałości badań, pobudzanie B+R biznesu lub wspieranie transferu technologii i komercjalizacji badań. Ponieważ poszczególne instrumenty często mają wiele celów, a wiele instrumentów służy osiągnięciu różnorodnych celów, zakres potencjalnych interakcji pomiędzy wszystkimi tymi instrumentami jest ogromny, a ich wpływ na ogólny wpływ polityki nielatwy do ustalenia<sup>63</sup>.

Pomiar wpływu badań uznaje się za trudny, zwłaszcza w odniesieniu do wykazania przyczynowości. Podkreśla się, że wybór metody oceny powinien zależeć od kontekstu. Aby pomiar miał silniejsze podstawy empiryczne, zaleca się systematyczne gromadzenie informacji o produktach, efektach i oddziaływaniach w bazach danych, takich jak amerykańska *STAR METRICS*, brytyjski *Research Outcomes System* lub brazylijska *Lattes Platform*. Większość ocen wpływu polityk naukowych i innowacyjnych szeroko wykorzystuje metody badań jakościowych, w tym ankietowanie beneficjentów i studia przypadków<sup>64</sup>.

Ocena wpływu (ang. *Impact Assessment*), także unijnej polityki naukowej, wpisana jest w procedury Unii Europejskiej. Unijna ocena wpływu składa się z trzech głównych faz, którym zwykle towarzyszą konsultacje z zainteresowanymi stronami<sup>65</sup>. Ocena wpływu badań jest najpowszechniej stosowana w krajach anglosaskich, przede wszystkim w Wielkiej Brytanii, oraz w krajach nordyckich<sup>66</sup>. Obserwuje się także wzmożone zainteresowanie oceną w krajach romańskich (Włochy, Francja, Hiszpania), w Belgii i Holandii oraz w krajach latynoamerykańskich (Brazylia).

<sup>62</sup> OECD, *Assessing the Socio-economic Impacts of Public R&D: Recent Practices and Perspectives*, [w:] *OECD Science, Technology and Industry Outlook 2008*.

<sup>63</sup> OECD, *Workshop On Assessing The Impact Of Science, Technology And Innovation Policy Instruments With Common Objectives* DSTI/STP/AH/A(2014)2.

<sup>64</sup> OECD, *Assessing The Impact Of State Interventions In Research – Techniques, Issues And Solutions*, DSTI/STP(2014)24.

<sup>65</sup> A. von Raggamby, *The Role of Tools in Impact Assessments*, [w:] OECD, *Conducting Sustainability Assessments*, 2010.

<sup>66</sup> OECD, *Framework For Monitoring And Impact Analysis: – Case Tekes, Finland*, TIP-RIHR WORKSHOP 15.16.9.2008.

W ocenie wpływu badań finansowanych ze źródeł publicznych przeważają analizy następujących rodzajów wpływów (tabela 2.4).

Tabela 2.4

#### Rodzaje wpływu badań

Wpływ	% analiz
Technologiczny	24%
Gospodarczy	86%
Społeczny (wliczając wpływ na środowisko)	29%

Źródło: H. Delanghe i P. Teirlinck, *Optimising the Policy Mix by the Development of a Common Methodology for the Assessment of (Socio-)Economic Impacts of RTDI Public Funding (CIA4OPM) Literature Review*, Brussels 2013.

W krajach unijnych stosuje się następujące metody oceny wpływu (tabela 2.5).

Tabela 2.5

#### Metody oceny wpływu

Metody zbierania danych		Metody analizy danych	
Istniejące badania ankietowe i bazy danych	67%	Analiza studiów przypadku	48%
Badania ankietowe uczestników	62%	Statystyka opisowa	48%
Wywiady	52%	Analizy ekonometryczne	38%
Badania dokumentacji	24%	Analiza dokumentów	38%
Dane z monitoringu	19%	Analiza wkładu / rezultatu	29%
Badania bibliometryczne i technometryczne	19%	Podejście kontrfaktyczne	24%
Grupy fokusowe, warsztaty, spotkania	19%	Metoda kosztów i korzyści	24%
Badania ankietowe nieuczestników	14%	Analiza kontekstu	19%
Inne metody	10%	Podejście grupy kontrolnej	19%
Panele ekspertów	10%	Porównanie „przed-po”	10%

Źródło: H. Delanghe i P. Teirlinck, *Optimising the Policy Mix by the Development*, op. cit.

W ocenie wpływu biorą udział następujące grupy (tabela 2.6).

Tabela 2.6

#### Grupy biorące udział w ocenie wpływu

Urzednicy	95%
-----------	-----

Zarządzający programem	48%
Analitycy polityki	48%
Beneficjenci	43%
Politycy	43%
Potencjalni użytkownicy interwencji	33%
Audytory – władze finansowe	14%
Obywatele	14%
Krajowi lub międzynarodowi sponsorzy programu	10%

Źródło: H. Delanghe i P. Teirlinck, *Optimising the Policy Mix by the Development*, op. cit.

**Typy wpływu.** Analizy wpływu badań (B+R) opierają się na: (a) taksonomii sfer oddziaływania B+R, (b) taksonomii wyników B+R, (c) taksonomii form oddziaływania, (d) metodach i technikach.

Najbardziej przemyślaną taksonomię typów sfer oddziaływania B+R zaproponowali Godin i Doré:

1. Wpływ naukowy: badania mają wpływ na późniejsze badania dzięki rozwojowi teorii, danych, podejść, metod, modeli, tematów i ustaleń. Mają one wpływ na powstawanie i rozwój dyscyplin naukowych, na badania interdyscyplinarne oraz na kształcenie. Bibliometria (oraz jej rozgałęzienia, takie jak webometria) pozwala mierzyć skalę tego wpływu.
2. Wpływ technologiczny: nowe produkty, procesy i usługi oraz techniczny *know-how* są częściowo efektem badań naukowych, tych dawniejszych oraz tych prowadzonych obecnie. Badania patentowe oraz badania działalności innowacyjnej pozwalają ocenić zakres tego wpływu.
3. Wpływ gospodarczy odnosi się do oddziaływania na sytuację budżetową firm, ich koszty operacyjne, przychody, zyski ze sprzedaży produktów i usług; na źródła finansowania, inwestycje i działalność produkcyjną oraz na rozwój nowych rynków; na poziomie zagregowanym odnosi się także do wpływu na wzrost gospodarczy i wzrost produktywności krajów i regionów.
4. Wpływ kulturalny odnosi się do tego, co (w krajach anglosaskich) nazywa się rozumieniem nauki przez społeczeństwo (ang. *public understanding of science*), oraz – przede wszystkim – odnosi się do wpływu na cztery rodzaje wiedzy: wiedzy co, dlaczego, jak i kto. Innymi słowy, chodzi tu o wpływ na wiedzę jednostek oraz rozumienie idei i rzeczywistości, a także wpływ na umiejętności intelektualne i praktyczne, na postawy, zainteresowania, wartości i przekonania. Źródłem wiedzy o zakresie tego wpływu (podobnie jak i kolejnych) są przede wszystkim badania jakościowe.
5. Wpływ społeczny: badania naukowe wpływają na dobrobyt, zachowania, praktyki i działania osób i grup, w tym na ich dobre samopoczucie i jakość życia. Dotyczy to także zwyczajów i nawyków: konsumpcji, pracy, seksu,

sportu, jedzenia. Badania mogą też przyczynić się do zmiany poglądów społeczeństwa oraz do unowocześnienia sposobów prowadzenia działalności gospodarczej.

6. Oddziaływanie polityczne: badania wpływają na sposób i zasady działania decydentów i polityk. Mogą one dostarczać dowodów wpływających na decyzje polityczne oraz przyczyniać się do wzrostu udziału obywateli w podejmowaniu decyzji naukowych i technologicznych. Skuteczna administracja i zarządzanie zależą częściowo od jakości wiedzy i informacji dostępnych dla polityków i decydentów. Niezależne badania mogą ulepszyć decyzje polityczne. Badania w takich dziedzinach jak ekonomia, prawo i nauki społeczne stanowią teoretyczne i empiryczne podstawy dla działań ministerstw. Ponadto, badania i ekspertyzy są zasobem, do którego rządy mogą sięgnąć, aby wiedzieć, jak reagować na nieprzewidziane zdarzenia.
7. Skutki organizacyjne odnoszą się do wpływu na działalność instytucji i organizacji: planowanie, organizację pracy, administrację, zasoby ludzkie itp.
8. Wpływ na zdrowie dotyczy oddziaływania badań na zdrowie publiczne, np. na długość życia, zapobieganie chorobom, system opieki zdrowotnej. Publiczne B+R może się przyczynić do poprawy zdrowia publicznego poprzez stworzenie nowych metod leczenia (np. urządzeń, leków i szczepionek) lub też wnosząc wkład do wiedzy, która wpływa na rozwój nowych metod leczenia i doradztwo w sprawach zdrowia dla rządów i innych organizacji.
9. Wpływ na środowisko odnosi się do wpływu badań na zarządzanie środowiskiem i zasobami naturalnymi oraz na przeciwdziałanie zanieczyszczeniu, a także na klimat i meteorologię.
10. Wpływy symboliczne dotyczą korzyści czerpanych przez uczelnie i instytuty badawcze z tytułu swojej reputacji naukowej.
11. Wpływ edukacyjny dotyczy wpływu badań naukowych na programy, narzędzia pedagogiczne, kwalifikacje, rynek pracy itp.<sup>67</sup>

Trzy pierwsze sfery oddziaływania są znane statystykom i uchwytne statystycznie, pozostałe – trudno mierzalne.

Z kolei najbardziej znaną taksonomię wyników B+R przedstawili Salter i Ben Martin – wyniki B+R to:

- wiedza skodyfikowana (publikacje, raporty techniczne, patenty, prototypy);
- umiejętność stawiania i rozwiązywania problemów;
- znajomość metod i technik badawczych;
- nowe prototypy, produkty i usługi;
- nowe firmy;
- nowa aparatura, metody i techniki;

---

<sup>67</sup> B. Godin i C. Doré, *Measuring the Impacts of Science: beyond the Economic Dimension*, Paper presented at the Helsinki Institute for Science and Technology Studies, HIST Lecture, 24 August 2007, Helsinki, Finland; OECD *Linking Science to Innovation, Economic Growth and Welfare: Scoping Paper*, 2013.

- wykształceni i wyszkoleni absolwenci szkół wyższych;
- sieci naukowe i zawodowe oraz przepływ informacji;
- wiedza społeczna, jak np. dostarczanie politykom ustaleń dla decyzji<sup>68</sup>.

Wyłącznie pierwszy typ wyników – a właściwie tylko publikacje i patenty – jest znany i wykorzystywany w statystyce.

Jeśli idzie o formy wpływu, stosuje się, po pierwsze, podział na produkty (rezultaty) wpływu, efekty i wpływy:

- produkty działalności badawczo-rozwojowej (publikacje, patenty, raporty, prototypy, produkty, procesy, usługi, standardy, cytowania, uzyskane nagrody, wiedza i umiejętności);
- efekty – produkty działalności badawczo-rozwojowej przekształcone w sektorze gospodarczym w nowe produkty, materiały, procesy, a w sektorze publicznym w nowe prawa, metody nauczania, leczenia, przesyłu energii, transportu itd.;
- wpływ – wpływ naukowy i technologiczny: nowa wiedza; kultura współpracy; powstanie sieci; reputacja; wpływ ekonomiczny: długofalowe efekty, które pociągnęły za sobą wzrost gospodarczy; konkurencyjność przemysłowa; innowacje; wzrost zatrudnienia; wpływ społeczny: poprawa jakości życia, spadek śmiertelności; rozwój społeczny; wpływ polityczny: wkład do polityk; zmiana ustaw<sup>69</sup>.

Koniecznym uzupełnieniem jest uwzględnienie wkładu (*input*).

Pytania i kryteria ewaluacji efektów społeczno-gospodarczych wynikają z porównania ze sobą pewnych podstawowych elementów interwencji publicznej, takich jak potrzeba (problem lub trudność dotycząca pewnej grupy, którą interwencja stara się rozwiązać albo przezwyciężyć), cel (*objective*, jasne i wyraźne wyartykułowanie zakładanych efektów), wkład (*input*, zasoby finansowe, ludzkie, materialne, organizacyjne i prawne mobilizowane na rzecz interwencji), produkt (*output*), czyli bezpośredni rezultat uzyskany wskutek interwencji, efekt (*outcome*), czyli zmiana społeczno-gospodarcza lub poznawcza wynikająca bezpośrednio lub pośrednio z interwencji, wpływ (*impact*), czyli bardziej długofalowe następstwa interwencji<sup>70</sup>.

Trzy elementy interwencji, które często najtrudniej odróżnić, dotyczą produktów, efektów i wpływu działania (*outputs*, *outcomes*, *impacts*).

## Tabela 2.7

<sup>68</sup> A.J. Salter i B. Martin, *The Economic Benefits of Public Funded Basic Research: A Critical Review*, „Research Policy” 2001, 30, 3, s. 509–532; A.J. Salter i B.R. Martin, *The Economic Benefits of Publicly Funded Basic Research*, presented at the conference The Impacts of Science, Vienna 2007.

<sup>69</sup> European Commission, *An RTD Evaluation Toolbox*, 2002; E. Geisler, *Integrated Figure of Merit of Public Sector Research Evaluation*, „Scientometrics” 1996, vol. 36, no. 3.

<sup>70</sup> European Commission, *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, 1999, vol. 6, Luxembourg; Ch. Pollit i G. Bouckaert, *Public Management Reform*, Oxford University Press, 2004.

### Produkty, efekty i wpływ działania (*outputs, outcomes, impacts*)

Produkty ( <i>outputs</i> )	Efekty ( <i>outcomes</i> )	Wpływ ( <i>impact</i> )
Co wytworzono dzięki podjętemu działaniu?	Jakie efekty uzyskano?	Co się wydarzyło w następstwie działania i osiągniętego rezultatu?
Na przykład wyroby takie jak software, narzędzia, patenty, publikacje, raporty, prototypy, pokaz działania, konferencje i inne formy upowszechnienia wiedzy, kwalifikowani ludzie.	Na przykład zmiany w kompetencjach firmy, zmiany w strategii firmy, tantiemy z licencji i patentów, zyski z inwestycji, mobilność personelu.	Na przykład podniesienie poziomu produkcji branży, poprawa efektywności gospodarczej, wzrost konkurencyjności, wzrost usług eksportowych.

Źródło: G. Becher i S. Kuhlmann, red., *Evaluation of technology policy programmes in Germany*, Springer, Dordrecht 1995.

Szczególnie trudno ustalić efekty i wpływy (*outcomes i impact*). Są one niezliczone, powiązane ze sobą w skomplikowany sposób, czasami niezamierzone, trudne do opisu i pomiaru, czasami pozytywne, czasami negatywne. Subiektywizm w ich ocenie jest nieunikniony. Z powodu wszystkich tych komplikacji nie sposób określić precyzyjnie, obiektywnie czy ilościowo, np. wkładu gospodarczego szkolnictwa wyższego<sup>71</sup>.

Stopniowo coraz większą wagę wiąże się z oceną „efektów” i „wpływów” badań i działalności naukowo-technicznej (w porównaniu do oceny „produktów” B+R). A zatem, choć takie produkty, jak artykuły w recenzowanych czasopiśmie, książki, materiały konferencyjne, patenty, wzory użytkowe czy oprogramowania itd. są ważnym elementem oceny, od ponad dekady punkt ciężkości ocen przesunął się w stronę oceny efektów i wpływów, za pomocą np. takich wskaźników, jak:

- liczba absolwentów (licencjatu, magisterium, studiów doktoranckich);
- zastosowania badań, np. w innowacjach technologicznych;
- liczba zawartych kontraktów;
- wskaźniki komercjalizacji (sprzedaż patentów i licencji, firmy „odpryskowe”);
- wskaźniki kompetencji eksperckich, doradczych i konsultacyjnych;
- mierniki wkładu (doradczego) do planowania oraz polityki – na szczeblu regionalnym, krajowym i międzynarodowym;
- mierniki wkładu do powszechnej edukacji (uczenie się przez całe życie, szkolne i pozaszkolne formy kształcenia naukowego);
- wskaźniki wkładu do życia kulturalnego i społecznego (m.in. współpracy lub prac na rzecz organizacji pozarządowych).

<sup>71</sup> G. Becher i S. Kuhlmann, red., *Evaluation of technology policy programmes in Germany*, op. cit.



Wskaźniki stosowane w amerykańskich ewaluacjach ukierunkowanych na B+R biznesu przedstawiają się w następujący sposób (tabela 2.8).

Tabela 2.8

**Wskaźniki stosowane w amerykańskich ewaluacjach  
ukierunkowanych na B+R biznesu**

Kategorie efektu	Typy skutków	Wkłady	Rezultaty (produkty)	Efekty	Wpływy (poza uczestnikami)
Miękkie skutki	Powiązania	Konsorcja. Partnerzy projektu. Prośby o wsparcie	Nowe sieci ze stałymi partnerami	Umowy o współpracę B+R	
	Kompetencje innowacyjne	Strategie komercjalizacji. Strategie własności intelektualnej. Plany marketingowe		Wzrost chęci i zdolności do poprawy wydajności	
	Efekty dźwigniowe z innymi funduszami		Łatwiejszy dostęp do kapitału ryzyka		
Skutki namacalne	Patenty i licencje		Patenty wyniki z projektu		
	B+R	Finansowanie nowych projektów. Wzrost efektywności B+R	Projekt korzysta ze zwiększenia wydatków na B+R. Nowe źródła finansowania		

Tabela 2.8 – cd.

Kategorie efektu	Typy skutków	Wkłady	Rezultaty (produkty)	Efekty	Wpływy (poza uczestnikami)
	Personel B+R	Ogólna liczba zatrudnionych w projekcie. Liczba naukowców i techników zatrudnionych w projekcie	Lepiej przeszkolony personel B+R	Wzrost liczby personelu. Wzrost liczby naukowców i techników	
	Innowacje produktowe	Cele kosztowe. Cele w zakresie efektywności	Liczba nowych produktów. Ogólna liczba produktów. Liczba projektów w stosunku do innowacyjnych produktów. Poprawa satysfakcji klienta		Dyfuzja wewnątrz branży. Dyfuzja pomiędzy branżami
	Innowacje procesowe	Cele kosztowe. Cele w zakresie efektywności.	Liczba nowych procesów. Ogólna liczba procesów. Liczba projektów w stosunku do innowacyjnych procesów. Poprawa satysfakcji klienta		Dyfuzja wewnątrz branży. Dyfuzja pomiędzy branżami
	Sprzedaż			Dodatkowa sprzedaż wynikająca z pracy nad projektem. Wejście na nowe rynki prywatne. Wejście na nowe rynki rządowe	

Kategorie efektu	Typy skutków	Wkłady	Rezultaty (produkty)	Efekty	Wpływy (poza uczestnikami)
	Nowe przedsięwzięcia biznesu			Liczba firm odpryskowych	
Skutki społeczno-gospodarcze	Wzrost konkurencyjności			Zwrot z inwestycji firmy	Wzrost jakości usług. Ekspansja rynkowa
	Zatrudnienie		Liczba miejsc pracy zachowanych w projekcie. Liczba miejsc pracy utworzonych w projekcie	Liczba miejsc pracy zachowanych w kraju. Liczba miejsc pracy utworzonych w kraju	Przyrost lub utrata miejsc pracy
	Efekty środowiska		Firmy uczestniczące w projekcie korzystające z bardziej ekologicznych technologii	Czystsze technologie stosowane przez inne firmy	
	Oddziaływanie społeczne				Wzrost z inwestycji podatnika; efekty odpryskowe

Źródło: *Supporting the monitoring and evaluation of innovation programmes*, 2006.

**Ocena wpływu jako analiza oparta na teorii.** Ocena wpływu wiąże się z ustanowieniem „teorii zmiany” lub „logiki interwencji”, czyli opisu kaskady przyczyn i skutków prowadzących od interwencji do jej pożądanym efektów. Istotą oceny wpływu jest stworzenie takiego domniemanego łańcucha przyczynowości od interwencji do oddziaływania oraz dokonanie pomiaru lub opisu zmian wywołanych wzdłuż tego łańcucha. Podejście to ma tę zaletę specyfiki i ostrości, ale także tendencję do ograniczenia stwierdzonych efektów do kategorii wcześniej ustalonych, z pominięciem nieoczekiwanych następstw, w tym przewrotnych lub niepożądanych, których projektanci interwencji nie przewidzieli.

Teoria zmiany opiera się na koncepcjach teoretycznych. Na przykład teoria zmiany oparta na idei systemów innowacji podkreśla rolę kontekstu oraz działań podejmowanych przez różnych interesariuszy, a podejście neoklasyczne skupia się na bodźcach i sygnałach ekonomicznych odbieranych przez indywidualne podmioty gospodarcze.

Pomimo wielu analiz powiązań badań i innowacji modele „wpływu” pozostają nadal liniowe. Idea „wpływu” ma oczywisty sens w modelu liniowym, gdzie oczekuje się, że zmiany w badaniach mogą być przyczyną zmian w społeczeństwie. Strzałka przyczynowości może jednak biec także w drugą stronę. Obecne teorie innowacji podkreślają nieliniowość procesu innowacji i jej zależność od otaczającego „systemu” innowacji, czyli instytucji, podmiotów i szerszego kontekstu społecznego, w którym przebiega innowacja. Sugeruje to istnienie różnorodności mechanizmów, które łączą działalność badawczą i zmiany społeczne. Idea „wpływu” jest więc problematyczna. Narzuca ona ideę liniowości na bardzo złożoną rzeczywistość i sugeruje, że to badania są w jakiś sposób przyczyną zmian w społeczeństwie i gospodarce, podczas gdy badania mogą być tyleż efektem, co przyczyną. Głównym problemem w analizie wpływu jest połączenie interwencji z jej efektami. Interwencje przebiegają jednak w dynamicznym środowisku. Zachodzi wiele zmian i nie zawsze jest oczywiste, co by się stało, gdyby do interwencji nie doszło<sup>72</sup>.

**Trudności oceny.** Często niełatwo opisać powiązania biegnące od „wkładów” (polityk) do „wyników” (efektów) w sposób liniowy; szczególnym wyzwaniem jest przeprowadzenie oceny ilościowej. Ekonometryczne specyfikacje przyczynowości cierpią na problem błędnego koła, gdy zachodzi sprzężenie zwrotne zależne od zmiennej niezależnej. Jest tak np. w ocenie wpływu polityki innowacyjnej na dobrobyt, produktywność lub wzrost PKB lub w ocenie wpływu polityki naukowej na wzrost gospodarczy<sup>73</sup>.

Trudno jest określać i mierzyć różnego typu pozanaukowe korzyści z inwestycji B+R. Przyczyny tego stanu wynikają z:

- braku prostej przekładni nauka – technika – ekonomia oraz nauka – technika – społeczeństwo; droga od odkrycia naukowego do innowacji rynkowej jest złożona i składa się z wielu faz, z których tylko jedną tworzą B+R;
- faktu, że efekty bezpośrednie współwystępują z reguły z pośrednimi, a mierzalne – z niemierzalnymi (np. pośrednim efektem badań kosmicznych jest rozwój nowych materiałów, a pośrednim efektem badań w dziedzinie farmacji – wzrost zdrowotności społeczeństwa);
- faktu, że efekty finansowe innowacji zależą od wielu czynników ubocznych, niezwiązanych z B+R, jak wielkość i struktura gospodarki, stopień specjalizacji naukowej i gospodarczej, poziom technologiczny kraju, stan infrastruktury bankowej i handlowej, charakter i wielkość rynku, ukształtowanie cen, wysokość stopy procentowej, poziom wykształcenia społeczeństwa, rozwój przedsiębiorczości itd.;

---

<sup>72</sup> OECD, *Assessing The Impact Of State Interventions In Research – Techniques, Issues And Solutions* DSTI/STP(2014)24.

<sup>73</sup> Ibidem.

- faktu, że konkretna innowacja nigdy nie polega na wykorzystaniu jednego elementu wiedzy, a niemal zawsze opiera się na wykorzystaniu wielu różnych rodzajów wiedzy;
- faktu, że ocena efektów bywa zależna od obranej perspektywy czasowej;
- różnicy w długości cykli technologicznych w różnych sektorach przemysłu (od kilkudziesięciu lat do kilkunastu miesięcy)<sup>74</sup>.

Ponadto:

- każde pole badań i sektor przemysłu produkcji i kanałów tworzy specyficzne kanały wpływu na użytkownika końcowego. To sprawia, że trudno jest opracować jednolite ramy oceny;
- badania podstawowe mogą mieć różne skutki, z których nie wszystkie da się łatwo określić;
- nietatwo lub nietanio jest dokonać identyfikacji wszystkich osób korzystających z wyników badań, zwłaszcza badań podstawowych;
- trudno jest zidentyfikować i opisać wszystkie mechanizmy przekazywania wyników badań do społeczeństwa. Badania opisały mechanizmy transferu między przedsiębiorstwami lub między uczelniami a przedsiębiorstwami. Są to głównie modele empiryczne, które często nie ujawniają pełnego wpływu badań na społeczeństwo;
- ze względu na brak potrzebnych kategorii beneficjentów, mechanizmów transferu i użytkowników końcowych trudno określić odpowiednie wskaźniki statystyczne dla pomiaru konkretnych wyników badań;
- badania międzynarodowych przepływów wiedzy dowodzą, że większy wpływ społeczny w danym kraju mogą wywrzeć wyniki zagranicznego B+R (szczególnie ucieleśnionego w importowanych towarach inwestycyjnych i konsumpcyjnych) niż krajowego B+R;
- najczęściej trudno ująć wpływ w wartościach monetarnych w celu zapewnienia porównywalności. Nawet jeśli można określić wpływ nieekonomiczny, trudno ustalić jego wartość pieniężną. Podejmowano próby określania wartości pieniężnej niektórych oddziaływań, np. ustalano oszczędności ekonomiczne związane ze zdrową populacją, ale próby te były zazwyczaj częściowe i subiektywne<sup>75</sup>.

---

<sup>74</sup> J. Moszczyński i J. Kozłowski, *Informacja w sprawie możliwości oszacowania nakładów i efektów polityki naukowej i naukowo-technicznej*, 1992, msp. niepubl.; OECD *Linking Science to Innovation, Economic Growth and Welfare: Scoping Paper*, 2013.

<sup>75</sup> OECD, *Linking Science to Innovation, Economic Growth and Welfare: Scoping Paper*, 2013; OECD, *Assessing the Socio-economic Impacts of Public R&D: Recent Practices and Perspectives*, [w:] *OECD Science, Technology and Industry Outlook 2008*.

## *Główne problemy oceny wpływu programów*

Główne problemy oceny wpływu programów ujmuje się w formie następujących pytań:

1. Jak przypisać interwencję do obserwowanych skutków?
2. Jak określić sytuację kontrfaktyczną, tzn. czy obserwowane skutki wystąpiłyby i tak?
3. Jak radzić sobie z opóźnieniem czasu między wynikami badań i skutkami?
4. Na czym skupić się w ocenie: na projektach badawczych czy na badaniach w ogóle?<sup>76</sup>

Na pytania te udziela się odpowiedzi dzięki następującym metodom (tabela 2.9).

**Tabela 2.9**

### **Ocena wpływu programów – pytania i metody badań**

Pytania	Metody odpowiedzi
Ile wydano dotychczas? Czy osiągnięty dotąd postęp odpowiada oczekiwaniom zgodnym z poziomem wydatków?	Analiza kosztów i korzyści
Jak środki finansowe zostały przekształcone w pożądane rezultaty i efekty?	<ul style="list-style-type: none"> <li>• <i>Peer review</i>; panel ekspertów</li> <li>• Studia przypadku</li> <li>• Badania ekonometryczne</li> </ul>
Czy badania w ramach programu miały wysoką jakość naukową? Czy były dobrze i efektywnie zarządzane?	<i>Peer review</i> ; panel ekspertów
Czy program wzmacnia sieci badawcze?	<ul style="list-style-type: none"> <li>• Analiza sieci</li> <li>• Metoda „przed-po”</li> </ul>
Jak działają mechanizmy zarządzania programem? Jak można je wzmocnić?	<ul style="list-style-type: none"> <li>• Monitoring</li> <li>• Metoda „przed-po”</li> </ul>
Jakie są skodyfikowane produkty wiedzy osiągnięte w ramach programu?	Badania ekonometryczne
Jaka jest efektywność programu w porównaniu z podobnymi programami?	Benchmarking
Czy uzyskane patenty są wartościowe? Do jakich należą grup technologicznych?	Analizy patentowe
W jakim stopniu produkty programu zostały skomercjalizowane?	Wskaźniki statystyczne
Jakie czynniki wpłynęły na adaptację przez przemysł technologii opracowanych w ramach programu?	Studia przypadku
Jakie są dotychczasowe korzyści i koszty rozwoju technologii?	Analiza kosztów i korzyści

<sup>76</sup> ESF, *The Challenges of Impact Assessment*, 2012.

Pytania	Metody odpowiedzi
Gdybyśmy mieli drugi raz okazję, czy jeszcze raz uruchomilibyśmy program lub inicjatywę?	Peer review; panel ekspertów
Jakie korzyści z programu odniosło społeczeństwo? Czy działania (np. program badawczy) pokazują „wartość za pieniądze”?	<ul style="list-style-type: none"> <li>• Modele ekonomiczne</li> <li>• Studia przypadku</li> </ul>
W jaki sposób badania prowadzone w ramach programu wpłynęły na polityków?	Metoda SIAMPI (Social Impact Assessment Methods)
Jak badania wpłynęły na poziom zdrowia społeczeństwa?	
Jak wiedza wytworzona w ramach programu jest wykorzystywana przez polityków?	Wywiady, analiza dokumentów

Źródło: ESF, *The Challenges of Impact Assessment*, 2012.

### **Ekonometryczna ocena wpływu B+R**

Badania ekonometryczne analizują wkład inwestycji B+R we wzrost gospodarczy, zarówno poprzez badania mikroekonometryczne, korzystające z danych na temat firm i produktywności przemysłu w celu oszacowania prywatnego i społecznego zwrotu z inwestycji w B+R, jak i poprzez badania makroekonometryczne, które szacują udział ogólnych inwestycji B+R w zagregowanej produktywności.

Badania mikroekonometryczne analizują wkład inwestycji w B+R we wzrost wydajności w firmach prywatnych, oceniają obecność efektów zewnętrznych (ang. *spillover effects*) oraz obliczają stopę społecznego zwrotu, czyli korzyści, jakie z prywatnych inwestycji w B+R odnoszą inne firmy w tej samej branży (lub także firmy w innych branżach). Mogą one zostać użyte do sprawdzenia, czy współpraca firm z publicznym sektorem badań wpływa na zyski lub obroty firm<sup>77</sup>.

Badania makroekonometryczne analizują ogólny wpływ B+R na wzrost PKB lub Współczynnika Produktywności Ogółem (TFP) oraz mierzą przepływy wiedzy (ang. *spillover effects*) do różnych firm i branż<sup>78</sup>. Stosują one różne modele oparte na funkcji produkcji – na najprostszym poziomie prace badawczo-rozwojowe mogą być dodawane do modelu regresji, obok innych środków produkcji, takich jak kapitał i praca. Podejście makroekonometryczne zastosowane do badania wpływu wydatków publicznych na B+R na wzrost gospodarczy pozwala uchwycić skutki dla całej gospodarki dzięki stosunkowo małej ilości danych<sup>79</sup>.

Ograniczeniem badań ekonometrycznych jest to, że opierając się na liniowym ujęciu innowacji, ignorują one powiązania między podmiotami B+R. Prócz tego,

<sup>77</sup> OECD RIHR, *Activity 1.2: Assessing The Role Of Impact In National Science And Research Systems*, 2013.

<sup>78</sup> OECD, *Assessing the Socio-economic Impacts of Public R&D: Recent Practices and Perspectives*, [w:] *OECD Science, Technology and Industry Outlook 2008*.

<sup>79</sup> OECD RIHR, *Activity 1.2*, op. cit.

choć wykazują one związki między zmiennymi, rzadko przedstawiają związki przyczynowe. Co więcej, skupiają się na powiązaniach między B+R a wzrostem produkcji i produktywności, wyłączając z analizy inne cele badań, takie jak bezpieczeństwo narodowe, bezpieczeństwo energetyczne, ochrona środowiska lub zdrowia, spójność społeczna<sup>80</sup>.

### *Nieekonometryczne oceny wpływu gospodarczego*

Inne, nieekonometryczne (ilościowe i jakościowe) metody oceny ekonomicznego zwrotu z inwestycji B+R (lub oceny wpływu czy korzyści) to:

- studia przypadku (ang. *case studies*). Metoda polega na badaniu konkretnego projektu lub grupy projektów badawczych. Koszty B+R projektu porównuje się z korzyściami (z reguły zmniejszeniem kosztów dla użytkownika) związanymi z komercjalizacją wyniku innowacji. Następnie stosuje się standardowe techniki oceny inwestycji dla określenia zwrotów, zarówno prywatnych, jak i publicznych<sup>81</sup>;
- metody retrospektywne (takie jak amerykańskie projekty HINSIGHT i TRACE);
- ocena kosztów i korzyści;
- panel ekspertów<sup>82</sup>;
- analizy sieci<sup>83</sup> i inne.

Prace ekonometryczne są obecnie uzupełniane przez rachunki PKB traktujące publiczne i prywatne inwestycje w badania i rozwój jako źródła inwestycji produkcyjnych. Włączenie B+R do rachunków narodowych wynika z przejścia od tradycyjnego traktowania B+R jako wydatków bieżących do postrzegania ich jako inwestycji w kapitał niematerialny, w zasoby wiedzy kraju.

Kwantyfikacja wpływu publicznych B+R lepiej udaje się na poziomie pojedynczych programów, agencji finansujących oraz organizacji wykonawczych, takich jak szkoły wyższe<sup>84</sup>.

Ważne kategorie oceny wpływu badań (oraz B+R) finansowanych ze źródeł publicznych, jakie rozwinęły się od lat 80., to „efekty – bezpośrednie i pośrednie” oraz „zwrot z inwestycji”.

Przez efekty bezpośrednie rozumie się odkrycia naukowe wykorzystywane w rozwoju nowego wyrobu lub nowych metod wytwarzania, czyli transfer z nauki

<sup>80</sup> OECD, *Assessing the Socio-economic Impacts of Public R&D*, op. cit.

<sup>81</sup> K. Smith, *Economic returns to R&D: Methods, Results and Challenges*, „SPSG Review Paper” 1991, no. 3.

<sup>82</sup> R.N. Kostoff, *Federal Research Impact Assessment: Axioms, Approaches, Applications*, „Scientometrics” 1995, vol. 34, no. 2.

<sup>83</sup> IPTS-Johanneum, *RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies*, 2002.

<sup>84</sup> Australian Government, Australian Research Council, *Assessing the wider benefits arising from university-based research: Discussion Paper*, June 2013.



do gospodarki informacji o wynikach badań naukowych (np. o promieniach X, laserze). Ten rodzaj korzyści był szeroko znany od lat, harmonizował z koncepcją liniowego modelu innowacji oraz stanowił uzasadnienie dla finansowania nauki w okresie, gdy była ona oceniana przez samych uczonych.

Efekty pośrednie, których znaczenie uświadomiono sobie w początku lat 90. pod wpływem nowego modelu innowacji, określono jako takie owoce badań naukowych, jak *know-how* (obeznanie z metodami i technikami, niekoniecznie dające się ująć w słowa), przeszkoleni naukowcy i inżynierowie, sieci zawodowe badaczy, narzędzia projektowania i techniki inżynieryjne (modelowanie, symulacje), nowa aparatura badawcza czy też „firmy odpryskowe”. Efekty pośrednie występują np. wówczas, gdy badacze wiedzę o ostatnich odkryciach uzyskaną dzięki własnym badaniom wykorzystują w doradztwie albo gdy *know-how* osiągnięty dzięki szkoleniom stosują w działalności innowacyjnej, albo też gdy aparatura skonstruowana na potrzeby badań znajduje nieoczekiwane zastosowania w produkcji itd., co we wszystkich tych przypadkach przyczynia się do wzrostu gospodarczego. Badanie PREST po raz pierwszy rozróżniło efekty bezpośrednie od pośrednich w roku 1993 i od tego czasu kategorie te stały się jednymi z podstawowych w ewaluacji B+R<sup>85</sup>. Wartość ich należy mierzyć liczbą ich zastosowań w metodologiach ewaluacji oraz ich płodnością. Stały się one punktem wyjścia dla wielu dalszych rozróżnień (tabela 2.10).

Tabela 2.10

**Efekty – bezpośrednie i pośrednie jako kategorie oceny wpływu badań  
(oraz B+R)**

Dziedzina wpływu	Efekty bezpośrednie		Efekty pośrednie	
	Krótkoterminowe	Długofalowe	Krótkoterminowe	Długofalowe
Nauka	Odkrycia naukowe	Korpus wiedzy	Rozwój kształcenia	Wpływ na gospodarkę
Gospodarka i społeczeństwo	Nowe technologie	Nowa wiedza techniczna	Poprawa produktywności	Poprawa konkurencyjności
Polityka	Lepsze rozumienie współzależności	Rozwiązywanie problemów	Lepsze rozumienie problemów	Lepsza ogólna satysfakcja

Źródło: T. Balogh, *From peer review to portfolio analysis. Evaluation of applied research and development*, mszp.

<sup>85</sup> W. Douglas i A.D. Rank, *Measuring the economic benefits of research and development: the current state of the art*, „Research Evaluation”, April 1998, vol. 7, no. 1.

Badając wpływ gospodarczy programów, zaleca się zwracanie uwagi na (tabela 2.11).

Tabela 2.11

## Matryca efektów

Bezpośrednie efekty społeczne (np. poprawa działania telekomunikacji, informacji meteorologicznych).	Bezpośrednie efekty przemysłowe (np. pobudzenie działalności gospodarczej, nowe stanowiska pracy).
Pośrednie efekty społeczne (np. redystrybucja, stabilizacja, handel).	Pośrednie efekty przemysłowe ( <i>spin-offs</i> <sup>a</sup> , <i>fall-outs</i> ).

<sup>a</sup> Badając „odpryski” (*spin offs*), zaleca się zwracanie uwagi na: efekty technologiczne („derywaty” z produktów powstałych w wyniku programu, nowe produkty, różnicowanie oferty, ulepszenia produktów), efekty komercyjne (współpraca międzynarodowa, nowe sieci sprzedaży, wykorzystanie programu w marketingu), efekty w dziedzinie organizacji i zarządzania (kontrola jakości, zarządzanie projektami, techniki produkcyjne), efekty w dziedzinie podniesienia wartości siły roboczej (powstanie masy krytycznej specjalistów, ulepszenie kompetencji siły roboczej).

Źródło: P. Cohendet, *Evaluating the industrial indirect effects of technology programmes: the case of the European Space Agency (ESA) programmes*, [w:] *Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997, s. 193.

Krzyżując pojęcia „bezpośrednich” i „pośrednich” korzyści z kategoriami znanych i nieznanymi sposobów wykorzystania B+R, uzyskuje się następujące sugestie badawcze, np.:

- bezpośrednie korzyści z zamierzonych sposobów wykorzystania wyników badań. Wyniki B+R to np. często infratechnologie (czyli te praktyki i techniki, metody pomiaru i testowania, oraz podstawowe dane naukowe i inżynierskie, dzięki którym wzrasta wydajność i skuteczność B+R i produkcji);
- bezpośrednie korzyści z niezamierzonych sposobów wykorzystania wyników badań. Wpływ na dalsze badania (które, w dalszej przyszłości, zaowocują innowacjami). Badania nad technologiami generycznymi. Badania, które nie odniosły sukcesu, a które niemniej zaowocowały i wartościową „wiedzę negatywną”, i podniesieniem kwalifikacji badawczych zespołu;
- pośrednie korzyści ze znanych sposobów wykorzystania kompetencji rozwiniętych w wyniku prac B+R. Przez „korzyści płynące z kompetencji” należy rozumieć korzyści płynące ze wzrostu wiedzy (innej niż odkrycia naukowe) uzyskanej przez badaczy w wyniku działalności badawczej (jak np. obozowanie ze stanem wiedzy); korzyści płynące ze wzrostu kompetencji badaczy w stosowaniu technik badawczych i rozwiązywaniu problemów technicznych<sup>86</sup>.

<sup>86</sup> D. Williams i A.D. Rank, *Measuring the economic benefits of research and development: the current state of the art*, „Research Evaluation”, April 1998, vol. 7, no. 1.

Zazwyczaj ewaluacje programów rządowych starają się dokumentować różne formy „korzyści”, „wpływów” i „efektów”, jednocześnie starając się odpowiedzieć na pytania: W jakim stopniu B+R finansowane przez państwo pobudza efekty gospodarcze? W jakim stopniu interwencje rządowe korygują ułomności rynku? (tabela 2.12).

Tabela 2.12

## Efekty gospodarcze B+R

Polityka publiczna	Mikroekonomia	Mezoekonomia	Makroekonomia	Teoria ekonomiczna
System	Czy polityki B+R pobudzają współpracę nauki z przemysłem?	Czy polityki B+R wspierają strukturalne dostosowanie?	Czy polityki B+R ulepszają efektywność gospodarczą?	Gospodarka narodowa
Programy	Jaki jest wpływ polityki B+R na powiązania między nauką a sektorami przemysłu?	Czy polityki B+R przyspieszają procesy dyfuzji?	Czy polityki B+R poprawiają przewagi konkurencyjne?	Przemysł
Projekty	Czy polityki B+R pobudzają prywatne B+R?	Czy polityki B+R przyspieszają procesy dyfuzji technologii?	Czy polityki B+R wpływają na procesy produkcji?	Firma

Źródło: H. Capron i B. van Potelsberge de la Potterie, *Public support to R&D Programmes*, [w:] *Policy evaluation in innovation and technology. Towards best practices*, 1997 OECD.

Oprócz wpływu, korzyści i efektu innymi pojęciami ogniskującymi ewaluacje są pojęcia „zwrotu” (z inwestycji).

Ekonomiści (a za nimi politycy) zazwyczaj dokonują rozróżnienia pomiędzy zwrotami „prywatnymi” i „społecznymi”. Prywatny zwrot to ten płynący do osoby lub firmy, która dokonała inwestycji. Dla firmy jest to przede wszystkim wzrastający strumień dochodów, wynikający z pomyślnej inwestycji lub projektu B+R. Społeczny zwrot to taki, który staje się własnością całego społeczeństwa. Innowacja może nie tylko przynosić zwrot osobie lub firmie, ale także (jeśli np. polega ona za zakupie nowej maszyny) może obniżać koszty wytwarzania innemu przedsiębiorstwu, ulepszać jakość wyrobów lub przynosić korzyści użytkownikom. Z punktu widzenia konsumentów B+R może prowadzić do innowacji produktowych, które obniżają koszty usług (np. ogrzewania mieszkań) lub też poprawiają jakość życia (np. w opiece lekarskiej). O zwrotach społecznych mówimy wówczas, gdy korzysta z nich nie tylko inwestor.

Większość analiz zwrotów społecznych koncentruje się na redukcji kosztów. Jeśli innowacja zmniejsza koszty produkcji, to uwalnia zasoby, które klienci

mogą użyć w innych celach. Załóżmy, że innowacja zmniejsza koszty produkcji żywności. Społeczeństwo może albo wytworzyć tę samą ilość żywności, ale dzięki mniejszym nakładom, używając zaoszczędzonych środków do wytworzenia innych dóbr, albo wytworzyć więcej dóbr dzięki tym samym co poprzednio nakładom. W obu przypadkach wzrasta realny dochód narodowy. Gdy więc wyniki B+R prowadzą do zmniejszenia wydatków produkcyjnych, powstają zwroty społeczne<sup>87</sup>.

### *Szacunki wpływu nieekonomicznego*

Literatura na temat wpływu pozaekonomicznych B+R jest znacznie uboższa i mniej solidna niż piśmiennictwo dotyczące skutków ekonomicznych. Według Godina oraz Doré dzieje się tak z trzech powodów:

1. Większość pomiarów nauki i badań przeprowadzono w kontekście gospodarczym.
2. Wymiar ekonomiczny jest często łatwiejszy do zmierzenia.
3. Większość wyników i wpływu nauki ma charakter niematerialny, rozproszony i często opóźniony. Choć również trudny do zmierzenia, ekonomiczny wymiar nauki i techniki łatwiej mierzyć i analizować<sup>88</sup>.

Niemniej jednak w ostatnich dekadach rządy i naukowcy zaczęli się interesować pomiarem skutków ekonomicznych publicznych inwestycji w B+R. Istnieje konsensus wśród naukowców, według którego jednym z pierwszych kroków na drodze pogłębiania wiedzy na temat wpływu pozapublicznego ekonomicznego B+R powinno być określenie ram łączących inwestycje w badania naukowe oraz dobrostan (ang. *well-being*) (tabela 2.13).

W ostatnich dekadach źródłem inspiracji dla wielu propozycji oceny wartości badań naukowych dla społeczeństwa stała się ocena wpływu społecznego (*Social Impact Assessment*). Ocena wpływu społecznego jest nastawiona przede wszystkim na badanie następstw programów społecznych. Bada ona konsekwencje dla ludzi wszelkich publicznych i prywatnych działań, które zmieniają sposób, w jaki ludzie żyją, pracują, bawią się, odnoszą się do siebie wzajemnie, organizują dla zaspokojenia swoich potrzeb oraz, ogólnie, radzą sobie jako członkowie społeczeństwa. Swoim zainteresowaniem obejmuje także wpływ kulturowy polegający na zmianie norm, wartości i wierzeń, które kierują i racjonalizują ich samopoznaniem jako jednostek i jako społeczeństwa<sup>89</sup>.

<sup>87</sup> K. Smith, *Economic returns to R&D: Methods, Results and Challenges*, „SPSG Review Paper” No. 3, November 1991.

<sup>88</sup> B. Godin i C. Doré, *Measuring the Impacts of Science: beyond the Economic Dimension*, Paper presented at the Helsinki Institute for Science and Technology Studies, HIST Lecture, 24 August 2007, Helsinki, Finland.

<sup>89</sup> Metody oceny społecznego wpływu zostały m.in. rozwinięte w ramach programu SIAMPI. SIAMPI to konsorcjum Królewskiej Holenderskiej Akademii Sztuki i Nauki (KNAW), Rathenau Institute (Holandia), CSIC (Hiszpania), MSH (Francja) oraz Uniwersytetu w Manchesterze (Wielka Brytania). Konsorcjum SIAMPI opracowuje metody oceny skutków społecznych projektów

Tabela 2.13

## Ramy analizy wpływu badań na dobrostan

B+R	Sektory	Przejawy dobrostanu
Wykonawcy (biznes, uczelnie itd.). Nauki (przyrodnicze i inżynierskie, społeczne i humanistyczne). Typy badań (podstawowe, stosowane).	Rząd (np. centralny, regionalny, władze miejskie). Biznes (np. telekomunikacja, elektronika). Organizacje <i>non profit</i> (np. uczelnie, organizacje pozarządowe, kościoły).	Ekonomiczne (np. konsumpcja, bogactwo, równość, bezpieczeństwo ekonomiczne). Społeczne (np. poprawa zdrowia, zmniejszenie ubóstwa). Środowiskowe (poprawa stanu zanieczyszczenia, wzrost bioróżnorodności).

Źródło: opracowanie własne na podstawie literatury przedmiotu.

Ewaluacja typu *Social Impact Assessment* korzysta najczęściej z matryc pozwalających na szacowanie zmian w populacji poddanej danemu programowi w trakcie jej realizacji. Zmienne rozumiane są jako zestawy wybranych wskaźników statystycznych (tabela 2.14).

Tabela 2.14

Matryca ewaluacji typu *Social Impact Assessment*

Zmienne	Przed	W trakcie	Po
Charakterystyka populacji			
Struktura wspólnoty			
Zasoby wspólnoty (polityczne i społeczne itd.).			

Źródło: *Guidelines and Principles for Social Impact Assessment*, maj 1994.

## 2.8. Dokonania w dziedzinie badań naukowych i technologicznych (kraju/regionu)

Osiągnięcia naukowe i technologiczne to rodzaj „niepełnej ewaluacji” – ewaluacji, która przedstawia tylko produkty badań, takie jak publikacje, cytowania i patenty, a pomija wkład, wpływ i oddziaływanie. W przeszłości raporty te – najczęściej w formie krajowych raportów wskaźnikowych – sporządzali bibliometry, np. holenderski Centre for Science and Technology Studies (CWTS), obecnie może je sporządzać każdy, kto ma dostęp do aplikacji indeksów cytowań, takich jak *InCites* lub *Scival*.

badawczych, programów badawczych oraz instrumentów finansowania badań w ramach projektu VII Programu Ramowego (2007–2013) SIAMPI SIAMPI <http://www.siampi.eu/>

Raporty wskaźnikowe oparte są na danych bibliometrycznych i patentowych. Najczęściej używane wskaźniki zawiera tabela 2.15.

Tabela 2.15

### Wskaźniki oparte na danych bibliometrycznych i patentowych

#### 1. Produkty badań (publikacje)

Publikacje według dziedzin i sektorów.

Udział w światowej produkcji naukowej według dziedzin i sektorów.

Wskaźniki aktywności w różnych dziedzinach nauki.

Stopień koncentracji tematycznej wyników badań.

Rozkład tematyczny i ewolucja wyników badań.

Trendy w publikacji czasopism naukowych.

#### 2. Wpływ badań

Znormalizowany (według dziedzin) pomiar wpływu działalności badawczej w różnych dziedzinach i sektorach.

Wzrost cytowań w różnych dziedzinach i sektorach.

Udział w światowej puli cytowań w różnych dziedzinach i sektorach.

Rozmieszczenie geograficzne instytucji o największym wpływie.

Trendy w zakresie cytowań w różnych dziedzinach i sektorach.

Przepływ cytowań na świecie. Kto kogo cytuje.

#### 3. Współpraca w zakresie badań

Udział w puli światowej publikacji według rodzaju współpracy, dziedzin nauki i sektorów.

Trendy w zakresie współpracy według dziedzin nauki i sektorów.

Wpływ produkcji naukowej i zakres międzynarodowej współpracy naukowej w przeliczeniu na jednego mieszkańca kraju.

Liczba publikacji w danym kraju według rodzaju instytucjonalnego współautorstwa (współautorstwo krajowe i międzynarodowe, tylko międzynarodowe, tylko krajowe) oraz według poziomu znormalizowanego wpływu (poniżej 1, od 1 do 1,75, powyżej 1,75).

#### 4. Doskonałość badawcza

Udział w puli światowej wysoko cytowanych publikacji.

Udział w puli światowej kwartyla różnych dziedzin i sektorów.

Udział w puli światowej 10% najczęściej cytowanych artykułów w danej dziedzinie.

Źródło: *Scoping the NESTI Contribution to the CSTP activity on the Impact of Science and Technology over the 2013–2014*, DSTI/EAS/STP/NESTI(2013)4.

## 2.9. Dyscypliny naukowe i technologiczne

Ewaluację dyscyplin naukowych na poziomie krajów przeprowadza się przede wszystkim w krajach nordyckich, w Niemczech, Irlandii, Holandii, Flandrii oraz w Wielkiej Brytanii i Stanach Zjednoczonych. Ewaluację przeprowadzają na ogół na zlecenie Rad Badawczych zagraniczni uczeni, którzy wizytują instytucje naukowe i przygotowują końcowy raport.

Na ogół ewaluację łączą metody ilościowe (bibliometria) i jakościowe (raporty „samooceny”, analizy dokumentów, wizytacje panelu ekspertów). W krajach nordyckich ewaluację dyscyplin przeprowadza się albo oddzielnie w poszczególnych krajach, albo jednocześnie w dwóch lub we wszystkich krajach. W Danii ewaluację przeprowadza, zleca i koordynuje Duńska Agencja ds. Nauki, Tech-

nologii i Innowacji, w Finlandii – Akademia Fińska; w Norwegii – Norweska Rada Badawcza; w Szwecji – odpowiednie dziedzinowe Rady Badawcze.

Zazwyczaj ocenia się jakość badań w danej dyscyplinie, jej słabe i mocne strony oraz potencjał rozwojowy. Sięga się po porównania między instytucjami oraz porównania międzynarodowe. Celem ewaluacji zazwyczaj jest poprawa międzynarodowej pozycji kraju.

Ewaluację prowadzi się nie tylko jako wkład do strategii finansowania lub do przeprowadzenia korekt w istniejących strategiach finansowania rad, ale również jako wkład do prac nad priorytetami uniwersytetów i ministerstw<sup>90</sup>.

Na przykład w Norwegii w ciągu 12 lat przeprowadzono następujących 16 ewaluacji:

- chemia (1997);
- nauki o Ziemi (1998);
- biologia, w tym biomedycyna (2000);
- pedagogika (2004);
- nauki technologiczne i inżynieryjne (2004);
- filologia skandynawska (2005);
- fizyka (2000);
- matematyka (2002);
- ICT (2002);
- językoznawstwo (2002);
- nauki polityczne (2002);
- medycyna i zdrowie (2004);
- farmakologia (2006);
- studia nad rozwojem (2007);
- ekonomia (2007);
- nauki historyczne (2008).

Norweska Rada Badawcza (NRB) wyznaczała grupy badawcze, które miały być poddane ewaluacji, decydowała o warunkach zamówienia, wybierała metodę, powoływała komisję przetargową. Raport końcowy przesyłano ocenianym. Mogli oni ustosunkować się do niego.

Efektami ewaluacji były:

- na poziomie ministerstwa – utworzenie norweskiego programu centrum doskonałości, nowe programy, wzrost nacisku na wspieranie przywództwa w badaniach;
- na poziomie rady – wprowadzenie w życie nowych instrumentów polityki oraz utworzenie specjalnych funduszy;
- na poziomie instytucji naukowych – reorganizacje, nowa współpraca badawcza<sup>91</sup>.

---

<sup>90</sup> European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies: A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes*, 2009.

<sup>91</sup> G.E.M Helgesen, *Evaluating Research Fields/Disciplines in Norway*, 2008.

SIWZ dla ewaluacji chemii obejmował krytyczny przegląd mocnych i słabych badań chemicznych w Norwegii; ocenę jakości naukowej badań podstawowych w chemii; identyfikację grup badawczych, które osiągnęły wysoki poziom międzynarodowy lub mają potencjał do osiągnięcia takiego poziomu; identyfikację obszarów badań mających znaczenie dla przyszłości kraju, które powinny zostać wzmocnione<sup>92</sup>.

Oprócz ewaluacji dyscyplin zarządzanych na szczeblu rządowym podejmują je także grupy uczelni z użyciem metod bibliometrycznych (*Scival Spotlight*) i statystycznych, z naciskiem na poszukiwanie doskonałości badawczej nie tyle w dyscyplinowych instytucjach naukowych, ile w zespołach badawczych, bez względu na ich dyscyplinowe afiliacje<sup>93</sup>.

Podejmuje się także próby ewaluacji dziedzin nauki na szczeblu europejskim<sup>94</sup> lub ewaluację „wyłaniających się” dyscyplin na poziomie kraju<sup>95</sup>.

Jednym z następstw ewaluacji dyscyplin jest wzrost zainteresowania zróżnicowaniem pojęcia jakości oraz zwrócenie uwagi na zróżnicowanie charakteru rezultatów badań w różnych dyscyplinach<sup>96</sup>.

## 2.10. Udział kraju w Programach Ramowych oraz innych programach EU

Ewaluację Programu Ramowego (PR) przeprowadza zarówno Unia Europejska (*ex ante*, *interim* oraz *ex post*), jak i – w zakresie swojego udziału w PR – poszczególne kraje (*ex post*). Oprócz nich przeprowadza się także ewaluacje działania Krajowych Punktów Kontaktowych<sup>97</sup>.

---

<sup>92</sup> European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies: A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes*, 2009.

<sup>93</sup> P. Wellings i R. Winzer, *Mapping Research Excellence: exploring the links between research excellence and research funding Policy*, The 1994 Group, September 2011.

<sup>94</sup> HERA, *Feasibility Study: The Evaluation And Benchmarking Of Humanities Research In Europe*, August 2007.

<sup>95</sup> Quick Scan, *Evaluation of Research in Context A Quick Scan of an Emerging Field*, 2007.

<sup>96</sup> J. Kekäle, *Conceptions Of Quality In Four Different Disciplines*, „Tertiary Education and Management” 2002, 8; M. Huang i Yu. Chang, *Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective*, „Journal of the American Society For Information Science And Technology” 2008, 59(11); N. Rons et al., *Research evaluation per discipline: a peer-review method and its outcomes*, 2007.

<sup>97</sup> Oprócz cytowanych niżej pozycji, korzystałem z następujących źródeł: A. Torras i Estruch, *Catalan companies in the European Framework Programme 2009; Effects of Swiss participation in EU Research Framework Programmes Interim Report 2010; Finland – Evaluation of Participation and Network in FP6 Tekes Programme Report 6/2008; Impact assessment of Framework Programme on French research and innovation system 2010; Instytut Badań nad Gospodarką Rynkową, Przyczyny sukcesów i niepowodzeń zespołów naukowych biorących udział w Szóstym Programie Ramowym – wnioski na przyszłość*, Warszawa, lipiec 2006; L. Vavříková, *Assessment of Fp6 In The Czech Republic. Impact Assessment of the 6th European RTD Framework Programme*, 2010; OPI *Ewaluacja udziału polskich zespołów badawczych w 6 Programie Ramowym – ocena efektywności finansowej*,



**Ewaluacja w UE.** Aż do Szóstego Programu Ramowego (6PR) ewaluacji PR dokonywano poprzez roczne panele niezależnych ekspertów, które oceniały jego wybrane obszary; system monitorowania 7PR oparto na zestawie wskaźników i zaprojektowano jako narzędzie do zarządzania wewnętrznego<sup>98</sup>. Kwestie reguluje m.in. rozporządzenie Parlamentu Europejskiego i Rady ustanawiające „Horyzont 2020”<sup>99</sup>.

Wśród dotychczasowych ewaluacji PR przeważały oceny wpływu badań<sup>100</sup>, w szczególności oparte na analizie dodatkowości, a także ewaluacje skupiające się na wpływie PR na międzynarodową współpracę badawczą<sup>101</sup> oraz analizy bibliometryczne dokonań<sup>102</sup>. W ocenie wpływu PR sięgnięto także po analizę semantyczną<sup>103</sup>.

**Ewaluacje udziału krajów w Programach Ramowych.** Ewaluacje udziału krajów w Programach Ramowych przeprowadza się w wielu krajach; np. w odniesieniu do PR VI, m.in. w takich krajach jak Austria, Belgia (Flandria i Bruksela), Cypr, Czechy, Dania, Estonia, Finlandia, Francja, Hiszpania, Irlandia, Islandia, Litwa, Łotwa, Niemcy, Norwegia, Szwajcaria, Szwecja, Turcja, Wielka Brytania<sup>104</sup>.

Ewaluacje albo dotyczą ogółu instytucji uczestniczących w PR, albo pewnej ich grupy, np. firm lub publicznego sektora nauki; albo wszystkich dyscyplin, albo tylko wybranych, takich jak np. teleinformatyka, energia, środowisko. Z reguły korzystają one z wielu metod ilościowych i jakościowych – np. z bibliometrii, ankiet, wywiadów.

Tytułem przykładu: w Hiszpanii ogólnym celem ewaluacji było dokonanie ilościowej i jakościowej oceny wpływu uczestnictwa w ramach Szóstego Pro-

---

*odpowiedniości, skuteczności i użyteczności* 2008; *Participation of German Higher Education Institutions In FP6* 2009; PRO INNO Europe, *Impact of publicly funded research on innovation. An analysis of European Framework Programmes* 2009; *Spanish National FP6 Impact Assessment Study* 2010; *Switzerland's Participation in the 6th EU Research FP* 2008; V. Albrecht i J. Vaněček, *Assessment of Participation of the Czech Republic in the EU Framework Programmes*, Technology Centre of the Academy of Sciences of the Czech Republic 2009.

<sup>98</sup> EC European Commission, *First FP7 Monitoring-Report* 2009.

<sup>99</sup> *Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające „Horyzont 2020” – program ramowy w zakresie badań naukowych i innowacji (2014–2020)*, Bruksela, dnia 30.11.2011 COM(2011) 809 końcowy 2011/0401 (COD).

<sup>100</sup> *Assessing the Socio-economic Impacts of the Framework Programme*, A study for DG Research by PREST, AUEB, BETA, ISI, Joanneum Research, IE HAS and Wise Guys, 2002.

<sup>101</sup> J. Edler, *The role of International Collaboration in the Framework Programme. Expert Analysis In Support Of The Ex Post Evaluation Of FP6*, 2008; F. Malerba, *Network Analysis of Research Collaborations*, Conference “Framework Programme Evaluation: Taking Stock and Moving Forward” Brussels, 18 June 2009.

<sup>102</sup> Th. Reuters, *Expert Group for the Interim Evaluation of Framework Programme 7. Bibliometric analysis* 2010.

<sup>103</sup> V. Duchêne i Ir. P. De Mazière, *Semantic analysis as tool to measure policy impact. The impact of FP-funded research in Social Sciences and Humanities (SSH) on EU policies*, June 18th, 2009.

<sup>104</sup> *Compilation of FP6 National Impact Studies*, April 2008.

gramu Ramowego (6PR) UE hiszpańskich uniwersytetów i publicznych instytutów badawczych. W badaniu zastosowano kilka kryteriów oceny, takie jak kariery naukowców, przebieg ich badań, publikacje, internacjonalizacja projektów badawczych, współpraca z sektorem przedsiębiorstw, poprawa infrastruktury badawczej itp. Ewaluacja miała trzy etapy. W pierwszym etapie sięgnięto po analizy ilościowe uzyskane za pomocą ankiet wysyłanych do naukowców uczestniczących w PR. Drugi etap polegał na przeprowadzeniu kilku analiz jakościowych za pomocą wywiadów z przedstawicielami instytucji naukowych. Trzeci etap oceny miał formę warsztatów z przedstawicielami uniwersytetów i publicznych instytutów badawczych. Warsztaty pozwoliły na dyskusję i analizę wyników poprzednich etapów<sup>105</sup>.

Z kolei ewaluacja brytyjska dotyczyła udziału kraju we wszystkich – do szóstego – Programach Ramowych. Miała ona dwa główne cele: określenie stopnia, do jakiego kolejne PR wpłynęły na wzmocnienie najsilniejszych brytyjskich pól badawczych, oraz zbadanie charakteru i zakresu, w jakim PR pomógł zainicjować nowe obszary badań. Najbardziej szczegółowo badano następujące kwestie:

- stopień zgodności między priorytetami PR a priorytetami brytyjskich strategii badawczych naukowych;
- zakres, w którym główne grupy badawcze z Wielkiej Brytanii były zaangażowane w PR, oraz powody braku zaangażowania;
- stopień, w jakim PR wspierają luki w finansowaniu nauki;
- wpływ PR na:
  - możliwości badawcze, umiejętności i kariery;
  - rozwój przemysłowy i konkurencyjność;
  - politykę Wielkiej Brytanii oraz krajowe finansowanie badań i rozwoju technologicznego;
  - współpracę brytyjskich naukowców z naukowcami za granicą;
- mechanizmy PR najsilniej związane z pozytywnym wpływem i korzyściami<sup>106</sup>.

**Ewaluacje Krajowych Punktów Kontaktowych.** W krajach uczestniczących w Programie Ramowym przeprowadza się także ewaluacje Krajowych Punktów Kontaktowych.

Na przykład w Polsce celem badania było dokonanie oceny funkcjonowania sieci Punktów Kontaktowych 7 Programu Ramowego w kontekście efektywności działań oraz dokonanie analizy porównawczej z systemami Punktów Kontaktowych w wybranych państwach członkowskich UE. W szczególności badanie postawiło przed sobą następujące zadania:

- przeprowadzenie analizy struktury sieci Punktów Kontaktowych w Polsce oraz analizy instytucji współpracujących w sieci na poziomie europejskim/krajowym i regionalnym, a także głównych przedsięwzięć z udziałem ośro-

<sup>105</sup> *Evaluation of the impact of the FP6 in the RTD public system in Spain*, 2010.

<sup>106</sup> *Technopolis, Impact of the EU RTD Framework Programme on the UK*, 2010.

- ków sieci – w tym przepływu informacji pomiędzy nimi, ze szczególnym uwzględnieniem roli i efektów pracy Krajowego Punktu Kontaktowego;
- dokonanie analizy usług/wsparcia udzielonego przez Regionalne, Branżowe oraz Lokalne Punkty Kontaktowe, w ramach działania sieci;
  - porównanie zakresu, zadań, struktury zatrudnienia i klientów Narodowych Punktów Kontaktowych w krajach takich jak: Polska, Austria, Czechy, Wielka Brytania, Irlandia;
  - przygotowanie wniosków i rekomendacji istotnych z punktu widzenia poprawy efektywności funkcjonowania sieci Punktów Kontaktowych w Polsce<sup>107</sup>.

### **2.11. Resorty ds. nauki oraz agencje ds. badań naukowych i/lub technologicznych (zarządzanie oraz systemy grantowe)**

Jak podkreślono, z reguły agencje finansujące badania naukowe przeprowadzają trzy rodzaje ewaluacji: bieżący monitoring oraz ewaluacje wewnętrzne i zewnętrzne. Ewaluacje zewnętrzne dotyczą albo realizowanych przez agencje programów, albo całokształtu ich działalności. Całościowe ewaluacje oceniają poszczególne programy pod kątem stopnia, w jakim realizują one statutową misję organizacji, oraz politykę agencji z punktu widzenia realizacji misji. Ponadto, z reguły, oceniają one także samą misję z punktu widzenia priorytetów rządowych oraz potrzeb nauki danego kraju. Przedmiotem oceny są także cele, założenia merytoryczne, zasady finansowania oraz zarządzanie programem – pod względem ich jakości, skuteczności i wpływu.

Ewaluacja pozwala agencji:

- wypracować podstawy merytoryczne dla rozwoju strategii;
- udokumentować praktyki finansowania i tym samym zapewnić przejrzystość w sprawach dotyczących pieniędzy podatników;
- podjąć decyzję o alokacji zasobów;
- wspierać proces uczenia się o systemie finansowania, co może prowadzić do ulepszenia programów;
- wykazać, że wykonywane badania oraz organizacje finansujące badania naukowe są „rozliczalne” (*accountable*);
- przekazywać zainteresowanym informacje o wpływie i wynikach programów rządowych;
- pozwolić beneficjentom wyrazić swoje opinie na temat systemu finansowania, procedur składania wniosków oraz warunków prowadzenia badań (np. przez wizyty w terenie, wywiady, badania ankietowe)<sup>108</sup>.

<sup>107</sup> PSDB, *Funkcjonowanie Sieci Punktów Kontaktowych w Polsce*, 2008.

<sup>108</sup> ESF, European Science Foundation, *A Guide To Evaluation Activities In Research Funding Agencies. Quality Assurance And Evaluation Guidelines*, 2012.

Przed kilkoma laty pomoc w przeprowadzaniu całościowych ewaluacji oferowało Forum Organizacji ds. Oceny Programów Finansowania i Programów Badawczych Europejskiej Fundacji Nauki<sup>109</sup>, a obecnie organizacja Science Europe<sup>110</sup>.

**Ewaluacja programów.** Wyróżnia się siedem typów programów:

1. Tryb responsywny – programy prowadzone na zasadach konkursowych, w których badacze w regularnych odstępach czasu mogą się ubiegać o sfinansowanie ich projektów we wszystkich (lub w wielu) dziedzinach.
2. Instrumenty rozwoju kariery – programy mające na celu przyciągnąć, rozwinać i zatrzymać utalentowanych badaczy. Często są one przeznaczone dla konkretnych obszarów badań lub dla pewnych etapów kariery.
3. Programy centrów doskonałości to zazwyczaj znaczne długoterminowe dotacje, oparte na zasadach konkursowych. Centra mogą mieć charakter międzyinstytucjonalnych sieci badawczych z badaczami, którzy pracują razem według wspólnie uzgodnionego programu pracy.
4. Programy tematyczne to programy, które określają finansowanie programów badawczych w ściśle określonym obszarze tematycznym.
5. Transfer wiedzy/ Współpraca z gospodarką – dotacje na transfer wiedzy i technologii.
6. Infrastruktura/ aparatura – programy przeznaczone na finansowanie dużego sprzętu badawczego i infrastruktury.
7. Inne<sup>111</sup>.

Każdy typ charakteryzuje się stosowaniem innych metod, zadawaniem innych pytań, sięganiem po odmienne wskaźniki, innym charakterem kontynuacji, innym okresem trwania oraz innymi kosztami (tabela 2.16).

**Tabela 2.16**

**Charakterystyki programów**

	<b>Responsywne/ kariery badawcze</b>	<b>Centra doskonałości</b>	<b>Tematyczne</b>
Metoda	Analiza dokumentów, aplikacji i nagród, analiza rezultatów, badania ankietowe, wywiady, grupy fokusowe.	Samoocena i międzynarodowy panel ekspertów. Literatura przedmiotu, analizy statystyczne, wywiady z zainteresowanymi stronami, analiza sieci, analiza bibliometryczna, porównania międzynarodowe.	Wywiady, statystyki, analizy dokumentów.

<sup>109</sup> <http://archives.esf.org/coordinating-research/mo-fora/evaluation-of-funding-schemes-and-research-programmes.html>.

<sup>110</sup> <http://www.scienceurope.org/policy/working-groups/research-policy-and-programme-evaluation/>.

<sup>111</sup> ESF, European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies. A report of the ESF Member Organisation Forum on Ex-Post. Evaluation of Funding Schemes and Research Programmes*, September 2009.

	<b>Responsywne/ kariery badawcze</b>	<b>Centra doskonałości</b>	<b>Tematyczne</b>
Pytania	Czy program osiągnął cele? Czy cele zachowują ważność? Czy warto kontynuować? Jak ulepszyć?	Czy program osiągnął cele? Czy cele zachowują ważność? Czy warto kontynuować? Jak ulepszyć?	Jaki jest odbiór programu w środowisku naukowym? Czy program osiągnął cele? Czy cele zachowują ważność? Czy warto kontynuować? Jak ulepszyć?
Wskaźniki	Ilościowe i jakościowe <ul style="list-style-type: none"> <li>• produkty, rezultaty i oddziaływania;</li> <li>• poziom zadowolenia (procedury, finansowanie).</li> </ul>	Ilościowe i jakościowe Liczba publikacji, cytowania, dorobek, budżety. Struktura współpracy, kwestie zarządzania i przywództwa, opinia zainteresowanych.	Ilościowe i jakościowe Wnioski i decyzje grantowe. Budżet. Porównanie z programami innych rad.

Źródło: ESF, European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies. A report of the ESF Member Organisation Forum on Ex-Post. Evaluation of Funding Schemes and Research Programmes*, September 2009.

**Ewaluacja agencji.** Agencje finansujące badania poddaje się od czasu do czasu ewaluacji celem oceny, czy dobrze spełniają one swoją funkcję w ekosystemie badań i innowacji.

Jako przykłady udanych ewaluacji przytacza się następujące:

- ewaluacja systemowa Niemieckiej Fundacji Badawczej (DFG) i Towarzystwa Maxa Plancka (MPG) w 1999 roku. Obie instytucje i MPG (DFG) poproszono o dostarczenie pisemnych odpowiedzi na szereg pytań sformułowanych przez panel ekspertów. Pytania te dotyczyły mocnych i słabych stron niemieckiego systemu badań oraz pozycji tych instytucji w ramach systemu. Inny zestaw pytań dotyczył strategii oraz środków zapewniania jakości. Po analizie odpowiedzi panel odwiedził DFG i MPG, odbył także wiele rozmów z przedstawicielami innych organizacji badawczych i odwiedził trzy uniwersytety;
- ewaluacja Rady Badań Naukowych Norwegii w 2001 roku. Celem uzyskania solidnej empirycznej podstawy oceny sięgnięto po wiele metod: po analizę bibliometryczną, badania ankietowe oraz wywiady z osobami pracującymi w RBN i z różnymi zainteresowanymi stronami. Zagraniczni naukowcy i administratorzy oceniali także oddzielnie każdy z działów badawczych RBN. Przygotowano szesnaście raportów;
- ewaluacja Austriackiego Funduszu Nauki w 2004 roku. Międzynarodowy zespół oceniający, składający się z 20 osób z instytutów naukowych i firm konsultingowych (Technopolis, Joanneum Research, WIFO, ETH Zurich [KOF] oraz Uniwersytet w Twente), pod kierunkiem Erika Arnolda, ocenił

rolę funduszu w austriackim systemie innowacji i w porównaniach międzynarodowych. Sięgnięto po szeroki zakres metod jakościowych i ilościowych<sup>112</sup>.

- Ewaluacja Australijskiej Rady Badawczej (ARC) polegała na oszacowaniu wpływu finansowania ARC na produktywność australijskiej gospodarki. W ten sposób obliczono stopę zwrotu z inwestycji rządowych. Na podstawie literatury australijskiej i międzynarodowej określono sześć głównych korzyści z badań: 1) tworzenie nowej wiedzy, 2) własność intelektualna, 3) lepsze kompetencje, 4) dostęp do międzynarodowych badań, 5) zdrowie i środowisko, 6) korzyści kulturalne. Na podstawie obliczeń korzyści płynących przez te sześć kanałów obliczono ogólną społeczną stopę zwrotu z inwestycji<sup>113</sup>.

## 2.12. Instytucje naukowe/dyscypliny/zespoły (finansowanie instytucjonalne oparte na dokonaniach)

Zasady finansowania instytucjonalnego badań naukowych różnią się pomiędzy krajami; bywają nimi:

- ogólne dotacje otrzymywane od ministerstwa edukacji lub władz regionalnych czy lokalnych na prowadzenie ogólnej działalności badawczej i dydaktycznej (General University Fund, GUF);
- środki pochodzące z kontraktów na prace B+R oraz dotacji celowych otrzymywanych od rządu i samorządu (środki publiczne finansowania bezpośredniego [*direct government funds*] oraz z innych źródeł zewnętrznych);
- dochody pochodzące z darowizn, akcji i udziałów oraz posiadanego majątku, a także nadwyżka uzyskana z wpływów ze sprzedaży usług niezwiązanych z pracami B+R, np. czesne wpłacane przez poszczególnych studentów, prenumerata czasopism, sprzedaż serum czy płodów rolnych. Są to „środki własne” uczelni<sup>114</sup>.

Udział finansowania instytucjonalnego badań w całości dotacji rządowych także różni się pomiędzy krajami; w Portugalii – jak się zdaje – wszystkie środki idą na finansowanie projektów, a w Niemczech, Izraelu i Nowej Zelandii dominuje finansowanie instytucjonalne. Różnią się także między krajami związki ewaluacji przeprowadzanej przez organa zapewniające dotacje instytucjonalne z poziomem finansowania; wyniki ewaluacji mogą określać wysokość dotacji

<sup>112</sup> ESF, European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies. A report of the ESF Member Organisation Forum on Ex-Post. Evaluation of Funding Schemes and Research Programmes*, September 2009.

<sup>113</sup> OECD, *Assessing The Socio-Economic Impacts Of Public Investment In R&D*, 2007.

<sup>114</sup> *Podręcznik Frascati. Proponowane Procedury Standardowe dla Badań Statystycznych w Zakresie Działalności Badawczo-Rozwojowej*, 2002, OECD/MNiSW 2009, tłum. D. Przepiórkowska; N. Frölich, *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_steeringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_steeringbynumbers.pdf)

(tak jest np. w Wielkiej Brytanii) lub być wyłącznie narzędziem zarządzania i mobilizacji placówek naukowych (tak jest np. w Niderlandach).

Ta pierwsza forma może być realizowana na rozmaity sposób; dwa najczęściej stosowane sposoby to (według J. Salmiego i A.M. Hauptmana):

1. Budżet negocjowany pomiędzy rządem (na poziomie ministerstwa lub agencji) a przedstawicielami uczelni, zazwyczaj na okres 3–4 lat. Negocjacje mogą być oparte przede wszystkim na kryterium „wkładu” (przede wszystkim na liczbie studentów) lub na kryterium „dokonań” (takich jak osiągnięcia naukowe, liczba absolwentów, zakres środków na badania uzyskanych na drodze konkursu). Zdarza się, że punktem wyjścia do negocjacji są sumy przekazywane w latach ubiegłych.
2. Finansowanie według algorytmu (*formula-based funding*) – wzór matematyczny dla obliczenia poziomu finansowania dla poszczególnych instytucji na podstawie zestawu wskaźników. Podobnie jak w przypadku budżetu negocjowanego formuła może opierać się przede wszystkim na kryteriach „wkładu” (np. liczba personelu, liczba studentów, liczba studentów zagranicznych) lub „dokonań” (np. uzyskane stopnie naukowe, liczba absolwentów, liczba publikacji i projektów badawczych, poziom kwalifikacji personelu). Może także uwzględniać priorytety, rządowe lub regionalne, takie jak np. wzrost zatrudnienia<sup>115</sup>.

Jedną z form finansowania według algorytmu są ogólnokrajowe systemy oceny instytucjonalnej badań prowadzonych w szkołach wyższych.

Ewaluacja ta:

- ma charakter *ex post*;
- opiera się zarówno na danych o wkładzie (np. o liczbie pracowników lub wielkości ubiegłorocznej dotacji), jak i na danych o produktach, procesach i oddziaływaniu badań naukowych;
- jej wyniki są wykorzystywane do określenia wielkości finansowania instytucji (np. szkoły wyższej lub jej wydziałów – w skali roku);
- ma wyłącznie (lub także) charakter zewnętrzny (w niektórych krajach ewaluacja zewnętrzna jest łączona z samoewaluacją);
- tworzy ogólnokrajowy system (a nie jest np. stosowana z własnej inicjatywy przez poszczególne instytucje naukowe)<sup>116</sup>.

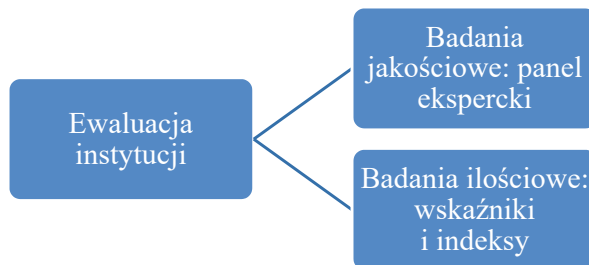
<sup>115</sup> J. Salmi i A.M. Hauptman, *Resource Allocation Mechanisms in Tertiary Education: A Typology and an Assessment*, [w:] *Global University Network for Innovation (ed.) Higher Education in the World 2006. The Financing of Universities*, Palgrave MacMillan, Basingstoke, U.K. 2006, cyt. za: *Funding Systems and their Effects on Higher Education Systems*, OECD EDU/WKP(2007)1; N. Frölich, *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_steeringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_steeringbynumbers.pdf); *Higher Education Funding – Overseas Models*, 2006, <http://www.parliament.nz/NR/rdonlyres/5694EE99-CA9E-499C-9E94-E40F1B114354/40871/0605HigherEducationFundingpdf3.pdf>.

<sup>116</sup> D. Hicks, *Overview of models of performance-based research funding systems*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.

**Metody.** W ewaluacji instytucjonalnej na ogół łączy się metody ilościowe i jakościowe. Dwa najczęściej stosowane łącznie ze sobą podejścia w ewaluacji instytucji to panele ekspertów oraz wskaźniki i indeksy (wskaźniki złożone) (rysunek 2.6).

Rysunek 2.6

### Podejścia w ewaluacji instytucji



Źródło: opracowanie własne.

Badania jakościowe w ewaluacji instytucji naukowych, dyscyplin i programów korzystają na ogół z następujących metod: 1) wywiadów indywidualnych i grupowych z pracownikami, menadżerami, partnerami oraz użytkownikami i klientami, 2) analizy dokumentów pisanych, w tym raportów rocznych, planów strategicznych, dzienników badań i zapisów realizacji programu, informacji o zasobach i finansach, czasami także osobnych opracowań bibliometrycznych, 3) badań ankietowych, 4) bezpośredniej obserwacji, 5) studiów przypadku. Ewaluatorzy przeprowadzają wizytację celem dokonania bezpośredniej obserwacji programu działania oraz czasami osobiście na krótko włączają się do badań w ramach „obserwacji uczestniczącej”. Rozmawiają z pracownikami na temat ich doświadczeń i poglądów. Zapisy tych wywiadów, obserwacji i dokumentów porządkują według głównych tematów, kategorii i przykładów, a następnie przedstawiają w raporcie.

Badania ilościowe – druga z podstaw oceny w ewaluacji instytucjonalnej to wskaźniki. W ramach ewaluacji może je zbierać komisja ekspertów (np. od ocenianych jednostek) albo odrębna placówka statystyczna, np. obserwatorium nauki i techniki. Najogólniej mówiąc, wskaźniki to miary (zazwyczaj ujęte w formie ilościowej) pewnej cechy działalności, które pozwalają na jej ocenę. Jednym z kluczowych zadań budowy systemu ewaluacji jest ich dobór i przetestowanie. Najważniejsze kwestie rozstrzygane podczas planowania ewaluacji to: szersze kategorie (wymiary, bloki), w ramach których grupuje się wskaźniki ilościowe, oraz oceny jakościowe ujęte w skali (np. jakość – wpływ – środowisko); wagi dla poszczególnych wskaźników; stopień i formy różnicowania kryteriów według ocenianych dyscyplin naukowych; skale ocen końcowych („kategorie”) – np. kategoria w skali pięciu stopni.



W ramach ewaluacji instytucjonalnej na podstawie oceny badaczy lub grup badawczych ocenia się instytuty, uczelnie lub ich części organizacyjne (wydziały, fakultety, kolegia). Jednostki oceny (np. poszczególni badacze lub grupy badawcze) nie muszą być tożsame z jednostkami, którym przyznaje się dotacje na podstawie ewaluacji<sup>117</sup>.

Nową i ciekawą propozycją, wysuniętą w ramach przygotowań do nowego systemu oceny w Wielkiej Brytanii – Research Assessment Framework – jest ocena dyscyplin (jednostek dyscyplinowych) w ramach poszczególnych uczelni. W przeciwieństwie do płynnych i krzyżujących się ze sobą grup badawczych oraz w przeciwieństwie do zróżnicowanych pod względem misji i dyscyplin instytutów i wydziałów, dyscypliny wewnątrz instytucji są względnie jednolite i porównywalne. Propozycja ta z jednej strony ułatwia przeprowadzenie ewaluacji, rozwiązując jeden z podstawowych problemów metodologicznych ewaluacji, a mianowicie radzenie sobie z (na ogół) wielkim zróżnicowaniem obszarów badawczych instytucji, z drugiej strony utrudnia korzystanie z ewaluacji jako podstawy określania poziomu finansowania.

Krajowe ewaluacje przeprowadza się przede wszystkim w krajach anglosaskich, Wielkiej Brytanii, Australii i Nowej Zelandii, a także w niektórych krajach skandynawskich (Danii i Norwegii) oraz w Niderlandach (tabela 2.17).

W większości państw krajowy system ewaluacji instytucjonalnej odnosi się wyłącznie do szkół wyższych (rządowe instytuty uczelniane są finansowane na innych zasadach, przez ministerstwa, których misję realizują).

W Wielkiej Brytanii działał najbardziej znany ogólnonarodowy program ewaluacji *The Research Assessment Exercise* (RAE), przeprowadzony w latach 1986, 1989, 1992, 1996, 2001 i 2008, został zastąpiony przez *The Research Excellence Framework* (REF). W Australii istniał Narodowy Program Ewaluacji *Research Quality Framework*, zastąpiony od 2008 roku przez *Excellence in Research for Australia*. Od roku 2006 program oceny instytucjonalnej służącej pobudzaniu badań na światowym poziomie z inicjatywy Ministerstwa Nauki, Rozwoju i Innowacji wprowadziła Dania. Dotąd badania były najczęściej finansowane na podstawie sum z lat poprzednich; stopniowo coraz większa część finansowania ma być przyznawana na podstawie ewaluacji<sup>118</sup>.

Podobnie jak w Wielkiej Brytanii, Australii i Danii, w Niderlandach obowiązuje ogólnokrajowy obligatoryjny system ewaluacji (oparty na Standardowym Protokole Ewaluacji, który obowiązywał w latach 2009–2015)<sup>119</sup>. W przeci-

<sup>117</sup> Ibidem.

<sup>118</sup> *A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process, A Pilot Study Commissioned by OCLC Research. Prepared by Key Perspectives Ltd Truro, UK, 2009.* Ocenę przeprowadza się w następujący sposób: kolegia rektorskie powołują 68 grup eksperckich, składających się z 350 badaczy, którym powierza się zadanie wyboru najważniejszych w ich dyscyplinie czasopism i wydawnictw.

<sup>119</sup> Opracowany wspólnie przez Królewską Niderlandzką Akademię Sztuk i Nauk (KNAW), Stowarzyszenie Uniwersytetów w Niderlandach (VSNU) oraz Niderlandzką Organizację Badań Naukowych (NOW).

wieństwie jednak do Wielkiej Brytanii, Australii i Danii ewaluacja niderlandzka ma charakter jakościowy i składa się z samooceny (wydziałów lub fakultetów) oraz zewnętrznego przeglądu, przy czym tę pierwszą przeprowadza się co trzy, a tę drugą – co sześć lat; samoocenę – raz równolegle, a raz w środku okresu pomiędzy przeglądem zewnętrznym. Ta podwójna ewaluacja nie ma charakteru konkurencji o środki, nie jest też źródłem porównań. Zewnętrzne ewaluacje są przeprowadzane przez niezależne komitety<sup>120</sup>.

Tabela 2.17

## Ewaluacje w wybranych krajach

Kraj	Norwegia	Norwegia	Niderlandy	Wielka Brytania	Australia
<b>Organizator</b>	Norweska Rada Badawcza	Norweskie Stowarzyszenie Instytucji Szkolnictwa Wyższego	Poszczególne uczelnie, w latach 1993–2002 stowarzyszenie uczelni VSNU	Rada Finansowania Szkolnictwa Wyższego (HEFCE)	Ministerstwo ds. Edukacji
<b>Przedmiot ewaluacji</b>	Pola badań	Produkty badań prowadzonych w uczelniach	Uczelnie same określają przedmiot ewaluacji, np. pole badań, wydział	Wydziały uniwersyteckie	Uniwersytety
<b>Główny cel</b>	Jakość badań	Podział środków	Jakość badań	Podział środków	Podział środków
<b>Metoda</b>	<i>Peer review</i> oparte na informacjach	Pomiar publikacji	<i>Peer review</i> i autoewaluacja	<i>Peer review</i>	Wskaźniki
<b>Częstotliwość</b>	Doraźnie	Corocznie	Samoewaluacja oraz ewaluacja zewnętrzna co sześć lat	Reserach Assessment Exercise (86, 89, 92, 96, 2001 i 08)	Corocznie
<b>Rezultaty</b>	Raport, a w jego rezultacie plany i negocjacje	Punkty za publikacje	Raport z oceną w skali 5-punktowej oparty na 4 kryteriach	Ocena w skali zmienianej w kolejnych cyklach	Podział środków finansowych

Źródło: H. Foss Hansen, *Research Evaluation: Methods, Practice and Experience*, Danish Agency for Science, Technology and Innovation, 2009.

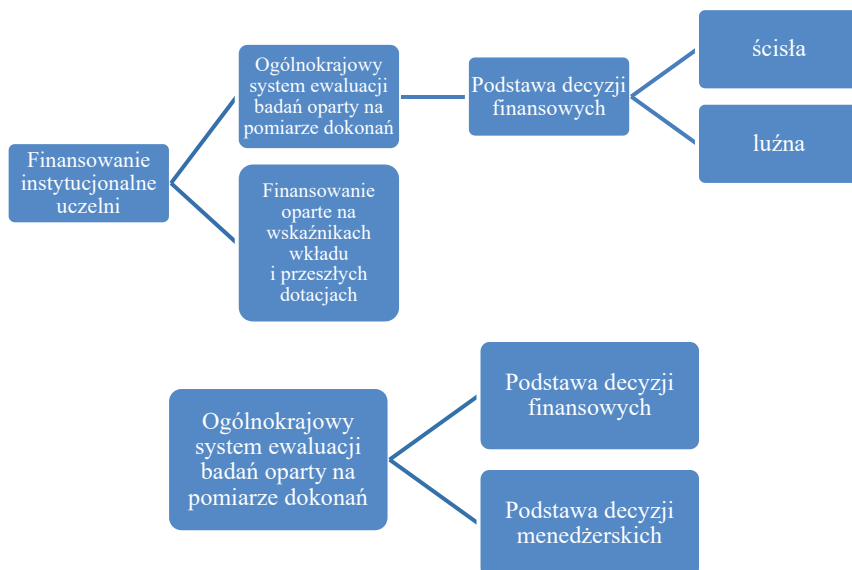
<sup>120</sup> *A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process*, A Pilot Study Commissioned by OCLC Research. Prepared by Key Perspectives Ltd Truro, UK, 2009.

W przeciwieństwie do wymienionych wyżej krajów w Irlandii nie ma jednego, narodowego programu ewaluacji instytucjonalnej szkół wyższych. Poszczególne uniwersytety organizują swoje własne ewaluacje, korzystając z zewnętrznych ekspertów, często zagranicznych. Podczas ewaluacji korzysta się z ustaleń rankingów. Kryteria oceny dotyczą poziomu administrowania, jakości kształcenia (oceniana m.in. na poziomie ocen studentów) oraz wartości badań naukowych<sup>121</sup>.

Opisane różnice można by sumarycznie i w uproszczeniu ująć w następujący sposób (rysunek 2.7).

Rysunek 2.7

### Krajowe systemy ewaluacji instytucjonalnej – najważniejsze różnice



Źródło: opracowanie własne.

Ogólnokrajowe programy ewaluacji instytucjonalnej uczelni, których zadaniem jest określenie poziomu finansowania badań (finansowanie oparte na ocenie dokonań), różnią się pod względem wpływu na wielkość funduszy. W Wielkiej Brytanii wydziały, które otrzymały najniższą kategorię, nie otrzymywały żadnych funduszy, nawet na podstawowe wynagrodzenie. W innych krajach jednak niska punktacja nie wpływa na wynagrodzenia i tylko w umiarkowanym stopniu rzutuje na całość finansowania. W Austrii 20% finansowania blokowego przy-

<sup>121</sup> *A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process*, A Pilot Study Commissioned by OCLC Research. Prepared by Key Perspectives Ltd Truro, UK, 2009. Zwolennicy tego systemu podkreślają, że brak ogólnokrajowego systemu ewaluacji umożliwia badaczom poświęcanie swojego czasu w sposób najbardziej zgodny z ich kompetencjami – dydaktycznymi, badawczymi lub administracyjnymi.

znaje się na podstawie dokonań w nauczaniu i badaniach. W Norwegii 40% ogółu finansowania blokowego przydziela się na podstawie dokonań (15% na podstawie dokonań badawczych, 25 – dydaktycznych)<sup>122</sup>.

Tabela 2.18

## Krajowe systemy finansowania szkół wyższych oparte na wynikach

Kraj	System	Rok wdrożenia/ rok istotnej modyfikacji
Austria		2002
Australia	Indeks złożony, Ramy Badań Jakości (RQF), Jakość Badań w Australii (ERA)	1995/obecnie
Dania	Wdrożenie modelu norweskiego	Obecnie
Finlandia		2010
Flandria	BOF-key	2003/2008
Hiszpania	CNEAI – Krajowa Komisja ds. Oceny	1989
Hong Kong	RAE	1993
Niemcy	Zróznicowane systemy stosowane przez landy	
Norwegia	Norweski model (nowy model finansowania oparty na wynikach badań prowadzonych na uniwersytetach)	2005
Nowa Zelandia	Finansowanie badań oparte na dokonaniach (PBRF)	2003/obecnie
Polska	Metoda parametryczna	1991(wstępna wersja)/ 1998>obecnie
Republika Czeska	Dwa modele finansowania badań	
Słowacja		1992/2002
Szwecja	Nowy model alokacji zasobów	2008
Wielka Brytania	RAE zostały przekształcone w REF – Ramy Osiągnięć Naukowych (Research Excellence Framework)	1986/obecnie
Włochy	Trzyletnia Ocena Badań, Valutazione Triennale della Ricerca (VTR)	Evaluation 2001–2003/ funded 2009

Źródło: D. Hicks, *Overview of Models of Performance-Based Research Funding Systems*, DSTI/STP/RIHR(2010)3; S. Box i E. Basti, *Performance-Based Funding for Public Research in Tertiary Education Institutions: Country Experiences. Summary of country questionnaire responses*, DSTI/STP/RIHR(2010)5.

<sup>122</sup> *Performance-based models for publicly funded research in tertiary education institutions* 2009, DSTI/STP/RIHR(2009)7.

Sarah Box i Ester Basri tak opisują kluczowe cechy finansowania opartego na ocenie dokonań w poszczególnych krajach (tabela 2.19).

Tabela 2.19

**Kluczowe cechy finansowania opartego na ocenie dokonań  
w poszczególnych krajach**

Częstotliwość oceny	Narzędzie finansowania	Wskaźniki	Budżet
<b>Australia</b>			
Coroczne dotacje.  Wskaźniki często uwzględniają wyniki z ostatnich dwóch lat.	Dotacje blokowe na badania szkolnictwa wyższego, w tym: – Joint Research Engagement (JRE) Initiative; – Research Infrastructure Block Grants (RIBG); – Sustainable ReRe Research Excellence in Universities (SRE).	Wskaźniki obejmują: 1. Dochody z badań otrzymanych przez uczelnie (średnio w ciągu dwóch ostatnich lat). 2. Zdolności instytucji do upowszechniania wyników badań w (głównie) recenzowanych publikacjach naukowych (średnio w ciągu dwóch ostatnich lat). Cztery kategorie: książki (ważone przez współczynnik 5), rozdziały, artykuły w czasopismach oraz referaty konferencyjne. 3. Ogół zajęć oferowanych studentom Institutions' Commonwealth Funded Higher Degree by Research (HDR) (ważone przez koszty kursu).  Liczba badaczy aktywnych naukowo (średnio w ciągu dwóch ostatnich lat).  Nowy system Excellence in Research for Australia (ERA) zostanie oparty na następujących wskaźnikach: a) jakość badań: ranking placówek, analiza cytowań, <i>peer review</i> , dochody z opartych na <i>peer review</i> badań australijskich i międzynarodowych; b) zakres działalności badawczej: ogół produktów badań, dochody z badań itd.; c) zastosowania badań: przychody z komercjalizacji badań i inne wskaźniki; d) uznanie: szereg miar prestiżu.	Fundusze są stałe i roczne dotacje oparte na względnych dokonaniach każdej uczelni (stosunku do innych).

Tabela 2.19 – cd.

Częstotliwość oceny	Narzędzie finansowania	Wskaźniki	Budżet
	Finansowanie blokowe oparte na wynikach dla wsparcia szkoleń studentów ubiegających się o stopień doktora i magistra w tym: <ul style="list-style-type: none"> <li>– Research Training Scheme</li> <li>– Commercialisation Training Scheme</li> <li>– Australian Postgraduate Awards</li> <li>– International Postgraduate Research Scholarships</li> </ul>	1. Uzyskane przez studentów stopnie w uczelniach (ważone wg poziomowi i koszt kursu, dane uśrednione w ciągu dwóch ostatnich lat).	
<b>Austria</b>			
Co trzy lata (zgodnie z cyklem trzyletniego systemu budżetowego publicznych wyższych uczelni).	Bezpośrednie finansowanie uczelni ze środków publicznych (GUF)	Główne wskaźniki to poziom finansowania zewnętrznego badań, ważony przez komisję ekspertów.	20% bezpośrednich środków publicznych uczelni wyższych (GUF).
<b>Belgia (wspólnota flamandzka)</b>			
Roczne	Finansowanie uczelni przez: <ul style="list-style-type: none"> <li>– BOF (Specialny Fundusz Badań);</li> <li>– IOF (Fundusz Badań Przemysłowych);</li> <li>– część badawcza transferu operacyjnego (<i>operational remittance</i>)</li> </ul>	Wskaźniki są obliczane dla poszczególnych instytucji jako część całości uzyskanej przez wszystkie instytucje. Obejmują one: <ol style="list-style-type: none"> <li>1. Dyplomy licencjackie i magisterskie.</li> <li>2. Doktoraty.</li> <li>3. Roczne operacje przelewu / liczbę personelu naukowego.</li> <li>4. Publikacje i cytowania.</li> <li>5. Dochód z zamówień przemysłowych.</li> <li>6. Dochody z Europejskiego Programu Ramowego.</li> <li>7. Patenty.</li> </ol>	W 2008 roku 75% funduszy na badania przekazanych przez rząd flamandzki było udzielonych w ramach finansowania opartego na dokonaniach.

Częstotliwość oceny	Narzędzie finansowania	Wskaźniki	Budżet
		8. Firmy-córki. Również stosowany miernik mobilności i różnorodności.	
<b>Czechy</b>			
Coroczne	Finansowanie przeznaczone na realizację konkretnych badań uniwersyteckich oraz wsparcie instytucjonalne; oba uwarunkowane wynikami	Dla specjalnego finansowania badań uniwersyteckich (badania prowadzone przez studentów) (Model 1) 1. Wskaźniki bibliometryczne. 2. Liczby doktorantów w akredytowanych programach studiów. 3. Liczba stopni naukowych magistra przyznawanych w poprzednim roku akademickim. 4. Liczba stopni doktora przyznanych w poprzednim roku akademickim.  Dla finansowania instytucjonalnego (Model 2): 1. Wskaźniki bibliometryczne.	W 2010 roku: specjalne badania uniwersyteckie – ok. 4% wszystkich wydatków rządowych na B+R. Jedna trzecia ogółu rządowych wydatków na B+R alokowana na podstawie oceny dokonań, w tym na uczelnie i organizacje pozauczelniane.
<b>Dania</b>			
Coroczne	Fundusz restrukturyzacyjny (fundusz uzupełniający Ogólny Fundusz Uniwersytecki, GUF)	Wskaźniki obejmują: 1. Wskaźnik bibliometryczny. 2. Zewnętrzne finansowanie uzyskane przez szkoły wyższe. 3. Liczbę prac doktorskich.	W 2010 r. 10% funduszu restrukturyzacji zostało przydzielone zgodnie z wynikami wskaźnika bibliometrycznego. Instytucje otrzymały równowartość ich udziału w puli publikacji naukowych.
<b>Finlandia</b>			
Coroczne	Finansowanie statutowe	Wskaźniki obejmują: 1. Nauczanie i badania liczone w osobolatach. 2. Liczba stopni doktorskich określonych w porozumieniu między ministerstwem i instytucjami. 3. Liczba uzyskanych stopni doktorskich. 4. Finansowania badań w zawodach na szczeblu krajowym (fundusze Akademii Finlandii, decyzje w sprawie finansowania centrów doskonałości, finansowanie TEKES).	Okolo 34% jest finansowane na podstawie oceny dokonań.

Tabela 2.19 – cd.

Częstotliwość oceny	Narzędzie finansowania	Wskaźniki	Budżet
		5. Publikacje naukowe (w recenzowanych czasopismach międzynarodowych oraz inne publikacje naukowe). 6. Umiejdzynarodowienie badań (udział międzynarodowego konkurencyjnego finansowania badań oraz zakres mobilności naukowców i nauczycieli).	
<b>Niemcy</b>			
Raz na okres budżetowy (albo każdego roku, albo w niektórych landach co dwa lata)	Alokacja funduszy instytucjonalnych na szczeblu landów	Wskaźniki różnią się w zależności od landu i rodzaju instytucji. Najczęściej używane wskaźniki to finansowania zewnętrzne oraz liczba zakończonych doktoratów. Niektóre landy używają wskaźnika dotyczącego współpracy, podczas gdy Bawaria i Berlin – liczby publikacji.	W zależności od landu, Berlin – 30% finansowania opartego na ocenie dokonań, Nadrenia Północna-Westfalia – 11%, Badenia-Wirtembergia i Hesja – 10%; Brandenburgia – 8%. Większość landów ogranicza możliwe straty uczelni przez uwzględnienie budżetu z poprzedniego roku.
<b>Norwegia</b>			
Roczne (na podstawie wyników z poprzednich lat)	Finansowanie oparte na ewaluacji dokonań (PBRF)	Cztery wskaźniki: 1. Punkty za publikacje. 2. Środki Programu Ramowego UE. 3. Środki Rady ds. Badań Naukowych Norwegii. 4. Liczba przyznanych stopni doktorskich.	Każdego roku 30% wszystkich środków dla uczelni przydziała się na podstawie oceny wyników edukacji i badań naukowych. W 2010 roku ok. 6% wyniosły środki na B+R.
<b>Nowa Zelandia</b>			
Coroczny przydział funduszy	Finansowanie oparte na ewaluacji dokonań (PBRF)	Trzy wskaźniki używane do określania poziomu finansowania: 1. Wyniki okresowe oceny jakości (procesu <i>peer review</i> ). 2. Liczba uzyskanych stopni badawczych (roczne). 3. Zewnętrzne dochody z badań (roczne).	Instytucje otrzymują przydziały na podstawie zmian w uzyskiwaniu stopni badawczych oraz zewnętrznych dochodów z badań.



Częstotliwość oceny	Narzędzie finansowania	Wskaźniki	Budżet
			W 2008 roku finansowanie oparte na wynikach wyniosło 11,8% publicznych funduszy uczelni.
<b>Polska</b>			
Ocena co 5 lat	Finansowanie instytucjonalne	Wskaźniki obejmują: 1. Publikacje recenzowane. 2. Monografie. 3. Międzynarodowe projekty badawcze. 4. Uprawnienia do udzielania stopni naukowych. 5. Nowe technologie, materiały, produkty, systemy, usługi, metody i oprogramowania. 6. Opatentowane wynalazki, prawa na wzory użytkowe, prawa autorskie.	W roku 2008 60% rządowych nakładów na B+R pochłonęło finansowanie instytucjonalne, częściowo oparte na ocenie dokonań.
<b>Szwecja</b>			
Coroczne	Finansowanie bezpośrednie	Wskaźniki: finansowanie zewnętrzne (50%) i wskaźniki bibliometryczne (50%).	Wszystkie nowe bezpośrednie finansowania od 2009 roku oraz 10% wszystkich instytucji bezpośredniego finansowania z poprzedniego roku jest rozdzielanych na podstawie dokonań.
<b>Wielka Brytania</b>			
RAE przeprowadza się co 5–7 lat, roczne wskaźniki aktywności	Finansowanie blokowe niekapitałowe (nieprzeznaczone na infrastrukturę)	<i>Peer review</i> w ramach Assessment Exercise (RAE) wsparte wskaźnikami, takimi jak: 1. Przychody z fundacji. 2. Ilość badań naukowych dla (przychody). 3. Rozmiary studiów podyplomowych.	RAE oraz inne systemy finansowania opartego na dokonaniach (finansowanie projektów) mają udział w ok. jednej trzeciej dotacji publicznej.

Źródło: S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

**Rola zleceniodawcy.** Kluczowe czynniki ewaluacji instytucjonalnej – jej zakres i przedmiot, czas i zasoby, uprawnienia, plan wykorzystania raportu, wybór ekspertów – albo ustala zleceniodawca (np. ministerstwo lub, jeszcze częściej, agencja finansująca badania), albo też pozostawia swobodę decyzji

realizatorom. Zleceniodawca (a dokładniej organ wewnętrzny powołany do przeprowadzenia ewaluacji) powołuje przewodniczących oraz sekretarzy paneli i określa główne cele ewaluacji (np. ulepszenie jakości i relewantności badań), procedury wizyt kontrolnych, oczekiwania co do raportu końcowego, zakres odpowiedzialności i uprawnienia paneli. Czasami ustala szczegółowe pytania i kryteria oceny (*terms of reference*), a czasami tylko ogólne wytyczne (np. że ewaluacja powinna brać pod uwagę jakość badań oraz ich ocenę dokonaną przez klientów i użytkowników), pozostawiając ich konkretyzację członkom zespołu. To drugie rozwiązanie ma wadę, gdyż pozostawia pole do dowolności, ale też ważną zaletę, gdyż umożliwia uwzględnienie w kryteriach oceny dyscyplinowych specyfik ocenianych jednostek, wliczając w nie prowadzenie badań inter-, multi- i transdyscyplinarnych.

Zazwyczaj zleceniodawca przygotowuje i ogłasza z góry plan wszystkich ewaluacji, które zamierza przeprowadzić w nadchodzącym cyklu. Zawiera on m.in. informacje o planowanych wizytacjach, dokumentacji, jaką powinni otrzymać ewaluowani, publikacji wyników oceny oraz sposobach kontynuacji (*follow-up*) przedsięwzięcia.

Poszczególne kraje różnią się między sobą pod względem częstotliwości przeprowadzania ewaluacji instytucjonalnej. Z zasady na im większą skalę jest zakrojona ewaluacja, im jest bardziej wnikliwa, czaso- i kosztochłonna, tym rzadziej się odbywa. W Wielkiej Brytanii przeprowadzano ją (w systemie RAE) co pięć, siedem lat, w Hiszpanii i w Holandii odbywa się ona co sześć lat, w Austrii i Słowacji co trzy lata, w Polsce nie rzadziej niż raz na pięć lat<sup>123</sup>. Z reguły jednak w krajach, w których pełne szczegółowe ewaluacje przeprowadza się co kilka lat, oceniane instytucje przesyłają co roku informacje, uwzględniane przy ustalaniu wysokości rocznej dotacji; jest tak np. w Wielkiej Brytanii i w Polsce.

Produktem ewaluacji są zarówno raporty komisji ewaluacyjnych, które dokonały oceny poszczególnych jednostek, jak i raport końcowy w danym cyklu ewaluacji.

Nigdy w szkołach wyższych finansowanie badań oparte na dokonaniach nie jest jedynym źródłem ich finansowania; udział tego rodzaju funduszy waha się od 2% (Norwegia) do ok. 25% (Wielka Brytania) ogółu publicznego finansowania. Znaczenia tego rodzaju systemu nie da się jednak mierzyć wyłącznie odsetkiem ogólnego budżetu, gdyż sukces w staraniach o dotację jest też np. źródłem prestiżu, procentującego w staraniach o granty i kontrakty<sup>124</sup>.

W USA nie ma krajowego systemu oceny badań naukowych, w ramach którego regularnie poddaje się ocenie badania *ex post*. Większość alokacji zasobów w Stanach Zjednoczonych odbywa się poprzez ocenę projektów i jednostek *ex ante*, a nie poprzez ocenę instytucji *ex post*. Nacisk położony jest na jakość

---

<sup>123</sup> D. Hicks, *Overview of Models of Performance-Based Research Funding Systems*, DSTI/STP/RIHR(2010)3.

<sup>124</sup> Ibidem.

akademicką, a w przypadku badań ukierunkowanych na zadania na przewidywane korzyści i dopasowanie projektów do misji instytucji<sup>125</sup>.

Różne kraje stosują rozmaite podejście co do decyzji, czy łączyć ocenę badań z oceną dydaktyki, oraz – skoro występuje trend odchodzenia od wyłącznie wewnątrz naukowych kryteriów ocen – czy i w jakim zakresie w poszerzonych kryteriach powinno znaleźć się miejsce dla dydaktyki (np. tylko kształcenia doktorantów czy szerszym). Niektóre kraje z zasady stosują finansowanie oparte na ocenie dokonań do finansowania dydaktyki, np. w Austrii system obejmuje zarówno elementy nauczania, jak i badań, w Finlandii finansowanie statutowe szkół wyższych zawiera komponent oparty na ocenie wyników nauczania (zarówno pod względem zakresu, jak i jakości i skuteczności). Jednak niektóre kraje, jak Belgia oraz (z pewnymi wyjątkami) Polska, finansują dydaktykę na podstawie miar instytucjonalnych, a nie ocen dokonań<sup>126</sup>. Podobnie w Anglii „finansowanie badań i finansowanie dydaktyki płynie oddzielnymi strumieniami. Finansowanie nauczania oblicza się przede wszystkim na podstawie liczby studentów kończących komponenty programów akademickich. Ponadto, niewielki procent całkowitej kwoty środków przeznaczają się na wynagrodzenie za realizację konkretnych rządowych inicjatyw”<sup>127</sup>.

Dylemat wyboru co do rodzaju powiązań pomiędzy wskaźnikami a decyzją co do poziomu finansowania instytucji jest różnie rozstrzygany w różnych krajach. W Wielkiej Brytanii (*Research Assessment Exercise*) kategoria, jaką w ewaluacji otrzymała uczelnia, w sposób jednoznaczny i automatyczny przesądza o poziomie jej finansowania. Częściej zachodzi sytuacja, w której powiązanie jest luźniejsze. Ocena, nieraz ujęta w kategorię, jest podstawą do negocjacji pomiędzy ministerstwem lub agencją a finansowaną przez nie placówką<sup>128</sup>. W Australii „*Australian Composite Index* służy do przydzielania dotacji blokowych; wydaje się, że szkoły wyższe są autonomiczne co do sposobu ich wykorzystania. Jednak dotacja jest podzielona na siedem części, z których każda jest opatrzona odrębnym zestawem szczegółowych wytycznych agencji finansowych, które w ten sposób zachowują duży stopień kontroli nad wydatkami. Podkreśla się zatem, że jedną ręką przyznaje się autonomię, a drugą – odbiera ją”<sup>129</sup>.

Podobnie stosuje się rozmaite rozwiązania dylematu autonomii i kontroli dotacji finansowania badań opartego na ocenie dokonań. Skala kontroli rządu nad sposobem wydatkowania rozciąga się od pozostawienia decyzji całkowicie w rękach instytucji naukowych po żądanie przestrzegania zobowiązań związanych

<sup>125</sup> OECD IPP, *Issue Brief: Research Organisation Evaluation*, 2014.

<sup>126</sup> S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

<sup>127</sup> N. Frølich, *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_steeringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_steeringbynumbers.pdf).

<sup>128</sup> *Measuring Government Activity*, OECD 2009, s. 12–13.

<sup>129</sup> D. Hicks, *Overview of Models of Performance-Based Research Funding Systems*, DSTI/STP/RIHR(2010)3.

z przekazywanymi środkami. W Australii rząd oczekuje, że uczelnie rozdziela środki przez określone przez siebie mechanizmy podziału. „W Niemczech, Danii, Norwegii i w Nowej Zelandii instytucje mogą swobodnie decydować, czy i komu przekazać otrzymane od rządu środki. W Austrii niektóre instytucje mogą (na różnych poziomach) dokonywać wewnętrznej alokacji funduszy według ustalonych wskaźników. W Finlandii podział funduszy uzyskanych na podstawie ewaluacji dokonań pozostaje w gestii uczelni, choć musi być dokonany zgodnie z celami uzgodnionymi z Ministerstwem Edukacji. W Szwecji oczekuje się, że instytucje będą dokonywać wewnętrznego podziału funduszy w taki sposób, aby stwarzać zachęty do publikowania dobrych prac, a jednocześnie do szukania zewnętrznych źródeł finansowania. W Wielkiej Brytanii uczelnie są autonomiczne w podejmowaniu decyzji w zakresie stosowania środków, ale oczekuje się, że będą premiować jakość badań”<sup>130</sup>.

**Oceny wartości ewaluacji instytucjonalnej.** W ocenach wartości ewaluacji instytucjonalnej podkreśla się, po pierwsze, że nie ma wprawdzie systemu idealnego, ale w sumie przyniosła ona więcej dobrych niż złych skutków. W brytyjskim RAE pozytywne skutki wynikały z dobrego zarządzania samych instytucji oraz z wkładu poszczególnych menadżerów i naukowców<sup>131</sup>.

Po drugie, zwraca się uwagę, że ocena ewaluacji instytucjonalnej nie jest prosta, gdyż tak jak w przypadku każdego badania wpływu pozostaje problem oddzielenia wpływu ewaluacji od działania innych czynników<sup>132</sup>.

Po trzecie, podkreśla się, że wpływ ewaluacji instytucjonalnej jest większy od przyznanych na jej podstawie kwot. Jest tak, gdy np. prawdopodobieństwo sukcesu wniosku o przyznanie dotacji na projekt badawczy wzrasta, gdy wnioskodawca pracuje na wyżej ocenionym wydziale<sup>133</sup>.

Po czwarte, uwypukla się fakt, że pojęcie efektywności badań naukowych, rozumiane zarówno jako efektywność naukowa (jak najwyższa jakość w stosunku do użytych zasobów), jak i efektywność społeczno-gospodarcza (jak najsilniejszy dodatni wpływ na otoczenie zewnętrzne nauki), jest nie tylko wielowymiarowe, ale także niejednoznaczne. Badania nie zawsze są oceniane dodatnio. Nowa wiedza może być krytyczna, może prowokować, a nawet zostać uznana za szkodliwą<sup>134</sup>.

---

<sup>130</sup> S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

<sup>131</sup> *The impact of the 1992 Research Assessment Exercise on higher education institutions in England*, [http://www.hefce.ac.uk/pubs/hefce/1997/m6\\_97.htm](http://www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm).

<sup>132</sup> *Performance-based models for publicly funded research in tertiary education institutions* 2009, DSTI/STP/RIHR(2009)7.

<sup>133</sup> D. Hicks, *Overview of models of performance-based research funding systems*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD Paryż 2010.

<sup>134</sup> H. Foss Hansen, *Performance indicators used in performance-based research funding*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD Paryż 2010

Po piąte, zwraca się uwagę, że podobnie jak w powiązaniach nauka-gospodarka (badań naukowych w sektorze publicznym oraz biznesu) dużym problemem jest pogodzenie sprzecznych oczekiwań naukowców i biznesmenów, tak i w ewaluacji trudną kwestią jest zbliżenie stanowisk polityków i ewaluatorów.

**Systemy ewaluacji instytucjonalnej: uboczne konsekwencje.** Badania następstw ewaluacji instytucjonalnej przyniosły wiele istotnych ustaleń. Należą do nich obserwacje:

1. Tendencji spadku wartości wskaźników oraz ogólniej „spadającego zwrotu z ewaluacji” w miarę upływu czasu. Spadek wartości wynika m.in. z takich przyczyn, jak:
  - zjawisko „ewaluacyjnej fetygi” (zmęczenie badaczy zbyt często powtarzającymi ewaluacjami, zwłaszcza jeśli nie mają one istotnego znaczenia jako źródła nauki lub decyzji finansowych);
  - przewrotne uczenie się (wybitne wyniki są efektem manipulacji danymi; np. gdy badacze szybko uczą się, jak obracać ewaluację na własną korzyść i ulepszają nie tyle badania naukowe, ile sposób zapewniania swoim instytucjom cech premiowanych w ewaluacji, np. poprzez publikowanie prac wyłącznie w czasopiśmie o wysokim IF, pomimo racji przemawiających za innym rozwiązaniem, dzielenia publikacji na „najmniejsze jednostki publikowalne”, zatrudnianie uczonych legitymujących się wielką liczbą cytowań w przeddzień zgłoszenia wykazu do komisji ewaluacyjnej<sup>135</sup>; gdy uwzględnia się tylko konferencje międzynarodowe, konferencjom krajowym nadaje się nazwę międzynarodowych i zaprasza kilku gości z zagranicy itd.)<sup>136</sup>.
2. Ubocznych ujemnych skutków nastawienia na realizację premiowanych celów, polegających na:
  - wzmacnianiu najlepszych ponad należne zasługi, uniemożliwianiu poprawy słabszych instytucji poprzez wstrzymywanie ich dotacji, zaostrzaniu nie-

<sup>135</sup> S. Van Thiel i F.L. Leeuw, *The Performance Paradox in the Public Sector*, „Public Performance & Management Review” 2002, vol. 25, no. 3, 267–281, cyt. za: L. Butler, *Impacts of Performance-Based Funding Systems: A Review of the concerns and the evidence*, DSTI/STP/RIHR(2010)6.

<sup>136</sup> W literaturze przedmiotu znane jest tzw. prawo Michaela Gibbonsa, które głosi, że nie ma takiego wskaźnika dokonań, którego by środowisko fizyków nie potrafiło obrócić na swoją korzyść przed rozpoczęciem kolejnego cyklu finansowania. Q. Schiermeire i J. Ockenden, *Perspectives of science in Central and Eastern Europe: emerging directions*, OS Press, 2001, s. 224. „Środowisko naukowe jest bardzo inteligentne i potrafi szybko reagować na bodźce i wskaźniki, jeśli są one związane z zasobami”, E. Basri i M. Cervantes, *Evaluation of Publicly Funded Research: Recent Trends And Perspectives*, DSTI/STP(2006)24. W szczególności podatny na manipulację jest pomiar cytowań uznawany za miarę wpływu naukowego. „Autorzy najprawdopodobniej cytują autorów, których znają; biorąc pod uwagę tendencję do cytowania kolegów i/lub publikacji w języku angielskim, czynnik reputacji oznacza, że pewni autorzy są częściej cytowani od innych. Może tak być ze względu na znaczenie ich prac, ale także ze względu na nieformalne sieci”. E. Hazelkorn, *Pros and cons of research assessment, lessons from rankings*, 2009, <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1000&context=dirrerep>

- zdrowej rywalizacji, podcinaniu motywacji; podkreśla się, że z założenia ocena dokonań premiuje doskonałość kosztem równości<sup>137</sup>;
- zaniedbywaniu innych ważnych celów, nieuwzględnionych wśród kryteriów ewaluacji (np. nacisk na liczbę publikacji powoduje podejmowanie łatwych do realizacji badań oraz rosnące rozdrabnianie publikacji kosztem podejmowania ważnych, długofalowych i ryzykownych badań oraz publikowania istotnych ustaleń; akcentowanie pomiaru badań prowadzi do lekceważenia prac rozwojowych, mających z reguły bardziej bezpośrednie znaczenie dla rozwoju gospodarki i społeczeństw; nacisk na liczbę cytowań skutkuje zapraszaniem do współpracy wyłącznie badaczy i instytucji legitymujących się już osiągniętą wysoką pozycją naukową kosztem włączania do badań młodych badaczy, zdobywających dopiero pozycję instytucji lub krajów spoza pierwszej ligi nauki; nacisk na badania pociąga za sobą zaniedbywanie dydaktyki; nacisk na publikacje w języku angielskim pociąga za sobą zaniedbywanie prac w języku ojczystym, upadek rodzimej terminologii oraz – niejednokrotnie – osłabienie pozabadawczego wpływu badań itd.);
  - takiej reakcji ocenianych na wskaźniki i inne metody oceny, która jest niezgodna z celami i intencjami ewaluacji (wskaźniki z natury mają ograniczenia i mogą prowadzić do nadmiernego nacisku na pewne działania kosztem innych. Na przykład system, który nagradza wyniki badań naukowych, a nie powiązania z przemysłem, może te powiązania osłabiać, w szczególności gdy zachęty działają aż do poziomu indywidualnego rozwoju zawodowego badacza<sup>138</sup>);
  - wykorzystywania wyników ewaluacji poza systemem ewaluacji (np. jako informacji, gdzie studiować)<sup>139</sup>.

### 3. Problemów stwarzanych przez wskaźniki ilościowe:

- mierzą tylko rzeczy mierzalne, a te niekoniecznie są najistotniejsze; jakość badań oraz ich efekty zależą od wielu czynników niezależnych od badaczy; pomiar efektów powinien być oparty na głębokim zrozumieniu zależności pomiędzy wkładem a efektami, o co trudno ze względu na wpływ innych czynników; wskaźniki ilościowe powinny być uzupełnione przez metody jakościowe, takie jak *peer review* czy studia przypadku;
- działania tzw. prawa Goodharta, które mówi, że gdy tylko wprowadza się pewien wskaźnik (jako pomoc np. w ewaluacji, zarządzaniu czy prowadzeniu polityki), wskaźnik ten szybko traci funkcję informacyjną, którą miał pełnić, gdyż zmienia zachowania ludzi, jakie miał obiektywnie mierzyć

<sup>137</sup> D. Hicks, *Overview of models of performance-based research funding systems*, w: *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010; E. Kalpazidou Schmidt, *Evaluation and science policy*, [w:] *The use of evaluations in Europe. Report from the European RTD Evaluation Network meeting*, red. K. Siune i E. Kalpazidou Schmidt, Report 2003/2, Aarhus, Danish Institute for Studies in Research and Research Policy, 2003.

<sup>138</sup> E. Basri i M. Cervantes, *Evaluation Of Publicly Funded Research: Recent Trends And Perspectives*, DSTI/STP(2006)24.

<sup>139</sup> Ibidem.

(np. wprowadzenie cytowań jako miary osiągnięć zmieniło praktyki cytowania, powodując wzrost samocytowań, wzajemnych cytowań lub nacisku ze strony redakcji czasopism na cytowanie tego właśnie czasopisma)<sup>140</sup>.

Z badań nad ubocznymi skutkami ewaluacji opartej na ocenie dokonań wyciąga się też wiele wskazań, m.in. dotyczących konieczności traktowania ewaluacji jako procesu uczenia się, zarówno dla ocenianych, jak i dla oceniających, jako okazji zdobycia użytecznych, skądinąd nieosiągalnych informacji, tak aby np. ewaluacja szkół wyższych nie zabijała kreatywności (np. wydział uniwersytecki powinien być oceniany jako ekosystem, habitat, środowisko, według jakości życia, szans zdobywania doświadczenia, troski o wzrost, a nie wyłącznie jako maszyna do wytwarzania oczekiwanych wyników)<sup>141</sup>.

**Wpływ systemu ewaluacji instytucjonalnej na cele i obszary.** Wpływ ewaluacji instytucjonalnej różni się w zależności od systemu. Niemniej, badania i analizy wskazały na wiele podobieństw co do charakteru oddziaływania. Spostrzeżenia te można uporządkować w następujący sposób – według obszarów rozpatrywanego wpływu.

**Uczelnie.** Całościową ocenę ewaluacji instytucjonalnej przeprowadził Marcel Herbst<sup>142</sup>. W Europie dominuje trend udzielania uczelniom przez rządy większej autonomii, ale pod warunkiem wzrostu ich uzależnienia poprzez wskaźniki dokonań, pisał Herbst. Można mieć wobec niego wiele zastrzeżeń. System ten zyskał uznanie jako dobre ogólne rozwiązanie, ale bardziej na wiarę niż wskutek głębszych analiz oraz szukania alternatyw. „Zamiast promować zdrową konkurencję pomiędzy uczelniami, próbuje się sztucznie tworzyć *quasi*-rynek w celu pchnięcia instytucji bliżej pozycji uznanych przez władze za pożądane”<sup>143</sup>. Udzielanie uniwersytetom większej autonomii pod warunkiem wiązania ich wskaźnikami dokonań pęta ich zdolność do samodzielnego zdobycia umiejętności efektywnego zarządzania, co z kolei staje się argumentem za utrzymaniem kontroli poprzez wskaźniki. System finansowania oparty na ocenie dokonań wprowadza się jako instytucję niby-rynku mającą na celu wyrównanie domniemanych ułomności szkół wyższych. Dobre uczelnie powinny otrzymywać więcej od gorszych, aby stały się bardziej konkurencyjne i pobudzały do działania te gorsze. Powinno się nagradzać efekt, a nie wkład. Otrzymując więcej pieniędzy, dobre uniwersytety ulepszają jednak badania i dydaktykę, wzmacniając swoją pozycję i zwiększając prestiż. Dzięki temu otrzymują więcej pieniędzy. Zatem w końcu i tak to nie

<sup>140</sup> [http://en.wikipedia.org/wiki/Goodhart's\\_law](http://en.wikipedia.org/wiki/Goodhart's_law).

<sup>141</sup> *Strengthening the Impact of Public Research Institutions: Evaluation of PRIs*, OECD DSTI/STP/RIHR(2009)6; P. Jedamus, M.W. Peterson i Associates, *Improving Academic Management. A Handbook of Planning and Institutional Research*, Jossey-Bess Publishers, 1989, s. 452–455; S. Lahlou, R. van der Meijden, M. Messu, G. Poquet i F. Prakke, *A Guideline for Survey-Techniques in Evaluation of Research*, 1992.

<sup>142</sup> M. Herbst, *Financing public universities: the case of performance funding*, Dordrecht, Springer, 2007.

<sup>143</sup> Ibidem.

efekt, ale wkład decyduje o ich osiągnięciach. Nie jest dowiedzione, że finansowanie oparte na dokonaniach pobudza dokonania bardziej niż finansowanie jako takie. Bardziej cementuje ono *status quo* niż prowadzi do zmiany. Domniemane ułomności uczelni są do pewnego stopnia wynikiem leku, który ma je uleczyć. System finansowania oparty na ocenie dokonań jest najskuteczniejszy wówczas, gdy w uczelniach najslabiej rozwinięte są wewnętrzne mechanizmy podziału funduszy. A z kolei te wewnętrzne mechanizmy alokacji odgrywają kluczową rolę w zarządzaniu oraz są konieczne dla uzyskania funduszy na nowe rozwojowe przedsięwzięcia. Systemy alokacji wspierające istniejące struktury utrudniają adaptację do złożonego i szybko zmieniającego się świata. System ten prowadzi do uniformizmu wprowadzanego w imię zapewnienia jakości, ale jakość zależy w większym stopniu od wewnętrznej woli jej osiągnięcia niż systemu, który ją wymusza, a od uniformizmu lepsze jest zróżnicowanie misji i zadań badawczych uniwersytetów w zależności od oczekiwań studentów oraz innych uwarunkowań zewnętrznych. Warunkiem podniesienia poziomu szkół wyższych jest udana współpraca między uczelniami zainicjowana i rozwijana przez nie same, np. współpraca w uruchomieniu wspólnego programu nauczania lub centrum badawczego. Narzucone odgórnie wskaźniki niekoniecznie jej sprzyjają. Najszybszy postęp badawczy następuje na obrzeżach ustabilizowanych dyscyplin. Nacisk na finansowanie dyscyplinowe może odwieść od bardziej ryzykownych, lecz obiecujących badań<sup>144</sup>.

**Jakość badań.** Sądzi się, choć nie jest to jednak dostatecznie udowodnione, że ewaluacja instytucjonalna prowadzi do poprawy jakości górnego poziomu średniego oraz redukuje badania o niskiej jakości, ale do pewnego stopnia eliminuje doskonałość w nauce<sup>145</sup>. Gdy celem ewaluacji jest przede wszystkim poprawa jakości badań, a także koncentracja zasobów u najlepszych, nie jest ona w stanie poprawić słabszych instytucji naukowych, a nawet może doprowadzić do upadku prowadzonych w nich badań<sup>146</sup>.

**Środowisko naukowe.** Ewaluacja instytucjonalna może mieć szkodliwy wpływ na tętniące życiem społeczności naukowe. Wzmacnia ona z reguły znaczenie konserwatywnej hierarchii naukowej, co hamuje powstawanie oryginalnych badań i nowych dziedzin oraz tłumi różnorodność i pluralizm. Złagodzić problem może istnienie różnych instytucji finansujących badania, ale nie zawsze, gdyż konserwatywna elita może kontrolować różne agencje<sup>147</sup>. Do nowatorstwa w nauce

---

<sup>144</sup> Ibidem.

<sup>145</sup> J. Gläser, G. Laudel, S. Hinze i L. Butler, *Impact of evaluation-based funding on the production of scientific knowledge* 2002, <http://www.laudel.info/pdf/02ExpertiseGlaeLauHinBut.pdf>.

<sup>146</sup> D. Hicks, *Overview of models of performance-based research funding systems*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.

<sup>147</sup> R. Whitley, *Chapter 1 Changing Governance of the Public Sciences*, [w:] *The Changing Governance of the Sciences, Sociology of the Sciences Yearbook*, Dordrecht 2008, Springer, cyt. za:



zniechęcają wskaźniki<sup>148</sup>. Ewaluacja instytucjonalna skutkuje unikaniem projektów o wysokim ryzyku oraz koncentracją na badaniach, w których można się spodziewać sukcesu w spełnieniu kryteriów finansowania<sup>149</sup>. Ponadto, sprzyja ona homogenizacji badań i eksperymentów poprzez nagradzanie bezpiecznych tematów i metod badań<sup>150</sup>. Podkreśla się także, że „system, w którym kluczowym kryterium są publikacje, promuje «inflację publikacji» i zachowanie status quo”<sup>151</sup>.

Jochen Gläser, Grit Laudel, Sybille Hinze, Linda Butler podnoszą pytanie, czy postęp badań zależy od „niezbędnej różnorodności”, tj. od minimalnego zakresu zróżnicowania zagadnień i perspektyw. Jeśli tak jest, utrzymywanie różnorodności może być ważniejsze niż promowanie wysokiej jakości. Koncentracja ewaluacji na kryteriach wysokiej jakości badań może doprowadzić do powstania luk w wiedzy naukowej, które nagle mogą się stać przeszkodą na drodze do dalszego rozwoju wiedzy. Ponieważ wydaje się, że na całym świecie zapanowała tendencja do wprowadzania ewaluacji instytucjonalnej, można się spodziewać ujednoczenia praktyk finansowania w skali międzynarodowej. Powstaje więc kolejne pytanie – jakie to pociąga za sobą skutki dla rozwoju nauki? Czy czasem nie zagraża pewnym ważnym niszą w produkcji wiedzy?<sup>152</sup>

**Nowatorstwo i interdyscyplinarność.** Sądzi się, że ewaluacja instytucjonalna faworyzuje badania „głównego nurtu” oraz bezpieczne tematy, a ujemnie wpływa na tzw. „badania błękitnego nieba”, czyli nowatorskie<sup>153</sup>. Ponadto, bardziej premiuje przeszłe osiągnięcia ustabilizowanych grup badawczych oraz ustabilizowanych dyscyplin, badania krótkoterminowe kosztem długofalowych, podejścia konwencjonalne kosztem nowatorskich, co bywa przeszkodą w kształtowaniu się nowych i obiecujących grup i pól badań<sup>154</sup>. Zwraca się też uwagę, że system zwiększa niechęć do podejmowania ryzyka w wyborze obszarów badawczych (np. wtedy, gdy wybór może wpłynąć na oczekiwany spadek liczby cytowań)

D. Hicks, *Overview of models of performance-based research funding systems*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD Paryż 2010.

<sup>148</sup> J. Taylor i R. Taylor, *Performance indicators in academia: An X-efficiency approach?*, „Australian Journal of Public Administration”, 2003, 62(2), cyt. za: N. Frølich, *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_steeringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_steeringbynumbers.pdf)

<sup>149</sup> I. Liefner, *Funding, resource allocation, and performance in higher education Systems*, „Higher Education” 2003, 46(4), s. 469–489, cyt. za: N. Frølich, *The politics of steering by numbers*, op. cit.

<sup>150</sup> A. Geuna i B.R. Martin, *University research evaluation and funding: An international comparison* „Minerva” 2003, 41(4), s. 277–304; N. Frølich, *The politics of steering by numbers*, op. cit.

<sup>151</sup> N. Frølich, *The politics of steering by numbers*, op. cit.

<sup>152</sup> J. Gläser, G. Laudel, S. Hinze i L. Butler, *Impact of evaluation-based funding on the production of scientific knowledge* 2002, <http://www.laudel.info/pdf/02ExpertiseGlaeLauHinBut.pdf>.

<sup>153</sup> L. Butler, *Impacts of performance-based research funding systems: A Review of the concerns and the evidence*, w: *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.

<sup>154</sup> *Performance-based models for publicly funded research in tertiary education institutions*, OECD Paryż 2010.

oraz zmniejsza zakres wielodyscyplinarności<sup>155</sup>. Ta ostatnia kwestia jest przedmiotem sporu, gdyż niektóre badania kwestionują ujemny wpływ systemu na badania interdyscyplinarne.

**Poziom czasopism naukowych.** Uważa się, że ewaluacja instytucjonalna wywiera dodatni wpływ na poziom czasopism, skłaniając wiele redakcji do spełniania warunków umożliwiających starania o wciągnięcie na listę Web of Science lub Scopusa<sup>156</sup>.

**Kadry.** Ocenia się, że ewaluacja instytucjonalna podnosi motywację badaczy, których publikacje i cytowania mają decydujący wpływ na punktację instytucji, ale osłabia motywację pozostałych; może wpływać ujemnie na motywację kobiet, początkujących badaczy oraz mniejszości etnicznych. Gdy podczas ewaluacji wszystkie publikacje autora za okres, który ona obejmuje (np. 5 lat), zalicza się na poczet instytucji, w której jest on zatrudniony w chwili przeprowadzania ewaluacji, niezależnie od jego wcześniejszych miejsc zatrudnienia, zachęca to bogatsze placówki, aby tuż przed jej rozpoczęciem „wykupować” wysoko wydajnych badaczy<sup>157</sup>.

**Plany badawcze.** Zwraca się też uwagę, że ewaluacja instytucjonalna ogranicza swobodę badawczą, podporządkowując myślenie o planach badawczych ustalonym raz na kilka lat zasadom i kategoriom<sup>158</sup>.

**Publikacje.** Jak wykazują badania przeprowadzone w Wielkiej Brytanii, Australii, Hiszpanii i Norwegii, ewaluacja instytucjonalna zwiększa wydajność publikacyjną badaczy. Badania nad wpływem na jakość publikacji (mierzoną przez pomiar cytowań) przyniosły jednak zróżnicowane ustalenia: podniosła ona poziom publikacji w Wielkiej Brytanii, ale obniżyła w Australii i Norwegii (skutek zasady „publikuj albo zgiń”). Badanie wpływu na miejsce wyboru publikacji wskazało, że często pod wpływem systemu badacze kierują artykuły nie do czasopism, w których chcieliby drukować, czytanych przez środowisko, tylko do tych mających wysokie notowania<sup>159</sup>.

**Dydaktyka.** Wpływ opisywanej ewaluacji „na dydaktykę jest trudny do oceny, ze względu na brak odpowiednich metod pomiaru. Jednak w wielu krajach wyrażano obawy, że nacisk na ilość i jakość badań może wywrzeć ujemny wpływ na poziom dydaktyki”<sup>160</sup>.

**Rozwój współpracy naukowej.** Wpływ ewaluacji instytucjonalnej na rozwój współpracy (wewnątrz krajowej i międzynarodowej) zależy m.in. od sposobu pomiaru publikacji opartych na współpracy autorów pochodzących z różnych instytucji. Sposób „częściowego naliczania” (*fractional counting*, gdy każdemu

---

<sup>155</sup> S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

<sup>156</sup> L. Butler, *Impacts of performance-based research funding systems*, op. cit.

<sup>157</sup> Ibidem.

<sup>158</sup> Ibidem.

<sup>159</sup> Ibidem.

<sup>160</sup> Ibidem.

autorowi przypisuje się część publikacji) zniechęca do wspólnego publikowania, a zasada „pełnego liczenia” (*whole counting*, gdy praca idzie na konto każdego autora) odwrotnie, wspiera współpracę<sup>161</sup>. Konsekwencją ewaluacji może być wzrost konkurencji pomiędzy instytucjami kosztem współpracy<sup>162</sup>. Zwłaszcza De Meis (et al.) argumentuje, że presja na publikowanie prowadzi do przesadnej konkurencji<sup>163</sup>.

**Rozwój współpracy z biznesem.** W pewnych krajach, jak np. w Nowej Zelandii, prywatne przedsiębiorstwa wysokiej techniki twierdzą, że ewaluacja instytucjonalna zniechęca szkoły wyższe do współpracy z biznesem oraz do transferu wiedzy do sektora prywatnego. Sugerują one, że „koncentracja na doskonałości akademickiej zniechęca pracowników sektora szkolnictwa wyższego do podejmowania badań aplikacyjnych, a pomiar publikacji stwarza barierę dla prowadzenia badań komercyjnych”<sup>164</sup>.

**Otoczenie zewnętrzne nauki.** Nacisk na pomiar wiedzy kodyfikowanej mierzonej poprzez wskaźniki bibliometryczne (pomiar publikacji w recenzowanych czasopismach, najlepiej objętych Web of Science) obniża (lub neguje) wartość raportów technicznych i technicznego *know-how*, które niejednokrotnie mają znacznie większy wpływ na zmiany poza sferą nauki<sup>165</sup>. W Wielkiej Brytanii podkreślano, że RAE ocenia zbyt nisko tematykę badań związanych z kształceniem zawodowym (np. prawo, edukacja, biznes i zarządzanie, ekonomia). W szczególności podnoszono obawy, że badania teoretyczne ocenia się znacznie wyżej niż badania zorientowane na praktykę<sup>166</sup>.

**Zarządzanie badaniami.** Bezdyskusyjnie dodatni jest wpływ ewaluacji instytucjonalnej na systemy informacji zarządczej. „Wzrost potrzeb w zakresie wiarygodnych i aktualnych danych spowodował, że wiele uczelni przeprojektowało swoje systemy informacji dotyczącej studentów, personelu, grantów, dochodów wewnętrznych, publikacji badawczych itd., w szczególności system pociągnął za sobą ulepszenie lub stworzenie uczelnianych repozytoriów publikacji (Australia i Wielka Brytania)”<sup>167</sup>. Równie silny i pozytywny jest jej wpływ na zarządzanie badaniami. Badania przeprowadzone w Wielkiej Brytanii dowiodły, że ewalu-

<sup>161</sup> Ibidem.

<sup>162</sup> S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

<sup>163</sup> L. de Meis, A. Velloso, D. Lannes, M.S. Carmo i C. de Meis, *The growing competition in Brazilian science: rites of passage, stress and burnout*, „Brazilian Journal of Medical and Biological Research” 2003, 36(9), s. 1135–1141, cyt. za: N. Frølich, *The politics of steering by numbers*, op. cit.

<sup>164</sup> S. Box i E. Basri, *Performance-Based Funding for Public Research in Tertiary Education Institutions*, op. cit.

<sup>165</sup> *The Evaluation of Scientific Research: Selected Experiences*, OECD DSTI/STP/SUR(97)9.

<sup>166</sup> *The impact of the 1992 Research Assessment Exercise on higher education institutions in England*, [http://www.hefce.ac.uk/pubs/hefce/1997/m6\\_97.htm](http://www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm).

<sup>167</sup> L. Butler, *Impacts of performance-based research funding systems*, op. cit.; *The impact of the 1992 Research Assessment Exercise on higher education institutions in England*, [http://www.hefce.ac.uk/pubs/hefce/1997/m6\\_97.htm](http://www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm).

acja ta (choć także inne czynniki, takie jak np. polityka rad ds. badań) stała się czynnikiem rozwoju planowania badań na uniwersytetach. Jednym z następstw systemu finansowania opartego na ewaluacji dokonań bywa restrukturyzacja uczelni, np. zamykanie pewnych wydziałów wykazujących słabe wyniki (lub odwrotnie, wzmacnianie ich, gdy problematykę badań uważa się za ważną)<sup>168</sup>.

**Profil badawczy (struktura dyscyplinowa, specjalizacja naukowa) kraju.** Jeśli idzie o wpływ na poszczególne dyscypliny, obawy (niedowiedzione jednak mocnymi argumentami) dotyczą wzmocnienia przez ewaluację instytucjonalną dyscyplin łatwiej poddających się pomiarom (nauki przyrodnicze, techniczne i medyczne) kosztem humanistyki. Podkreśla się, że nie zawsze służy ona badaniom stosowanym, ukierunkowanym na rozwiązanie konkretnych kwestii i zorientowanym na cele praktyczne, gdyż badania tego typu są publikowane w formie ekspertyz i „szarej literatury”<sup>169</sup>.

### 2.13. Instytucje naukowe – ewaluacje indywidualne

Instytucje badawcze publicznego sektora nauki są to albo organizacje badawcze zajmujące się badaniami podstawowymi, takie jak francuski CNRS lub niemiecki Max Planck Gesellschaft, albo organizacje ukierunkowane na misję, takie jak instytuty przemysłowe, rolnicze lub medyczne<sup>170</sup>, albo wreszcie organizacje, które są częścią wydziałów uniwersyteckich lub działają jako międzywydziałowe centra badawcze.

**Instytuty uczelniane.** Z reguły w różnych krajach współistnieją różne typy (zewnątrzny i wewnętrzny) procesu ewaluacji uczelni. W pewnych krajach istnieją opisane wcześniej „krajowe systemy ewaluacji” (np. w Wielkiej Brytanii i w Australii<sup>171</sup>), podczas gdy w innych ewaluacja instytucjonalna jest sprawą wewnętrzną samych uczelni (np. w Irlandii); w niektórych wreszcie krajach oba systemy ocen współistnieją ze sobą: np. w Niderlandach co sześć lat (naprzemienienie) przeprowadza się zarówno wewnętrzne, jak i zewnętrzne ewaluacje, przy czym wewnętrzne organizuje się trzy lata (w połowie cyklu) po zewnętrznych.

---

<sup>168</sup> Ibidem.

<sup>169</sup> Ibidem.

<sup>170</sup> Pozauczelniane publiczne instytucje naukowe definiuje się jako krajowe podmioty, niezależnie od ich statusu prawnego (podstawy w ramach prawa publicznego lub prywatnego), których głównym celem jest prowadzenie badań podstawowych, badań przemysłowych, prac rozwojowych, szkolenia, doradztwa i usług, a także rozpowszechnianie wyników badań poprzez szkolenia, publikacje i transfer technologii, których zyski (jeśli istnieją) są reinwestowane w działalność badawczą, rozpowszechnianie jej wyników lub szkolenia, oraz które są w całości lub w znacznej części własnością publiczną, i/lub są finansowane przede wszystkim ze środków publicznych poprzez finansowanie instytucjonalne (granty blokowe) lub poprzez badania oparte na umowie, tak aby przede wszystkim realizować misję publiczną. OECD *Strengthening the Impact of Public Research Institutions. Evaluation of PRIs*, 2009.

<sup>171</sup> *Research Excellence Framework* w Wielkiej Brytanii, *Research Quality Framework Initiative* w Australii.

W krajach, w których współistnieją oba typy ewaluacji, ewaluacje wewnętrzne są traktowane jako przygotowanie do zewnętrznych, zarówno pod względem oceny działania uczelni, jak i organizacji oraz przebiegu procesu oceny<sup>172</sup>.

Instituty uniwersyteckie w coraz większej liczbie państw oceniane są na szczeblu krajowym przez (omówione w poprzednim rozdziale) ewaluacje oparte na pomiarze dokonań. Innym trendem w ewaluacji instytucji i grup badawczych są ewaluacje stosowane jako podstawa przydzielania dużych wieloletnich dotacji dla centrów doskonałości (Inżynieryjne Centra Badawcze NSF od 1980 roku, Sieci Centrów Doskonałości w Kanadzie, australijskie Co-operative Research Centres, szwedzkie Centra Kompetencji, holenderskie the Dutch Top Technology Institutes, austriackie K-centres, niemieckie Centra Kompetencji Nanotechnologii)<sup>173</sup>.

**Pozauczelniane publiczne instytucje naukowe (PPIN)** są silnie zróżnicowane pod względem misji, funkcji, statusu prawnego, specjalizacji, powiązań z otoczeniem oraz źródeł finansowania<sup>174</sup>. Zazwyczaj zasady ich ewaluacji różnią się od zasad stosowanych w ocenie badań prowadzonych w uczelniach, są bardziej ukierunkowane na ocenę (nie tylko) produktów, takich jak publikacje naukowe, ale także efektów i wpływów. Ogólnie rzecz biorąc, ewaluacje instytucji pozauczelnianych przeprowadza się, aby lepiej zrozumieć skalę, charakter i determinanty zwrotu z inwestycji, a także aby poznać wszystkie niezamierzone skutki tych inwestycji. Celem ewaluacji jest poprawa zarządzania i finansowania PPIN oraz uczenie się z sukcesów i porażek. Kluczowe pytania zadawane podczas ewaluacji dotyczą efektywności i skuteczności instytucji i ich programów, uzasadnień interwencji rządu oraz tego, czy poprzez finansowanie rządu osiągnięto dodatkowość.

Ewaluacje przeprowadza się na różnych poziomach: na poziomie instytucji, ale także na poziomie programów, typów działalności lub (rzadko) na poziomie całego sektora. Najczęściej ewaluację przeprowadzają firmy konsultacyjne, rzadziej indywidualni zagraniczni eksperci.

Najważniejsze pytanie zadawane podczas ewaluacji dotyczy wpływu lub wartości dodanej (instytucji, programu lub sektora), rozumianych różnie, ale często w sposób szeroki, z uwzględnieniem oddziaływania na społeczeństwo i gospodarkę. Inne pytania dotyczą rezultatów (publikacje, patenty), skuteczności, transferu wiedzy, planowania strategicznego, zarządzania.

---

<sup>172</sup> Sądzi się, że brak zewnętrznego „krajowego systemu ewaluacji”, z jednoczesną silną kulturą wewnętrzną ewaluacji, może w pewnych krajach mieć dobre strony, gdyż pozwala na większą elastyczność badaczy i instytucji oraz na lepszy podział zadań; i tak, badacze utalentowani dydaktycznie mogą poświęcać więcej czasu dydaktyce, menadżersko – kierowaniu projektami itd.

<sup>173</sup> Opracowane na podstawie: OECD *Evaluation Of Publicly Funded Research: Recent Trends And Perspectives*, 2006; OECD *IPP Issue Brief: Research Organisation Evaluation*, 2014.

<sup>174</sup> Te ostatnie obejmują tak różne składniki, jak finansowanie instytucjonalne nieobwarowane warunkami (lub tylko ogólnymi) i nieukierunkowane na realizację określonego projektu lub programu, finansowanie projektów, dotacje rządowe i samorządowe na wykonywanie określonych usług, takich jak np. testowanie, kontrakty rządowe i samorządowe udzielane w wyniku przetargu, kontrakty ze strony biznesu, inne (np. dochody z tytułu wynajmu pomieszczeń lub produkcji na małą skalę). Ibidem.

Przeprowadzając ewaluacje, najczęściej sięga się po ocenę jakościową (wywiady i kwestionariusze do zainteresowanych stron), a także po analizę kosztów i korzyści oraz analizę danych. Niekiedy korzysta się ze wskaźników, które z natury informują o przeszłości, rzadziej uwzględnia się w ewaluacji przyszłe skutki badań<sup>175</sup>.

W pewnych krajach – np. we Flandrii (Belgia), Finlandii, Włoszech, Japonii, Nowej Zelandii i Norwegii – okresowe ewaluacje publicznych pozauczelnianych instytucji naukowych zostały wpisane do cyklu polityki. W Norwegii program finansowania instytucjonalnego pozauczelnianych instytucji naukowych (ok. 10% budżetu nauki) oparty został na wskaźnikach dokonań, takich jak m.in. liczba publikacji, współpraca z uczelniami, skala przychodów z grantów Rady ds. Badań oraz skala przychodów z zagranicy. W Finlandii, ustalając poziom finansowania, bierze się pod uwagę takie czynniki, jak wpływ społeczny, wydajność operacyjną, jakość zarządzania, produkty. W Japonii instytuty są zobowiązane do corocznego przesyłania raportów dokumentujących uzyskane osiągnięcia; pewne kategorie instytutów podlegają przeprowadzanym co kilka lat ewaluacjom.

Ewaluacje bywają albo jednorazowe, dokonywane z pewnej szczególnej okazji, albo periodyczne.

Jako przykład jednorazowej ewaluacji przytacza się ewaluacje w Max Planck Gesellschaft w procesie tworzenia nowego instytutu. Wiązą się one z wyborem komitetu *peer review* przyszłego dyrektora, któremu następnie powierza się zadanie utworzenia instytutu; ma on istnieć aż do jego emerytury. Model ten realizuje także Riken w Japonii.

Model periodyczny stosuje się we Francji. Tego rodzaju ewaluacje służą wspieraniu zmian strukturalnych w ramach jednostek. Instytuty INSERM powołuje się na 12 lat; coroczna otwarta rozmowa ma wspomagać odnowę.

Dominuje trend przechodzenia z ewaluacji jednorazowych na periodyczne. Ten trend zmierza w dwóch kierunkach:

1. Umieszczanie ewaluacji w pracach nad budową strategii. Typowy przykład to INSERM i hiszpańska agencja badań CSIC. Instytuty obu organizacji opracowują oddolne plany strategiczne według jednolitego formatu. Panele tematyczne dokonują przeglądu planów, przeprowadzają wysłuchania ich dyrektorów oraz formułują zalecenia.
2. Międzydziedzinowe podejście do przydzielania dotacji podstawowych. Przykładem tego podejścia jest niemieckie Towarzystwo Helmholtza. Towarzystwo wprowadziło konkurencyjny proces oparty na finansowaniu programów ukierunkowanych. Interdyscyplinarne programy są poddawane ocenie paneli składających się z ośmiu do dziesięciu członków, którzy spotykają się, aby przez kilka dni studiować pisemne propozycje. Następnie Senacka Komisja Towarzystwa przeprowadza ocenę porównawczą

<sup>175</sup> OECD, *Strengthening the Impact of Public Research Institutions. Evaluation of PRIs*, 2009.

programów. Zapewnia się stały monitoring wewnętrzny oraz okresowe przeglądy badań. Podejście to jest charakterystyczne dla nowego typu zarządzania, nie według dziedzin lub regionów, tylko według zbiorów działań.

Przejście do okresowych ewaluacji nastąpiło m.in. przed laty w Norwegii, gdy większość pozauczelnianych instytutów połączono w ramach Norweskiej Rady ds. Badań, oraz w Japonii, kiedy japońskie badawcze instytuty przemysłowe połączyły się w nowo utworzonym Instytucie Zaawansowanej Przemysłowej Nauki i Technologii (AIST). Innym przykładem jest niemieckie Stowarzyszenie Leibniza, która przejęła od Rady Nauki (*Wissenschaftsrat*) odpowiedzialność za ewaluację swoich 80 instytutów, współfinansowanych przez rząd federalny i rządy krajów związkowych. W Stowarzyszeniu Leibniza proces oceny składa się z dwóch faz. W pierwszej z nich ewaluację przeprowadza panel ekspertów opierający się na odpowiedziach na kwestionariusz oraz wizytach terenowych. Sprawozdania panelu są podstawą drugiej fazy, w której komitet ewaluacyjny dokonuje oceny porównawczej instytutów<sup>176</sup>.

#### **2.14. Narzędzia *policy intelligence*, takie jak ewaluacja, foresight, mapowanie technologii itd.**

Od czasu do czasu stosuje się ewaluacje narzędzi *policy intelligence* – form wsparcia informacyjno-analitycznego polityki naukowej i innowacyjnej. Przeprowadzono np. kilka ewaluacji foresightu. Foresight to metoda prognozowania polegająca na „dyskusji nad przyszłością w gronie przedstawicieli decydentów (władzy publicznej), środowisk naukowych, przemysłu, mediów, organizacji pozarządowych i opinii publicznej, przy czym nie chodzi o dokładność prognozy, lecz o uświadomienie perspektyw i przygotowanie do zmian”<sup>177</sup>. Podczas ewaluacji ocenia się m.in. efektywność narzędzi i to, czy osiągają one założone cele<sup>178</sup>.

W Polsce sporządzono ekspertyzę *Badanie ewaluacyjne realizowanych w Polsce projektów foresight (2010)*<sup>179</sup>.

---

<sup>176</sup> R. Hofer, *Draft Questionnaire on Public Research Institutes*, OECD, 2010; E. Geisler, *Integrated Figure of Merit of Public Sector Research Evaluation*, „Scientometrics” 1996, vol. 36, no. 3; R. Miller i A. Manseau, *Bibliometric indicators and the competitive environment of R&D laboratories*, „Scientometrics” 1996, vol. 36, no. 3.

<sup>177</sup> *Foresight*, Wikipedia 2014.

<sup>178</sup> K. Cuhls, *Evaluation of Foresight Activities*, Fraunhofer Institute Systems and Innovation Research.

<sup>179</sup> Politechnika Białostocka, *Ekspertyza Badanie ewaluacyjne realizowanych w Polsce projektów foresight*, Białystok, listopad 2010.

### 3. Problemy ewaluacji nauki

#### 3.1. Funkcje

Ewaluacje, wśród nich ewaluacja instytucji naukowych finansowanych ze źródeł publicznych, pełnią rozliczne funkcje. Można je przedstawić w pięciu grupach, są to funkcje:

- oceny;
- sterowania i koordynacji;
- informacyjno-edukacyjna;
- zarządcza;
- symboliczna.

W formie rozwiniętej funkcje te można ująć następująco (tabela 2.20).

**Tabela 2.20**

#### **Funkcje ewaluacji instytucji naukowych**

<ul style="list-style-type: none"> <li>• monitoringu, audytu, benchmarkingu, kontroli jakości;</li> </ul>
<ul style="list-style-type: none"> <li>• alokacyjna (podstawa decyzji finansowych: co finansować, kiedy zwiększyć, kiedy zmniejszyć, a kiedy zaprzestać finansowania);</li> <li>• administracyjna (narzędzie władzy);</li> <li>• koordynacyjna w relacjach pryncypał-agent (parlament-rząd, rząd-ministerstwo, ministerstwo-agencja, agencja-instytucje naukowe);</li> <li>• legitymizacji i uzasadnienia interwencji (w oczach „podatnika” i „obywatela”, dla rządu i w oczach parlamentu);</li> <li>• rozliczenia (<i>accountability</i>);</li> <li>• negocjacyjna (ustanawianie i modyfikacja systemu ewaluacji to pole negocjacji pomiędzy środowiskiem naukowym a politykami i urzędnikami reprezentującymi interes publiczny);</li> </ul>
<ul style="list-style-type: none"> <li>• informacyjna (źródło informacji o: efektywności planu ewaluacji, jego realizacji, ocenianych jednostkach itd.);</li> <li>• poznawcza (rozumienie obszaru ewaluacji, przewidywanie trendów);</li> <li>• edukacyjna (uczenie się, jak lepiej: prowadzić ewaluację, robić politykę, zarządzać badaniami; podstawa uczenia się polityków, urzędników, menadżerów projektów, badaczy, ewaluatorów);</li> </ul>
<ul style="list-style-type: none"> <li>• zarządcza na szczeblu politycznym (planowanie i podział zasobów, ustalanie priorytetów, wkład informacyjny do strategii oraz formułowania i oceny polityk);</li> <li>• zarządcza na szczeblu wykonawczym (źródło informacji służącej ulepszeniu strategii i zarządzania);</li> </ul>
<ul style="list-style-type: none"> <li>• symboliczna (źródło prestiżu i reputacji)<sup>a</sup>.</li> </ul>

<sup>a</sup> Źródła: własne propozycje; K. Olejniczak, *Teoretyczne podstawy ewaluacji ex post*, [w:] *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007, s. 19–21; *Ewaluacja w Administracji Publicznej Funkcje, Standardy i Warunki Stosowania*; [http://www.pte.org.pl/repository/files/PTE/Ewaluacja\\_w\\_administracji\\_i\\_publicznej.pdf](http://www.pte.org.pl/repository/files/PTE/Ewaluacja_w_administracji_i_publicznej.pdf); P. Cunningham, *Principles of the design and implementation of evaluations to maximise usefulness in different contexts*, 2010; L. Georghiou, *Evaluation in the Balance: Matching Methods and Policy Needs*, paper presented at The Evaluation Conference, Bruxelles, Nov. 10–12, 2001. Dostępne: <http://les>.



man.ac.uk/PREST; R. Barré, *Evaluation: a process for strategic debate*, 2002; B. Godin, *Are Statistics Really Useful? Myths and Politics of Science and Technology Indicators*, Project on the History and Sociology of S&T Statistics, Working Paper no. 20, 2002; tenże, *The Who, What, Why and How of S&T Measurement*, Project on the History and Sociology of S&T Statistics, Working Paper no. 26, [http://www.csiic.ca/PDF/Godin\\_26\\_a.pdf](http://www.csiic.ca/PDF/Godin_26_a.pdf), *Le Banquet* (Revue du CERAP), 19–20, January 2004; M. Trow, *Trust, Markets and Accountability in Higher Education: A Comparative Perspective*, „Higher Education Policy” 1996, 9, 4, s. 309–324; J. Van Steen, *The use of S&T Indicators in science policy.. how can they matter?*, „Research Evaluation” 1995, vol. 15, no. 2.

Źródło: opracowanie własne.

Dobra ewaluacja instytucjonalna to taka, która uwzględnia i pozwala osiągnąć skuteczność każdej funkcji. Poszczególne funkcje ewaluacji łączą się ze sobą; pomiędzy niektórymi z nich istnieją napięcia (np. funkcji rozliczania, z istotnym elementem kontroli – rozliczenie z efektów, błędów, opóźnień przed opinią publiczną – z funkcją uczenia się, która „działa najlepiej w sytuacji otwartej, szczerzej i autokrytycznej dyskusji o problemach i blokadach”)<sup>180</sup>.

### 3.2. Pytania i kryteria

Biorąc pod uwagę olbrzymie zróżnicowanie ewaluacji, trudno ustalić listę najważniejszych pytań, jakie zadaje. Weźmy pod uwagę kryteria stosowane przez ewaluację efektów społeczno-gospodarczych działalności B+R.

Pytania i kryteria wynikają z porównania ze sobą pewnych podstawowych elementów interwencji publicznej, takich jak potrzeba (problem lub trudność dotycząca pewną grupę, którą interwencja stara się rozwiązać albo przezwyciężyć), cel (*objective*, jasne i wyraźne wyartykułowanie zakładanych efektów), wkład (*input*, zasoby finansowe, ludzkie, materialne, organizacyjne i prawne mobilizowane na rzecz interwencji), produkt (*output*), czyli bezpośredni rezultat uzyskany wskutek interwencji, efekt (*outcome*), czyli zmiana społeczno-gospodarcza lub poznawcza wynikająca bezpośrednio lub pośrednio z interwencji, wpływ (*impact*), czyli bardziej długofalowe następstwa interwencji<sup>181</sup>. Trzy elementy interwencji, które często najtrudniej odróżnić, dotyczą wytworów, efektów i wpływu działania (*outputs, outcomes, impacts*).

<sup>180</sup> K. Olejniczak, *Teoretyczne podstawy ewaluacji ex post*, op. cit., s. 21.

<sup>181</sup> Por. *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, vol. 6, European Commission, Luxembourg, 1999; Ch. Pollit i G. Bouckaert, *Public Management Reform*, Oxford University Press.

Tabela 2.21

Efekty interwencji *outputs, outcomes, impacts*

Produkty ( <i>outputs</i> )	Efekty ( <i>outcomes</i> )	Wpływ ( <i>impact</i> )
Co wytworzono dzięki podjętemu działaniu?	Jakie efekty uzyskano?	Co się wydarzyło w następstwie działania i osiągniętego rezultatu?
Wyroby takie jak np. software, narzędzia, patenty, publikacje, raporty, prototypy, pokaz działania, konferencje i inne formy upowszechnienia wiedzy, kwalifikowani ludzie.	Zmiany w kompetencjach firmy, zmiany w strategii firmy, tantiemy z licencji i patentów, zyski z inwestycji, mobilność personelu.	Podniesienie poziomu produkcji branży, poprawa efektywności gospodarczej, wzrost konkurencyjności, wzrost usług eksportowych.

Źródło: G. Becher i S. Kuhlmann, red., *Evaluation of technology policy programmes in Germany*, op. cit.

Szczególnie trudno ustalić efekty (*outcomes*). Są one niezliczone, powiązane ze sobą w skomplikowany sposób, czasami niezamierzone, trudne do opisu i pomiaru, czasami pozytywne, czasami negatywne. Subiektywizm w ich ocenie jest nieunikniony. Z powodu wszystkich tych komplikacji nie sposób określić ich precyzyjnie, obiektywnie czy ilościowo, np. wkład gospodarczy szkolnictwa wyższego<sup>182</sup>. Metodologia ewaluacji rozróżnia bardzo wiele typów efektów (w szerszym znaczeniu tego słowa), takich jak np. efekty pośrednie i bezpośrednie, „efekt dźwigni” (*leverage effect*), efekt po stronie popytu i po stronie podaży, efekt mnożnikowy, sieciowy, strukturyzujący, efekty synergii, takie jak wewnętrzna i zewnętrzna spójność, komplementarność.

Pytania i kryteria ewaluacji dotyczą relacji pomiędzy tymi podstawowymi jednostkami, choć jak dotąd brak jednej standardowej wykładni. Na przykład pytanie o odpowiedniość (*relevance*) tłumaczy się (najczęściej) jako relację pomiędzy celami i potrzebami/problemami, pytanie o efektywność dotyczy porównania celu z efektami, pytanie o wydajność – porównania wkładu z produktem, pytanie o użyteczność – potrzeby/problemu z wpływem itd.<sup>183</sup> Trzeba jednak pamiętać, że sama tylko ewaluacja działalności B+R finansowanej ze źródeł publicznych jest dziedziną tak obszerną i zróżnicowaną, że poniższa lista nie uwzględnia wszystkich pytań, jakie zadaje się w jej trakcie, i odwrotnie, rzadko kiedy podczas ewaluacji stawia się wszystkie z przedstawionych tu pytań (tabela 2.22).

<sup>182</sup> G. Becher i S. Kuhlmann, red., *Evaluation of technology policy programmes in Germany*, op. cit.

<sup>183</sup> Ch. Pollit i G. Bouckaert, *Public Management Reform*, Oxford University Press.

Tabela 2.22

## Pytania i kryteria ewaluacji

Kryterium		Pytanie
Stosowność	<i>Appropriateness</i>	Co należy robić? Czy jest rzeczą właściwą robić to, co robimy?
Odpowiedniość	<i>Relevance</i>	Czy cele interwencji rzeczywiście służą rozwiązaniu problemu, jaki został przed nią postawiony? Czy działania organizacji są zgodne z jej misją?
Zdatność	<i>Eligibility</i>	Czy przedmiot interwencji (region, projekt lub grupa ludzi) posiada cechy kwalifikujące do interwencji publicznej?
Subsydiarność	<i>Subsidiarity</i>	Dlaczego problem nie może zostać rozwiązany na niższym szczeblu polityki/zarządzania?
Ekonomiczność	<i>Economy</i>	Czy działanie okazało się tańsze niż zakładaliśmy? Czy kosztowało to mniej niż oczekiwano?
Efektywność	<i>Effectiveness</i>	Czy działanie spełniło nasze oczekiwania? Czy efekty odpowiadają zakładanym celom?
Wydajność	<i>Efficiency</i>	Czy uzyskany produkt jest wart zainwestowanego weń wkładu?
Sprawność procesowa	<i>Process efficiency</i>	Czy działa to jak należy? Czy działanie przebiegło sprawnie?
Umiejętności	<i>Capacities, capabilities</i>	Jakie kwalifikacje posiadają wykonawcy i czy są one na miarę wykonywanych przez nich zadań?
Skuteczność	<i>Efficacy</i>	Jaki jest zwrot z inwestycji ( <i>return on investment, ROI</i> )? Czy uzyskany zwrot z inwestycji odpowiada zakładanemu?
Jakość	<i>Quality</i>	Jak wartościowy jest wynik ( <i>output</i> )? Jak dobre są efekty?
Użyteczność	<i>Utility</i>	Czy wpływ interwencji rozwiązał problem lub zaspokoił potrzebę, jaka ją zrodziła?
Trwałość	<i>Sustainability</i>	Czy efekty trwają w horyzoncie średnio- i długoterminowym?
Dodatkowość	<i>Additionality</i>	Co się wydarzyło poza i ponad to, co by się wydarzyło bez podjętego działania?

Tabela 2.22 – cd.

Kryterium		Pytanie
Zastąpienie	<i>Displacement</i>	Co się nie stało, a co by zaszło, gdyby nie oceniane działania? Co się stało, a nie zaszłoby, gdyby nie interwencja (np. nowe miejsca pracy powołane dzięki programowi, które wyniknęły z likwidacji miejsc pracy w innym regionie)?
Ulepszenie procesu	<i>Process improvement</i>	W jaki sposób mogliśmy robić to lepiej?
Strategia	<i>Strategy</i>	Co powinniśmy robić następnie?

Źródło: E. Arnold i K. Guy, *Technology diffusion programmes and the challenge for evaluation*, [w:] *Policy evaluation in innovation and technology. Towards the best practices*, OECD, 1997; K. Guy, *Beyond Socio-economic Impact Evaluation: The challenge for the next millennium*, [w:] *Socio-economic Impact Evaluation Workshop*, Helsinki, 26–27 November 1999, Helsinki 2000; Ch. Pollit i G. Bouckaert, *Public Management Reform*, Oxford University Press 2000; *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, vol. 6, European Commission, Luxembourg 1999.

Ważne (szczególnie w ewaluacjach przeprowadzanych w Unii Europejskiej) są pytania dotyczące dodatkowości (*additionality*), czyli pytania o różnicę, jaką wprowadza działanie (jak np. przekazanie dotacji). W tradycyjnym podejściu dodatkowość to problem wkładu (jak dużo firmy wydają na B+R w relacji do wydatków rządowych).

Najczęściej zadaje się pytania dotyczące produktu, procesu (działań), efektu i wpływu oraz powiązań: potrzeb z produktem/efektem/wpływem (użyteczność), potrzeb z celem (stosowność), celów z produktem/efektem/wpływem (skuteczność), działań z produktem/efektem/wpływem (efektywność).

Rozszerzona lista pytań dotyczy m.in. następujących kwestii (tabela 2.23).

Tabela 2.23

## Pytania i kryteria ewaluacji – lista rozszerzona

Kryterium	Pytanie
Stosowność	Czy podjęcie działania było uzasadnione?
Odpowiedniość	Czy cele działania służyły rozwiązaniu problemu, jaki został przed nim postawiony?
Ekonomiczność	Czy działanie okazało się tańsze niż zakładaliśmy? Czy kosztowało mniej niż oczekiwano?
Efektywność (wydajność)	Czy działanie spełniło nasze oczekiwania?
Produktywność	Czy produkt (efekt, wpływ) wart był zainwestowanego weń wkładu?

Kryterium	Pytanie
Sprawność procesowa	Czy działanie przebiegło sprawnie?
Umiejętności	Jakie kwalifikacje posiadali wykonawcy i czy były one na miarę wykonywanych przez nich zadań?
Skuteczność	Czy cele programu zostały osiągnięte? Jaki jest zwrot z inwestycji ( <i>return on investment, ROI</i> )? Czy uzyskany zwrot z inwestycji odpowiada zakładanemu?
Jakość	Jak wartościowy jest wynik ( <i>output</i> )? Jak dobre są efekty?
Użyteczność	Czy produkt-efekty-wpływ działania rozwiązał problem lub zaspokoił potrzebę?
Trwałość	Czy efekty trwają w horyzoncie średnio- i długoterminowym?
Dodatkowość	Co się wydarzyło poza i ponad to, co by się wydarzyło bez podjętego działania?
Zastąpienie	Co się nie stało, a co by zaszło, gdyby nie oceniane działania?
Ulepszenia procesu	W jaki sposób można by to robić lepiej?

Źródło: E. Arnold i K. Guy, *Technology diffusion programmes and the challenge for evaluation*, op. cit.; K. Guy, *Beyond Socio-economic Impact Evaluation: The challenge for the next millennium*, op. cit.; Ch. Pollit i G. Bouckaert, *Public Management Reform*, op. cit.; *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, op. cit.

### 3.3. Dobre wzorce

Wiele publikacji stara się sformułować wzory dobrej ewaluacji lub udzielić rad i wskazań, jak przeprowadzać ewaluację skutecznie. Podkreśla się np., że standaryzacja procedur ewaluacji zmniejsza jej czas i koszty oraz ułatwia porównania pomiędzy programami, że ewaluacja powinna być traktowana jako proces uczenia się, zarówno dla ocenianych, jak i dla oceniających, jako okazja zdobycia użytecznych, skądinąd nieosiągalnych informacji, że jej wyniki powinny by powiązane z procesem podejmowania decyzji, że np. ewaluacja szkół wyższych nie powinna zabijać kreatywności (np. wydział uniwersytecki powinien być oceniany jako ekosystem, habitat, środowisko, według jakości życia, szans zdobywania doświadczenia, troski o wzrost, a nie jako maszyna do wytwarzania oczekiwanych wyników) itd.<sup>184</sup> Zwraca się uwagę, że ewaluacje obok zamierzonego wpływu pociągają za sobą także skutki uboczne (np. wzmacniają najlepszych ponad należne zasługi, zaostrzają zatargi, podcinają

<sup>184</sup> P. Jedamus, M.W. Peterson i Associates, *Improving Academic Management. A Handbook of Planning and Institutional research*, Jossey-Bess Publishers, 1989, s. 452–455; S. Lahlou, R. van der Meijden, M. Messu, G. Poquet i F. Prakke, *A Guideline for Survey-Techniques in Evaluation of Research*, 1992.

motywacje). Wskaźniki ewaluacji powodują, że celem ocenianych staje się nie tyle osiągnięcie stanu postulowanego, ile ich formalne wypełnienie<sup>185</sup>.

Dzięki dotychczasowym praktykom ewaluacji zebrano niezliczone doświadczenia. Jedno z nich stwierdza „malejący zwrot” z ewaluacji określonego typu (pomimo stosowania mechanizmów autokorekcyjnych). Gdy dany typ ewaluacji jest powtarzany, maleje szansa na niespodziewane lekcje, nawet gdy nowe okoliczności przynoszą ze sobą nowe wyzwania<sup>186</sup>.

Pod wieloma względami ewaluacja badań, zwłaszcza podstawowych zorientowanych poznawczo, jest trudniejsza niż ocena innych działań ze względu na wielki stopień niepewności i ryzyka, jaki się z nimi wiąże. Oczekiwana od badacza oryginalność utrudnia sformułowanie zasad oceny badań prowadzonych „na granicach poznania”. W żartobliwym powiedzeniu, że Kopernik i młody Einstein nie uzyskaliby nigdy od współczesnych grantu, kryje się wiele prawdy. Craig Venture nie dokonałby tylu odkryć genomu człowieka, gdyby sam nie finansował badań, którymi kierował.

Życie jest prostsze dla ewaluatora, gdy przedmiotem jego zainteresowania jest prosty program, ściśle określony w czasie i z jasno zarysowanymi celami. Jednak np. koncepcja Europejskiej Przestrzeni Badawczej narzuca nowe wyzwania ewaluacji, takie jak choćby większy nacisk na wartości społeczne i etyczne oraz większy nacisk na aspekty międzynarodowe (np. powiązania nauki z przemysłem nie tylko w granicach narodowych, rozwój B+R dla przyciągania i zatrzymywania zagranicznych inwestycji). Te nowe wyzwania zmuszają do adaptacji metod zarządzania przemysłem (benchmarking)<sup>187</sup>.

Istnieje wiele (oficjalnych i nieoficjalnych) zaleceń, jak przeprowadzać dobre ewaluacje i z jakimi problemami etycznymi spotykają się ewaluatorzy.

Ellen Hazelkorn sugeruje np., że oceny badań naukowych powinny:

- „łączyć oceny oparte na wskaźnikach ilościowych z danymi jakościowymi, np. *peer review* i/lub ocenami użytkownika końcowego;
- uznawać istnienie istotnych różnic pomiędzy dyscyplinami badawczymi. Artykuły w recenzowanych pismach są głównym kanałem publikacji praktycznie wszystkich dyscyplin naukowych. Jednak złożoność wiedzy doprowadziła do powstania zróżnicowanego zestawu form prezentacji i upowszechnienia: nagrań audiowizualnych, programów komputerowych i baz danych, raportów i rysunków technicznych, projektów i prototypów, patentów lub

---

<sup>185</sup> E. Kalpazidou Schmidt, *Evaluation and science policy*, [w:] *The use of Evaluation in Europe*, 2002.

<sup>186</sup> L. George, *Issues in the evaluation of innovation and technology policy*, [w:] *Policy evaluation in innovation and technology. Towards best practices*, 1997 OECD.

<sup>187</sup> L. Georghiou, *Evaluation of research and innovation policy in Europe – new policies, new frameworks?*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001, [http://cspso.org/legacy/library/091110F6VH\\_lib\\_Gustonbhwshop.pdf](http://cspso.org/legacy/library/091110F6VH_lib_Gustonbhwshop.pdf), s. 65.

praw do nowych metod uprawy roślin, dzieł sztuki, analiz politycznych, opisu przypadków prawnych, map, tłumaczeń i edycji najważniejszych utworów według standardów akademickich itd.;

- uwzględniać oceny wpływu i korzyści. Badania nie istnieją w izolacji. Ich wpływ w znacznym stopniu różni się od dyscypliny. Wskaźniki mogą obejmować zatrudnienie absolwentów, liczbę założonych spółek i liczbę osób w nich zatrudnionych, zmiany wprowadzone w polityce i ustawodawstwie, zmniejszenie zanieczyszczeń i poprawę opieki zdrowotnej. Wskaźniki oceny interesariuszy wskazują, w jaki sposób badania są oceniane przez społeczeństwo. Mogą one obejmować np. prestiżowe nagrody krajowe i międzynarodowe;
- obejmować samoocenę<sup>188</sup>.

### 3.4. Podejścia jakościowe i ilościowe w ewaluacji

Ewaluacja z reguły łączy w sobie dwa podejścia: analizę wskaźników z ewaluacją panelu ekspertów, korzystającą z takich metod, jak badania ankietowe, analiza dokumentów (np. raportów rocznych instytutów), wizyty *in situ*, indywidualne lub grupowe wywiady z badaczami, administratorami i klientami. Stosunek pomiędzy oboma podejściami bywa różny – albo ocena ekspertów jest włączona do wskaźnika złożonego jako jeden z jej elementów, albo też (część) ostateczny werdykt należy do ewaluatorów, którzy obok wizytacji i innych metod biorą pod uwagę wyniki wskaźników dokonają<sup>189</sup>.

**Ewaluacja panelu ekspertów** – to ocena dokonywana przez zespół badaczy (lub także przedstawicieli praktyki). Ewaluację panelu ekspertów można uznać za rezultat połączenia tradycyjnego *peer review* (poziom mikro) z oceną polityki publicznej. W przeciwieństwie do tradycyjnego *peer review*, które ma na celu ocenę badań na poziomie mikro (artykuły, wnioski lub wnioskodawcy), ewaluacja panelu ekspertów jest przeprowadzana na poziomie mezo (instytucji, programu lub dyscypliny) oraz makro (krajowym)<sup>190</sup>.

Z kolei **wskaźniki** (*performance indicators*) definiuje się jako takie miary pewnej cechy działalności, które pozwalają na jej ocenę. Wskaźniki mogą mieć formę proporcji (np. stosunku liczby badaczy do liczby ich publikacji w recenzowanych czasopismach), procentu osiągnięcia założonego celu, miejsca w rankingu itd.<sup>191</sup> Podstawowy wymóg stawiany wskaźnikom to zrozumienie kontekstu,

---

<sup>188</sup> E. Hazelkorn, *Pros and cons of research assessment, lessons from rankings*, 2009, <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1000&context=dirrerep>

<sup>189</sup> Por. J. Edler, *Evaluation in Science and Technology Policy. Introduction*, Manchester 2010.

<sup>190</sup> L. Langfeldt, *Decision-making in expert panels evaluating research. Constraints, processes and bias*, NIFU Oslo 2002, [www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf](http://www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf)

<sup>191</sup> M. Herbst, *Financing public universities: the case of performance funding*, Dordrecht, Springer 2007, s. 66.

celów, wkładu, działań, wyników oraz zachodzących między nimi powiązań (tzw. logika interwencji)<sup>192</sup>.

Wyróżnia się różne grupy wskaźników ewaluacji. Jedna z nich obejmuje wskaźniki widziane pod kątem złożoności (wskaźniki pojedyncze oraz wskaźniki proporcji). Inna zawiera w sobie wskaźniki dzielone według elementu ewaluacji (wskaźniki wkładu, procesu, produktu, efektu i wpływu, do których dodaje się czasem wskaźnik kontekstu). Jeszcze inna grupa, obejmująca tylko wskaźniki proporcji, zróżnicowana jest według pytań stawianych w ewaluacji (np. użyteczność, stosowność, skuteczność itd.). Wskaźniki grupuje się też według szerszych kategorii, takich jak np. jakość, wpływ i efektywność badań. Można też zbudować zestaw, w którym znalazłyby się wskaźniki:

- wartości absolutnych (np. liczby publikacji i cytowań);
- proporcji (liczby cytowań w stosunku do liczby pracowników);
- udziału (np. w krajowej puli publikacji z danej dziedziny);
- pozycji (np. w rankingu);
- procentu realizacji celu;
- sumaryczne i złożone (takie jak np. *Impact Factor* czasopism, *Indeks Hirscha*, wskaźniki złożone, tzw. *composite indicators*, oraz wskaźniki określane przez panele *peer review* oceniające np. wydziały, najczęściej w formie pozycji skali). Najczęściej stosowane (różnie definiowane) wskaźniki zawiera tabela 2.24.

Tabela 2.24

## Grupy wskaźników ewaluacji

Pojedyncze wskaźniki		
<b>Wskaźnik wkładu</b>	Wykorzystane zasoby finansowe, kadrowe i infrastruktury	
<b>Wskaźniki procesu</b>	Przebieg działalności badawczo-rozwojowej, naukowo-technicznej i innowacyjnej	
<b>Wskaźnik produktu</b>	Uzyskane produkty – dobra i usługi	
<b>Wskaźnik efektu</b>	Bezpośrednie skutki produktu	
<b>Wskaźnik wpływu</b>	Długofalowe następstwa produktu	
<b>Wskaźnik środowiska</b>	Czynniki kontekstu, niezależne od działania, które miały wpływ na produkt, efekt i wpływ	
Wskaźniki proporcji		
<b>Użyteczność</b>	Potrzeba/produkt-efekty-wpływ	Ważne i wiarygodne tylko w takiej mierze, w jakiej istnieją jasne powiązania przyczynowo-skutkowe
<b>Stosowność</b>	Potrzeba/cel	

<sup>192</sup> L. Georghiou, *Evaluation in the Balance: Matching Methods and Policy Needs*, paper presented at The Evaluation Conference, Bruxelles, Nov. 10–12, 2001, <http://les.man.ac.uk/PREST>.



Wskaźniki proporcji cd.		
<b>Efektywność (wydajność)</b>	Wkład/produkt lub działania/produkt-efekty-wpływ	Ważne i wiarygodne tylko w takiej mierze, w jakiej istnieją jasne powiązania przyczynowo-skutkowe
<b>Skuteczność</b>	Cel/produkt-efekty-wpływ	
<b>Oplacalność</b>	Koszt/wpływ	

Źródło: opracowano na podstawie: *Measuring Government Activity*, OECD 2009, s. 16; H. Foss Hansen, *Research Evaluation: Methods, Practice and Experience*, Danish Agency for Science, Technology and Innovation, bmw, 2009; Ch. Pollit i G. Bouckaert, *Public Management Reform*, Oxford University Press 2000, s. 13; *Ewaluacja w administracji publicznej. Funkcje, standardy i warunki stosowania*, www.pte.org.pl/.../files/.../Ewaluacja\_w\_administracji\_publicznej.pdf.

Wśród wielu taksonomii wskaźników Hanne Foss Hansen proponuje podział na trzy grupy wskaźników wykorzystywanych w finansowaniu instytucji naukowych: 1) wskaźniki mające na celu pomiar badań, wkładu, procesu, struktur, produktu, efektu, wpływu; 2) sumaryczne wskaźniki, takie jak np. *Impact Factor* czasopism lub *Indeks Hirscha*, oraz 3) wskaźniki określane przez panele *peer review* oceniające np. wydziały<sup>193</sup>. Wskaźniki pierwszego typu Hansen dzieli na wskaźniki wkładu, procesu, struktury oraz rezultatów, dzieląc te ostatnie na wskaźniki produktu i efektu, a także opisuje ich zalety i ograniczenia.

Zgodnie z klasyfikacją przedstawioną we wcześniejszej części opracowania proponuję dokonać pewnej dekompozycji propozycji Hansena i wskaźniki struktury przenieść odpowiednio do grupy miar wkładu, procesu i efektu. Po takim zabiegu propozycja autora przedstawiałaby się w następujący sposób (tabela 2.25).

Tabela 2.25

## Wskaźniki wkładu

Wskaźniki wkładu	Zalety	Ograniczenia
Finansowanie zewnętrzne	Wskaźniki zdolności przyciągania zewnętrznego finansowania można mierzyć jako kwotę przychodów zewnętrznych z badań, także w przeliczeniu na FTE (przelicznik czasu pracy) i/lub jako liczbę i procent przychodów uzyskanych w konkursach (źródła krajowe i zagraniczne, w tym dotacje od rządu w porównaniu ze środkami przemysłu lub fundacji). Wskaźniki finansowania zewnętrznego mówią coś o konkurencyjności instytucji na rynkach finansowych.	Konkurencyjność to nie to samo co jakość. Reputacja i sieci powiązań również mają znaczenie. Ponadto poziom finansowania zewnętrznego jest bardzo zróżnicowany w zależności od dyscyplin. Różnice w poziomie finansowania zewnętrznego w połączeniu z różnicami w profilach instytucji poważnie ograniczają możliwości rzetelnych porównań pomiędzy instytucjami.

<sup>193</sup> H. Foss Hansen, *Performance indicators used in performance-based research funding*, DSTI/STP/RIHR(2010)4.

Tabela 2.25 – cd.

Wskaźniki wkładu	Zalety	Ograniczenia
Rekrutacja doktorantów i nauczycieli akademickich	Wskaźniki zdolności przyciągania studentów i pracowników naukowych mogą coś powiedzieć o konkurencyjności instytucji i jej absolwentów na rynku pracy. W zależności od celu można szacować liczbę lub udział wysoko wykwalifikowanych i / lub zagranicznych kandydatów.	Na atrakcyjność pracy na uniwersytecie ma wpływ wiele innych czynników, nie tylko jego atrakcyjność badawcza. Instytucje naukowe konkurują nie tylko z innymi instytucjami badawczymi, ale także z innymi ścieżkami kariery.
Aktywni badawczo pracownicy	Rozmiar badań aktywnych badaczy (powyżej określonego progu aktywności w danym okresie).	Brak jasnych definicji progu.
Liczba doktorantów	Liczba doktorantów jest wskaźnikiem zdolności do przyciągania studentów.	Różne dyscypliny mogą przyznawać różny status doktoratom.
Infrastruktura badawcza i urządzenia	Wyposażenie laboratorium, biblioteka (książki i dostęp do czasopism elektronicznych), urządzenia, personel pomocniczy, ogólne warunki pracy itp.	Trudny do porównań.

Źródło: opracowano na podstawie: H. Foss Hansen, *Performance indicators used in performance-based research funding*, DSTI/STP/RIHR(2010)4 (propozycja zmodyfikowana).

Tabela 2.26

## Wskaźniki procesu

Wskaźniki procesu	Zalety	Ograniczenia
Seminaria i konferencje	Liczba zorganizowanych seminariów i konferencji, jak również udziału w zewnętrznych konferencjach.	Liczba konferencji może być odbiciem turystyki naukowej.
Zaproszeni <i>keynote speakers</i>	Liczba zaproszeń <i>keynote speakers</i> na krajowe i międzynarodowe konferencje jako miara jakości, wpływu i znaczenia.	Liczba zaproszeń dla <i>keynote speakers</i> może odzwierciedlać gęstość sieci powiązań, a nie jakość. Nie ma możliwości porównania pomiędzy dyscyplinami.
Stanowiska wizytujących badaczy	Liczba odwiedzających może być potraktowana jako <i>proxy</i> prestiżu w środowisku naukowym.	Przyjazdy i wyjazdy wizytujących badaczy mogą odzwierciedlać spotkania w ramach sieci, a nie być wyrazem szacunku. Nie ma możliwości porównania pomiędzy dyscyplinami.

Wskaźniki procesu	Zalety	Ograniczenia
Współpraca i partnerstwo w dziedzinie badań	Gdy badania naukowe są prowadzone w coraz większym stopniu we współpracy krajowych i międzynarodowych grup badawczych, ważny staje się pomiar współpracy (np. krajowej, międzynarodowej, interdyscyplinarnej).	Różne są typy współpracy. Może ona być luźna lub intensywna, może obejmować różne instytucje. Wzory współpracy i publikacji różnią się w poszczególnych dziedzinach.

Źródło: opracowano na podstawie: H. Foss Hansen, *Performance indicators used in performance-based research funding*, DSTI/STP/RIHR(2010)4 (propozycja zmodyfikowana).

Tabela 2.27

## Wskaźniki produktu

Wskaźniki produktu	Zalety	Ograniczenia
Publikacje	Publikacje mają kluczowe znaczenie dla postępu w nauce. Pomiar publikacji w przeliczeniu na etaty <i>FTE</i> daje możliwość porównań. W zależności od celu można liczyć określone rodzaje publikacji, np. artykuły opublikowane w wysokiej rangi czasopismach.	Nacisk na jakość i wydajność. Ocena i ranking czasopism nie są jednoznaczne. W różnych dyscyplinach dominują różne typy publikacji (np. artykuły, książki, materiały konferencyjne).
Wytwory nietekstowe	W niektórych dziedzinach ważne są wytwory nietekstowe, takie jak dzieła sztuki, muzyki lub wykonania.	Ze względu na zróżnicowany charakter tych wytworów niełatwo je mierzyć.
Liczba doktorantów i odsetek badaczy uzyskujących tytuł doktora.	Nowe generacje naukowców są niezbędne dla dalszego postępu w badaniach naukowych. Pomiar absolwentów studiów doktoranckich można uzupełnić przez pomiar udziału absolwentów studiów doktoranckich kończących prace w odpowiednim czasie.	Różne dyscypliny mogą przyznawać różny status doktoratom, a poziom kończenia może się różnić w zależności od dyscypliny.
<i>Public outreach</i>	Można opracować pomiar obecności badaczy w społeczeństwie, np. w mediach.	Obecność w mediach może być bardzo luźno związana z aktywnością badawczą.

Źródło: opracowano na podstawie: H. Foss Hansen, *Performance indicators used in performance-based research funding*, DSTI/STP/RIHR(2010)4 (propozycja zmodyfikowana).

Tabela 2.28

## Wskaźniki efektu

Wskaźniki efektu	Zalety	Ograniczenia
Cytowania	Cytaty mówią coś o wpływie naukowym i widoczności. Bazy danych, takie jak Web of Science, Scopus oraz Google Scholar umożliwiają pomiar cytowań.	Liczba cytowań nie jest doskonałym wskaźnikiem jakości badań. Nie wszystkie dyscypliny są równie dobrze reprezentowane w indeksach cytowań. Dotyczy to zwłaszcza humanistyki, części nauk społecznych i techniki.
Liczba nagród	Wskaźnik jakości badań i wpływu.	Brak międzynarodowo uzgodnionej równoważności, z wyjątkiem Nagrody Nobla. Brak możliwości porównania wielu dyscyplin.
Zatrudnienia absolwentów studiów doktoranckich	Zatrudnienie absolwentów studiów doktorskich w przemyśle może być wskaźnikiem kompetencji absolwentów i wkładu badań naukowych do przemysłu i rozwoju społecznego.	Poziom zatrudnienia zależy także od innych czynników, np. gospodarki regionalnej lub krajowej.
Transfer wiedzy i komercjalizacja własności intelektualnej opartej na badaniach.	Miara wielkości przychodów uzyskanych z patentów, licencji lub nowo powstających firm. Ważne ogniwo między IP, komercjalizacją i korzyściami gospodarczymi.	Patenty są złym wskaźnikiem IP, zależnym zarówno od dziedziny, jak i krajowego kontekstu.
Ocena użytkownika końcowego	Raporty na zlecenie, doradztwo i kontrakty zewnętrzne są miarą gotowości podmiotów zewnętrznych ponoszenia opłat za badania. Takie miary, np. liczone jako kwota i procent finansowania od użytkowników końcowych (np. przemysłu, rządu, społeczności lokalnych), są wskaźnikiem przewidywanego wkładu do innowacji.	Zróżnicowanie dyscyplinowe.
Renoma	Na stanowiska redaktora czasopisma naukowego, udział w redakcji i w Komitecie Naukowym traktowane są jako wskaźniki renomy badacza w środowisku naukowym.	Wartość wskaźnika może być oznaką udziału w sieciach, a nie tylko być miernikiem uznania.

Źródło: opracowano na podstawie: H. Foss Hansen, *Performance indicators used in performance-based research funding*, DSTI/STP/RIHR(2010)4 (proponycja zmodyfikowana).

Wskaźniki Hanne Foss Hansena można uzupełnić o (pominięte przez niego) miary wpływu według propozycji przedłożonej w ramach prac przygotowawczych nad brytyjskim Research Excellence Framework (tabela 2.29).

Tabela 2.29

## Wskaźniki wpływu

Typ wpływu	Możliwe wskaźniki
Kształcenie wysoko wykwalifikowanych ludzi	Mobilność pomiędzy akademią a gospodarką. Zatrudnienie postdoków w przemyśle lub w firmach odpryskowych.
Tworzenie nowych przedsiębiorstw, usprawnienie istniejących, komercjalizacja nowych produktów lub procesów	Kontrakty badawcze i przychody z przemysłu. Badania we współpracy z przemysłem (mierzone np. liczbą wspólnych publikacji). Przychody z własności intelektualnej. Wzrastające obroty lub obniżone koszty poszczególnych przedsiębiorstw czy branż. Miary sukcesu nowych produktów/usług (np. wzrost przychodów). Miary sukcesu w zakładaniu firm odpryskowych (np. wzrost przychodów lub zatrudnienia). Patenty udzielone lub licencje. Mobilność personelu pomiędzy akademią a gospodarką.
Przyciąganie inwestycji B+R z rynku globalnego	Przychody z badań na rzecz zagranicznych przedsiębiorstw. Wspólne badania z zagranicznymi przedsiębiorstwami.
Ulepszenie polityki lub usług publicznych	Przychody z badań na rzecz organizacji rządowych. Zmiany wprowadzone do prawodawstwa i regulacji rządowych. Miary ulepszonych usług publicznych (np. wzrost wskaźników kompetencji czytania i liczenia). Wymiana personelu z organizacjami rządowymi. Udział w komitetach doradczych / politycznych. Wpływ na debatę polityki publicznej (np. mierzona poprzez cytowania przez NGOs i media).
Ulepszona opieka zdrowotna	Przychody z badań uzyskanych z fundacji wspierających badania medyczne. Miary ulepszonych usług badawczych (np. liczba uratowanych, obniżony poziom zakażeń). Zmiany w szkoleniu klinicznym, praktykach lub wskazaniach. Rozwój nowych i ulepszonych leków, metod diagnostyki i terapii; liczba zaawansowanych prób klinicznych. Udział w komitetach doradczych polityki zdrowotnej. Zmiany w zachowaniach publicznych (np. zmniejszenie palenia).

Tabela 2.29 – cd.

Typ wpływu	Możliwe wskaźniki
Postęp w kierunku zrównoważonego rozwoju, wliczając w to środowisko	Zastosowanie rozwiązań (nowych technologii, zmian zachowań itd.). Miary poprawy jakości środowiska (np. redukcja zanieczyszczeń, odnowa zasobów naturalnych).
Wzrost bogactwa kulturowego, wliczając w to bogactwa związane z nauką i techniką	Wzrost społecznego zainteresowania nauką i badaniami (np. na podstawie badań ankietowych). Zmiany w społecznych postawach do nauki (np. na podstawie badań ankietowych). Wzrost uznania dla dziedzictwa kulturowego (np. na podstawie badań ankietowych). Uczestnictwo w inicjatywach (np. wystawach, audycjach). Pozytywne recenzje lub miary uczestnictwa w inicjatywach.
Ulepszenie dobrobytu społecznego, spójności społecznej lub bezpieczeństwa narodowego	Zastosowanie nowych idei na rzecz ulepszenia równości, włączenia i spójności. Miary ulepszonej równości społecznej, włączenia lub spójności (np. ulepszone wskaźniki edukacji grup upośledzonych lub wzrost odsetka wyborców spośród grup cechujących się dotąd niskim poziomem uczestnictwa w wyborach).
Inne wskaźniki jakości życia	Propozycje

Źródło: *REF impact pilot exercise. Guidance on submissions*, November 2009, <http://www.hefce.ac.uk/research/ref/impact/>.

**Potrzeba komplementarności.** Uważa się, że kryteria ocen, jakiegokolwiek by były, włącznie z głównym kryterium „jakości badań”, są zbyt złożone, aby mogły być ujęte wyłącznie za pomocą wskaźników ilościowych, chwytających jedynie pewne (mieralne) aspekty badań. Z drugiej jednak strony, podkreśla się, że wysoka jakość badań, która nie przekłada się na wskaźniki ilościowe (jak np. wysoką liczbę cytowań), budzi poważne wątpliwości i wymaga wyjaśnień. Wskaźniki ilościowe służą zatem jako wsparcie i udokumentowanie ocen jakościowych i nie powinny mieć rozstrzygającego znaczenia.

Jednoczesna krytyka obu podejść – ilościowego i jakościowego – skłania do szukania sposobów ich wzajemnego uzupełniania i kontroli.

W odniesieniu do miar ilościowych tematem debaty jest to, czy mierzyć liczbę badań (liczbę publikacji), czy ich jakość (np. wyłącznie publikacje recenzowane, liczbę cytowań, wartość czasopism mierzoną przez *Impact Factor* [IF] itd.). Każdy wskaźnik jakości badań ma swoje ograniczenia (np. problem auto-cytowań, istnienie kół wzajemnych cytowań itd.).

Wartość wskaźników bibliometrycznych różni się w zależności od dyscyplin. Wskaźniki te nie nadają się do porównań międzydyscyplinowych, a w przypadku

wąskich lub lokalnych pól badań są w ogóle nieprzydatne. Żadne wskaźniki statystyczne, bez względu na ich charakter, nie wyjaśniają, dlaczego dana instytucja uzyskała dany wynik – ma wysoki poziom lub jest słaba. Tu konieczne są osądy lub odrębne badania. Podobnie, same tylko i wyłącznie wskaźniki bibliometryczne nie powinny być podstawą decyzji co do poziomu finansowania lub dalszych losów (wzmocnienie lub likwidacja) placówek. Konieczne jest także uwzględnienie potrzeb kraju i priorytetów rządowych.

Wskaźniki bibliometryczne ludzą obiektywizmem i jako takie bywają przeciwstawiane *peer review*. Przeciwwstawiając je ocenom jakościowym, zapomina się, że same one mają charakter ujętych ilościowo i zagregowanych decyzji jakościowych – czy cytować daną publikację, czy jej nie cytować. Cytuje się z rozmaitych przyczyn, a badania nad przyczynami cytowania, choć zaowocowały wieloma publikacjami, nie przyniosły jednoznacznych konkluzji. Bibliometrię, w ogóle, uznaje się bardzo „miękką” dyscyplinę.

Zdarza się, że wyjątkowo wysoko cytowana publikacja jednego z badaczy podciąga ogólną ocenę instytucji ponad jej rzeczywistą wartość – choć np. badacz był tylko jednym z wielu jej współautorów, a ponadto aktualnie nie pracuje w instytucji, którego nazwę podał.

Stosowanie wskaźnika IF w ocenie instytucji jest łatwe, ale pomija fakt dysproporcjonalnego rozkładu cytowań prac umieszczonych w danym czasopiśmie. To nie tyle czasopismo „podciągają” liczbę cytowań publikacji, ile suma cytowań publikacji decyduje o IF. Wartość wskaźnika IF dla pozycji czasopism, autorów oraz instytucji powoduje, że bywa poddawany on manipulacji (np. czasopisma nalegają na cytowanie zamieszczanych w nich prac)<sup>194</sup>.

Trzeba jednak podkreślić, że w przeciwieństwie do *peer review*, sumaryczne wskaźniki cytowań są oparte na decyzjach środowiska naukowego, a nie decyzjach wąskiej grupy recenzentów i panelistów. Pomimo czynnika „mody” lub istnienia „kół wzajemnych cytowań” liczba cytowań jest w większym stopniu efektem niezależnych wyborów wielu badaczy niż decyzji nielicznych uczonych, działających nieraz pod wpływem „myślenia grupowego”<sup>195</sup>.

Stosowanie wskaźników bibliometrycznych różni się w zależności od kraju. I tak, w Stanach Zjednoczonych mają one małe znaczenie, pomimo uznania dla miar dokonań jako środka obiektywizacji decyzji budżetowych. W Finlandii

---

<sup>194</sup> Trzeba jednak pamiętać, że wskaźnik IF jest tylko najbardziej popularnym, ale nie jedynym wskaźnikiem mierzącym jakość i wpływ czasopism. W ostatnich latach jesteśmy świadkami rozkwitu nowych propozycji – zarówno w odniesieniu do baz Thomson-Reuters, jak i Elsevier: *5-Year IF*, *Eigenfactor™*, *Article Influence™ Score*, *SJR*, *SNIP*, *Mean Normalised Journal Score MNJS*, *Total Normalised Journal Score TNJS* itd.; S. Haustein, *Multidimensional Journal Evaluation* oraz A.F.J. van Raan i inni, *The new set of bibliometric indicators of CWTS*, w: *11<sup>th</sup> International Conference on Science and Technology Indicators “Creating Value for Users”* 9–11 September 2010. Book of Abstracts; <http://info.scopus.com/journalmetrics/>

<sup>195</sup> *Performance-based models for publicly funded research in tertiary education institutions* 2009, OECD DSTI/STP/RIHR(2009)7.

i w Polsce natomiast *Impact Factor* czasopism „został niemal kanonizowany jako podstawa ocen”<sup>196</sup>.

Krytyka w stosunku do wskaźników ilościowych zbiega się z krytyką wobec ocen jakościowych (sądów), uważanych za subiektywne i uznaniowe. Uznaje się, że w *peer review* łatwiej o zgodę, czego nie finansować, niż o konsensus co do tego, co powinno zostać sfinansowane.

Czasami podkreśla się, że choć *peer review* może i jest najlepszym sposobem oceny propozycji badawczych, kandydatur na stanowiska lub publikacji, ale niekoniecznie dobrze sprawdza się na wyższych poziomach agregacji. A wskaźniki bibliometryczne odwrotnie, w zasadzie nie powinny być stosowane do oceny jednostek, ale są bardziej wiarygodne do oceny wydziałów, dyscyplin lub krajów<sup>197</sup>.

Za remedium na wady każdej z dwóch metod uznaje się tę przeciwną; zaleca się doskonalenie metod ilościowych i jakościowych ewaluacji, ostrożność w ich stosowaniu, uczenie się na doświadczeniach oraz ewaluację ewaluacji. W odniesieniu do metod ilościowych podkreśla się zatem, że nigdy nie powinno się polegać na jednym wskaźniku, analizę danych powinni przeprowadzać biegli eksperci bibliometrii; w stosunku do metod ilościowych – zwraca się uwagę na potrzebę wzmacniania opinii przez „twarde dane”, a wobec całej ewaluacji – zaleca się łączenie obu podejść<sup>198</sup>. Nieraz przywołuje się badania, które stwierdzają dość silną korelację różnych miar – np. liczby publikacji i liczby cytowań, a także wskaźników bibliometrycznych oraz ocen jakościowych<sup>199</sup>.

### 3.5. Ewaluacja a formy pokrewne

Ewaluację odróżnia się od form pokrewnych: monitoringu, audytu, sprawozdawczości itd. Różnice pomiędzy ewaluacją a niektórymi innymi mechanizmami ujmuje tabela 2.30.

Audyt ma dłuższą tradycję od ewaluacji. Jego celem jest sprawdzenie zgodności z prawem oraz prawidłowości gospodarki finansowej, ustalenie, czy przepisy i procedury były przestrzegane. Ewaluacja jest z reguły bardziej elastyczna.

---

<sup>196</sup> P. Weingart, *Impact of bibliometrics upon the science system: inadvertent consequences?*, „Scientometrics”, January 2005, vol. 62, no. 1.

<sup>197</sup> Na obronę wskaźników bibliometrycznych trzeba jednak powiedzieć, że stale wzrasta wiedza o słabych i silnych stronach bibliometrii, a dzięki rywalizacji pomiędzy Web of Science (Thomson Reuters) oraz Scopus (Elsevier) ciągle ulepszane są indeksy cytowań.

<sup>198</sup> *Performance-based models for publicly funded research in tertiary education institutions* 2009, OECD DSTI/STP/RIHR(2009)7.

<sup>199</sup> A. Andras i B. Charlton, *Evaluating universities using simple scientometric research output metrics: Total citation counts per university for a retrospective seven-year rolling sample*, „Science and Public Policy” 2007, vol. 34 (8), s. 555–563; H.F. Moed, *The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review*, „Science and Public Policy” 2007, vol. 34 (8), s. 575–583, cyt. za: *Performance-based models for publicly funded research in tertiary education institutions* 2009, DSTI/STP/RIHR(2009)7.



Wykonawcami jej są często nie tylko upoważnieni licencjonowani eksperci, ale także interesariusze. Procedury ewaluacji zależą z reguły w większym stopniu od wstępnych ustaleń wykonawców<sup>200</sup>.

Tabela 2.30

### Różnica pomiędzy ewaluacją a innymi mechanizmami opartymi na feedbacku

Różnica pomiędzy ewaluacją a innymi mechanizmami opartymi na feedbacku	
Opracowania naukowe	Ewaluacja kładzie nacisk na praktyczne wykorzystanie informacji.
Tradycyjny audyt	Ewaluacja ocenia wydatki publiczne z szerszego punktu widzenia (bierze także pod uwagę pytania, czy cele programu są właściwe oraz czy zostały skutecznie i efektywnie osiągnięte).
Monitoring	Ewaluację przeprowadza się jako przedsięwzięcie jednorazowe. Podczas ewaluacji poszukuje się bardziej wnikliwych informacji dotyczących badanego programu. Przy tym istnienie systemu stałego monitoringu to często podstawa pomyślnej ewaluacji.
Pomiar dokonań ( <i>Performance indicators</i> )	Ewaluacja bada nieraz oceniany obiekt za pomocą tych samych kategorii (np. jakość, efektywność), ale stara się osiągnąć więcej, próbując znaleźć wyjaśnienie dla osiągniętych wyników i zrozumieć logikę działania publicznego.
Benchmarking	Benchmarking ma węższe znaczenie. Jest jedną z metod oceny organizacji. Ewaluacja może sięgać po tę metodę.
Analiza polityki	Ewaluacja kładzie nacisk na ocenę <i>ex post</i> . Analizę polityczną określa się nieraz jako ewaluację <i>ex ante</i> w badaniu przyszłych opcji politycznych.

Źródło: OECD *Improving Evaluation Practices. Best Practice Guidelines for Evaluation and Background Paper*, PUMA/PAC(99)1, por. Kawalec, s. 13–14.

W przeciwieństwie do audytu ewaluacja jest wykorzystywana bardziej jako narzędzie zarządzania programami niż dla publicznej *accountability*. Mimo tych wszystkich różnic granice między tymi dwiema formami kontroli zacierają się. Tradycyjny audyt finansowy jest często uzupełniany o audyt efektywności. Audyt dokonań (*performance audit*) jest bliski ewaluacji<sup>201</sup>. Monitoring, wyczerpujące i regularne sprawdzanie zasobów, efektów i rezultatów interwencji publicznej, oparty jest na takich źródłach informacji jak raporty, przeglądy, bilanse, wskaźniki itp. Monitoring przeprowadzają najczęściej organizacje odpowiedzialne za inter-

<sup>200</sup> W. Krull, *Current practice and guidelines in a selection of industrialized countries*, [w:] *Evaluation of Research and Development. Current Practice and Guidelines. Synthesis report*, eds. W. Krull, D. Sensi i D. Sotiriou, Luxembourg: Commission of the European Communities Directorate-General Telecommunications, Information Industries and Innovation, 1991.

<sup>201</sup> *Improve Evaluation Practices*, op. cit.; E. Arnold i K. Guy, *Technology diffusion programmes and the challenge for evaluation*, w: *Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997.

wencję publiczną. Podobnie jak ewaluacja, monitoring jest traktowany jako narzędzie doskonalenia działalności<sup>202</sup>. Ewaluacja może korzystać z benchmarkingu. *Benchmark* to po angielsku miara, kryterium. Benchmarking to „miarowanie”. Menadżerowie nieraz rozumieją benchmarking jako technikę bliską „mapowania procesu”, tj. analizowania procesów pracy i ich ulepszania przez porównywanie ze wzorcowymi praktykami procesów pracy w innych organizacjach<sup>203</sup>.

Analiza polityki to ocena bieżącej polityki w porównaniu z możliwymi alternatywami. Produkt analizy politycznej to porada wskazująca na prawdopodobne skutki proponowanych działań: wprowadzenie interwencji A spowoduje takie to a takie konsekwencje. Mówiąc bardziej szczegółowo, celem analizy polityki jest określenie, które z różnych alternatywnych polityk pozwolą najlepiej osiągnąć dany zestaw celów<sup>204</sup>. Analiza polityki jest bliska ewaluacji *ex ante* programów i instrumentów polityki.

### 3.6. Wielowymiarowość ocen

Przejście od „jednej skali” do jednoczesnego stosowania wielu skal zachodzi w wielu dziedzinach, zarówno w diagnostyce, jak i w ewaluacjach. W pomiarze inteligencji powszechnie znane jest porzucenie pojęcia inteligencji definiowanej przez pomiar umiejętności językowych, logicznych i matematycznych (IQ Lewisa Termana) na rzecz pojęcia wielości inteligencji lub inteligencji wielorakiej (koncepty Daniela Golemana oraz Howarda Gardnera), na podstawie których zbudowano odpowiednio zróżnicowane testy.

W dziedzinie ewaluacji jednym z najlepiej znanych przejawów omawianej zmiany jest proponowane przejście od wykorzystywania jednej skali do stosowania wielu skal w rankingach szkół wyższych. Promotorem zmiany jest unijny projekt metodologiczny *Design and Testing the Feasibility of a Multi-dimensional Global University Ranking*<sup>205</sup>. Ze względu na pokrewieństwo z problematyką ewaluacji instytucjonalnej warto w kilku zdaniach opisać (za cytowanym dokumentem) przesłanki i rozwiązania proponowane w wielowymiarowych rankingach uczelni.

Główną przesłanką projektu wielowymiarowego światowego rankingu uczelni jest stwierdzenie, z jednej strony, rosnącej konieczności uzyskania aktualnych informacji na temat szkolnictwa wyższego w skali międzynarodowej,

a z drugiej strony faktu, że systemy szkolnictwa wyższego i badań naukowych

---

<sup>202</sup> Por. *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, vol. 6, European Commission, Luxembourg, 1999.

<sup>203</sup> PUMA/PAC(95)21

<sup>204</sup> Wikipedia 2013.

<sup>205</sup> Podrozdział cytuje fragmenty: *U-Multirank. Interim Progress Report Design Phase of the Project 'Design and Testing the Feasibility of a Multi-dimensional Global University Ranking'*, CHERPA-Network January 2010, [www.u-multirank.eu/UMR\\_IR\\_0110.pdf](http://www.u-multirank.eu/UMR_IR_0110.pdf)

stają się coraz bardziej złożone i (na pierwszy rzut oka) mniej zrozumiałe dla zainteresowanych. Prawo Wymaganego Zróżnicowania Asbhy'ego, stwierdza dalej cytowany dokument, uświadamia, że im bardziej złożone stały się systemy szkolnictwa wyższego, tym bardziej złożony musi stać się nasz sposób patrzenia na nie. Innymi słowy: jeśli proste systemy wyższego kształcenia mogą być ujęte w prostej tabeli rankingowej, obecnie międzynarodowa opinia na temat szkolnictwa wyższego wymaga bardziej skomplikowanych narzędzi. Nie ma sensu porównywać wszystkich instytucji, niezależnie od ich misji, profili i właściwości. Wybór porównywalnych instytucji na podstawie wcześniejszej klasyfikacji musi być podstawą ewaluacji. Zaawansowane narzędzia oceny są zatem bardziej niż kiedykolwiek pilnie potrzebne. Tym bardziej, że rola szkolnictwa wyższego w społeczeństwie zwiększa się, co oznacza, że w powiązania ze szkolnictwem wyższym i badaniami wchodzi więcej interesariuszy, którzy mogą nie znać wystarczająco szkolnictwa wyższego i badań naukowych.

Globalizacja prowadzi do wzrostu presji konkurencyjnej na instytucje szkolnictwa wyższego, czyli tzw. „wyścigu reputacji”, opartego na rankingach, w którym miarą są niemal wyłącznie (łatwiejsze do pomiaru) badania naukowe, stwierdza dalej cytowany raport. Ranking to jednowymiarowa, porządkowa lista, począwszy od „najlepszej” do „najgorszej” uczelni, przypisująca każdej oddzielnej pozycję w równej odległości od siebie (od 1 do np. 500). Większość istniejących rankingów szkół wyższych ma formę tabeli. Rankingi pozwalają porównać uczelnie, porządkując je według stopnia, w jakim spełniają określone kryteria, przede wszystkim oparte na pomiarze liczby publikacji i cytowań oraz liczby Nagród Nobla. Wszystkie istniejące tabele ligowe szkół wyższych uwypuklają znaczenie badań naukowych, ponieważ można je najłatwiej zmierzyć. Inne funkcje uczelni (dydaktyka, „trzecia misja”) i inne charakterystyki (np. orientacja międzynarodowa) nie są uwzględniane. Opierając się na wąskiej bazie danych, rankingi stwarzają jednocześnie wrażenie, że dokonują wszechstronnej oceny jakości całych uczelni. Są to głównie rankingi całych instytucji, tzn. porównujące całe instytucje we wszystkich dziedzinach, z pominięciem ich wewnętrznego zróżnicowania. Rankingi sugerują, że istnieje tylko jeden model uczelni, który się liczy: duży uniwersytet badawczy.

Ma to ujemny wpływ na różnorodność uczelni. Taka jednostronna konkurencja zagraża nie tylko działaniom podejmowanym przez uczelnie w innych dziedzinach, takich jak dydaktyka, innowacje, wkład w rozwój regionalny, kształcenie ustawiczne itp., ale także uczelniom o odmiennych misjach i profilach. Efektem tego nacisku jest wzmocnienie pionowej stratyfikacji („lepszy” lub „gorszy” prestiż lub działania), a nie dywersyfikacji poziomej (różnice w misji instytucjonalnych i profilach). Innymi słowy, wzmacnia się nie tyle różnorodność i specjalizację, ile hierarchię. Stworzenie narzędzi oceny premiujących różnorodność oraz różnych form wybitności szkół wyższych może być krokiem w stronę stworzenia bardziej zróżnicowanej struktury bodźców.

Czasami stosowane w rankingu porównanie całych uczelni może być celowe. Najczęściej jednak szkoły wyższe różnią się bardzo w poszczególnych dziedzinach. Tylko bardzo mała liczba „światowej klasy” uczelni jest dobra we wszystkich dziedzinach. Ranking całej instytucji zaciera te różnice. Najlepsza i najbardziej realistyczna strategia to skoncentrowanie wysiłków w ograniczonej liczbie dziedzin. Nawet te nieliczne uczelnie, które mieszczą się wśród najlepszych w wielu dziedzinach, są lepsze w jednych, a gorsze w innych. A wielu zainteresowanych chciałoby wiedzieć o pozycji szkoły w konkretnej dyscyplinie.

Klasyczny ranking (np. szanghajski, Times Higher Education Supplement) oparty jest na jednym wskaźniku złożonym. Można taki wskaźnik złożony traktować jako „destylację rzeczywistości do formy możliwej do zarządzania”. Ale w tym samym czasie tkwi w nich niebezpieczeństwo nadmiernego uproszczenia skomplikowanych realiów i obliczenia średnich, które wprowadzają w błąd. W nowych rankingach użytkownik korzystający z wersji internetowej może sam określić wagi przypisane do poszczególnych wskaźników. W zasadzie, wyznaczanie wag powinno się opierać na dobrze uzasadnionej koncepcji. W praktyce, rankingi szkół wyższych przyznają wagi, nie korzystając z żadnej szeroko akceptowanej teorii lub dobrych argumentów empirycznych. Ponadto, różni użytkownicy mają odmienne priorytety i preferencje w porównywaniu uczelni. Wskaźnik złożony o stałych arbitralnie wybranych wagach oznacza protekcyjny stosunek do użytkowników rankingu.

### 3.7. Problem specyfiki dyscyplinowej w ewaluacji nauki

Tendencja do zwracania większej uwagi w ewaluacji badań naukowych oraz w zasadach finansowania na „produkty”, „efekty” i „wpływy” oraz szerszego uwzględniania opinii użytkowników pociągnęła za sobą wzrost zainteresowania kwestią odmienności poszczególnych dyscyplin. Do lat 70. miernikiem badań była przede wszystkim ich jakość wyrażająca się liczbą publikacji w wysoko notowanych wydawnictwach, co prowadziło do ujednoczenia sposobu traktowania różnych dyscyplin. Od tego czasu zaczęto stopniowo zwracać coraz większą uwagę na fakt, że poszczególne dyscypliny i pola badań różnią się od siebie, mają odmienne cykle badawcze, „daty ważności”, cele, produkty, efekty, metody komunikacji oraz sposoby interakcji z otoczeniem. Badacze w różnych dyscyplinach inaczej określają jakość badań i ich oryginalność. Uznano, że ewaluację należy projektować oraz przeprowadzać tylko z uwzględnieniem różnic oraz na podstawie ich dogłębnego zrozumienia<sup>206</sup>.

---

<sup>206</sup> *Evaluation of Research in Context. A Quick Scan of an Emerging Field*, 2007, Rathenau Instituut – Department Science System Assessment, [http://www.allea.org/Content/ALLEA/WG%20Evaluating/NL\\_Quick\\_Scan.pdf](http://www.allea.org/Content/ALLEA/WG%20Evaluating/NL_Quick_Scan.pdf); M. Lamont, *How Professors Think: Inside The Curious World of Academic Judgment*, 1st Edition, Cambridge, MA: Harvard University Press, 2009; M. Huang i Yu. Chang, *Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective*, „Journal of the American Society for Information Science and Technology”

Najważniejsze cechy **zróźnicowania dyscyplin naukowych** zależą od następujących cech:

- charakteru wiedzy („twarda” i „mięka”; podstawowa i stosowana; naukowa i inżynierska; teoretyczna, doświadczalna, symulacyjna, aplikacyjna itd.);
- charakteru rozwoju dyscypliny i pola badań (faza w „cyklu życia” obszaru badań oraz tempo postępu badań);
- pozycji względem innych pól i dyscyplin;
- dyscyplinowego, względnie interdyscyplinarnego charakteru;
- nastawienia (teoretyczne, doświadczalne, metodologiczne lub aplikacyjne);
- miejsca w „głównym nurcie badań” lub poza nim;
- stopnia ryzyka i prawdopodobieństwa sukcesu (zakłada się, że im ambitniejsza próba przesunięcia frontu wiedzy, tym większe ryzyko porażki);
- horyzontu czasowego (problemy badawcze różnią się pod względem czasu potrzebnego na ich rozwiązanie; rozwiązanie niektórych może wymagać długich obserwacji, podczas gdy inne mogą być rozwiązane dzięki eksperymentom przeprowadzonym w ciągu jednej doby)<sup>207</sup>.

Tradycyjnie odróżnia się dyscypliny „twarde” i „ściste”, takie jak nauki przyrodnicze i techniczne, oraz „miękkie”, jak humanistyka i nauki społeczne. Nauki przyrodnicze badają zjawiska przyrody, humanistyka i nauki społeczne – zachowania ludzi. Nauki społeczne, a w jeszcze większym stopniu humanistyka, odwołują się do wyobraźni i intuicji, choć z drugiej strony w coraz wyższym stopniu sięgają po techniki analizy statystycznej (np. językoznawstwo). Antropologia i archeologia, badając artefakty, korzystają z dorobku chemii, fizyki i biologii. Nauki przyrodnicze charakteryzuje względnie dobrze rozwinięta struktura teoretyczna, kumulatywność wiedzy, znaczenie pomiaru i eksperymentu. Nauki społeczne i humanistyczne – zainteresowanie tym, co poszczególne, konkretne, jakościowe, większa swoboda w wyborze metody. Wpływ nauk przyrodniczych na społeczne i humanistyczne zmniejsza dzielące je różnice.

Pozycja poszczególnych dyscyplin w sieci nauki zmienia się. „Nowa historiografia” zbliżyła badania historyczne do nauk społecznych. Architektura była niegdyś gałęzią sztuki, aby stać się częścią inżynierii oraz – ostatnio – ponownie dziedziną sztuki.

Niekiedy wyróżnia się jeszcze inne bieguny i mówi o dyscyplinach „kiergentnych”, których cechą mają być stabilne elity, metody i procedury, oraz dyscyplinach „dywergentnych”, tolerujących znacznie większy poziom badawczych odchyleń.

2008, 59(11), s. 1819–1828; J. Kekäle, *Conceptions Of Quality In Four Different Disciplines*, 2004, [www.springerlink.com/index/QQTP97CFBQF9YA44.pdf](http://www.springerlink.com/index/QQTP97CFBQF9YA44.pdf).

<sup>207</sup> Między innymi J. Gläser, G. Laudel, S. Hinze i L. Butler, *Impact of evaluation-based funding on the production of scientific knowledge* 2002, <http://www.laudel.info/pdf/02ExpertiseGlaeLauHinBut.pdf>.

Kiedy indziej wyodrębnia się dyscypliny „miejskie”, zajmujące wąski obszar intelektualnego terytorium (w sensie zagadnień i metod, choć niekoniecznie liczby badaczy), lecz charakteryzujące się intensywnością powiązań i szybkim postępem badań, oraz dyscypliny „wiejskie”, o szerokich, nieokreślonych granicach, długofalowych projektach badawczych i wolniejszym przesuwaniu się frontu badań.

Inne spektrum określa kontekstowość i uniwersalność. „Nauki humanistyczne i społeczne mają charakter bardziej kontekstowy niż nauki ścisłe, techniczne czy nauki o życiu, które ze swej natury mają bardziej uniwersalny charakter. Często badania prowadzone w humanistyce i naukach społecznych odpowiadają na konkretne potrzeby i służą rozwiązywaniu problemów społecznych czy edukacyjnych, a ich efekt stanowią monografie czy podręczniki dotyczące aktualnych polskich problemów społecznych”. Charakteryzuje je „<lokalność> zastosowań i wpływu na odbiorców”<sup>208</sup>.

Jeszcze inne spektrum wyznaczają cele i motywy prowadzonych badań. Kilkanaście lat temu w miejsce taksonomii liniowej badań (podstawowe > stosowane > prace rozwojowe) zaproponowano taksonomię matrycową, uwzględniającą zarówno pobudki, jak i cele badań, lepiej oddającą ich złożoność (tabela 2.31).

Tabela 2.31

## Taksonomia matrycowa badań

Wzgląd na rozwój czystej wiedzy	Wysoki	<b>Kwadrat Nielsa Bohra</b> Czyste badania podstawowe. Chęć zrozumienia praw natury bez względu na zastosowania.	<b>Kwadrat Ludwika Pasteura</b> Badania podstawowe inspirowane przez użytkownika. Chęć zrozumienia praw natury ze względu na zastosowania.
	Niski		<b>Kwadrat Tomasza Edisona</b> Czyste badania stosowane. Wzgląd na zastosowania bez chęci zrozumienia praw natury.
		Niski	Wysoki
Wzgląd na bezpośrednie zastosowania			

Źródło: opracowano na podstawie: D.E. Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation*, Washington, D.C., Brookings Institution Press, 1997.

Ważnym elementem różnicującym dyscypliny i pola badań jest ich faza rozwojowa.

Narodziny, wzrost i spadek zainteresowania dyscypliną lub polem badań można najczęściej przedstawić w formie krzywej S (tabela 2.32).

<sup>208</sup> J. Kekäle, *Conceptions Of Quality In Four Different Disciplines*, 2004, [www.springerlink.com/index/QQTP97CFBQF9YA44.pdf](http://www.springerlink.com/index/QQTP97CFBQF9YA44.pdf).

Tabela 2.32

## Krzywa S

	Faza 1	Faza 2	Faza 3	Faza 4
Kontekst poznawczy	Formułowanie paradygmatu	Nauka normalna. Konstruktywne zastosowania	Spadek produktywności. Spadająca liczba zastosowań	Wyczerpanie
Orientacja metodologiczna	Oryginalność – filozoficzna – programowa	Weryfikacja	Spójność	Apologetyczna – filozoficzna kontrowersja
Piśmiennictwo	Innowacyjne dokumenty, reprinty	Artykuły	Podręczniki, czasopisma dyscyplinowe	Bibliografie
Struktura społeczna	Brak	Niewidzialne kolegia	Formalne grupy i towarzystwa	Grupy rezydualne
Formy instytucjonalne	Nieformalne	Małe sympozja	Kongresy i formalne spotkania	Instytucjonalizacja (wydział uniwersytecki)

Źródło: M. de May, *The cognitive paradigm: an integrated understanding of scientific development*, University of Chicago Press, Chicago 1992, s. 150.

Dostrzeżenie różnic pomiędzy dyscyplinami i polami badań wiązało się z zakwestionowaniem znaczenia „metody naukowej”.

Uczeni postrzegają naukę w sposób zmitologizowany, poprzez tworzone przez nich samych mity, pisze Henry H. Bauer w książce *Scientific Literacy and the Myth of the Scientific Method* (1994). Najważniejszy to mit „metody naukowej”, który mówi, że badania naukowe polegają na stosowaniu jednej ogólnej metody w różnych dyscyplinach i specjalnościach. Metoda ta to systematyczne, kontrolowane obserwacje lub doświadczenia, których wyniki prowadzą do hipotez, falsyfikowanych w toku dalszej pracy, prowadzących z kolei do wiarygodnych teorii. Fakty dowodzą jednak, że często doświadczenia nie są w stanie potwierdzić prawidłowych obliczeń teoretycznych. Jak pisał Artur Eddington, dobrze nie pokładać zbyt wielkiego zaufania w wynikach obserwacji, dopóki nie zostaną one potwierdzone przez teorię.

Mit metody naukowej zakłada jednorodność badań naukowych, pisze autor. Jednak to, co rozumiemy pod słowem „nauka”, to zbiór bardzo różnych typów działalności. Nie tylko chemicy, fizycy, geolodzy, biologowie lub astronomowie to odrębne „plemiona”, porozumiewające się różnymi językami, mające różny stosunek do matematyki czy rozważań spekulacyjnych, ale także każda dyscyplina to tylko luźna federacja różnych „plemion”: plemię chemików dzieli się jeszcze na szczepy analityków, organików, nieorganików, chemików fizycznych, elektrochemików, chemików teoretycznych czy chemików polimeru. Ponadto,

poprzez ten podział przebiega jeszcze dalszy, na chemików teoretyków i chemików eksperymentalistów.

Naukowcy różnią się pod każdym możliwym względem: częstotliwości publikacji, sposobu korzystania z pracy studentów i techników, podejścia do czasu (czas zmian w kosmosie astronoma jest różny od czasu przebiegu ewolucji istot żywych ewolucjonisty, czasu przekształceń geologicznych geologa lub czasu, w jakim zachodzą zjawiska badane w doświadczeniach fizycznych). Różnice pomiędzy przedstawicielami różnych dyscyplin wykraczają poza problemy teorii, metody i słownictwa i dotyczą także rodzaju wyobraźni, a nawet cech kulturowych<sup>209</sup>.

Najłatwiej wskazać, w jaki sposób różnice pomiędzy dyscyplinami powinny wpłynąć na oceniane produkty. I tak, charakterystyczne produkty badań ICT to m.in. oprogramowania, patenty, licencje, nowa aparatura naukowa; produkty badań inżynierskich to m.in. patenty, prototypy, raporty techniczne, a nawet *know-how*; produkty architektury to m.in. budynki, projekty, prototypy, instalacje, wystawy, oprogramowania, a także np. nowy język kodu w projektach. Od rodzaju dyscypliny zależy okres cytowań – od krótkiego (genetyka) po bardzo długi (paleontologia, systematyka) – oraz ich intensywność<sup>210</sup>; w niektórych naukach, jak nauki społeczne, a szczególnie humanistyczne, ocena cytowań powinna być stosowana z najwyższą ostrożnością. Międzynarodowe indeksy cytowań mogą być wspomagane przez krajowe indeksy (np. Arton – Polska Literatura Humanistyczna<sup>211</sup>).

Trudniej określić, w jaki sposób różnice pomiędzy dyscyplinami powinny wpłynąć na inne elementy ewaluacji. Nie sposób zbudować solidnego systemu ewaluacji instytucjonalnej bez rozległej wiedzy o specyficznych cechach poszczególnych dyscyplin.

Pomijając sztuki piękne, humanistykę i nauki społeczne, najważniejsza różnica, która powinna zostać wzięta pod uwagę przy projektowaniu nowego systemu ewaluacji, dotyczy różnic pomiędzy badaniami naukowymi a technologicznymi.

Rozróżnienie pomiędzy nauką a technologią nie jest stałą niezmienną, niezależną od epok, dyscyplin i grup technologicznych oraz wyznawanych poglądów<sup>212</sup>. Od lat budzi ono zainteresowanie badaczy. Przeważa pogląd, że są to

---

<sup>209</sup> H.H. Bauer, *Scientific literacy and the myth of the scientific method*, University of Illinois Press, Urbana 1992.

<sup>210</sup> Badania podstawowe, w tym systematyczne i paleontologiczne, mają długą «datę ważności». Ich wyniki są często cytowane przez wiele lat (często nawet kilkadziesiąt), ponieważ nie dezaktualizują się tak szybko jak np. badania molekularne czy z biologii doświadczalnej. Dlatego też, aby wyrównać szanse badań podstawowych z jednej strony oraz badań molekularnych i doświadczalnych z drugiej strony, należy brać pod uwagę długoletnie okresy cytowalności w przypadku badań podstawowych.

<sup>211</sup> <http://arton.bg.us.edu.pl/arton/arton.php>.

<sup>212</sup> Wielu pionierów nowożytnej nauki podkreślało jej technologiczny aspekt (Francis Bacon). Wielokrotnie w dziejach dokonywano konwersji zagadnień technicznych na język nauki, robił to Arystoteles oraz rzesze badaczy, poczynając od XVII wieku.



odrębne, na wpół autonomiczne dziedziny, jednak od XIX wieku silnie wzajemnie ze sobą powiązane, zwłaszcza w pewnych fazach rozwoju technologii (np. w początkach technologii opartych na nauce, jak np. elektronice, lub też w fazach późniejszych, gdy dojrzałe technologie przechodzą okres rewolucji naukowej, jak stało się z biotechnologią) lub dyscypliny naukowej (np. w początkach dyscyplin opartych na technologiach, jak np. termodynamika, lub też w fazach zaawansowanych, gdy dyscyplina dojrzewała do zasilania technologii, jak stało się z fizyką i chemią w II połowie XIX wieku, a z biologią w II połowie XX wieku). Łącznikiem między obiema dziedzinami jest bardzo często aparatura<sup>213</sup>.

Nauka i technologia różnią się przede wszystkim pod względem celu i orientacji. Celem nauki jest poznanie i zrozumienie świata, podczas gdy w technologii cele te są podporządkowane praktycznym, takim jak np. zabijanie albo zapobieganie chorobom, kontrola natury i społeczeństwa poprzez produkcję artefaktów oraz stosowanie narzędzi intelektualnych. Ponadto, różnią się one pod względem cech poznawczych. Prace technologiczne skupione są wokół projektowania; naukowe – poznawania. W technologicznych większą rolę odgrywa praktyczna intuicja i doświadczenie, w naukowych – świadoma analiza i kalkulacja<sup>214</sup>.

Wzrost świadomości różnic pomiędzy dyscyplinami naukowymi przebiegał równoległe z umysławianiem sobie wielowymiarowości form wyników badań i prac rozwojowych. Uznano, że wyniki B+R to nie tylko wiedza skodyfikowana (publikacje, raporty techniczne, patenty), a także (rysunek 2.8).

Z kolei umysłowanie wielowymiarowości B+R łączyło się ze zwróceniem większej niż dotąd uwagi na wielokierunkowość misji instytutów badawczych (rysunek 2.9).

Wiązało się także z częstszym niż dotąd podkreśleniem wag i znaczeniem tzw. trzeciej misji uczelni (obok dydaktyki i badań naukowych) (rysunek 2.10).

Jednocześnie wskazano również na wieloznaczność najważniejszych istotnych cech badań naukowych, jak m.in. ich oryginalności. Pojęcie oryginalności uznano za wieloznaczne, gdyż jest ono stosowane w odniesieniu do rozmaitych atrybutów badań (rysunek 2.11).

---

<sup>213</sup> L. Velho, *Knowledge frontiers: public sector research and industrial innovation in biotechnology, engineering ceramics, and parallel computing*, Clarendon Press, New York, Oxford University Press, Oxford 1995, cyt. za: W. Faulkner, *Conceptualizing Knowledge Used in Innovation: A Second Look at the Science-Technology Distinction and Industrial Innovation*, „Science, Technology, & Human Values”, Autumn 1994, vol. 19, Issue 4, s. 425–458.

<sup>214</sup> W. Faulkner, *Conceptualizing Knowledge Used in Innovation*, op. cit.; A. Cordero, *On the growing complementarity of science and technology*, „PHIL & TECH” 4:2 Winter 1998, <http://scholar.lib.vt.edu/ejournals/SPT/v4n2/pdf/CORDERO.PDF>.

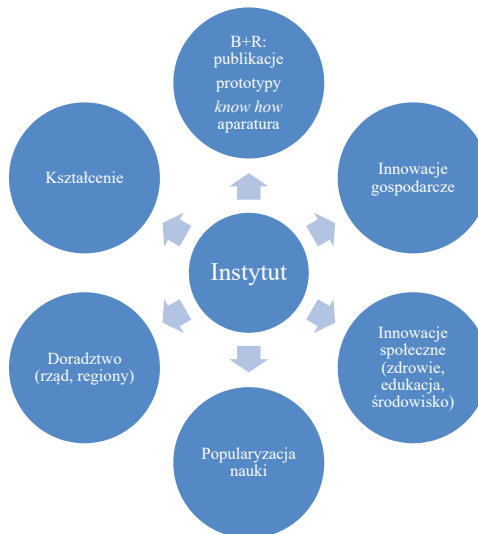
Rysunek 2.8



Źródło: A.J. Salter i B. Martin, *The Economic Benefits of Public Funded Basic Research: A Critical Review*, „Research Policy” 2001, 30, no. 3, s. 509–532.

Rysunek 2.9

### Wielokierunkowość misji instytucji badawczych



Źródło: A. Rip i H.C. Marais, *Assessing University Research under Condition of Changing Knowledge Production*, SAUVACA Publication Series 1998/2, s. 14.

Rysunek 2.10

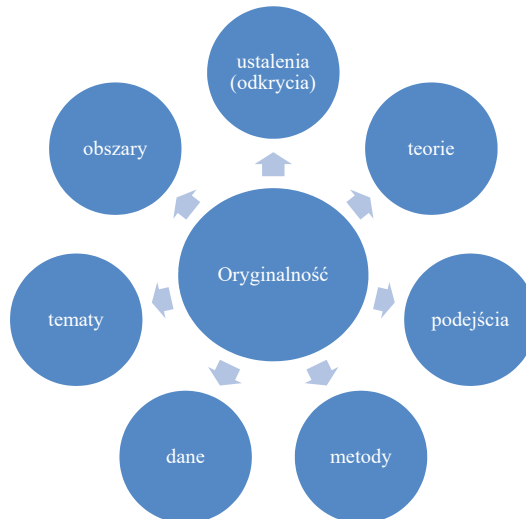
## Trzecia misja uczelni



Źródło: A.M. Inzelt, *Towards Indicators on University Third Mission*, PRIME – PhD Winter School, Rome, 5–9 November 2007.

Rysunek 2.11

## Oryginalność jako atrybut badań



Źródło: J. Guetzkow, M. Lamont i G. Mallard, *What Is Originality in the Humanities and the Social Sciences?*, „American Sociological Review”, April 2004, vol. 69, no. 2.

### 3.8. Ewolucja ewaluacji nauki

„Powojenna ekspansja B+R finansowanych ze źródeł publicznych odbywała się zrazu na podstawie modelu innowacji, który podkreślał raczej rolę «wkładu» niż «rezultatu» i wskutek tego traktował innowacje jako «czarną skrzynkę»», piszą Amir Piric i Neville Reeve<sup>215</sup>. Uzasadnienie inwestycji publicznych w B+R wahało się od podkreślania ich funkcji prestiżowej i strategicznej (militarnej) do kładzenia nacisku na konkretne projekty „wielkiej skali”. Niewiele jednak robiono dla tak dziś zdawałoby się oczywistych kwestii, jak metody ustalania priorytetów, mechanizmy podziału funduszy czy też pomiar ich efektów i szerszego oddziaływania. Prowadzone od lat 50. badania nad rolą technologii we wzroście gospodarczym oraz nad charakterem zmiany technologicznej jako głównego czynnika wzrostu uchyliły nieco wieka tej czarnej skrzynki. Wskutek presji finansowej (od lat 70.) oraz radykalnej zmiany roli i miejsca sektora publicznego (od lat 80.) argumenty mówiące, dlaczego rządy powinny inwestować w badania naukowe, jak powinny to robić oraz co społeczeństwo powinno uzyskiwać w zamian za zgodę na te inwestycje, stały się znacznie ważniejsze niż dawniej. Badania naukowe były, po raz pierwszy na taką skalę, przedmiotem oceny z punktu widzenia efektywności, skuteczności, przejrzystości itd. To z kolei pociągnęło za sobą poddawanie ich ewaluacjom, mającym na celu określenie zarówno ich kosztów, jak i korzyści<sup>216</sup>.

Ewaluacja badań naukowych i technologicznych po II wojnie światowej odzwierciedla dwa zjawiska: zmiany postrzegania funkcji nauki i jej powiązań z techniką oraz zmiany sposobu działania organizacji rządowych i całego sektora publicznego.

Choć inwestycje publiczne w B+R po roku 1945 miały stosunkowo duży udział w budżetach państw rozwiniętych, problem badania społeczno-gospodarczego wpływu B+R dla uzyskania przesłanek co do decyzji jest znacznie świeższej daty.

W czasie II wojny światowej osiągnięcia nauki uświadomiły jej ogromną rolę. Lata 50. i 60. to lata wielkiego rozwoju badań naukowych prowadzonych na uniwersytetach i w laboratoriach rządowych. U podstaw tego rozwoju leżała wiara w sprawczą rolę nauki dla powojennej odbudowy i wzrostu gospodarczego. Decyzje polityczne w stosunku do nauki opierano na liniowym modelu innowacji. Konsekwencje tego modelu były proste: chcesz więcej innowacji, finansuj więcej nauki<sup>217</sup>. Nawet laboratoria rządowe były wówczas na ogół autonomiczne w swych decyzjach co do badań naukowych. Badacze żywili przekonania, że badania nie powinny podlegać pomiarowi, gdyż są one niemierzalne. W okresie

---

<sup>215</sup> A. Piric i N. Reeve, *Evaluation of public investment in R&D – towards a contingency analysis w: Policy evaluation in innovation and technology. Towards best practices*, 1997 OECD.

<sup>216</sup> Ibidem.

<sup>217</sup> L. Georghiou, *Research evaluation in European national science and technology system*, „Science & Public Policy” 1995, vol. 5, no. 1.

„oddolnej” nauki publicznej ocenę i autoryzację badań pozostawiano samym laboratoriom. W rezultacie, przede wszystkim zajmowały się one badaniami podstawowymi, publikowały, organizowały konferencje i sporządzały raporty techniczne. Nie więcej niż 15% laboratoriów było aktywnych rynkowo. Gdy powstały warunki zachęcające do komercjalizacji prac, okazało się, nie były one wcale magazynem gotowych technologii<sup>218</sup>.

W dekadach tych procedury ewaluacji oparte były na założeniu, że nauka jest (w mniejszym lub większym stopniu) „oderwana” od społeczeństwa. Jest to odrębna działalność społeczna z własnymi normami i wartościami; jej uprawianie i ocena nie powinny być krępowane przez ingerencję z zewnątrz. Stąd nacisk na ocenę „jakości naukowej” badań. Skoro nauka to odrębna działalność społeczna, to powiązania pomiędzy nauką a społeczeństwem mają charakter jednokierunkowy (od nauki do społeczeństwa). Założenia te zakładały zatem rozdział „nauki” i „społeczeństwa”. Wobec tego procedury ewaluacji miały odzwierciedlać tę dychotomię (np. „wewnętrzne” i „zewnętrzne” kryteria oceny)<sup>219</sup>. Kryteria ewaluacji badań oraz jej przeprowadzanie pozostawiono samemu środowisku naukowemu. *Peer review* (czyli ocena środowiskowa) było podstawowym mechanizmem selekcji w dyscyplinach naukowych i całej nauce. Ocen dokonywano na podstawie kryterium kreatywności, innowacyjności i głębi metodologicznej, a nie znaczenia społecznego<sup>220</sup>.

Lata 70. to początki presji, by poddać środowiska naukowe ocenie publicznej. Państwo zaczęło przejawiać większą inicjatywę w stosunku do nauki<sup>221</sup>. Kryteria ewaluacji, ewaluatorzy oraz procedury ewaluacji przestały być wyłączną domeną badaczy i zostały otwarte na negocjacje z politykami, urzędnikami ministerialnymi, ekonomistami, ekspertami zarządzania. Ewaluacja przebiegała jednak nadal wewnątrz organizacji badawczych, takich jak wyższe uczelnie, oraz była przeprowadzana przez samych badaczy<sup>222</sup>. Tradycyjna ewaluacja *peer review* została uzupełniona o inne formy, np. o ewaluację efektywności organizacji naukowych. Kryteria nienaukowe (społeczne i ekonomiczne) zostały włączone do ewaluacji. „Zmodyfikowane *peer review*” (*modified peer review*) oraz „ocena wartości” (*merit review*) wykorzystywały *peer review* do stawiania szerszych pytań dotyczących wpływu społeczno-gospodarczego podejmowanych działań<sup>223</sup>. Źródłem zmian w ewaluacji badań w latach 70. były nowe trendy w administracji

<sup>218</sup> M. Papadakis, *The Economic Impacts of Public Science Can Be Measured*, „The Scientist” October 27, 1997, s. 8.

<sup>219</sup> J. Spaapen i Ch. Sylvian, *Assessing the value of research for society*, „Research Evaluation” 1993, nr 2.

<sup>220</sup> Ibidem.

<sup>221</sup> A. Rip i H.C. Marais, *Assessing university research under conditions of changing knowledge production*, Pretoria 1998.

<sup>222</sup> L.F. Frederiksen, *The Agora and the Role of Research Evaluation*, Copenhagen Business School, Internet.

<sup>223</sup> L. Georghiou, *Research evaluation in European national science and technology system*, „Science & Public Policy” 1995, vol. 5, no. 1.

publicznej (nowe zarządzanie publiczne). Celem ewaluacji była teraz kontrola wydatków publicznych (lub stwarzanie wrażenia, że wydatki są pod kontrolą)<sup>224</sup>.

W latach 80. opisane tendencje uległy pogłębieniu. W tych to latach rozwinęto na dużą skalę systematyczną ewaluację działań rządu na rzecz rozwoju badań naukowych i technologicznych<sup>225</sup>. Postęp w badaniach i jakość badań nie były teraz traktowane jako cele same w sobie, lecz jako narzędzia osiągnięcia celów społecznych<sup>226</sup>.

Motorami zmian były nadal zmiany administracji publicznej oraz zmiany funkcji N+T w społeczeństwie. Wielką rolę w upowszechnieniu ewaluacji nauki odegrało nowe zarządzanie publiczne (*New Public Management*), kładące nacisk na wskaźniki dokonań i na ewaluacje.

Na zmiany charakteru ewaluacji badań naukowych i technologicznych miały też wpływ przemiany zachodzące w charakterze innowacji. W związku z upowszechnianiem nowego paradygmatu techniczno-gospodarczego, opartego na ICT, w latach tych wzrosło rynkowe ukierunkowanie badań i nasiliły się powiązania nauki i przemysłu<sup>227</sup>.

Zmianom zachodzącym w procesach gospodarczych towarzyszyły zmiany w koncepcjach ekonomicznych, w koncepcjach polityki naukowo-technicznej oraz w metodologii ewaluacji. Na przykład w Niemczech w latach 80. polityka N+T, która dotąd koncentrowała się na wspieraniu badań i prac rozwojowych, zwróciła większą uwagę na późniejsze fazy procesu innowacji, takie jak uruchomienie produkcji, wprowadzenie produktu na rynek, dyfuzja technologicznych innowacji, usługi oraz infrastruktura wspierająca innowacje. W rezultacie zmienił się z czasem charakter ewaluacji, które z początku dotyczyły jedynie B+R, a potem w coraz większym stopniu innych aspektów innowacji<sup>228</sup>.

Wszystkie te zmiany wiązały się z kolei z przemianami sposobu widzenia działalności innowacyjnej. W latach 50. zgodnie z modelem „technologicznego pchnięcia” sądzono, że to podaż nowych (opartych na badaniach naukowych) technologii jest głównym źródłem zmiany technologicznej. Później kładziono coraz większy nacisk na rolę rynku w innowacjach. Doprowadziło to do nowego modelu innowacji, „ciągnięcia przez potrzeby”. W połowie lat 80. przedstawiono

---

<sup>224</sup> L.F. Frederiksen, *The Agora and the Role of Research Evaluation*, op. cit.; M. Power, *The Audit Society. Rituals of Verification*, Oxford University Press, Oxford 1997.

<sup>225</sup> T. Luukkonen i B. Stähle, *Evaluation of research fields. Scientists' Views*, 1993; T. Luukkonen, *The increasing professionalisation of the evaluation of mission-oriented research in Finland: implications for the evaluation process w: Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997; F. Meyer-Krahmer i Th. Reiss, *Ex ante evaluation and technology assessment – two emerging elements of technology policy evaluation*, „Research Evaluation”, April 1992, vol. 2, no. 1.

<sup>226</sup> L.F. Frederiksen, *The Agora and the Role of Research Evaluation*, op. cit.

<sup>227</sup> J. Spaapen i Ch. Sylvian, *Assessing the value of research for society*, „Research Evaluation” 1993, no. 2, s. 117–126.

<sup>228</sup> F. Meyer-Krahmer i Th. Reiss, *Ex ante evaluation and technology assessment – two emerging elements of technology policy evaluation*, op. cit.

argumenty przemawiające za sprzężeniami pomiędzy nauką, techniką i rynkiem. Model sprzężeń zawierał w sobie mechanizmy pchania i ciągnięcia łączącego naukę i rynek poprzez stadia inżynierii, rozwoju technologicznego, produkcji, marketingu i sprzedaży<sup>229</sup>. Model sprzężeń stał się inspiracją zarówno dla strategii rozwoju nauki i techniki, jak i ich ewaluacji. Pod wpływem nowych koncepcji od połowy lat 80. w istotny sposób ulepszono metodologię pomiaru ekonomicznych korzyści B+R (czyli wpływ B+R na tworzenie dochodu narodowego). Zwrócono zwłaszcza większą uwagę na pośrednie korzyści z B+R, a w szczególności na pomiar korzyści kompetencyjnych (wykorzystywania w gospodarce kompetencji wykształconych w trakcie przeprowadzania badań).

### 3.9. Modernizacja administracji rządowych jako źródło zmian ewaluacji nauki

Wprowadzenie systematycznej ewaluacji programów i instrumentów politycznych w sferze N+T oraz organizacji badawczych, finansowanych z budżetu państwa, jest następstwem modernizacji administracji rządowych oraz przemian funkcji badań naukowych i technologicznych, jakie zaszły w II połowie XX wieku.

W ostatnich dwóch dekadach w krajach OECD punkt ciężkości procesu budżetowego i zarządzania (oba procesy były ze sobą powiązane) przesunął się od pomiaru „wkładu” (*input*) do pomiaru produktu (*output*), efektu (*outcome*) i wpływu (*impact*). W praktyce procesu budżetowego (projektowania, realizacji i oceny budżetu) nastąpiło przejście od budżetu jako środka kontroli legalności i celowości wydatków publicznych do budżetu jako narzędzia wspierania zarządzania i efektywności realizacji programu sektora publicznego. Reformy budżetu państwa przeprowadzono np. w Australii (1985), Austrii (1987), Danii (1983), Finlandii (1977 i 1990)<sup>230</sup>. Zmiany te dalekie są od zakończenia i następują w różnym tempie w różnych krajach.

Zmiany systemu budżetu wiązały się z ogólną reformą administracji publicznej, wprowadzaniem zarządzania zorientowanego na dokonania (*performance management*). Często zmiany te podsumowuje się jako przejście od rządzenia (*governing*) do współzarządzania (*governance*). Angielskie słowo *governance*, które robi zawrotną karierę w naukach politycznych i w naukach o zarządzaniu, nie ma dobrego polskiego odpowiednika. Brak polskiego odpowiednika<sup>231</sup> (tak jak i brak np. odpowiednika dla *accountability*) sam w sobie jest charakterystyczny, świadcząc, że w Polsce przeważają jak dotąd tradycyjne układy zachowań społecznych i organizacyjnych. *Governance* opisywano i oceniano

---

<sup>229</sup> L. Georghiou, *Research evaluation in European national science and technology system*, op. cit.

<sup>230</sup> *Budgeting for results*, OECD 1995.

<sup>231</sup> Proponuje się ostatnio „współzarządzanie” jako polski odpowiednik „governance”. Por. J. Hausner, B. Jessop i S. Mazur, *Governance Wybór tekstów klasycznych*, Scholar, Warszawa 2017.

z wielu perspektyw badawczych. Pojęcie to powstało w badaniach urbanistycznych w kontekście skutecznego sposobu radzenia sobie z rozwiązywaniem problemów na szczeblu lokalnym. Później rozszerzono jego znaczenie, tak by objęło nawet „rządzenie” sprawami globalnymi. Powstanie jego było świadectwem kryzysu rządzenia. Jak ujął to Daniel Bell, zorientowano się, że państwo jest zbyt wielkie dla małych problemów i zbyt małe dla wielkich. Rządy państw, rządy regionalne i lokalne oraz władze związków państw (Unia Europejska) i organizacji międzynarodowych, takich jak Bank Światowy, dokonały reformy procesu budżetowego i zarządzania pod presją okoliczności. Dowodzą, że tradycyjny rząd, z charakterystycznym dla siebie naciskiem na hierarchię, pionowy i odgórny przepływ decyzji, biurokrację i interwencjonizm, jest źle dostosowany do szybko zmieniającego się środowiska gospodarczego, społecznego i kulturalnego, w którym działa coraz więcej autonomicznych sił (np. koncerny), aktorzy sceny publicznej są coraz gęściej powiązani ze sobą w sieci, zagadnienia stają się coraz bardziej złożone, a żądania partycypacji obywatelskiej – coraz bardziej donośne<sup>232</sup>.

Reforma jest konieczna jako reakcja na stopniową utratę przez rządy monopolu władzy. Na szczeblu międzynarodowym rząd pozbywa się części swoich uprawnień w zamian za większe poczucie bezpieczeństwa, współpracy i ekonomii skali. Na szczeblu krajowym oddaje władzę różnym grupom interesu, w miarę jak spada skuteczność jego działania oraz w miarę jak zmiana technologiczna (np. Internet) daje obywatelom więcej możliwości współdecydowania. Rząd często odkrywa, że pozbycie się części władzy zwiększa efektywność jego działań. Formy cedowania władzy obejmują przesuwanie uprawnień na tzw. *quango* (*quasi-autonomous non-governmental organizations*), tworzenie *quasi*-rynków (rząd płaci za usługi, ale dostarczają je wyłonione w przetargach organizacje), prywatyzację, partnerstwo publiczno-prywatne oraz sieci polityczne. Ta ostatnia forma świadczy, że wśród rosnącej grupy podmiotów mających aspiracje i możliwości działania politycznego (już nie tylko związki zawodowe i stowarzyszenia pracodawców, ale także organizacje pozarządowe, wspólnoty lokalne, grupy ekologiczne itd.) rząd ma działać nie jak „król zwierząt”, tylko jako „ułatwicz” i katalizator rozmaitych przedsięwzięć (szukać konsensusu, prowadzić konsultacje, koordynować, sugerować i współzawodniczyć)<sup>233</sup>.

Reformy administracji publicznej pobudzone były przez ideologię neoliberalną oraz stymulowane koncepcyjnie przez OECD (PUMA) i inne organizacje międzynarodowe (UNDP, UNESCO, Bank Światowy, Międzynarodowy Fundusz Walutowy). Poszły one w wielu kierunkach, np. modelu rynkowego (rozległa prywatyzacja i zmniejszenie sektora publicznego), państwa partycypacyjnego, kładącego nacisk na uczestnictwo obywateli w rządzeniu, czy też rządu ela-

---

<sup>232</sup> A. Kazancigil, *Governance and science: market-like modes of managing society and producing knowledge*, „International Social Science Journal”, March 1998.

<sup>233</sup> *Government of the Future*, OECD 2001.



stycznego (takiego, który da szansę urzędnikom o twórczych pomysłach i silnej motywacji)<sup>234</sup>. Nurt reformy zarządzania publicznego (*Public Management Reform*) zrodził kilka szeroko komentowanych i tłumaczonych na różne języki publikacji, jak *Rządzić Inaczej* oraz tzw. raport Al Gore'a pt. *Tworząc rząd, który pracuje lepiej i kosztuje mniej: od biurokracji do rezultatów* (1993, oryg. *Creating a Government that Works Better and Costs Less: From Red Tape to Results*) oraz kilka raportów organizacji międzynarodowych, ogłaszanych drukiem od połowy lat 90. Wszystkie te prace (nie negując wcześniejszych zasad – „rządów prawa” i „służby społecznej”) kładły nacisk na takie wartości, jak wydajność i skuteczność, *accountability*, przejrzystość, decentralizacja, współuczestnictwo, prywatyzacja funkcji rządowych. Remedium na spadek sprawności działania administracji publicznych miało być przesunięcie punktu ciężkości z hierarchii na powiązania poziome, partnerstwo, szukanie konsensusu i negocjacje<sup>235</sup>.

W formie uproszczonego schematu można przeciwieństwa pomiędzy starym i nowym systemem ująć w następujący sposób (tabela 2.33).

Tabela 2.33

## Stary i nowy system zarządzania publicznego – różnice

Stary system	Nowy system
Stabilne i proste środowisko	Szybko zmieniające się i złożone środowisko
Rozmiary / skala	Szybkość / responsywność
Szttywność	Plastyczność
Ukierunkowanie na proces	Ukierunkowanie na rezultaty
Konsensus	Konstruktywna debata
Hierarchia	Przywództwo
Legalizm	Zarządzanie
Zgodność z przepisami	Efektywność
Polityka responsywna	Polityka aktywna i inicjująca
Zasady i procedury	Rezultaty
Nakazowo-kontrolny styl zarządzania	Techniki zaczerpnięte z biznesu
Zarządzanie odgórne	W większym stopniu zarządzanie oparte na współpracy i pracy zespołowej
Nadzór	Wskaźniki dokonań

<sup>234</sup> Ch. Pollit, *Public Management Reform*, „OECD Journal on Budgeting” 2003, vol. 3, no. 19, Issue 3.

<sup>235</sup> A. Kazancigil, *Governance and science: market-like modes of managing society and producing knowledge*, „International Social Science Journal”, March 1998.

Tabela 2.33 – cd.

Stary system	Nowy system
Szczegółowe sztywne przepisy i kontrola finansowa	Zarządzanie dla rezultatów i kontrola oparta na efektywności
Kontrola wkładu ( <i>input</i> ), <i>ex ante</i>	Pomiar rezultatów i efektów ( <i>outcome</i> , <i>output</i> )
Nacisk na proces	Nacisk na rezultaty
Zarządzanie i proces budżetowy skupione wokół wkładu	Zarządzanie i proces budżetowy skupione wokół rezultatów
Zarządzanie i proces budżetowy skupione wokół zakresu i skali	Zarządzanie i proces budżetowy skupione wokół szybkości i responsywności
Hierarchiczny, nakazowo-kontrolny model zarządzania	Przywództwo, szukanie ugody, konsultacje, koordynacja, dialog, negocjacje
Monopol	Współzawodniczenie z innymi centrami władzy
Sztwywność	Plastyczność
Zarządzanie kierowane przez procedury	Zarządzanie kierowane przez rezultaty
Zarządzanie zintegrowane werkalnie	Zarządzanie zintegrowane wirtualnie
Urzędnicy	Menadżerowie i przywódcy
Zgodność z obowiązującymi przepisami i procedurami	Skuteczność i efektywność, odpowiedzialność za rezultaty
Finansowanie budżetowe jako dystrybucja	Finansowanie budżetowe jako inwestycja

Źródło: opracowane na podstawie publikacji OECD.

Rozrost sektora publicznego, zmiana struktury wydatków publicznych (np. zwiększenie udziału „twardych” wydatków na służbę zdrowia, spowodowanych starzeniem się społeczeństw), zaniepokojenie wzrostem zadłużenia budżetowego oraz skutkami nierównowagi budżetowej wpłynęły na zmianę w podejściu rządów do wydatków publicznych. W późnych latach 70. i we wczesnych latach 80., po załamaniu się powojennej prosperity, zrozumiano potrzebę wprowadzenia nowego systemu budżetu, który pozwalałby na lepsze równoważenie celów w stosunku do możliwości finansowych rządów i lepiej łączyłby proces budżetowy z innymi funkcjami kierowniczymi. Tę nową formę budżetu nazwano *performance budget*, *performance-based budgeting* (w języku polskim: budżet zadaniowy).

Tradycyjny roczny budżet faworyzował wydatki krótkoterminowe. Skłaniał do tego, by brać pod uwagę wyłącznie skutki finansowe decyzji alokacyjnych. Środki przydziela nie ze względu na wagę celów i zadań oraz wnioski płynące z analizy politycznej i ewaluacji, tylko z punktu widzenia organizacji wykonawczych. Decyzje co do podziału funduszy opierały się nie na np. ocenie krańcowej użyteczności alternatywnych opcji ich wykorzystania, ale na histo-

rycznych podziałach kwot oraz na presji interesariuszy. Organizacje wykonawcze wywierały nieustanny nacisk na budżet, dyktowany bardziej potrzebą zwiększenia ich własnego znaczenia niż koniecznością realizacji ważnych zadań<sup>236</sup>. Sama tylko konstrukcja budżetu wystarczyła np. do wytłumaczenia, dlaczego instytucje naukowe wyciągały stale ręce pod dodatkowe środki. Działo się tak m.in. dlatego, że szansa stworzenia czegoś nowego zależała od nowych funduszy, gdyż budżet tylko z największym trudem znosił redystrybucję funduszy rozdzielonych w przeszłości<sup>237</sup>. Także sama tylko konstrukcja budżetu wyjaśniała, dlaczego w krajach słabiej rozwiniętych instrumenty polityczne (np. stosowane do rozwoju nauki) były z reguły proste, nieodróżnicowane, statyczne oraz pasywne<sup>238</sup>.

Tworząc tradycyjny budżet, zakładano, że każdy dotychczasowy wysiłek powinien być kontynuowany, a organizacja utrzymana, nawet gdy oddala się od swojej misji i marnotrawi publiczne środki (jeśli tylko formalnie pozostaje w zgodzie z przepisami księgowania). Gdy brakowało celów jako punktów odniesienia, nie sposób było w praktyce proponować alternatywnych rozwiązań podziału środków. W takiej sytuacji zapisy budżetowe okazywały się z reguły repetycją rozwiązań przyjętych w przeszłości. Brak elastyczności pociągał za sobą szczególnie oplakane skutki w tak dynamicznie rozwijających się dziedzinach, jak rozwój nauki i techniki. Kontrola realizacji konwencjonalnego budżetu miała wyłącznie formę audytu księgowości. Było to związane z faktem, że agencje wykonawcze i sami wykonawcy nie byli w stanie odnieść swoich wydatków do żadnego celu. W tych warunkach możliwa była tylko wąsko księgowo-prawnicza ocena wyników. Rząd nie był zatem w stanie utrzymywać merytorycznej (a nie czysto formalnej) kontroli nad wydatkami. Sztywność budżetu w zderzeniu ze złożonymi i szybko zmieniającymi się problemami rodziła pokusę tworzenia procedur koordynacyjnych oraz funduszy specjalnych. Często jednak systemy koordynacji były zbyt słabe, by oprzeć się naciskowi wykonawców, a fundusze specjalne stwarzały szerokie pole do nadużyć oraz do niekontrolowanego wpływu pieniędzy budżetowych<sup>239</sup>.

Proste porównanie chronologii wprowadzania nowych zasad z wprowadzaniem nowych form ewaluacji B+R (koniec lat 70. i lata 80.) ujawnia zadziwiająca zbieżność. Także zaskakujące jest porównanie terminologii nowego zarządzania i procesu budżetowego z terminologią stosowaną w ewaluacji społeczno-gospodarczej (np. *input*, *output*, *outcome*, *impact*). Nie są to zbieżności przypadkowe, tylko przyczynowe.

---

<sup>236</sup> *Manual on the budgeting of scientific and technological activities. Integrating the 'science and technology' function in the general State budget*, UNESCO 1984.

<sup>237</sup> H. Vessuri, *Science, politics, and democratic participation in policy-making: a Latin American view*, „Technology in Society” 2003, vol. 25, no. 2, s. 265–266.

<sup>238</sup> F.R. Sagasti, *The science and technology policy instruments project*, „Science and Public Policy” 1979.

<sup>239</sup> *Manual on the budgeting of scientific and technological activities. Integrating the 'science and technology' function in the general State budget*, UNESCO 1984.

Skoro jednak ministerstwa i agencje rządowe są nie po to, aby wydawać pieniądze, należy wymagać od nich, aby opisały i określiły ilościowo, co chcą osiągnąć lub co osiągnęły za pieniądze podatnika. Zarządzanie i proces budżetowy wprowadza zatem kontrolę wydatków, której elementami są (tabela 2.34).

Tabela 2.34

#### Elementy kontroli wydatków budżetowych

Wartość za pieniądze ( <i>value for money</i> )	Produkty ( <i>outcomes</i> )	Efektywność ( <i>effectiveness</i> )
	Efekty ( <i>outputs</i> )	
	Proces ( <i>process</i> )	Wydajność ( <i>efficiency</i> )
	Wkłady ( <i>inputs</i> )	Ekonomiczność ( <i>economy</i> )
	Koszty ( <i>costs</i> )	

Źródło: J. Kromann Kristensen, W. Groszyk i B. Bühler, *Outcome focused management and budgeting*, „OECD Journal on Budgeting” 2002, vol. 1, no. 20, Issue 4.

Nie wszystkie jednak kraje wprowadziły budżet zadaniowy (w innych krajach kontrolę wydatków przeprowadza się dzięki ewaluacjom), a te, które taki budżet wprowadziły, traktują tę formę budżetu jako materiał konsultacyjny, a nie jako podstawę rozliczania wydatków.

### 3.10. Rozliczalność

Jednym z kluczowych słów w działalności sektora publicznego II połowy XX wieku stało się słowo *accountability*. Nie ma ono jednoznacznego odpowiednika w języku polskim. Najczęściej *accountable* tłumaczy się jako odpowiedzialny, zobowiązany do wyliczenia się, wytłumaczalny, dający się usprawiedliwić<sup>240</sup>. *Accountable* odnosi się do sytuacji, w której np. administrator bezpośrednio odpowiedzialny przed radą nadzorczą musi okazać, że potrafił wywiązać się z powierzonych mu obowiązków, a z powierzonych mu rzeczy zrobił mądry użytek<sup>241</sup>. *Accountable* (rozliczane) są organizacje i ludzie, a nie programy czy instrumenty polityczne. Mówiąc opisowo, *accountability* to zobowiązanie stron uczestniczących we wprowadzaniu w życie interwencji publicznej do tego, aby dostarczyć zwierzchnim władzom politycznym oraz opinii publicznej informacji i wyjaśnień dotyczących oczekiwanych i aktualnych wyników działań. Osią informacji i wyjaśnień ma być sposób wykorzystania zasobów publicznych<sup>242</sup>.

<sup>240</sup> J. i H. Jaślanowie, *Słownik terminologii prawniczej i ekonomicznej angielsko-polski*, Warszawa 1991.

<sup>241</sup> *The Penguin Modern Guide to Synonyms and Related Words*, London 1971, s. 493.

<sup>242</sup> Por. *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, vol. 6, European Commission, Luxembourg, 1999.

Brak polskiego odpowiednika *accountability* rozumianego w sposób używany obecnie w krajach OECD świadczy, że sektor publiczny w Polsce nie działa jeszcze według zasad obowiązujących w państwach rozwiniętych. „Rozliczalność” nadal rozumiane jest u nas w kategoriach zgodności z prawem, a w szczególności z przepisami finansowymi, a nie zgodności z misją organizacji, mandatem społecznym i zasadami skutecznego działania. Ponadto, definiowane jest ono w sposób wąski i negatywny (kontrola i kara), podczas gdy dziś w *accountability* nacisk położony jest bardziej na aspekt pozytywny (pomiar dokonań, informowanie opinii publicznej o osiągniętych celach i rezultatach). *Accountability* w dzisiejszym rozumieniu jest jednak efektem zmian w rozumieniu zadań sektora publicznego w zachodnim świecie. Przed kilkoma dekadami słowo to także nie miało dzisiejszego znaczenia. Dawniej, tak jak dziś w Polsce „rozliczanie”, dotyczyło tylko zgodności z przepisami i spraw finansowych, a nie skuteczności i efektywności działania. Zmiana znaczenia odzwierciedliła przejście od administracji „nakazu i kontroli” do nowoczesnej administracji.

Podczas gdy w Polsce „rozliczanie” nie nabrało jeszcze dzisiejszego znaczenia *accountability*, w krajach OECD słowo to rozszerza swój zakres, a zmiana ta odzwierciedla kolejną ewolucję sektora publicznego. W ostatnich dekadach *accountability* odnosi się do odpowiedzialności urzędników rządowych przed zwierzchnikami, zwierzchników przed ministrami, ministrów przed parlamentem, a parlamentu przed wyborcami. Wraz z ewolucją systemu rządzenia ten prosty model jest coraz częściej w wieloraki sposób modyfikowany. *Accountability* oznacza coraz częściej „rozliczanie” nie tylko przed zwierzchnikiem, ale także przed interesariuszami i opinią publiczną, niekoniecznie bezpośrednio, ale bardziej długofalowe, nie tyle oparte na kontroli zgodności wyników z zakładanymi celami, ile na osiągnięciu bardziej długofalowych pozytywnych efektów. W korporacjach obejmuje ono przewidziane prawem konsultacje z interesariuszami, a w demokracji partycypacyjnej – udział obywateli w sprawowaniu władzy. Zmiana znaczenia i praktyk „rozliczania” jest reakcją na sytuację, w której tradycyjne metody administracji często zawodzą w zarządzaniu współczesnymi instytucjami rządowymi, instytucje dostarczające usług publicznych funkcjonują nieraz skuteczniej, gdy są zwolnione z codziennego nadzoru, a w szybko zmieniającym się świecie stwierdza się potrzebę większej organizacyjnej elastyczności<sup>243</sup>.

W miarę integracji z Europą i światem brak znajomości sensu oraz praktyk *accountability* w znaczeniu stosowanym w większości państw EU i OECD powoduje wiele kolizji i zamieszania w Polsce. Wąskie, NIK-owskie rozumienie „rozliczania” pociąga za sobą wysiłki organizacji publicznych na rzecz osiągania czysto formalnej zgodności z przepisami zamiast osiągania efektów satysfakcjonujących społeczeństwo i stanowiących wkład w rozwój kraju. Formalna zgodność z przepisami może kryć olbrzymie marnotrawstwo. Brak praktyk *accountability* pozbawia często organizacje publiczne wielu szans. Uniwersytety

<sup>243</sup> OECD *Choice of policy instruments*, PUMA/SBO(97)9.

tracą okazję pozyskania międzynarodowych funduszy, nie mogą przedstawić planów strategicznych. Ministerstwa tracą wiarygodność na forum międzynarodowym, nie mogą legitymować się zapisaną w dokumentach wizją polityczną i priorytetami.

### 3.11. Trendy w rozwoju ewaluacji nauki

Wbrew potocznym opiniom, jak podkreśla Chris Freeman, „sama z siebie nauka wcale nie pobudza wzrostu gospodarczego ani też nie służy jakimkolwiek celom społecznym (takim jak ochrona środowiska) lub wojskowym. Ściśle biorąc, wyniki prac badawczych i rozwojowych to tylko strumień informacji: artykułów, raportów badawczych, książek. W najlepszym razie, prototypów lub wstępnych realizacji urzędów. Wystarczałyby to, gdyby nauka była uprawiana jedynie dla niej samej. Wszelako dla wywarcia wpływu na gospodarkę, jakość życia lub stan obronności kraju, same tylko wyniki nie są wystarczające. Co więcej, gdyby miały one pozostać w stadium szkiców, prototypów lub modeli i nigdy nie przejść do etapu wdrożeń, z punktu widzenia polityki naukowej byłyby niemal całkowicie bezużyteczne”<sup>244</sup>. Gdy dawniej ewaluacja koncentrowała się na jakości badań, jej przedmiotem mogły być tylko produkty badań naukowych, takie jak – przede wszystkim – publikacje w renomowanych czasopismach międzynarodowych. Z chwilą rozszerzenia zakresu jej zainteresowań musi ona jednak także uwzględniać najważniejsze uwarunkowania oraz formy konwersji „strumienia informacji” na strumień decyzji i inicjatyw decydujących o poprawie dobrobytu i wzroście gospodarczym<sup>245</sup>.

Tendencja zmiany punktu ciężkości w ewaluacji badań – od pomiaru wkładu do pomiaru „produktu”, „efektu” i „wpływu” – wynika zarówno z przenikania idei tzw. Nowego Zarządzania Publicznego<sup>246</sup>, jak i przemian praktyk w działalności badawczej, innowacyjnej i gospodarczej (wzrost rynkowego ukierunkowania badań i nasilenie powiązań nauki z gospodarką oraz sferą publiczną i społeczną, m.in. wskutek wpływu ICT).

Nowe zarządzanie publiczne (*New Public Management*) kładzie nacisk na wskaźniki dokonań i na ewaluacje. W ostatnich dekadach w krajach OECD punkt ciężkości procesu budżetowego i zarządzania przesunął się z finansowania

<sup>244</sup> Ch. Freeman, *Science and economy at the national level*, [w:] *The Economics of Hope, Essays on technical change, economic growth and the environment*, London–New York, Printer 1992, s. 31–49.

<sup>245</sup> Na przykład w Niderlandach w ocenie instytucji bierze się pod uwagę nie tylko same produkty badań, ale także zarządzanie badaniami, strategie rozwoju badań, urzędnicy badawcze, studia doktoranckie oraz społeczną relewantność badań. *Standard Evaluation Protocol (SEP) 2009–2015. Protocol For Research Assessment In The Netherlands*, VSNU, KNAW and NWO 2009 <http://www.uu.nl/NL/onderzoek/HetUtrechtseonderzoek/onderzoeksbeoordelingen/Documents/SEP2009-2015.pdf>.

<sup>246</sup> *Funding Systems and their Effects on Higher Education Systems*, OECD EDU/WKP(2007)1.

opartego na wskaźnikach „wkładu” (*input*) oraz bazującego na kwotach z lat ubiegłych na finansowanie oparte na pomiarze dokonań – produktu (*output*), efektu (*outcome*) i wpływu (*impact*)<sup>247</sup>.

„Państwo w coraz większym zakresie rozdziela środki dla uczelni na podstawie wskaźników dokonań. Poprawa wydajności, «rozliczalności» (*accountability*) i jakości w szkolnictwie wyższym to główne cele i uzasadnienia dla tego typu finansowania”<sup>248</sup>.

W finansowaniu dydaktyki w szkolnictwie wyższym następowało przejście od finansowania zorientowanego na „wkład” (mierzony np. przez liczbę studentów lub powierzchnię budynków uczelnianych) do finansowania ukierunkowanego na „efekty” (np. przez liczbę absolwentów), a także od finansowania podażowego (określanego przez ministerstwa lub agencje) do finansowania popytowego (w którym nauczanie jest finansowane przez studentów bądź w formie opłat, bądź voucherów)<sup>249</sup>. Hasłami reform stały się: wzrost autonomii i odpowiedzialności uczelni, poprawa efektywności i wydajności, międzynarodowa konkurencyjność, poprawa jakości, orientacja na klienta, bliższa współpraca i większe środki od sektora prywatnego, wdrożenie procesu bolońskiego<sup>250</sup>. Najczęściej przejście było powolne i często polegało na łączeniu tradycyjnych i nowych form finansowania<sup>251</sup>.

Podobne tendencje zarysowały się w ewaluacji badań naukowych. Już w latach 70. kryteria i procedury ewaluacji przestały być domeną samych badaczy i zostały otwarte na negocjacje z politykami, urzędnikami administracji publicznej, ekonomistami, ekspertami zarządzania<sup>252</sup>. W latach 80. tendencje

<sup>247</sup> *Outcome focused management and budgeting*, OECD, PUMA/SBO 2001. Innymi, powiązаныmi czynnikami zmiany form finansowania stały się koncepcje współzarządzania (*governance*), nowej ekonomii instytucjonalnej (m.in. teorie własności intelektualnej, kosztów transakcji, związków pryncypał-agent) oraz rozwijania strategii politycznych. *Funding Systems and their Effects on Higher Education Systems*, OECD EDU/WKP(2007)1

<sup>248</sup> N. Frølich, *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_steeringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_steeringbynumbers.pdf).

<sup>249</sup> F. Kaiser, H. Vossensteyn i J. Koelman, *Public funding of higher education. A comparative study of funding mechanisms in ten countries*, 2002, CHEPS, <http://doc.utwente.nl/45235/>.

<sup>250</sup> *Funding Systems and their Effects on Higher Education Systems*, OECD EDU/WKP(2007)1.

<sup>251</sup> Na podstawie literatury przedmiotu Diana Hicks podkreśla, że obecne systemy finansowania badań naukowych na uczelniach można traktować jako realizację sześciu kluczowych idei bliskich nowemu zarządzaniu publicznemu: zwiększenie wydajności – finansowanie badań naukowych na podstawie wyników zwiększa wydajność; zastąpienie tradycyjnego „nadzoru i kontroli” systemem „zachęt rynkowych”; większe uwzględnienie potrzeb obywateli; decentralizacja oparta na przekonaniu, że programy są bardziej elastyczne i efektywne, gdy są bliższe świadczeniu usług (zarządzane strategicznie uczelni, zamiast podejmowania wszystkich kluczowych decyzji przez ministerstwa i agencje); przejście od świadczenia do kontraktowania usług przez rządy (np. rząd jako nabywca „usług edukacyjnych” w Australii); skupienie się na wynikach, a nie procesach i strukturach. D. Hicks, *Overview of Models of Performance-Based Research Funding Systems*, DSTI/STP/RIHR(2010)3.

<sup>252</sup> L. Georghiou, *Research evaluation in European national science and technology system*, op. cit.

te uległy pogłębieniu. Rozwinięto wówczas na dużą skalę systematyczną ewaluację działań rządu na rzecz rozwoju badań naukowych i technologicznych<sup>253</sup>. Postęp w badaniach i jakość badań nie były teraz traktowane jako cele same w sobie, lecz jako narzędzia osiągnięcia celów społecznych<sup>254</sup>. Coraz ważniejszymi kryteriami oceny stały się „korzyści”, „wpływy”, „efekty” oraz „zwrot z inwestycji”. Pojęcia te zakorzeniły się też w koncepcjach ekonomicznych (np. *spillover effect*, *external effect*), metodologiach ewaluacji oraz w żargonie politycznym (uzasadnienia budżetu oraz różnych form interwencji rządowych).

Rozwijane w ostatnich latach formy ewaluacji nauki charakteryzują się czterema cechami:

- opierają się na nowym pojęciu „wpływu”, odchodząc od jego wąsko ekonomicznego rozumienia i kładąc jeszcze silniejszy niż dotąd nacisk na (szeroko rozumiane) wpływy społeczne i kulturowe inwestycji w B+R;
- podkreślają one wielkie zróżnicowanie nie tylko dyscyplin i pól badań (pod względem tematyki, metodologii, etosu zawodowego itd.), ale także rodzaj powiązań, jakie łączą każde z nich ze światem zewnętrznym;
- biorą pod uwagę wielowymiarowość ocen;
- bazują na odmiennej organizacji ewaluacji, w której biorą udział nie tylko eksperci, ale także różnego typu interesariusze, włącznie z obywatelami;
- w większej mierzy uwzględniają mierniki współpracy, przede wszystkim współpracy międzynarodowej badań interdyscyplinarnych oraz badań w ujawniających się (*emerging*) dyscyplinach<sup>255</sup>.

### Prace cytowane

*A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process*, A Pilot Study Commissioned by OCLC Research. Prepared by Key Perspectives Ltd Truro, UK, 2009.

Academy of Finland, *National criteria for research infrastructures*, 3.4.2017, [http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/firi/firi\\_national\\_criteria\\_ri\\_2017.pdf](http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/firi/firi_national_criteria_ri_2017.pdf)

Albrecht V. i Vaněček J., *Assessment of Participation of the Czech Republic in the EU Framework Programmes*, Technology Centre of the Academy of Sciences of the Czech Republic 2009.

<sup>253</sup> T. Luukkonen i B. Stähle, *Evaluation of research fields. Scientists' Views*, 1993; T. Luukkonen, *The increasing professionalisation of the evaluation of mission-oriented research in Finland: implications for the evaluation process w: Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997; F. Meyer-Krahmer i Th. Reiss, *Ex ante evaluation and technology assessment – two emerging elements of technology policy evaluation*, „Research Evaluation”, April 1992, vol. 2, no. 1.

<sup>254</sup> L.F. Frederiksen, *The Agora and the Role of Research Evaluation*, Copenhagen Business School, Internet.

<sup>255</sup> Między innymi *The Evaluation of Scientific Research: Selected Experiences*, OECD DSTI/STP/SUR(97)9.



- Andras A. i Charlton B., *Evaluating universities using simple scientometric research output metrics: Total citation counts per university for a retrospective seven-year rolling sample*, „Science and Public Policy” 2007, vol. 34 (8), s. 555–563.
- Arnold E. i Guy K., *Technology diffusion programmes and the challenge for evaluation*, [w:] *Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997.
- Assessing the Socio-economic Impacts of the Framework Programme*, A study for DG Research by PREST, AUEB, BETA, ISI, Joanneum Research, IE HAS and Wise Guys, 2002.
- Australian Government, Australian Research Council, *Assessing the wider benefits arising from university-based research: Discussion Paper*, June 2013.
- Aymerich M. et al., *Measuring the payback of research activities: A feasible ex-post evaluation methodology in epidemiology and public health*, „Social Science & Medicine” 2012, 75, s. 505–510.
- Balogh T., *From peer review to portfolio analysis. Evaluation of applied research and development*, msp.
- Barré R., *Evaluation: a process for strategic debate*, 2002.
- Basri E. i Cervantes M., *Evaluation Of Publicly Funded Research: Recent Trends And Perspectives*, DSTI/STP(2006)24.
- Bauer H.H., *Scientific literacy and the myth of the scientific method*, University of Illinois Press, Urbana 1992.
- BBSRC, *Evaluation of BBSRC's Research Equipment Initiative appendices*, 2012.
- Becher G. i Kuhlmann S., red., *Evaluation of technology policy programmes in Germany*, Springer, Dordrecht 1995.
- Benchmarking Of National Policies Public And Private Investments In R&D*. Final Report Expert Group, OECD June 2002.
- Bendyk E., *Makroskop*, <https://edwinbendyk.wordpress.com/tag/recenzje/>.
- Berkenkotter C., *The Power and the Perils of Peer Review*, „Rhetoric Review”, Spring 1995, vol. 13, no. 2.
- Box S. i Basri E., *Performance-Based Funding for Public Research in Tertiary Education Institutions: Country Experiences. Summary of country questionnaire responses*, DSTI/STP/RIHR(2010)5.
- Bozeman B. i Melkers J., red., *Evaluating R&D Impacts: Methods and Practice*, Boston, MA, Springer 1993.
- Budgeting for results*, OECD 1995.
- Butler L., *Impacts of performance-based research funding systems: A Review of the concerns and the evidence*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.
- Calvert J. i Martin B., *Science funding: Europe*, [w:] *International Encyclopaedia of the Social and Behavioural Sciences*.
- Campbell D.F.J., *The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria*, [w:] *Learning from Science and Technology Policy Evaluation*, red. P. Shapira, S. Kuhlmann, Cheltenham, UK, Northampton, MA, USA: Edward Elgar, 2003.
- Capron H. i van Potelsberge de la Potterie B., *Public support to R&D Programmes*, [w:] *Policy evaluation in innovation and technology. Towards best practices*, 1997 OECD.

- Cervantes M., *CSTP work on Evaluation and Impact Assessment: Key Findings*. Presentation at the 1stRIHR Meeting 18–19 May 2009.
- Chubin D.E. i Hackett E.J., *Peerless Science. Peer review and US Science Policy*, 1990.
- Cohendet P., *Evaluating the industrial indirect effects of technology programmes: the case of the European Space Agency (ESA) programmes*, [w:] *Policy evaluation in innovation and technology. Towards the best practices*, OECD 1997.
- Compilation of FP6 National Impact Studies*, April 2008.
- Cordero A., *On the growing complementarity of science and technology*, „Phil & Tech” 4, 2, Winter 1998, <http://scholar.lib.vt.edu/ejournals/SPT/v4n2/pdf/CORDERO.PDF>.
- Cuhls K., *Evaluation of Foresight Activities*, Fraunhofer Institute Systems and Innovation Research.
- Cunningham P., Edler J., Flanagan K. i Larédo Ph., *Innovation policy mix and instrument interaction: a review*, NESTA, November 2013.
- Cunningham P., *Principles of the design and implementation of evaluations to maximise usefulness in different contexts*, 2010.
- Dahler-Larsen P, *The Evaluation Society*, Stanford, Calif.: Stanford Business Books, 2001.
- de May M., *The cognitive paradigm: an integrated understanding of scientific development*, Chicago, University of Chicago Press, 1992.
- de Meis L., Velloso A., Lannes D., Carmo M.S. i de Meis C, *The growing competition in Brazilian science: rites of passage, stress and burnout*, „Brazilian Journal of Medical and Biological Research” 2003, 36(9), s. 1135–1141.
- Delanghe H. i Teirlinck P., *Optimising the Policy Mix by the Development of a Common Methodology for the Assessment of (Socio-)Economic Impacts of RTDI Public Funding (CIA4OPM) Literature Review*, Brussels 2013.
- Designing Policy Mixes: Enhancing Innovation System Performance and R&D Investment Levels*, March A study funded by the European Commission – DG Research 2009.
- Diane H. i Krzys S., *Peer Review in Academic Promotion and Publishing: Its Meaning, Locus, and Future*, Center for Studies in Higher Education, Berkeley 2011.
- Dietz J.S., *Building a social capital model of research development: the case of the Experimental Program to Stimulate Competitive Research*, „Science and Public Policy” 2000, vol. 27, no. 2.
- Douglas W. i Rank A.D., *Measuring the economic benefits of research and development: the current state of the art*, „Research Evaluation”, April 1998, vol. 7, no. 1.
- Drumond R., *Misconduct and journal peer review*, [w:] *Peer Review In Health Sciences*, eds. F. Godlee, i T. Jefferson, 2nd ed., BMJ Books, London 2003.
- Duchêne V. i De Mazière Ir.P., *Semantic analysis as tool to measure policy impact. The impact of FP-funded research in Social Sciences and Humanities (SSH) on EU policies*, June 18th, 2009.
- EC European Commission, *First FP7 Monitoring-Report*, 2009.
- Edler J., *Evaluation in Science and Technology Policy. Introduction*, Manchester 2010.
- Edler J., *The role of International Collaboration in the Framework Programme. Expert Analysis In Support Of The Ex Post Evaluation Of FP6*, 2008.
- Effects of Swiss participation in EU Research Framework Programmes*, Interim Report 2010.

- Ekiert K., *Ewaluacja w administracji publicznej. Funkcje, standardy i warunki stosowania*, Rządowe Centrum Studiów Strategicznych, Warszawa 2004.
- Ellison G., *Is Peer Review In Decline?* NBER Working Paper Series, 2007.
- Erno-Kjølhede E. et al., *Managing university research in the triple helix*, „Science and Public Policy”, February 2001, vol. 28, no. 1.
- ESF, European Science Foundation, *A Guide To Evaluation Activities In Research Funding Agencies. Quality Assurance And Evaluation Guidelines*, 2012.
- ESF, European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies. A report of the ESF Member Organisation Forum on Ex-Post. Evaluation of Funding Schemes and Research Programmes*, September 2009.
- ESF, *The Challenges of Impact Assessment*, 2012.
- ESRC Economic and Social Research Council, *Economic Impact Evaluation of the Economic and Social Data Service*, 2012.
- EU, *Community Support for Research Infrastructures in the Sixth Framework Programme Evaluation of pertinence and impact*, 2009.
- European Commission, *An RTD Evaluation Toolbox*, 2002.
- European Commission, *Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, 1999, vol. 6, Luxembourg.
- European Science Foundation, *Evaluation in National Research Funding Agencies: approaches, experiences and case studies: A report of the ESF Member Organisation Forum on Ex-Post Evaluation of Funding Schemes and Research Programmes*, 2009.
- Ewaluacja instytucji naukowych w Polsce w świetle porównań międzynarodowych i konsultacji*, Departament Strategii Ministerstwa Nauki i Szkolnictwa Wyższego, Warszawa.
- Ewaluacja instytucji publicznego sektora nauki oparta na pomiarze dokonań – charakter, pochwały, krytyka*, „Nauka Polska. Jej Potrzeby, Organizacja i Rozwój” 2013.
- Ewaluacja w Administracji Publicznej Funkcje, Standardy i Warunki Stosowania*, [http://www.pte.org.pl/repository/files/PTE/Ewaluacja\\_w\\_administracji\\_publicznej.pdf](http://www.pte.org.pl/repository/files/PTE/Ewaluacja_w_administracji_publicznej.pdf).
- Evaluating socio-economic programmes. Glossary of 300 concepts and technical terms*, vol. 6, European Commission, Luxembourg 1999.
- Evaluation of Research in Context. A Quick Scan of an Emerging Field*, 2007, Rathenau Instituut – Department Science System Assessment, [http://www.allea.org/Content/ALLEA/WG%20Evaluating/NL\\_Quick\\_Scan.pdf](http://www.allea.org/Content/ALLEA/WG%20Evaluating/NL_Quick_Scan.pdf).
- Evaluation of the impact of the FP6 in the RTD public system in Spain*, 2010.
- Falagas M.E., *Peer review in open access scientific journals*, „Open Medicine” 2007, vol. 1, no. 1.
- Faulkner W., *Conceptualizing Knowledge Used in Innovation: A Second Look at the Science-Technology Distinction and Industrial Innovation*, „Science, Technology, & Human Values”, Autumn 1994, vol. 19, Issue 4, s. 425–458.
- FCT, *Evaluation Guidelines*, 2013.
- Feller I., *The academic policy analyst as reporter: the who, what and how of evaluating science and technology programs*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer

- Institute for Systems and Innovation Research, bmw, 2001, [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhworkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhworkshop.pdf).
- Finland – *Evaluation of Participation and Network in FP6*, Tekes Programme Report 6/2008.
- FIRI, *Evaluation criteria for research infrastructure*, cyt. za: *Applied Multivariate Research. Design and Interpretation*, Sage 2006.
- Foresight, Wikipedia 2014.
- Foss Hansen H., *Performance indicators used in performance-based research funding*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.
- Foss Hansen H., *Research Evaluation: Methods, Practice and Experience*, Danish Agency for Science, Technology and Innovation, bmw, 2009.
- Frederiksen L.F., *The Agora and the Role of Research Evaluation*, Copenhagen Business School, Internet.
- Freeman Ch., *Science and economy at the national level*, [w:] *The Economics of Hope, Essays on technical change, economic growth and the environment*, London–New York, Pinter 1992.
- Frølich N., *The politics of steering by numbers, Debating performance-based funding in Europe*, NIFU-STEP 2008, [http://www.nordforsk.org/\\_img/nifu\\_step\\_rapport\\_stereringbynumbers.pdf](http://www.nordforsk.org/_img/nifu_step_rapport_stereringbynumbers.pdf).
- Fundacja Nauki Polskiej, [http://www.fnp.org.pl/o\\_fundacji/ewaluacja-programow-fundacji/](http://www.fnp.org.pl/o_fundacji/ewaluacja-programow-fundacji/).
- Funding Systems and their Effects on Higher Education Systems*, OECD EDU/WKP(2007)1.
- Geisler E., *Integrated Figure of Merit of Public Sector Research Evaluation*, „Scientometrics” 1996, vol. 36, no. 3.
- George L., *Issues in the evaluation of innovation and technology policy*, [w:] *Policy evaluation in innovation and technology. Towards best practices*, OECD, 1997.
- Georghiou L., Amanatidou E., Belitz H., Cruz L., Edler J., Edquist C., Granstrand O., Guinet J., Leprince E., Orsenigo L., Rigby J., Romanainen J., Stampfer M. i van den Biesen J., *Improving the Effectiveness of Direct Public Support Measures to Stimulate Private Investment in Research*, Report of the ETAN Working Group on Direct Measures for Directorate General Research, *European Commission*, 2002.
- Georghiou L., *Evaluating Foresight and Lessons for Its Future Impact*, The Second International Conference On Technology Foresight – Tokyo, 27–28 Feb. 2003, Dostępne: <http://www.nistep.go.jp/IC/ic030227/pdf/p6-1.pdf>.
- Georghiou L., *Evaluation in the Balance: Matching Methods and Policy Needs*, paper presented at The Evaluation Conference, Bruxelles, Nov. 10–12, 2001. Dostępne: <http://les.man.ac.uk/PREST>.
- Georghiou L., *Evaluation of research and innovation policy in Europe – new policies, new frameworks?* [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001. Dostęp: [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhworkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhworkshop.pdf).

- Georghiou L., *Research evaluation in European national science and technology system*, „Science and Public Policy” 1995, vol. 5, no. 1.
- Geuna A. i Martin B.R., *University research evaluation and funding: An international comparison*, „Minerva” 2003, 41(4), s. 277–304.
- Gläser J., Laudel G., Hinze S. i Butler L., *Impact of evaluation-based funding on the production of scientific knowledge*, 2002, <http://www.laudel.info/pdf/02Expertise-GlaeLauHinBut.pdf>.
- Godin B., *Are Statistics Really Useful? Myths and Politics of Science and Technology Indicators*, Project on the History and Sociology of S&T Statistics, Working Paper no. 20, 2002.
- Godin B., *The Who, What, Why and How of S&T Measurement*, Project on the History and Sociology of S&T Statistics, Working Paper no. 26, [http://www.csiic.ca/PDF/Godin\\_26\\_a.pdf](http://www.csiic.ca/PDF/Godin_26_a.pdf), *Le Banquet* (Revue du CERAP), 19–20, January 2004.
- Godin B. i Doré C., *Measuring the Impacts of Science: beyond the Economic Dimension*, Paper presented at the Helsinki Institute for Science and Technology Studies, HIST Lecture, 24 August 2007.
- Government of the Future*, OECD 2001.
- Grant w opalach*, „Forum Akademickie” 2000, z. 1.
- Gro Emmertsen Lund, *Fifth-Generation Evaluation*, Internet [dostęp: 27.07.2017].
- Gruss P., *Peer review at the Max Planck Society*, w: *Science between evaluation and innovation: a conference on peer review*, DFG 2002.
- Guetzkow J., Lamont M. i Mallard G., *What Is Originality in the Humanities and the Social Sciences?*, „American Sociological Review”, Apr. 2004, vol. 69, no. 2.
- Guidelines and Principles for Social Impact Assessment*, maj 1994.
- Guston D.H., *The expanding role of peer review processes in the United States*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, bmw, 2001, Dostęp: [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhwkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhwkshop.pdf).
- Guy K., *Beyond Socio-economic Impact Evaluation: The challenge for the next millennium*, [w:] *Socio-economic Impact Evaluation Workshop*, 26–27 November 1999, Helsinki 2000.
- Hausner J., Jessop B. i Mazur S., *Governance. Wybór tekstów klasycznych*, Scholar, Warszawa 2017.
- Haustein S., *Multidimensional Journal Evaluation*, [w:] 11<sup>th</sup> International Conference on Science and Technology Indicators „Creating Value for Users” 9–11 September 2010. Book of Abstracts, <http://info.scopus.com/journalmetrics/>.
- Hazelkorn E., *Pros and cons of research assessment, lessons from rankings*, 2009, <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1000&context=dirrerep>.
- Helgesen G.E.M., *Evaluating Research Fields/Disciplines in Norway*, 2008.
- HERA, *Feasibility Study: The Evaluation and Benchmarking of Humanities Research in Europe*, August 2007.
- Herbst M., *Financing public universities: the case of performance funding*, Dordrecht, Springer 2007.

- Hicks D., *Overview of models of performance-based research funding systems*, [w:] *Performance-based Funding for Public Research in Tertiary Education Institutions*, OECD, Paryż 2010.
- Hicks D., *Overview of Models of Performance-Based Research Funding Systems*, DSTI/STP/RIHR(2010)3.
- Higher Education Funding – Overseas Models*, 2006, <http://www.parliament.nz/NR/rdonlyres/5694EE99-CA9E-499C-9E94-E40F1B114354/40871/0605HigherEducationFundingpdf3.pdf>.
- Hofer R., *Draft Questionnaire on Public Research Institutes*, OECD, 2010.
- Hornbostel S. i Olbrecht M., *Peer Review In Der DFG: Die Fachkollegiaten*, iFQ-Working Paper no. 2, November 2007.
- Hoskin K.W. i Macve R.H., *Accounting And The Examination: A Genealogy Of Disciplinary Power*, „Accounting Organizations and Society” 1986, vol. 11, no. 2.
- Hoskin K., Macve R. i Stone J., *The Historical Genesis of Modern Business and Military Strategy: 1850–1950*, Interdisciplinary Perspectives on Accounting Conference, Manchester, 7–9 July, 1997.
- House of Commons. Science and Technology Committee, *Peer review in scientific publications*, Eighth Report of Session 2010–12, 2011.
- Howlett M. i Rayner J., *Design Principles for Policy Mixes: Cohesion and Coherence in ‘New Governance Arrangements’*, „Policy and Society” 01/2007, 26(4), s. 1–18. <http://archives.esf.org/coordinating-research/mo-fora/evaluation-of-funding-schemes-and-research-programmes.html>.
- <http://arton.bg.us.edu.pl/arton/arton.php>.
- [http://en.wikipedia.org/wiki/Goodhart's\\_law](http://en.wikipedia.org/wiki/Goodhart's_law).
- <http://si-per.eu/Search/MainSearch>.
- <http://www.scienceurope.org/policy/working-groups/research-policy-and-programme-evaluation/>.
- <https://rio.jrc.ec.europa.eu/en/policy-support-facility/peer-reviews>.
- Huang M. i Chang Yu, *Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective*, „Journal Of The American Society For Information Science And Technology” 2008, 59(11), s. 1819–1828.
- Impact assessment of Framework Programme on French research and innovation system* 2010.
- Instytut Badań nad Gospodarką Rynkową, *Przyczyny sukcesów i niepowodzeń zespołów naukowych biorących udział w Szóstym Programie Ramowym – wnioski na przyszłość*, Warszawa, lipiec 2006.
- Inzelt A.M., *Towards Indicators on University Third Mission*, PRIME – PhD Winter School, Rome, 5–9 November 2007.
- IPP, Innovation Policy Platform, *Policy goals and means*, 2014-01-13.
- IPTS-Johanneum, *RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies*, 2002.
- Jaślan J. i Jaślan H., *Słownik terminologii prawniczej i ekonomicznej angielsko-polski*, Warszawa 1991.
- Jedamus P., Peterson M.W. i Associates, *Improving Academic Management. A Handbook of Planning and Institutional Research*, Jossey-Bess Publishers, 1989.

- Kaiser F., Vossensteyn H. i Koelman J., *Public funding of higher education. A comparative study of funding mechanisms in ten countries*, 2002, CHEPS, <http://doc.utwente.nl/45235/>.
- Kalpazidou Schmidt E., *Evaluation and science policy*, [w:] *The use of evaluations in Europe. Report from the European RTD Evaluation Network meeting*, ed. K. Siune i E. Kalpazidou Schmidt, Report 2003/2, Aarhus, Danish Institute for Studies in Research and Research Policy, 2003.
- Kalpazidou Schmidt E. i Siune K., *The Use Of R&D Evaluations In European Science Policy*, [www.cfa.au.dk/Staff/eks.htm](http://www.cfa.au.dk/Staff/eks.htm).
- Kazancigil A., *Governance and science: market-like modes of managing society and producing knowledge*, „International Social Science Journal”, March 1998.
- Kekäle J., *Conceptions Of Quality In Four Different Disciplines*, „Tertiary Education and Management” 2002, 8.
- Kostoff R.N., *Federal Research Impact Assessment: Axioms, Approaches, Applications*, „Scientometrics” 1995, vol. 34, no. 2.
- Kromann Kristensen J., Groszyk W. i Bühler B., *Outcome focused management and budgeting*, „OECD Journal on Budgeting” 2002, no. 20, vol. 1, Issue 4.
- Krull W., *Current practice and guidelines in a selection of industrialized countries*, [w:] *Evaluation of Research and Development. Current Practice and Guidelines. Synthesis report*, eds. W. Krull, D. Sensi i D. Sotiriou, Luxembourg: Commission of the European Communities Directorate-General Telecommunications, Information Industries and Innovation, 1991.
- Krull W., Sensi D. i Sotiriou D. eds., *Evaluation of Research and Development. Current Practice and Guidelines. Synthesis report*, Luxembourg: Commission of the European Communities Directorate-General Telecommunications, Information Industries and Innovation, 1991.
- Kuhlmann S., *Evaluation as a source of `strategic intelligence*, [w:] *Learning from Science and Technology Policy Evaluation*, Proceedings from the 2000 US-EU Workshop on Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, red. P. Shapira, S. Kuhlmann, Georgia Tech Research Corporation and the Fraunhofer Institute for Systems and Innovation Research, 2001, Dostęp: [http://cspo.org/legacy/library/091110F6VH\\_lib\\_Gustonbhwkshop.pdf](http://cspo.org/legacy/library/091110F6VH_lib_Gustonbhwkshop.pdf).
- Lahlou S., van der Meijden R., Messu M., Poquet G. i Prakke F., *A Guideline for Survey-Techniques in Evaluation of Research*, 1992.
- Lamont M., *How Professors Think: Inside The Curious World of Academic Judgment*, 1st Edition, Harvard University Press, Cambridge 2009.
- Langfeldt L., *Decision-making in expert panels evaluating research. Constraints, processes and bias*, NIFU Oslo 2002, [www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf](http://www.statsvitenskap.uio.no/fag/polit/disputas/fulltxt/langfeldt.pdf).
- Liefner I., *Funding, resource allocation, and performance in higher education Systems*, „Higher Education” 2003, 46(4), s. 469–489.
- Luukkonen T., *The increasing professionalisation of the evaluation of mission-oriented research in Finland: implications for the evaluation process*, [w:] *Policy evaluation in innovation and technology. Towards the best practices*, OECD, 1997.
- Luukkonen T. i Ståhle B., *Evaluation of research fields. Scientists' Views*, 1993.

- Malerba F., *Network Analysis of Research Collaborations*, Conference “Framework Programme Evaluation: Taking Stock and Moving Forward” Brussels, 18 June 2009.
- Malone R.E., *Should peer review be an open process?*, „Journal Of Emergency Nursing” 1999, 25.
- Mandviwalla M, Patnayakuni R. i Schuff D., *Improving the peer review process with information technology*, „Decision Support Systems” 2008, 46.
- Manual on the budgeting of scientific and technological activities. Integrating the ‘science and technology’ function in the general State budget*, UNESCO 1984.
- Marsh H.W., Jayasinghe U.W. i Bond N.W., *Improving the Peer-Review Process for Grant Applications. Reliability, Validity, Bias, and Generalizability*, „American Psychologist”, April 2008, vol. 63, no. 3.
- McQueen A. i Beard A., *A Creative Approach to Assessing Proposals*, University Of Malta, The Edward De Bono Institute For The Design And Development Of Thinking, *The Fifth International Conference On Creative Thinking: Designing Future Possibilities* 21–22nd June 2004.
- Measuring Government Activity*, OECD 2009.
- Merlin D., *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press 1991.
- Meyer-Krahmer F. i Reiss Th., *Ex ante evaluation and technology assessment – two emerging elements of technology policy evaluation*, „Research Evaluation”, April 1992, vol. 2 no. 1.
- Miller R. i Manseau A., *Bibliometric indicators and the competitive environment of R&D laboratories*, „Scientometrics” 1996, vol. 36, no. 3.
- Moed H.F., *The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review*, „Science and Public Policy” 2007, vol. 34 (8), s. 575–583.
- Moszczyński J. i Kozłowski J., *Informacja w sprawie możliwości oszacowania nakładów i efektów polityki naukowej i naukowo-technicznej*, 1992, niepubl. msp.
- MRC Medical Research Council 2011, *Measuring the link between research and economic impact*.
- Mulligan A., *Is peer review in crisis?*, „Oral Oncology” 2005, 41.
- Nedeva M., *Peer Review: omnipresence vs. omnipotence*, Evaluation of Science and Innovation Policies, 17–21 May 2010, Manchester.
- Nielsen M., *Three myths about scientific peer review*, Blog January 8, 2009, <http://michaelnielsen.org/blog/three-myths-about-scientific-peer-review/>.
- Nightingale P. i Scott A., *Peer review and the relevance gap: ten suggestions for policy-makers*, „Science and Public Policy”, October 2007, 34(8).
- Noveck B.S., *Peer to Patent. Collective Intelligence, Open Review and Patent Reform*, 2006.
- OECD, *Assessing The Impact Of State Interventions In Research – Techniques, Issues And Solutions*, DSTI/STP(2014)24.
- OECD, *Assessing the Socio-economic Impacts of Public Investment In R&D*, 2007.
- OECD, *Assessing the Socio-economic Impacts of Public R&D: Recent Practices and Perspectives*, [w:] *OECD Science, Technology and Industry Outlook 2008*.
- OECD, *Choice of policy instruments*, PUMA/SBO(97)9.
- OECD, *Evaluation Of Publicly Funded Research: Recent Trends And Perspectives*, 2006.



- OECD, *Framework For Monitoring And Impact Analysis: – Case Tekes, Finland*, TIP-RIHR WORKSHOP 15.16.9.2008.
- OECD, *Improving Evaluation Practices. Best Practice Guidelines for Evaluation and Background Paper*, PUMA/PAC(99)1.
- OECD, *Improving Policy Instruments Through Impact Assessment*, Sigma Paper: No. 31, CCNM/SIGMA/PUMA(2001)1.
- OECD IPP, *Innovation Policy Platform Issue Brief Peer Review*, 2014.
- OECD IPP, *Issue Brief: Research Organisation Evaluation*, 2014.
- OECD IPP, *Innovation Policy Platform, Issue Brief: Public Sector Research Funding*.
- OECD, *Linking Science to Innovation, Economic Growth and Welfare: Scoping Paper*, 2013.
- OECD, *Peer Review: Its Uses, Demands, and Issues*, 2007.
- OECD, *Policy Mix for Innovation In Poland. Key Issues and Recommendations*, 2007.
- OECD, *Public Funding Of R&D: Emerging Policy Issues*, 13–14 March 2001.
- OECD RIHR, *Activity 1.2: Assessing The Role of Impact In National Science and Research Systems* 2013.
- OECD, *Strengthening the Impact of Public Research Institutions. Evaluation of PRIs*, 2009.
- OECD, *The Evaluation of Scientific Research. Selected Experiences*, Paryż 1997.
- OECD, *Workshop on Assessing The Impact of Science, Technology and Innovation Policy Instruments with Common Objectives* DSTI/STP/AH/A(2014)2.
- Olejniczak K., *Teoretyczne podstawy ewaluacji ex post*, [w:] *Ewaluacja ex-post. Teoria i praktyka badawcza*, PARP, Warszawa 2007, s. 19–21.
- OPI, *Ewaluacja udziału polskich zespołów badawczych w 6 Programie Ramowym – ocena efektywności finansowej, odpowiedniości, skuteczności i użyteczności*, 2008.
- Optimizing the research and innovation policy mix: The practice and challenges of impact assessment in Europe*, 2011.
- Outcome focused management and budgeting*, OECD, PUMA/SBO 2001.
- Papadakis M., *The Economic Impacts of Public Science Can Be Measured*, „The Scientist” October 27, 1997.
- Participation of German Higher Education Institutions In FP6* 2009.
- Peer Review: the challenges for the humanities and social sciences. A British Academy Report*, The British Academy 2007.
- Performance-based models for publicly funded research in tertiary education institutions* 2009, DSTI/STP/RIHR(2009)7.
- Performance-based models for publicly funded research in tertiary education institutions*, OECD, Paryż 2010.
- Piric A. i Reeve N., *Evaluation of public investment in R&D – towards a contingency analysis* [w:] *Policy evaluation in innovation and technology. Towards best practices*, 1997 OECD.
- Podręcznik Frascati. Proponowane Procedury Standardowe dla Badań Statystycznych w Zakresie Działalności Badawczo-Rozwojowej*, 2002, OECD/MNiSW 2009, tłum. D. Przepiórkowska.
- Policy Mixes for R&D in Europe*, UNU-MERIT, A study commissioned by the European Commission – Directorate-General for Research, April 2009.
- Policy-Mix Methodological Report*, 2006.

- Politechnika Białostocka, Ekspertyza *Badanie Ewaluacyjne Realizowanych w Polsce Projektów Foresight*, Białystok, listopad 2010.
- Pollit Ch., *Public Management Reform*, „OECD Journal on Budgeting” 2003, vol. 3, no. 19, Issue 3.
- Pollit Ch. i Bouckaert G., *Public Management Reform*, Oxford University Press 2000.
- Pollit Ch. i Bouckaert G., *Public Management Reform*, Oxford University Press, 2004.
- Porter A.L. i Rossini F.A., *Peer Review of Interdisciplinary Research Proposals*, „Science, Technology, & Human Values”, Summer 1985, vol. 10, no. 3.
- Power M., *The Audit Society. Rituals of Verification*, Oxford University Press, Oxford 1997.
- PRO INNO Europe, *Impact of publicly funded research on innovation. An analysis of European Framework Programmes 2009*.
- PSDB, *Funkcjonowanie Sieci Punktów Kontaktowych w Polsce*, 2008.
- Quick Scan, *Evaluation of Research in Context A Quick Scan of an Emerging Field*, 2007.
- Raising EU R&D Intensity. Report to the European Commission by an Independent Expert Group Improving the Effectiveness of the Mix of Public Support Mechanisms for Private Sector Research and Development*, European Commission 2003.
- REF impact pilot exercise. Guidance on submissions*, November 2009, <http://www.hefce.ac.uk/research/ref/impact/>.
- Reuters Th., *Expert Group for the Interim Evaluation of Framework Programme 7. Bibliometric analysis 2010*.
- Rip A. i Marais H.C., *Assessing University Research under Condition of Changing Knowledge Production*, SAUVACA Publication Series 1998/2.
- Roche D., *La France des Lumières*, Paris: A. Fayard 1993. Tłum. ang. Arthur Goldhammer, *France in the Enlightenment*, Cambridge University Press, Cambridge 1997.
- Rons N. et al., *Research evaluation per discipline: a peer-review method and its outcomes*, 2007.
- Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające „Horyzont 2020” – program ramowy w zakresie badań naukowych i innowacji (2014–2020)*, Bruksela, dnia 30.11.2011 COM(2011) 809 końcowy 2011/0401 (COD).
- Sagasti F.R., *The science and technology policy instruments project*, „Science and Public Policy” 1979.
- Salmi J. i Hauptman A.M., *Resource Allocation Mechanisms in Tertiary Education: A Typology and an Assessment*, [w:] *Global University Network for Innovation (ed.) Higher Education in the World 2006. The Financing of Universities*, Palgrave MacMillan, Basingstoke, U.K. 2006.
- Salter A.J. i Martin B., *The Economic Benefits of Public Funded Basic Research: A Critical Review*, „Research Policy” 2001, 30, no. 3.
- Schiermeire Q. i Ockenden J., *Perspectives of science in Central and Eastern Europe: emerging directions*, OS Press, 2001.
- Science between evaluation and innovation: a conference on peer review*, DFG 2002.
- Scoping the NESTI Contribution to the CSTP activity on the Impact of Science and Technology over the 2013–2014*, DSTI/EAS/STP/NESTI(2013)4.
- Scott A., *Peer review and the relevance of science*, „Futures” 2007, 39.
- SIAMPI <http://www.siampi.eu/>.

- Smit Ch., *Peer Review: Time for a Change?*, „BioScience”, September 2006, vol. 56, no. 9.
- Smith K., *Economic returns to R&D: Methods, Results and Challenges*, SPSG Review Paper, 1991, No. 3.
- Spaapen J. i Sylvian Ch., *Assessing the value of research for society*, „Research Evaluation” 1993, no. 2, s. 117–126.
- Spanish National FP6 Impact Assessment Study* 2010.
- Standard Evaluation Protocol (SEP) 2009–2015. Protocol For Research Assessment In The Netherlands*, VSNU, KNAW and NWO 2009 <http://www.uu.nl/NL/onderzoek/HetUtrechtseonderzoek/onderzoeksbeoordelingen/Documents/SEP2009-2015.pdf>.
- Stokes D.E., *Pasteur’s Quadrant: Basic Science and Technological Innovation*, Washington, D.C., Brookings Institution Press, 1997.
- Strengthening the Impact of Public Research Institutions: Evaluation of PRIs*, OECD DSTI/STP/RIHR(2009)6.
- Supporting the monitoring and evaluation of innovation programmes* 2006.
- Switzerland’s Participation in the 6th EU Research FP* 2008.
- Taylor J. i Taylor R., *Performance indicators in academia: An X-efficiency approach?* „Australian Journal of Public Administration” 2003, 62(2).
- Technopolis, *Evaluation of the design and implementation of Estonian RTDI policy: implications for policy planning*, Final Report, December 2005.
- Technopolis, *Impact of the EU RTD Framework Programme on the UK*, 2010.
- The Evaluation of Scientific Research: Selected Experiences*, OECD DSTI/STP/SUR(97)9.
- The impact of the 1992 Research Assessment Exercise on higher education institutions in England*, [http://www.hefce.ac.uk/pubs/hefce/1997/m6\\_97.htm](http://www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm).
- The Penguin Modern Guide to Synonyms and Related Words*, London 1971.
- Torras A. Estruch, *Catalan companies in the European Framework Programme* 2009.
- Trow M., *Trust, Markets and Accountability in Higher Education: A Comparative Perspective*, „Higher Education Policy” 1996, 9, 4, s. 309–324.
- U-Multirank. *Interim Progress Report Design Phase of the Project ‘Design and Testing the Feasibility of a Multi-dimensional Global University Ranking’*, CHERPA-Network January 2010, [www.u-multirank.eu/UMR\\_IR\\_0110.pdf](http://www.u-multirank.eu/UMR_IR_0110.pdf)
- Van der Wall E., *Peer review under review: room for improvement?*, Neth Heart J. 2009 May; 17(5).
- Van Raan A.F.J. et al., *The new set of bibliometric indicators of CWTS*, [w:] *11<sup>th</sup> International Conference on Science and Technology Indicators “Creating Value for Users”* 9–11 September 2010. Book of Abstracts; <http://info.scopus.com/journalmetrics/>.
- Van Steen J., *The use of S&T Indicators in science policy, how can they matter?*, „Research Evaluation” 1995, vol. 5, no. 2.
- Van Thiel S. i Leeuw F.L., *The Performance Paradox in the Public Sector*, „Public Performance & Management Review” 2002, vol. 25, no. 3
- Vavříková L., *Assessment Of Fp6 In The Czech Republic. Impact Assessment of the 6th European RTD Framework Programme*, 2010.
- Velho L., *Knowledge frontiers: public sector research and industrial innovation in biotechnology, engineering ceramics, and parallel computing*, Clarendon Press, New York, Oxford University Press, Oxford 1995.

- Vessuri H., *Science, politics, and democratic participation in policy-making: a Latin American view*, „Technology in Society” 2003, vol. 25, no. 2, s. 265–266.
- Von Raggamby A., *The Role of Tools in Impact Assessments*, [w:] OECD, *Conducting Sustainability Assessments*, 2010.
- Weingart P., *Impact of bibliometrics upon the science system: inadvertent consequences?*, „Scientometrics”, January 2005, vol. 62, no. 1.
- Wellings P. i Winzer P., *Mapping Research Excellence: exploring the links between research excellence and research funding Policy*, The 1994 Group, September 2011.
- Whitley R., *Chapter 1 Changing Governance of the Public Sciences*, [w:] *The Changing Governance of the Sciences, Sociology of the Sciences Yearbook*, Dordrecht, Springer 2008.
- Williams D. i Rank A.D., *Measuring the economic benefits of research and development: the current state of the art*, „Research Evaluation”, April 1998, vol. 7, no. 1.
- Wissenschaftsrat, *Evaluation of large infrastructure projects*, 2013.
- Wood F.Q., *The Peer Review Process*, Canberra Australian Govt. Pub. Service, 1997.
- Wood F.Q.i Meek V.L., red., *Research Grants. Management and Funding*, Australian Academy of Science Canberra 1993.
- Ziman J., *Prometheus Bound. Science in a dynamic steady state*, Cambridge Cambridge University Press 1994.