# Pułapki pomiaru i iluzja sukcesu w świetle książki Setha Godina The Knot Problems Can Be Solved


**AUTOR**

Fundacja Dobre Państwo



## STRESZCZENIE / ABSTRACT

Artykuł analizuje problem utożsamiania rzeczywistych celów z ich niedoskonałymi wskaźnikami zastępczymi (false proxy). Autor ostrzega przed sytuacją, w której narzędzia pomiarowe przestają służyć rozpoznawaniu rzeczywistości, a zaczynają samodzielnie kształtować hierarchię celów i systemy bodźców. Przywołując prawo Goodharta oraz teorię skal Stevensa, tekst wskazuje na ryzyko optymalizacji parametrów, które nie przekładają się na realną wartość społeczną czy kompetencyjną. Kluczowym problemem jest błędna operacjonalizacja zjawisk społecznych, gdzie korelacja statystyczna jest mylnie brana za dowód trafności pomiaru. W efekcie organizacje mogą wykazywać wzrost sprawności w obszarach nieistotnych, podczas gdy rzeczywista użyteczność ich działań maleje.

The article analyzes the problem of equating real goals with their imperfect proxy indicators (false proxies). The author warns against a situation in which measurement tools cease to serve as means of recognizing reality and instead begin to independently shape the hierarchy of goals and incentive systems. Drawing on Goodhart's Law and Stevens' power law, the text points to the risk of optimizing parameters that do not translate into real social or competency value. The key issue is the flawed operationalization of social phenomena, where statistical correlation is mistaken for evidence of measurement validity. As a result, organizations may demonstrate increased efficiency in irrelevant areas while the actual utility of their actions declines.


Artykuł analizuje niebezpieczeństwo poznawczego i instytucjonalnego uwikłania, które powstaje w momencie utożsamienia niedoskonałych mierników (tzw. false proxies) z rzeczywistymi celami organizacji. Autor stawia tezę, że zgodnie z prawem Goodharta i

prawem Campbella, wskaźnik przestaje być rzetelnym narzędziem obserwacji, gdy staje się celem sterowania lub podstawą systemu nagród. Mechanizm ten prowadzi do systemowej deformacji zachowań: uczestnicy organizacji, zamiast dążyć do realizacji misji (np. realnej poprawy zdrowia pacjentów czy kompetencji uczniów), optymalizują swoje działania pod kątem uzyskania korzystnych wyników w mierzalnych parametrach. Tekst dowodzi, że presja na efektywność formalną często prowadzi do paradoksalnego spadku użyteczności społecznej działań, gdyż to, co łatwo policzalne, wypiera to, co istotne, lecz trudne do zmierzenia. Rozwiązaniem tego problemu jest przejście od jednowymiarowych rankingów ku wielowymiarowej ocenie rezultatów, która rozróżnia skuteczność operacyjną od aksjologicznej.

**SŁOWA KLUCZOWE**

prawo Goodharta | false proxy | trafność konstruktu | prawo Campbella | teoria relacji pryncypał-agent | paradoks Simpsona | cream skimming | sieć nomologiczna | operacjonalizacja zjawisk społecznych | endogeniczność pomiaru | model Avedisa Donabediana | wartość dodana w edukacji | selekcja wartości ekstremalnych | regresja do średniej

# Wskaźniki zastępcze nie są tożsame z rzeczywistymi celami

Błędne wskaźniki sukcesu – epistemologia pomiaru, prawo Goodharta, teoria bodźców i instytucjonalna produkcja pozornej skuteczności

Jednym z paradoksów współczesnej cywilizacji organizacyjnej pozostaje możliwość systematycznego zwiększania sprawności działań, których rzeczywista użyteczność społeczna nie wzrasta, a w pewnych okolicznościach nawet maleje. Instytucje dysponują dziś niezwykle rozwiniętymi narzędziami gromadzenia danych, analizowania procesów i monitorowania zachowań, lecz wzrost zdolności pomiarowych nie gwarantuje proporcjonalnego rozwoju wiedzy o znaczeniu obserwowanych zjawisk.

Można bowiem precyzyjnie mierzyć cechę, która pozostaje jedynie pośrednio związana z przedmiotem rzeczywistego zainteresowania, poprawnie obliczać wskaźnik oparty na wadliwym

założeniu teoretycznym albo optymalizować określony parametr w sposób prowadzący do pogorszenia innych, znacznie istotniejszych właściwości systemu. W rezultacie powstaje szczególna postać instytucjonalnego uwikłania poznawczego, w której rozwój procedur ewaluacyjnych sprzyja utrwalaniu błędnego rozumienia sukcesu. Paradoks ten staje się szczególnie niebezpieczny wtedy, gdy wskaźnik przestaje służyć rozpoznawaniu rzeczywistości, a zaczyna samodzielnie określać hierarchię celów, strukturę bodźców ekonomicznych, sposób rozdzielania zasobów oraz kryteria oceny człowieka i organizacji. Autor książki ujął ten problem poprzez rozróżnienie pomiędzy rzeczywistym celem działania a jego niedoskonałym wskaźnikiem zastępczym, określanym jako false proxy.

Szczególnej uwagi wymagają przypadki utożsamiania znaczenia działalności z wielkością zgromadzonego finansowania, oceniania rozwoju ucznia wyłącznie na podstawie ocen, przyjmowania liczby odbiorców reklamy za równoważnik jej skuteczności czy traktowania pozycji książki na liście sprzedaży jako rozstrzygającego dowodu jej wartości intelektualnej. W każdym z tych przykładów obserwowany parametr posiada określoną wartość informacyjną, lecz jego znaczenie zależy od rzeczywistej relacji z celem nadrzędnym. Środki pozyskane przez organizację mogą umożliwiać finansowanie działań, ale same w sobie nie dowodzą poprawy położenia beneficjentów; wynik szkolny może informować o opanowaniu części materiału, lecz nie stanowi wyczerpującego obrazu kompetencji poznawczych.

Zasięg reklamy bywa koniecznym warunkiem oddziaływania na odbiorców, ale nie przesądza o zmianie ich zachowania. Wskaźnik staje się poznawczo niebezpieczny dopiero wówczas, gdy jego częściowa trafność zostaje błędnie przekształcona w roszczenie do reprezentowania całej wartości badanego zjawiska. Zasadniczym punktem wyjścia do naukowego rozpoznania tego mechanizmu jest filozofia pomiaru. Pomiar bowiem nie stanowi prostego przeniesienia rzeczywistości do postaci liczbowej, lecz wymaga uprzedniego określenia, jaka właściwość ma zostać przedstawiona za pomocą określonej skali, jakie relacje empiryczne uzasadniają zastosowaną procedurę i jakie wnioski można prawomocnie wyprowadzać z wyników.

W klasycznej teorii pomiaru kluczowe miejsce zajmuje praca Stanleya Smitha Stevensa On the Theory of Scales of Measurement (1946), w której przedstawiono rozróżnienie skal nominalnych, porządkowych, interwałowych i ilorazowych. Choć klasyfikacja ta nie rozwiązuje wszystkich współczesnych problemów teorii pomiaru, a jej statystyczne konsekwencje pozostają przedmiotem dyskusji, ujawnia ona zasadniczą różnicę między przypisywaniem obserwacjom liczb a uprawnionym wykonywaniem na tych liczbach określonych operacji.

Ranking może poprawnie porządkować obiekty według przyjętego kryterium, nie uprawniając jednak do stwierdzenia, że różnica między pierwszą a drugą pozycją jest równoważna różnicy

między dziewiątą a dziesiątą. Podobnie ocena punktowa może odzwierciedlać wynik testu, lecz bez dodatkowego uzasadnienia nie pozwala interpretować każdego przyrostu punktów jako identycznej zmiany rzeczywistej kompetencji.

Problem ten nabiera większego znaczenia w odniesieniu do zjawisk społecznych i organizacyjnych, które nie są bezpośrednio obserwowalne tak jak długość przedmiotu czy masa substancji. Zaufanie, jakość życia, sprawiedliwość proceduralna czy zdolność organizacji do uczenia się wymagają konstrukcji teoretycznej i empirycznej operacjonalizacji. W praktyce badawczej wykorzystuje się zatem zestawy pytań, obserwowane zachowania czy dane administracyjne, zakładając związek tych wskaźników z badaną właściwością. Sam fakt istnienia korelacji statystycznej nie rozstrzyga jednak o trafności takiej reprezentacji. Wskaźnik może reagować zarówno na zmianę zjawiska, jak i na czynniki uboczne. Jeżeli placówka edukacyjna uzyskuje wyższe wyniki egzaminacyjne po zmianie metod przygotowania uczniów, należy ustalić, czy poprawa odzwierciedla szerszy rozwój kompetencji, czy przede wszystkim wzrost znajomości formatu zadań.

Analogicznie wzrost deklarowanego zadowolenia z usług publicznych może wynikać z rzeczywistej poprawy doświadczeń obywateli, lecz może również pozostawać związany ze zmianą struktury respondentów, sposobu przeprowadzania badania albo oczekiwań dotyczących tego, co należy uznać za usługę dobrej jakości.

## Trafność konstruktu jako fundament rzetelnego pomiaru

Kluczową rolę w rozwiązaniu tego problemu odgrywa koncepcja trafności konstruktu, rozwinięta przez Lee Cronbacha i Paula Meehla w pracy Construct Validity in Psychological Tests z 1955 roku. Autorzy zakwestionowali możliwość sprowadzenia trafności narzędzia pomiarowego do pojedynczego współczynnika korelacji z wybranym kryterium; zaproponowali zamiast tego interpretację wyników w ramach szerszej sieci nomologicznej. Sieć ta obejmuje relacje między konstruktami teoretycznymi, obserwowalnymi zmiennymi a przewidywaniami poddawanymi empirycznej weryfikacji. Jeśli test ma mierzyć kompetencję poznawczą, należy wykazać, że jego wyniki pozostają w uzasadnionych relacjach z innymi przejawami tej kompetencji i nie są wyjaśniane wyłącznie przez czynniki uboczne, jak choćby znajomość konkretnego formatu testu. Trafność nie jest przy tym raz na zawsze uzyskanej właściwości samego zestawu pytań – zależy bowiem od interpretacji wyników oraz celu, któremu mają one służyć. Instrument przydatny do przesiewowej oceny zjawiska nie musi posiadać dostatecznej trafności, by rozstrzygać o indywidualnych prawach, dostępie do świadczeń czy przyszłości zawodowej konkretnej osoby.

Szczególnie istotna jest późniejsza polemika Denny'ego Borsbooma, Gideona Mellenbergha i Jaapa van Heerdena z artykułu The Concept of Validity z 2004 roku. Autorzy zaproponowali podejście, według którego trafność pomiaru wymaga istnienia mierzonej właściwości oraz przyczynowego wpływu jej zmienności na uzyskiwane wyniki. Perspektywa ta przesuwa punkt ciężkości z samego współwystępowania obserwacji na mechanizm, za którego pośrednictwem dana właściwość jest reprezentowana przez wynik pomiarowy. Choć stanowisko to nie rozstrzyga wszystkich problemów filozofii pomiaru, ujawnia istotną trudność metodologiczną: wskaźnik może być silnie skorelowany z właściwym celem, mimo że jego zmiana wywołana bezpośrednią manipulacją nie prowadzi do zmiany tego celu. Właśnie ta różnica między relacją statystyczną a przyczynową stanowi jedno z głównych źródeł błędów w projektowaniu systemów oceny. Rozróżnienie to można zilustrować przykładem działalności organizacji społecznej.

## Miernik przestaje być rzetelny, gdy staje się celem

Jeśli organizacje dysponujące większym finansowaniem przeciętnie osiągają lepsze rezultaty, nie oznacza to automatycznie, że każda metoda zwiększania funduszy prowadzi do proporcjonalnej poprawy realizacji misji. Zależność ta może wynikać z wielkości organizacji, jakości zarządzania, rozpoznawalności, dostępu do kompetencji lub specyfiki rozwiązywanych problemów. Choć wzrost przychodów może umożliwić rozwój wartościowej działalności, może on zostać pochłonięty przez koszty pozyskiwania kolejnych środków, administrację lub przedsięwzięcia podporządkowane wymaganiom grantodawców.

Dlatego ocena rzeczywistego sukcesu wymaga rozpoznania łańcucha przyczynowego łączącego zasoby finansowe z konkretnymi działaniami, rezultatami bezpośrednimi i trwałą zmianą społeczną. Pomylenie zasobu z rezultatem prowadzi do szczególnie charakterystycznej deformacji organizacyjnej: instytucja zaczyna traktować zdolność pozyskiwania środków jako dowód wartości własnego oddziaływania, podczas gdy samo oddziaływanie pozostaje odrębnym przedmiotem badania. W rozwoju ekonomii politycznej i nauk o zarządzaniu podobna trudność została dostrzeżona w kontekście wykorzystywania wskaźników statystycznych do sterowania zachowaniami uczestników systemu. Charles Goodhart, analizując w latach siedemdziesiątych XX wieku problemy prowadzenia polityki pieniężnej, zwrócił uwagę na nietrwałość zależności empirycznych traktowanych jako bezpośrednie instrumenty regulacji. Historyczny związek między określoną miarą podaży pieniądza a innymi zmiennymi makroekonomicznymi może ulec zmianie, gdyby podmioty gospodarcze zaczęły dostosowywać swoje postępowanie do nowej reguły polityki pieniężnej.

Późniejsza popularna formuła prawa Goodharta – według której wskaźnik przestaje być dobrym miernikiem, gdy staje się celem – stanowi uproszczoną parafrazę szerszej intuicji dotyczącej niestabilności regularności statystycznych poddanych presji sterowania. Warto zachować tę różnicę, gdyż pierwotny problem nie dotyczył wyłącznie świadomego manipulowania danymi, lecz endogenicznej zmiany zachowań uczestników rynku reagujących na sposób wykorzystania konkretnego parametru. Miernik trafnie opisujący rzeczywistość przed interwencją może utracić wartość prognostyczną właśnie dlatego, że stał się podstawą podejmowania decyzji. W kontekście polityki pieniężnej mechanizm ten wyjaśnia ograniczenia prostych reguł sterowania gospodarką w oparciu o historyczne zależności. Jeśli bank centralny przyjmie określony agregat pieniężny jako główny cel operacyjny, instytucje finansowe i uczestnicy rynku mogą zmienić sposób zarządzania aktywami, konstruowania instrumentów finansowych czy utrzymywania płynności. W rezultacie wcześniej stabilna relacja między wybranym agregatem a innymi zmiennymi gospodarczymi może ulec osłabieniu. Nie oznacza to jednak, że każdy agregat pieniężny traci znaczenie lub że modele makroekonomiczne stają się bezużyteczne.

Wskazuje to raczej na cechę systemów refleksyjnych: uczestnicy nie są biernymi obiektami pomiaru, lecz poznają reguły i dostosowują do nich swoje decyzje. Właśnie dlatego miernik użyteczny jako narzędzie obserwacji nie musi być równie skuteczny jako bezpośredni cel sterowania. Przejście od zastosowania diagnostycznego do regulacyjnego zmienia bowiem nie tylko funkcję wskaźnika, ale i strukturę zachowań, które ten wskaźnik ma reprezentować. Zbliżoną, lecz szerszą problematykę rozwinął Donald T.

## Instytucjonalna presja przekształca pomiar w narzędzie kształtowania rzeczywistości

Koncepcję tę rozwinął Campbell w pracy Assessing the Impact of Planned Social Change z 1979 roku. Autor zajmował się metodologią ewaluacji programów społecznych oraz trudnościami związanymi z przypisywaniem obserwowanych zmian konkretnym interwencjom. Jego analiza doprowadziła do rozpoznania szczególnego zagrożenia, jakie niesie wykorzystywanie ilościowych wskaźników społecznych jako podstawy decyzji administracyjnych i instytucjonalnych. Im większe znaczenie parametr uzyskuje dla podziału zasobów, oceny organizacji czy pozycji jednostek, tym silniejsza staje się presja prowadząca do zniekształcania procesu pomiarowego lub zachowań, które ten pomiar miał badać. Prawo Campbella nie oznacza jednak, że wskaźniki ilościowe są z natury niewiarygodne, ani że ich stosowanie nieuchronnie prowadzi do nadużyć.

Właściwa interpretacja tego prawa dotyczy ryzyka rosnącego wraz z konsekwencjami decyzyjnymi przypisanymi miernikowi — szczególnie gdy uczestnicy systemu mogą wpływać na wynik oceny. W takich warunkach pomiar przestaje być neutralnym sposobem obserwacji rzeczywistości, a staje się elementem jej instytucjonalnego kształtowania. Goodhart i Campbell rozpoznali zatem dwa częściowo pokrywające się mechanizmy, których nie należy jednak całkowicie utożsamiać.

W pierwszym przypadku kluczowa jest niestabilność relacji statystycznych wynikająca z wykorzystania ich do sterowania systemem. W drugim akcent pada na presję korupcyjną i zniekształcenia procesów społecznych, wywołane nadaniem wskaźnikowi rangi narzędzia oceny i kontroli. Oba ujęcia można połączyć w szerszej teorii endogeniczności pomiaru, zgodnie z którą instytucjonalne zastosowanie wskaźnika przekształca warunki jego późniejszego wytwarzania. Należy przy tym odróżnić rzeczywistą manipulację danymi od sytuacji, w której zachowanie uczestników zmienia się zgodnie z formalnymi regułami systemu, lecz w sposób nieprzewidziany przez projektantów. Szkoła fałszująca wyniki egzaminacyjne dopuszcza się bowiem czynu jakościowo innego niż szkoła, która legalnie zwiększa liczbę ćwiczeń pod konkretny format testu.

Obie praktyki mogą zniekształcać interpretację rezultatów, lecz mają odmienny status prawny i moralny. Trafna diagnoza powinna zatem rozpoznawać konkretny mechanizm, zamiast posługiwać się jedną kategorią "grania pod wskaźnik" dla wszystkich form adaptacji instytucjonalnej. Współczesna refleksja nad prawem Goodharta pozwala na wskazanie bardziej szczegółowych źródeł tych zniekształceń. David Manheim i Scott Garrabrant w pracy Categorizing Variants of Goodhart's Law z 2018 roku zaproponowali rozróżnienie między odmianami regresyjnymi, ekstremalnymi, przyczynowymi i adwersarialnymi. Klasyfikacja ta jest propozycją analityczną z literatury dotyczącej optymalizacji i bezpieczeństwa sztucznej inteligencji, a nie empirycznie potwierdzoną, wyczerpującą taksonomią wszystkich błędów pomiarowych. Jej znaczenie polega na ukazaniu, że rozbieżność między celem a wskaźnikiem może powstawać na kilka odmiennych sposobów, z których każdy wymaga innego rodzaju zabezpieczenia. Przykładowo, regresyjna odmiana prawa Goodharta wynika z niedoskonałego związku między wskaźnikiem a właściwą wartością.

## Statystyczne i przyczynowe pułapki selekcji na podstawie mierników

Jeśli miernik stanowi sumę informacji o badanej właściwości oraz składnika przypadkowego lub ubocznego, wybieranie jednostek osiągających najwyższe wartości może prowadzić do nadreprezentacji przypadków, w których korzystny wynik został częściowo wytworzony przez czynnik niezwiązany z celem. W statystyce zjawisko to wiąże się z selekcją wartości ekstremalnych i

regresją do średniej. Jeżeli badacz lub organizacja wybiera najlepsze rezultaty na podstawie jednorazowego testu, część pozornie wyjątkowych osiągnięć może wynikać z losowej zmienności, sprzyjających okoliczności albo błędu pomiarowego. Powtórny pomiar może wówczas przynieść słabsze wyniki bez rzeczywistego pogorszenia kompetencji badanych osób. Analogicznie fundusz inwestycyjny wybierający zarządzających wyłącznie na podstawie najwyższych historycznych stóp zwrotu naraża się na pomylenie trwałej umiejętności z wpływem korzystnego przypadku.

Mechanizm ten nie wymaga oszustwa ani świadomego dostosowywania zachowań do miernika; wynika on z samej statystycznej struktury selekcji. Odmiana ekstremalna występuje wtedy, gdy wskaźnik zachowuje użyteczność w zwykłych warunkach, ale zostaje wykorzystany do oceny przypadków tak odległych od pierwotnego zakresu obserwacji, że wcześniejszy związek przestaje być wiarygodny. Model prognostyczny może poprawnie przewidywać wyniki typowych gospodarstw domowych, nie zachowując jednak podobnej trafności wobec przypadków o wyjątkowych właściwościach ekonomicznych. System oceniania projektów może poprawnie różnicować przedsięwzięcia spełniające standardowe wymagania, lecz nie radzić sobie z rozwiązaniami radykalnie nowatorskimi, których wartości nie da się przedstawić za pomocą historycznie utrwalonych kryteriów.

W algorytmicznej optymalizacji podobne ryzyko pojawia się wtedy, gdy model poszukuje rozwiązań maksymalizujących ocenę poza rozkładem przykładów, na których został wytrenowany. Nie jest to klasyczny błąd losowy, lecz problem ekstrapolacji i ograniczonej ważności empirycznej modelu. Środkiem zaradczym staje się określanie zakresu zastosowania wskaźnika, badanie niepewności predykcji oraz ograniczanie decyzji opartych na wynikach pochodzących z obszarów niedostatecznie reprezentowanych w danych. Przyczynowa odmiana zniekształcenia pojawia się natomiast wtedy, gdy podmiot wykorzystuje obserwowaną korelację tak, jakby reprezentowała mechanizm, który można bezpośrednio uruchomić przez manipulowanie miernikiem. Wysoka jakość edukacji może sprzyjać osiąganiu wysokich wyników egzaminacyjnych, ale zwiększanie czasu poświęcanego na ćwiczenie samych testów nie musi prowadzić do proporcjonalnego rozwoju wszystkich kompetencji składających się na jakość edukacji.

Dobrze prosperujące przedsiębiorstwa mogą posiadać znaczną rozpoznawalność marki, lecz finansowanie działań zwiększających tę rozpoznawalność nie zagwarantuje poprawy jakości oferowanych produktów. Organizacje skutecznie realizujące misję społeczną mogą przyciągać darczyńców, ale nadmierna koncentracja na pozyskiwaniu darowizn nie musi prowadzić do równie wysokiego wzrostu społecznej użyteczności przedsięwzięcia.

# Mechanizmy manipulacji i pułapka selektywnej mierzalności

W każdym przypadku związek między celem a jego wskaźnikiem może mieć kierunek przyczynowy inny niż założył decydent lub wynikać z oddziaływania czynników zewnętrznych. Kluczowe staje się zatem odróżnienie predykcyjnej przydatności zmiennej od skuteczności interwencji mającej na celu zmianę jej wartości. Odmiana adwersarialna obejmuje sytuacje, w których uczestnicy świadomie wykorzystują słabości systemu oceny — manipulując danymi, sposobem klasyfikowania zdarzeń lub samym zachowaniem — aby uzyskać korzystną ocenę bez proporcjonalnego wzrostu rzeczywistej wartości. W systemach informatycznych przejawia się to poprzez manipulowanie danymi wejściowymi, w gospodarce przez selektywne przedstawianie wyników, a w administracji przez dostosowywanie rejestracji spraw do parametrów podlegających kontroli.

Zastosowanie kategorii adwersarialności wymaga jednak rozpoznania zamiaru i możliwości strategicznego oddziaływania na system. Nie każda rozbieżność między wskaźnikiem a celem wynika ze świadomej manipulacji; niektóre niekorzystne rezultaty powstają bez udziału osób zainteresowanych wykorzystaniem luk w pomiarze. W tym sensie teoria błędnych wskaźników nie może sprowadzać się do moralizatorskiej krytyki osób osiągających wysokie wyniki w ramach istniejących reguł. Powinna przede wszystkim badać, dlaczego konkretne reguły dopuszczają powstawanie rezultatów sprzecznych z misją systemu.

Ekonomia organizacji dostarcza dodatkowego uzasadnienia dla takiej interpretacji poprzez teorię relacji pryncypał-agent. W klasycznym modelu agencji podmiot zlecający działanie nie posiada pełnej wiedzy o wysiłku, kompetencjach czy rzeczywistych intencjach wykonawcy. Konstruuje więc mechanizmy kontraktowe, które mają zbliżyć interesy obu stron. Problem komplikuje się jednak, gdy praca obejmuje wiele zadań różniących się stopniem mierzalności i wpływem na wynik końcowy.

Bengt Holmström i Paul Milgrom w pracy z 1991 roku pt. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design" wykazali znaczenie tego zjawiska dla projektowania wynagrodzeń, podziału obowiązków i struktury własności. Jeśli część kluczowych czynności jest łatwo mierzalna, a pozostałe trudniej zaobserwować, silne uzależnienie zapłaty od pierwszej grupy może prowadzić do przesunięcia wysiłku z zadań niemierzonych ku zadaniom premiowanym. Paradoksalnie zatem zwiększenie intensywności bodźca ekonomicznego może pogorszyć realizację nadrzędnego celu organizacyjnego.

Znaczenie tej teorii obrazuje przykład nauczyciela, którego obowiązki obejmują zarówno przygotowanie uczniów do egzaminów, jak i rozwijanie ich samodzielności myślenia, kompetencji społecznych czy zdolności pracy zespołowej. Jeśli system wynagradzania i oceny zawodowej zostanie silnie powiązany z wynikami standaryzowanych testów, nauczyciel zyskuje ekonomiczny i reputacyjny motyw do poświęcania większej ilości czasu na działania bezpośrednio poprawiające te wyniki. Nie musi to oznaczać pogorszenia pozostałych kompetencji, o ile przygotowanie egzaminacyjne pozostaje spójne z programem kształcenia. Ryzyko pojawia się jednak wtedy, gdy rozwijanie umiejętności trudniejszych do pomiaru wymaga tych samych ograniczonych zasobów czasu i uwagi. Silne premiowanie jednego rezultatu zmienia wówczas relatywną opłacalność poszczególnych czynności; system może odnotować poprawę w obszarze monitorowanym kosztem zaniedbania innych aspektów pracy. Holmström i Milgrom wskazywali, że w określonych warunkach słabsze bodźce finansowe lub większy udział wynagrodzenia stałego mogą być bardziej uzasadnione niż silne premie oparte na wąskim wskaźniku, właśnie by chronić jakość zadań trudnych do obserwacji.

## Sprzeczność między nagradzanymi wskaźnikami a deklarowanymi celami

Podobny mechanizm występuje w pracy lekarzy, prawników, urzędników, menedżerów i pracowników organizacji społecznych, gdyż każda z tych profesji obejmuje działania o odmiennych właściwościach pomiarowych. Pracę lekarza można częściowo oceniać przez liczbę przeprowadzonych konsultacji, lecz parametr ten nie oddaje w pełni trafności diagnostycznej, jakości komunikacji, adekwatności leczenia czy bezpieczeństwa pacjentów. Działalność kancelarii można mierzyć liczbą prowadzonych spraw albo wartością uzyskanych roszczeń, ale takie wskaźniki nie muszą odzwierciedlać jakości doradztwa, skuteczności zapobiegania sporom i ochrony interesów klienta. W administracji łatwo zmierzyć liczbę wydanych decyzji, lecz sama wydajność ilościowa nie przesądza o ich zgodności z prawem, zrozumiałości i odporności na skuteczne zaskarżenie. Organizacja społeczna może wykazywać rosnącą liczbę beneficjentów, ale liczba osób objętych działaniem nie musi być równoważna liczbie osób, których położenie rzeczywiście uległo trwałej poprawie. W każdym przypadku zastosowanie pojedynczego wskaźnika do całościowej oceny pracy wielozadaniowej niesie ryzyko zmiany struktury wysiłku – nawet jeśli uczestnicy zachowują pełną formalną zgodność z przyjętymi regułami.

Z tym problemem bezpośrednio łączy się klasyczna analiza Stevena Kerra, opublikowana w 1975 roku pod tytułem On the Folly of Rewarding A, While Hoping for B. Autor zwrócił uwagę na

sprzeczność, która pojawia się, gdy organizacja deklaruje oczekiwanie określonego zachowania, lecz faktycznie nagradza inne. Instytucja może wymagać współpracy, a jednocześnie premiować wyłącznie indywidualną rywalizację; oczekiwać długofalowej orientacji na rozwój, a zarazem rozliczać pracowników z krótkookresowych rezultatów; deklarować tolerancję dla uzasadnionego eksperymentowania, lecz sankcjonować każde niepowodzenie bez względu na jakość wcześniejszej decyzji.

Praca Kerra ma charakter przede wszystkim analizy organizacyjnej i teoretycznej, a nie uniwersalnego dowodu eksperymentalnego, że każdy bodziec wywołuje przewidywaną reakcję. Pozwala jednak sformułować ważną hipotezę: jeżeli system nagród pozostaje sprzeczny z deklarowaną misją, zachowania uczestników mogą w większym stopniu odpowiadać temu, co jest rzeczywiście premiowane, niż temu, czego instytucja oficjalnie oczekuje.

W świetle teorii Kerra problem błędnych wskaźników przestaje być wyłącznie zagadnieniem epistemologicznym, ponieważ staje się również problemem normatywnej architektury organizacji. Wskaźnik nie służy już jedynie opisywaniu zachowań, lecz wskazuje uczestnikom, które z nich są nagradzane, uznawane za profesjonalne i społecznie wartościowe. Jeżeli przedsiębiorstwo premiuje liczbę zawartych transakcji, może sprzyjać zwiększeniu aktywności sprzedażowej, ale jednocześnie wytwarzać presję na zawieranie umów niedostatecznie odpowiadających potrzebom klientów. Jeśli organizacja medialna wynagradza wyłącznie zdolność przyciągania uwagi, może to prowadzić do wzrostu liczby treści wywołujących silne reakcje emocjonalne, niezależnie od ich jakości informacyjnej. Z kolei jeżeli instytucja akademicka nagradza przede wszystkim liczbę publikacji, może zwiększać intensywność aktywności wydawniczej, ale jednocześnie osłabiać bodźce do podejmowania badań długotrwałych, trudnych i obciążonych wysokim ryzykiem nieuzyskania wyników nadających się do szybkiej publikacji.

Nie istnieje przy tym automatyczna zależność między ilością a niską jakością, ponieważ wysoka produktywność może współistnieć z wysoką wartością pracy. Istotne pozostaje jednak ustalenie, czy konstrukcja systemu oceny premiuje rzeczywiste osiągnięcia, czy przede wszystkim zachowania pozwalające uzyskiwać korzystne wartości wybranych parametrów.

## Przemieszczenie celów i iluzja skuteczności formalnej

W naukach społecznych podobny proces zidentyfikował Robert King Merton, nazywając go przemieszczeniem celów (goal displacement). Analizując biurokrację i jej dysfunkcje w pracy "Bureaucratic Structure and Personality" z 1940 roku, wskazywał on na ryzyko sytuacji, w której reguły i procedury – stworzone przecież jako środki do celu – zyskują samodzielną wartość

normatywną. Wówczas posłuszeństwo wobec reguły może stać się przedmiotem szczególnego uznania, nawet jeśli jej mechaniczne zastosowanie utrudnia osiągnięcie nadrzędnego celu organizacji. Nie oznacza to, że formalne procedury są niepożądane; zapewniają one przecież przewidywalność, równe traktowanie i ochronę przed arbitralnością. Problem pojawia się jednak wtedy, gdy ich przestrzeganie zostaje odłączone od refleksji nad funkcją, której miały służyć, a instytucja traci zdolność rozpoznawania sytuacji wymagających modyfikacji reguł lub innego sposobu ich stosowania. W takim układzie przemieszczenie celów nie musi wynikać ze świadomej manipulacji czy konfliktu interesów.

Może ono rozwijać się jako przewidywalna konsekwencja socjalizacji organizacyjnej, w której zachowania pierwotnie instrumentalne zostają uwewnętrznione jako wartości same w sobie. Kluczową konsekwencją tych rozważań jest konieczność rozróżnienia skuteczności formalnej, operacyjnej, przyczynowej i aksjologicznej. Skuteczność formalna to wykonanie czynności zgodnie z procedurami. Skuteczność operacyjna odnosi się do osiągnięcia bezpośredniego, mierzalnego rezultatu. Skuteczność przyczynowa wymaga wykazania, że rezultat wystąpił dzięki konkretnemu działaniu, a nie tylko równocześnie z nim. Skuteczność aksjologiczna dotyczy natomiast zgodności rzeczywistych konsekwencji z wartościami, które uzasadniały ustanowienie celu. Kategorie te stanowią autorską syntezę porządkującą, a nie powszechnie przyjętą klasyfikację naukową.

Ich analityczna użyteczność pozwala wyjaśnić, dlaczego organizacja może wykazać pełną realizację planu operacyjnego, mimo braku dowodów na rozwiązanie problemu społecznego. Formalne przeprowadzenie szkolenia nie dowodzi nabycia kompetencji, ich nabycie nie gwarantuje wykorzystania w praktyce, a samo wykorzystanie nie przesądza automatycznie o trwałej poprawie sytuacji odbiorców. Wnioskowanie o skuteczności wymaga zatem analizy kolejnych relacji przyczynowych, a nie prostego utożsamiania wykonanej czynności z osiągnięciem wartości nadrzędnej. Przeniesienie tych rozważań na grunt edukacji ujawnia złożone konsekwencje łączenia jakości kształcenia z wynikami testów wysokiej stawki. W klasycznej psychometrii standaryzowany sprawdzian jest narzędziem pozyskiwania informacji o umiejętnościach ucznia, a jego wartość zależy od trafności zadań, rzetelności wyników, poziomu trudności i możliwości porównywania rezultatów.

Sytuacja zmienia się jednak, gdy wyniki egzaminacyjne zaczynają bezpośrednio decydować o finansowaniu placówek, ocenie nauczycieli lub prestiżu instytucji. Powstaje presja na maksymalizację rezultatów testowych, co prowadzi do zwiększenia czasu poświęcanego na ćwiczenia odpowiadające formatowi sprawdzianu. Nie każda taka praktyka jest szkodliwa – znajomość wymagań pomaga uczniom efektywniej wykorzystać nabyte kompetencje. Problem pojawia się wtedy, gdy działania służące poprawie wyniku stają się niezależne od rozwoju

umiejętności, których wynik miał być wskaźnikiem. Daniel Koretz w analizach rozliczalności edukacyjnej wskazywał właśnie na zjawisko inflacji wyników, w którym wzrost punktacji przewyższa rzeczywisty przyrost kompetencji mierzony niezależnymi narzędziami. Przykład ten pozwala precyzyjnie odróżnić rzeczywiste uczenie się od wzrostu skuteczności w rozwiązywaniu konkretnych zadań. Jeśli uczeń osiąga lepszy wynik dzięki głębszemu zrozumieniu matematyki, poprawa wskaźnika odzwierciedla realną zmianę kompetencji.

Jeśli jednak wyższa punktacja wynika głównie z opanowania wąskiego zestawu schematów, ma ona ograniczoną wartość dla zadań wymagających przenoszenia wiedzy do nowych sytuacji. W psychologii uczenia się odpowiada temu zagadnienie transferu – zdolności wykorzystania umiejętności poza pierwotnym kontekstem. Transfer bliski dotyczy sytuacji podobnych do tych, w których kompetencja została wykształcona; transfer daleki odnosi się do zastosowań bardziej odmiennych. Spór o zakres i warunki transferu pozostaje istotną częścią badań nad uczeniem się, a jego wyniki nakazują ostrożność wobec twierdzeń, że sukces w jednym formacie zadania automatycznie świadczy o szerokiej kompetencji poznawczej. Z perspektywy teorii pomiaru kluczowe jest zatem stosowanie narzędzi badających zarówno opanowanie treści programowych, jak i zdolność wykorzystania wiedzy tam, gdzie rozwiązanie nie opiera się wyłącznie na wyuczonym schemacie. Istotnego rozwinięcia dostarcza koncepcja trafności zbieżnej i różnicowej Donalda Campbella i Donalda Fiske'a z 1959 roku.

Zaproponowana przez nich multitrait-multimethod matrix umożliwia analizę, czy odmienne narzędzia mierzące tę samą właściwość dostarczają wyników spójnych, a jednocześnie potrafią odróżnić tę właściwość od innych konstruktów.

## Pułapka rankingów i błąd selekcji w ocenie jakości

Podejście to ujawnia zależność wyniku od zastosowanej metody badania. Jeżeli dwa kwestionariusze deklaratywne wykazują wysoki związek statystyczny, może on częściowo wynikać ze wspólnego sposobu pozyskiwania odpowiedzi, a nie wyłącznie z rzeczywistego podobieństwa mierzonych właściwości. Dlatego wiarygodna ocena jakości określonego zjawiska wymaga zestawienia danych administracyjnych, obserwacji zachowań, badań ankietowych i wyników uzyskiwanych w niezależnych warunkach. Szczególnego znaczenia nabiera to w naukach o organizacji, gdzie wysoka zgodność wielu wskaźników pochodzących z tego samego systemu sprawozdawczego nie musi oznaczać równie wysokiej zgodności z rzeczywistym doświadczeniem uczestników. Przedstawiona perspektywa pozwala lepiej rozpoznać ograniczenia rankingów szkolnych i akademickich. Ranking jest uporządkowaniem jednostek według określonej reguły

agregacji danych, której konstrukcja wymaga wyboru zmiennych, przypisania im wag, ustalenia sposobu normalizacji i określenia zasad traktowania informacji niepełnych. Każdy z tych etapów wiąże się z decyzjami wpływającymi na ostateczną pozycję ocenianej instytucji.

Jeżeli ranking uniwersytetów premiuje głównie rozpoznawalność naukową i liczbę cytowań, dostarcza on użytecznej informacji o aspektach działalności badawczej, lecz nie musi trafnie oddawać jakości kształcenia, znaczenia lokalnego czy dostępności edukacji. Z kolei ranking szkół średnich oparty przede wszystkim na wynikach egzaminacyjnych może odzwierciedlać jakość nauczania, ale również selektywność przyjmowania uczniów, ich wcześniejsze przygotowanie, pozycję społeczno-ekonomiczną rodzin i dostępność dodatkowych zasobów. W takim układzie wysoka pozycja instytucji nie pozwala automatycznie ustalić, jaką część osiągnięć uczniów można przypisać działaniu samej placówki. Powstaje bowiem zasadnicza różnica między poziomem obserwowanego rezultatu a przyczynowym wkładem organizacji w jego wytworzenie.

Zagadnienie to autor książki przedstawia w języku ekonomii edukacji poprzez odróżnienie selekcji od wartości dodanej. Szkoła przyjmująca uczniów o bardzo wysokich kompetencjach początkowych może uzyskiwać znakomite średnie wyniki końcowe, nawet jeśli jej rzeczywisty wpływ na rozwój poszczególnych osób nie przewyższa wpływu innych placówek. Instytucja pracująca z uczniami z istotnymi zaległościami może natomiast osiągać wyniki przeciętnie niższe, mimo że przyrost kompetencji jej wychowanków jest znaczny.

Modele wartości dodanej zmierzają do ograniczenia tego problemu przez uwzględnianie wcześniejszych rezultatów i innych właściwości jednostek, jednak one również podlegają błędom pomiarowym, niepełnej kontroli czynników zakłócających i niestabilności oszacowań. Dlatego ani surowy wynik końcowy, ani szacowany przyrost kompetencji nie powinny być automatycznie utożsamiane z pełną wartością edukacyjną placówki. Trafna ocena wymaga uwzględnienia celu badania, warunków funkcjonowania instytucji oraz stopnia niepewności związanych z wybranymi miarami. Problem rankingu nie polega zatem wyłącznie na jego potencjalnej niesprawiedliwości, ale także na metodologicznej skłonności do redukowania złożonego procesu do porównania jednowymiarowego.

Szczególnie interesującą odmianę tego zagadnienia przedstawia paradoks Simpsona, w którym zależność obserwowana na poziomie danych zagregowanych może przyjmować odmienny, a nawet przeciwny kierunek po podziale zbioru na odpowiednio wyodrębnione grupy. W ocenie instytucji zjawisko to występuje wtedy, gdy porównywane organizacje obsługują populacje istotnie różniące się pod względem ryzyka, potrzeb albo warunków początkowych. Szpital przyjmujący więcej pacjentów w ciężkim stanie może wykazywać wyższą ogólną śmiertelność niż placówka zajmująca

się przypadkami mniej skomplikowanymi, mimo że jakość leczenia w poszczególnych kategoriach ryzyka pozostaje porównywalna albo korzystniejsza.

Podobnie placówka edukacyjna obsługująca uczniów o odmiennym przygotowaniu początkowym może uzyskiwać wyniki trudne do uczciwego porównania z instytucjami posiadającymi bardziej selektywną populację. Nie oznacza to, że dane zagregowane są zawsze błędne ani że poprawne wnioskowanie wymaga nieograniczonego dzielenia populacji na coraz mniejsze grupy. Wskazuje jednak na potrzebę rozpoznawania struktury badanych zbiorowości oraz odróżniania różnic wynikających z jakości działania od tych wynikających z odmiennego składu populacji. W ekonomii i statystyce problem ten łączy się z zagadnieniem doboru próby, endogeniczności oraz identyfikacji przyczynowej.

Jeżeli organizacja może wpływać na skład ocenianej populacji, silne powiązanie nagród z wynikiem końcowym tworzy bodźce do selekcjonowania przypadków łatwiejszych i unikania tych, których obsługa wiąże się z większym ryzykiem niepowodzenia. W ochronie zdrowia podobne ryzyko występuje, gdy system oceny premiuje niską śmiertelność bez uwzględnienia ciężkości przypadków. W edukacji może to prowadzić do preferowania kandydatów o wysokich kompetencjach początkowych, a w pomocy społecznej do koncentrowania działań na osobach, których sytuację najłatwiej poprawić, kosztem tych potrzebujących intensywnego i długotrwałego wsparcia. Zjawisko to bywa określane mianem cream skimming, czyli selekcjonowania przypadków najbardziej obiecujących z punktu widzenia oczekiwanego wyniku.

## Wielowymiarowa ocena rezultatów zapobiega deformacji jakości opieki

Nie każda specjalizacja instytucji stanowi przejaw niewłaściwej selekcji, gdyż organizowanie usług dla konkretnych grup może wręcz podnieść jakość ich realizacji. Problem pojawia się wtedy, gdy konstrukcja systemu oceny tworzy bodźce do unikania osób najbardziej potrzebujących pomocy, choć to właśnie ich sytuacja uzasadnia istnienie programu.

Istotną ilustrację wielowymiarowości oceny rezultatów stanowi amerykański Hospital Readmissions Reduction Program, wprowadzony w ramach reformy ochrony zdrowia z 2010 roku (mechanizmy finansowe zaczęły oddziaływać na szpitale w 2012 roku). Program powiązał konsekwencje finansowe z ponadprzeciętnymi wskaźnikami ponownych hospitalizacji w ciągu trzydziestu dni w określonych grupach pacjentów. Uzasadniano to przekonaniem, że część rehospitalizacji można ograniczyć poprzez lepszą jakość leczenia, koordynację opieki i odpowiednie

przygotowanie pacjentów do funkcjonowania poza szpitalem. Wskaźnik ponownych przyjęć ma zatem wartość informacyjną, jednak jego obniżenie nie jest samo w sobie wystarczającym dowodem poprawy zdrowia populacji. Ponowna hospitalizacja może bowiem wynikać zarówno z błędów w poprzednim leczeniu, jak i z rzeczywistej potrzeby medycznej. W konsekwencji ograniczanie rehospitalizacji jest społecznie korzystne tylko wtedy, gdy dotyczy przyjęć możliwych do uniknięcia bez narażania bezpieczeństwa pacjentów.

Badania nad tym programem ujawniły trudności w interpretowaniu zmian wskaźników jako jednoznacznego dowodu na poprawę lub pogorszenie opieki. W analizie opublikowanej w 2018 roku w JAMA opisano spadek częstości ponownych hospitalizacji po wdrożeniu programu, ale jednocześnie rozważano ryzyko wystąpienia niekorzystnych skutków w zakresie śmiertelności, szczególnie w wybranych kategoriach chorób.

Autorzy analizy wskazywali, że wcześniejsze badania były niejednoznaczne, a wykazanie związku przyczynowego wymaga uwzględnienia zmian w charakterystyce pacjentów czy sposobie kodowania rozpoznań. Sam spadek rehospitalizacji nie pozwala zatem stwierdzić, że mniejszy wskaźnik zawsze oznacza poprawę kliniczną, ale nie dowodzi też automatycznie wzrostu śmiertelności. Wartość metodologiczna tego przykładu polega na konieczności jednoczesnego badania rezultatów bezpośrednich, skutków niezamierzonych i zależności między różnymi wymiarami jakości opieki. Analiza zdrowotna wymaga odróżnienia procesu leczenia od jego wpływu na rzeczywisty dobrostan pacjenta. Liczba badań czy procedur nie jest automatyczną miarą jakości, gdyż niektóre działania mogą być zbędne lub zbyt ryzykowne. Z drugiej strony ograniczenie świadczeń nie musi oznaczać poprawy efektywności, jeśli prowadzi do opóźnienia diagnozy.

W medycynie wypracowano zatem rozróżnienie między wskaźnikami struktury, procesu i wyników, znane m.in. z modelu Avedisa Donabediana. Struktura dotyczy zasobów i organizacji, proces – faktycznie wykonywanych czynności, a wyniki – konsekwencji zdrowotnych. Żaden z tych wymiarów nie jest wystarczający: odpowiednia struktura nie gwarantuje właściwego procesu, a poprawnie wykonany proces nie zawsze przynosi oczekiwany wynik ze względu na biologiczną zmienność i niezależne czynniki ryzyka. Zestawienie tych poziomów pozwala jednak uniknąć błędu polegającego na utożsamianiu nakładów z rezultatami lub rezultatów z całkowitą jakością działalności medycznej. Z perspektywy prawa i administracji publicznej problem wskaźników przybiera inną postać, gdyż skuteczność instytucjonalna musi być oceniana nie tylko przez rezultat materialny, lecz także przez legalność, bezstronność, proporcjonalność i poszanowanie gwarancji proceduralnych.

## Wartość aksjologiczna wykracza poza mierzalne rezultaty

Sąd rozpoznający większą liczbę spraw może przyczyniać się do ograniczenia zaległości, jednak sama liczba zakończonych postępowań nie rozstrzyga o jakości orzeczeń. Organ administracji wydający decyzje szybciej niż dotychczas może poprawiać dostępność usług publicznych, lecz przyspieszenie to nie powinno odbywać się kosztem rzeczywistej możliwości uczestniczenia strony w postępowaniu i przedstawiania argumentów. Z kolei instytucja kontrolna może zwiększać liczbę wykrywanych nieprawidłowości, przy czym wyższy wskaźnik stwierdzonych naruszeń może oznaczać zarówno wzrost ich rzeczywistej częstotliwości, jak i po prostu skuteczniejsze ich ujawnianie.

W każdym z tych przypadków pomiar wymaga odróżnienia rezultatu postępowania od funkcji, której to postępowanie miało służyć. Jeżeli system premiuje wyłącznie szybkość i liczbę rozstrzygnięć, może nieodpowiednio uwzględniać koszty proceduralne ponoszone przez strony oraz znaczenie stabilności, przewidywalności i prawidłowego uzasadniania decyzji. Szczególnie problematyczne pozostaje utożsamianie odsetka korzystnych rozstrzygnięć z jakością działalności profesjonalnej. Kancelaria prawna może wykazywać wysoki udział wygranych spraw, jeśli przyjmuje przede wszystkim przypadki o znacznym prawdopodobieństwie powodzenia.

Inny podmiot może osiągać niższy udział sukcesów, ponieważ podejmuje się reprezentowania osób znajdujących się w trudniejszej sytuacji dowodowej albo prawnej. Analogicznie skuteczność instytucji ochrony praw może być przedstawiana przez liczbę pozytywnie zakończonych postępowań, choć istotna część jej społecznej wartości może polegać na zapobieganiu naruszeniom, podnoszeniu standardów działania albo budowaniu zdolności obywateli do korzystania z własnych uprawnień. Działania prewencyjne są szczególnie trudne do oceny, gdyż ich sukces może polegać na niewystąpieniu zdarzenia, które bez interwencji prawdopodobnie miałoby miejsce. W takim przypadku sama obserwacja liczby zdarzeń nie dostarcza pełnej informacji o skuteczności, ponieważ konieczne staje się oszacowanie kontrfaktycznego poziomu ryzyka. Z perspektywy metodologicznej jest to jeden z najważniejszych powodów, dla których organizacji świadczących usługi publiczne nie powinno się oceniać wyłącznie na podstawie liczby wykonanych czynności.

> W świecie zdominowanym przez dyktat danych ryzykujemy, że stworzymy instytucje perfekcyjnie sprawne w generowaniu liczb, lecz całkowicie bezużyteczne w realizacji swoich misji. Czy potrafimy jeszcze odróżnić sukces mierzony w tabelach od rzeczywistej wartości, która z natury wymyka się każdej skali? Być może największym błędem

> współczesnego zarządzania jest wiara, że to, czego nie da się zmierzyć, nie istnieje lub nie ma znaczenia.

## Inspiracje

[1] 


**"The Knot Problems Can Be Solved"** Seth Godin

*Hachette UK | ISBN: 9781399645072*

**Seth Godin** (ur. 10 lipca 1960) to amerykański autor, przedsiębiorca i mówca publiczny, powszechnie znany ze swoich wpływowych koncepcji z zakresu marketingu, przywództwa i rozpowszechniania idei. Po uzyskaniu tytułu licencjata z informatyki i filozofii na Uniwersytecie Tufts w 1982 roku oraz tytułu MBA w Stanford Graduate School of Business w 1984 roku, założył **Seth Godin** Productions, jedną z pierwszych firm zajmujących się pakowaniem książek. W 1995 roku założył Yoyodyne, internetową firmę marketingu bezpośredniego, przejętą później przez Yahoo!, w której Godin pełnił funkcję wiceprezesa ds. marketingu bezpośredniego. Jako płodny autor i bloger, Godin spopularyzował przełomowe zasady biznesowe, takie jak marketing za przyzwoleniem i wyróżnianie się. Założył również platformę współpracy Squidoo oraz internetowe warsztaty przywódcze altMBA. Jego prace głęboko zmieniły współczesne praktyki w zakresie angażowania klientów, pozycjonowania marki i przywództwa organizacyjnego.

***Seth Godin*** *(born July 10, 1960) is an American author, entrepreneur, and public speaker widely recognized for his influential concepts in marketing, leadership, and the spread of ideas. After earning a Bachelor of Science in computer science and philosophy from Tufts University in 1982 and an MBA from Stanford Graduate School of Business in 1984, he founded* ***Seth Godin*** *Productions, an early book packaging company. In 1995, he launched Yoyodyne, an online direct marketing firm later acquired by Yahoo!, where Godin served as Vice President of Direct Marketing. A prolific author and daily blogger, Godin popularized landmark business principles including permission marketing and remarkability. He also founded the collaborative platform Squidoo and the altMBA online leadership workshop. His writings have profoundly transformed modern practices in customer engagement, brand positioning, and organizational leadership.*

*altMBA*

## Literatura uzupełniająca

### Książki

## Książki autora i dzieła przywołane w tekście:




**[1] "This Is Strategy"**

Seth Godin

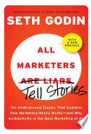
2024 | Simon and Schuster | ISBN: 9798893310306




**[2] "Whatcha Gonna Do With That Duck?"**

Seth Godin

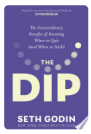
2013 | Penguin UK | ISBN: 9780670923007




**[3] "All Marketers are Liars"**

Seth Godin

2009 | Penguin | ISBN: 9781101184547




**[4] "The Dip"**

Seth Godin

2011 | Hachette UK | ISBN: 9780748129270



**[5] "We are All Weird"**

Seth Godin

2011 | Amazon Publishing | ISBN: 9781936719228




**[6] "Purple Cow"**

Seth Godin

2005 | Penguin UK | ISBN: 9780141924861

## Książki autorów przywoływanych i literatura pokrewna:

**[7] "Measuring the Mind"**

Denny Borsboom

ISBN: 9781107151956




**[8] "The sociology of science"**

Robert King Merton

ISBN: 9788131608609




**[9] "On the Theory of Scales of Measurement"**

Stanley Smith Stevens

1946



**[10] "The Travels and Adventures of Serendipity"**

Robert King Merton

2004 | Princeton University Press | ISBN: 9780691117546



**[11] "Social Theory and Social Structure"**

Robert King Merton



**[12] "RSSB Teaching Associate Recruitment Exam 2026 Accountancy & Business Statistics (ABST) Paper II Chapterwise MCQ Bank | 3000+ MCQs with Detailed Explanations | Based on Latest RSSB Syllabus"**

Nodia

2026 | Nodia Learning


## Artykuły naukowe

### Prace autorów przywoływanych w tekście:


**[1] "On the Theory of Scales of Measurement"**

S. S. Stevens

1946 | Science | DOI: 10.1126/science.103.2684.677

Źródło

**[2] "ON THE PROBLEM OF SCALES FOR THE MEASUREMENT OF PSYCHOLOGICAL MAGNITUDES"**

S. Stevens

2006

Źródło

**[3] "Construct validity in psychological tests."**

Lee J. Cronbach, Paul E. Meehl

1955 | Psychological Bulletin | DOI: 10.1037/h0040957

Źródło

**[4] "Beyond the two disciplines of scientific psychology."**

Lee J. Cronbach

1975 | American Psychologist | DOI: 10.1037/h0076829

Źródło

**[5]** **"Assessing similarity between profiles."**

Lee J. Cronbach, Goldine C. Gleser

1953 | Psychological Bulletin | DOI: 10.1037/h0057173

Źródło

**[6]** **"Clinical versus statistical prediction: A theoretical analysis and a review of the evidence."**

Paul E. Meehl

1954 | University of Minnesota Press eBooks | DOI: 10.1037/11281-000

Źródło

**[7]** **"Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy."**

William M. Grove, Paul E. Meehl

1996 | Psychology Public Policy and Law | DOI: 10.1037/1076-8971.2.2.293

Źródło

**[8]** **"8. On A Distinction Between Hypothetical Constructs And Intervening Variables"**

Kenneth MacCorquodale, Paul E. Meehl

1948 | University of Minnesota Press eBooks | DOI: 10.5749/9780816683406-010

Źródło

**[9]** **"The Concept of Validity."**

Denny Borsboom, Gideon J. Mellenbergh, Jaap van Heerden

2004 | Psychological Review | DOI: 10.1037/0033-295x.111.4.1061

Źródło

**[10]** **"Item bias and item response theory"**

Gideon J. Mellenbergh

1989 | International Journal of Educational Research | DOI: 10.1016/0883-0355(89)90002-5

Źródło

**[11]** **"Contingency Table Models for Assessing Item Bias"**

Gideon J. Mellenbergh

1982 | Journal of Educational Statistics | DOI: 10.3102/10769986007002105

Źródło

**[12]** **"QUALIDEM: development and evaluation of a dementia specific quality of life instrument––validation"**

Teake P. Ettema, Rose-Marie Dröes, J. de Lange, Gideon J. Mellenbergh, Miel W. Ribbe

2006 | International Journal of Geriatric Psychiatry | DOI: 10.1002/gps.1692

Źródło

**[13]** **"Problems of Monetary Management: The UK Experience"**

Charles Goodhart

1984 | DOI: 10.1007/978-1-349-17295-5_4

Źródło

**[14]** **"A Formal Approach to the Problem of Logical Non-Omniscience"**

Scott Garrabrant, Tsvi Benson-Tilsen, Andrew Critch, Nate Soares, Jessica Taylor

2017 | Electronic Proceedings in Theoretical Computer Science | DOI: 10.4204/eptcs.251.16

Źródło

**[15]** **"Substitutes for leadership: Their meaning and measurement"**

Steven Kerr, John M. Jermier

1978 | Organizational Behavior and Human Performance | DOI: 10.1016/0030-5073(78)90023-5

Źródło

**[16]** **"Toward a contingency theory of leadership based upon the consideration and initiating structure literature"**

Steven Kerr, Chester A. Schriesheim, Charles J. Murphy, Ralph M. Stogdill

1974 | Organizational Behavior and Human Performance | DOI: 10.1016/0030-5073(74)90037-3

Źródło

**[17]** **"Severity of omicron variant of concern and effectiveness of vaccine boosters against symptomatic disease in Scotland (EAVE II): a national cohort study with nested test-negative design"**

Aziz Sheikh, Steven Kerr, Mark E. J. Woolhouse, Jim McMenamin, Chris Robertson

2022 | The Lancet Infectious Diseases | DOI: 10.1016/s1473-3099(22)00141-4

Źródło

**[18] "Severe COVID-19 outcomes after full vaccination of primary schedule and initial boosters: pooled analysis of national prospective cohort studies of 30 million individuals in England, Northern Ireland, Scotland, and Wales"**

Utkarsh Agrawal, Stuart Bedston, Colin McCowan, Jason Lee Oke, Lynsey Patterson, Steven Kerr

2022 | The Lancet | DOI: 10.1016/s0140-6736(22)01656-7

Źródło

**[19] "Evaluating Value-Added Models for Teacher Accountability"**

Daniel F. McCaffrey, J. R. Lockwood, Daniel M. Koretz, Laura S. Hamilton

2003 | PsycEXTRA Dataset | DOI: 10.1037/e658712010-001

Źródło

**[20] "Models for Value-Added Modeling of Teacher Effects"**

Daniel F. McCaffrey, J. R. Lockwood, Daniel M. Koretz, Thomas A. Louis, Laura S. Hamilton

2004 | Journal of Educational and Behavioral Statistics | DOI: 10.3102/10769986029001067

Źródło

**[21] "Quality Control in the Development and Use of Performance Assessments"**

Stephen B. Dunbar, Daniel M. Koretz, Hiram D. Hoover

1991 | Applied Measurement in Education | DOI: 10.1207/s15324818ame0404_3

Źródło

**[22] "The Vemont Portfolio Assessment Program: Findings and Implications"**

Daniel M. Koretz, Brian M. Stecher, Stephen P. Klein, Daniel F. McCaffrey

1994 | Educational Measurement Issues and Practice | DOI: 10.1111/j.1745-3992.1994.tb00443.x

Źródło

### Artykuły naukowe autora:

**[23] "Permission Marketing: Turning Strangers Into Friends And Friends Into Customers"**

Seth Godin

1999

Źródło

**[24] "Tribes: We Need You to Lead Us"**

Seth Godin

2008

Źródło

**[25]** **"Purple Cow: Transform Your Business by Being Remarkable"**

Seth Godin

2003

Źródło

**[26]** **"All Marketers Are Liars: The Power of Telling Authentic Stories in a Low-Trust World"**

Seth Godin

2005

Źródło

**[27]** **"Linchpin: Are You Indispensable?"**

Seth Godin, Jessica Hagy, Hugh MacLeod

2010

Źródło

**[28]** **"Unleashing the Idea Virus"**

Seth Godin, Malcolm Gladwell

2000

Źródło

**[29]** **"The Guerrilla Marketing Handbook"**

Jay Conrad Levinson, Seth Godin

1994

Źródło

**[30]** **"Unleashing the ideavirus : stop marketing at people! turn your ideas into epidemics by helping your customers do the marketing for you"**

Seth Godin

2001 | Hyperion eBooks

Źródło

**[31]** **"The Future of the Library: What Is a Public Library For?"**

Seth Godin

2016 | Public Library Quarterly | DOI: 10.1080/01616846.2016.1245008

Źródło

[32] **"Survival Is Not Enough"**

Seth Godin

2002

Źródło

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: fbd89edf-3a20-4e5a-bb0b-198e9eae489d