

Uniwersytet Warszawski

Wydział Zarządzania

mgr Katarzyna Kinga Kowalczyk

Rola recenzji w zarządzaniu środkami na badania naukowe

Autoreferat rozprawy doktorskiej

Promotor: prof. zw. dr hab. Grażyna Wieczorkowska – Wierbińska

Promotor pomocniczy: dr Grzegorz Król

Słowa kluczowe: peer-review, recenzenci, ewaluacja, styl oceny, efekt kolejności,
efekt halo, overflow

Warszawa, 2019

Spis treści

Wprowadzenie.....	3
Struktura pracy	4
Cel pracy	5
Opis zbiorów danych.....	8
Najważniejsze wyniki.....	9
Zadanie badawcze 1. Różnice między recenzentami. Czy można zredukować efekt kolejności?9	
Zadanie badawcze 2. Czy można zredukować efekt HALO?.....	11
Zadanie badawcze 3. Wpływ przeciążenia poznawczego	12
Rekomendacje dla procesu zarządzania	14

Wprowadzenie

Recenzje stanowią fundament pracy naukowej. Proces recenzowania stanowi podstawę takich działań w świecie naukowym jak np. ewaluacja uczelni przez Państwową Komisję Akredytacyjną czy procedur awansowych pracowników naukowych. Od przychylności recenzji zależy otrzymanie grantu Narodowego Centrum Nauki, Narodowego Centrum Badań i Rozwoju czy Fundacji Nauki Polskiej, publikacja monografii czy artykułu w czasopiśmie naukowym.

Tempo kariery naukowca uzależnione jest od pozycji czasopism, w których publikuje. W Polsce artykuł w periodyku znajdującym się na liście A, często otwiera szybką drogę do naukowej kariery oraz wydatnie zwiększa szansę uzyskania naukowego grantu. W wyniku działania tego mechanizmu, recenzje wywierają olbrzymi wpływ na decyzje o podziale środków na badania naukowe.

Analiza danych pochodzących z jednego z konkursów dla naukowców, otrzymanych od jednej z czołowych polskich instytucji przyznającej granty pokazała, że styl oceny recenzenta (surowy lub pobłażliwy oraz fakt wykorzystywania lub nie, pełnego spektrum dostępnej skali ocen) istotnie wpływa na przyznawane wnioskowi punkty.

Dla każdej z sześciu dyscyplin naukowych wszystkie konkurujące o dofinansowanie projekty oceniało trzech recenzentów. Przeprowadzone analizy pokazały, że aż w 4 z 6 dziedzin recenzenci istotnie różnili się ocenami (jeden z recenzentów oceniał inaczej niż pozostali – tj. stawiał niższe lub wyższe oceny, mniej lub bardziej różnicował wymiary oceny)¹.

Z tego powodu, w celu zachowania obiektywności oceny i zapewnienia równych szans wnioskodawcom na uzyskanie grantu na badania, styl oceniania recenzenta powinien być kontrolowany, czego się niestety nie robi. Publikowane są jedynie wskazówki jak należy tworzyć rzetelne recenzje², nie przeprowadza się natomiast systematycznych ocen pracy recenzentów. Jednym z powodów takiego stanu może być to, że nie są dostępne wiarygodne narzędzia służące do pomiaru jakości recenzji³.

Nakłady na naukę, badania naukowe i prace rozwojowe (B+R) w Polsce powiększają się systematycznie. Przykładowo, w 2005 r. wynosiły 5.575 mln zł, w 2010 r. wyniosły 10.416,2 mln zł, osiągając poziom 16,168 mln zł w 2014 roku.

¹ Michałowicz, 2013, 2016

² np. Wydany w 2011 r. przez Ministerstwo Nauki i Szkolnictwa wyższego dokument „Dobre praktyki w procedurach recenzyjnych w nauce”

³ Van Rooyen i inni, 1999

Wzrost nakładów nie musi się przekładać bezpośrednio na naukowy czy gospodarczy sukces, ponieważ ważna jest trafna selekcja projektów przeznaczonych do finansowania.

Wadliwy system podziału dostępnych środków, został wskazany w końcowych wnioskach po kontroli Najwyższej Izby Kontroli jako jedna z przyczyn słabości systemu finansowania działalności naukowej i wdrożeniowej. Pomimo tego, że kontrola miała miejsce kilka lat temu, to wiele zawartych w końcowym raporcie konkluzji nie straciło na aktualności. Sposób dokonywania recenzji stanowiący integralną część systemu podziału środków finansowych, ma wpływ na stopień efektywności tego systemu.

Przebieg procesu recenzowania może się różnić, w zależności od realizowanego celu.

Uniwersalne są jednak związane z nim problemy, do których zaliczyć można m.in.⁴:

- występowanie różnic pomiędzy indywidualnymi stylami oceniania wybranych recenzentów (tendencja do nadmiernej surowości lub pobłażliwości, małe różnicowanie ocen, trzymanie się środka skali) - jeden punkt może przesądzić o otrzymaniu dofinansowania badań, pozytywna recenzja umożliwi publikację w prestiżowym czasopiśmie, co z kolei pozwoli na uzyskanie nagrody finansowej;
- różny stopień przeciążenia recenzenta innymi niż sama recenzja obowiązkami lub zwiększenie liczby koniecznych do dokonania recenzji;
- silne zróżnicowanie w poziomie motywacji recenzentów (np. traktowanie sporządzania recenzji jako konieczności);
- sposób organizacji procesu recenzji, który jest w stanie wzmacniać podatność na zniekształcenia ocen, np. wystąpienie efektu kolejności.

Struktura pracy

Praca ma charakter empiryczny i składa się z następujących części:

- W części teoretycznej omówiono m.in. cele i krytykę recenzji naukowych (*peer-review*), psychologiczny model procesu recenzji, błędy poznawcze towarzyszące procesowi recenzji, problem przeciążenia informacyjnego (*overflow*). Część teoretyczną zamyka opis zadań badawczych i hipotezy.

⁴Badania własne

- Część empiryczną otwiera opis analizowanych zbiorów danych, opis wyników badań ilościowych (eksperymentalnych, korelacyjnych) i jakościowych (wywiadów). Część empiryczną zamykają rekomendacje dla procesu recenzowania.
- Część uzupełniająca - zawiera materiały dodatkowe, które zaburzałyby płynność wyводу, ale mogą być ważne dla osób zainteresowanych szczegółami analiz.

Praca powstała dzięki wsparciu ekspertów, pracowników uczelni wyższych z Polski i zagranicy, z którymi przeprowadzone zostały wywiady. Ze względu na konieczność zachowania anonimowości, ich nazwiska (i afiliacje) nie zostały wymienione w rozprawie. Ich wkład jednak jest nie do przecenienia i dlatego w tym miejscu autorka składa im podziękowania.

Część badań została zrealizowana w ramach projektu naukowego NCN pn. „NCN Preludium nr UMO-2016/21/N/HS4/00528 nt.: „Recenzowanie w czasach zalewu (*overflow*), konsekwencje dla zarządzania środkami na badania naukowe”, którego autorka pracy jest kierownikiem.

Cel pracy

Rozprawa doktorska ma na celu wskazanie niektórych ograniczeń procesu recenzowania naukowego (*peer-review*) artykułów naukowych i projektów grantowych. Przedstawiona oraz wsparta analizami zostanie teza, zgodnie z którą, efektywność i obiektywność procesu recenzowania może zostać podniesiona dzięki analizom pracy ekspertów. W części empirycznej zostały przetestowane rozwiązania, które mogłyby zmniejszyć zniekształcenia towarzyszące pracy recenzentów podczas dokonywania ocen. Jest to szczególnie ważne w czasach przeciążenia informacyjnego (*overflow*).

Opierając się na przeglądzie literatury, w badaniach skoncentrowano się na problemach sformułowanych w postaci zadań badawczych:

Zadanie badawcze 1: Różnice w stylach oceny recenzentów (stopień surowości, różnicowanie ocen różnych obiektów) zostały stwierdzone w wielu badaniach i zreplicowane w badaniach przedstawionych w rozprawie. Wyniki analiz nakazywałyby, aby konkurujące między sobą projekty, teksty naukowe były oceniane przez ten sam zestaw recenzentów. Taki sposób oceniania jest np. stosowany przez Fundację na rzecz Nauki Polskiej w programie „Start”. W proponowanym schemacie oceny „surowy” recenzent będzie w tym samym stopniu zaniżał oceny wszystkich

konkurujących między sobą uczestników. Ale ocena serii nie rozwiązuje wszystkich problemów, ponieważ pojawia się stwierdzany w wielu badaniach efekt kolejności - szczególnie w sytuacji, w której recenzent nie może wracać do wcześniej ocenionych obiektów. Pierwszym problemem badawczym było sprawdzenie czy wprowadzenie przerw może wyeliminować efekt kolejności w ocenianiu serii obiektów.

Zadanie badawcze 2: Od recenzentów wymaga się nie tylko decyzji akceptującej lub odrzucającej dany projekt/publikację, ale także oceny obiektów na wielu wymiarach /kryteriach częściowych oceny. U podstaw tej praktyki leży przyjmowane milcząco założenie, że istnienie ocen częściowych zobiektywizuje oceny, a obiekty będą opisywane w postaci wielowymiarowych profili. W praktyce, jeśli nawet wymiarom przypisuje się zróżnicowane wagi (patrz ocena 673 abstraktów) to decyzje podejmuje się na podstawie średniej. Konieczność dokonywania wielu ocen częściowych jest natomiast dużym obciążeniem poznawczym dla oceniającego – nic więc dziwnego, że w badaniach wykazuje się niezmiennie efekt HALO polegający na bardzo wysokim skorelowaniu ocen częściowych **przypisanie obiektowi jednej ważnej pozytywnej lub negatywnej cechy wpływa na skłonność do przypisywania innych cech zgodnych ewaluatywnie**⁵.

Na siłę efektu HALO może mieć wpływ sposób oceniania serii obiektów na kilku wymiarach częściowych, który można przeprowadzić w sposób **obiektyowy** (najpierw obiekt 1 na wszystkich wymiarach, potem obiekt 2...) lub **wymiarowy** (najpierw wszystkie obiekty na wymiarze 1, potem wszystkie obiekty na wymiarze 2...).

Celem tego zadania badawczego jest przetestowanie hipotezy mówiącej, że ocenianie wymiarowe w odróżnieniu od oceniania obiektowego zmniejszy siłę efektu HALO.

Zadanie badawcze 3. Trzecie zadanie poświęcone jest sprawdzeniu czy standardy publikacyjne wprowadzone w erze przedkomputerowej powodują dodatkowe i niepotrzebne przeciążenia informacyjnego (*overflow*) i w ten sposób wpływają na oceny recenzenta.

Zadanie badawcze 4. W rozprawie zamieszczono także omówienie wywiadów przeprowadzonych z 35 doświadczonymi recenzentami (w tym 15 z wiodących uniwersytetów zachodnich) z różnych dyscyplin naukowych. Wszystkie wywiady zostały poddane transkrypcji (w sumie 170 stron tekstu) i zostały przedstawione syntetycznie w części trzeciej.

Część empiryczna pracy składa się z opisu badań jakościowych (wywiady z recenzentami) i ilościowych.

⁵ Brzezińska, Brzeziński, Eliaż 2004

Na dane ilościowe składają się analizy danych zastanych (673 abstraktów konferencyjnych) i przeprowadzone przez Autorkę 4 badania eksperymentalne, w których uczestniczyło w sumie 912 osób).

W sumie testowano następujące hipotezy:

H1a	Abstrakty konferencyjne oceniane na początku otrzymują niższe oceny niż pozostałe abstrakty.
H1b	Dobre projekty oceniane na początku otrzymują niższe oceny niż te same projekty oceniane na końcu serii.
H1c	Słabe projekty oceniane na początku otrzymują wyższe oceny niż te same projekty oceniane na końcu serii.
H1d	Projekty oceniane na początku otrzymują mniej skrajnych ocen niż projekty oceniane na końcu serii.
H1e	Zastosowanie przerw spowoduje zminimalizowanie wystąpienia efektu kolejności.

H2	Zastosowanie wymiarowego w miejsce obiektowego sposobu oceny spowoduje redukcję efektu halo.
----	--

H3a	Przypisy środkowe powodują większe rozproszenie uwagi „recenzentów” niż przypisy dolne.
H3b	Na ocenę projektu wpływa stan psychoenergetyczny „recenzenta”: dodatnio poziom motywacji, ujemnie: stopień zmęczenia i stopień rozproszenia.
H3c	Osoby oceniające projekty opisane w standardzie przypisów środkowych deklarują większe rozproszenie, niż ci, którzy oceniali projekty opisane w standardzie przypisów dolnych. Efekt jest tym silniejszy, im silniejsza jest u „recenzenta” tendencja do sekwencyjnego przetwarzania informacji.

Opis zbiorów danych

Zbiór A – 33 recenzentów

Zbiór A zawiera oceny 673 abstraktów konferencyjnych zgłoszonych na zagraniczną konferencję dotyczącą zdrowia publicznego. Wszystkie abstrakty były oceniane niezależnie przez 3 recenzentów, każdy z 33 recenzentów oceniał średnio 61 abstraktów (od 17 do 88). Recenzenci byli losowani niezależnie, przez system komputerowy, oddzielnie dla każdego z abstraktów. Ewaluatorzy oceniali prace na 6 wymiarach, opisanych na skali 0 do 10 każdy. Oceny częściowe zostały przekształcone w średnią ogólną z następującymi wagami: jakość (waga 0,1), wartość teoretyczna lub praktyczna (0,1), oryginalność (0,1), zgodność z tematem konferencji (0,1), jakość prezentacji (0,1), ocena podsumowująca i rekomendacja (0,5).

Zbiór APA - 69 doktorantów

Zbiór APA zawiera wyniki badania, w którym uczestniczyło 76 doktorantów (72% kobiet) w wieku od 26 do 46 lat ($M=29,43$; $SD=5,17$). Siedmiu uczestników zostało wykluczonych z badania ze względu na zbyt krótki czas czytania poleceń, aby mogli je zrozumieć. Finalna próba składała się z 69 uczestników (71,5% kobiet) wieku 22-46 lat ($M=29,5$; $SD=5,24$).

Zbiór AR – 678 absolwentów szkół wyższych

Zbiór AR zawiera wyniki badania przeprowadzonego na ogólnopolskiej, wylosowanej z panelu Ariadna próbie 667 respondentów (62,5% kobiet) w wieku od 28 do 42 lat ($M=34,35$; $SD=3,0$) z wyższym wykształceniem.

Zbiór H – 46 studentów wcielających się w rolę recenzentów

Zbiór H zawiera wyniki badania, w którym uczestniczyło 81 studentów (69% kobiet) w wieku od 17 do 27 lat ($M=19$; $SD=2$). Z badania zdecydowano się wykluczyć odpowiedzi 23 osób, które nie udzieliły poprawnych odpowiedzi na pytania kontrolne, oraz 12 osób, których językiem ojczystym był rosyjski/ukraiński, ponieważ ich odpowiedzi świadczyły o niewystarczającej znajomości języka polskiego. W ostatecznej próbie zostało 46 osób (67% kobiet) w wieku 17-27 ($M=19,52$; $SD=1,46$).

Zbiór E –88 studentów wcielających się w rolę recenzentów

Zbiór E zawiera wyniki badania, w którym uczestniczyło 88 studentów (60% kobiet) w wieku od 19 do 29 lat ($M=21,26$; $SD=1,09$).

Zbiór W- 35 recenzentów

Zbiór zawiera wywiady przeprowadzonymi z polskimi (20) i zagranicznymi (15) recenzentami, reprezentującymi zarówno nauki społeczne (np.: zarządzanie, psychologia), jak i ścisłe (np. fizyka). Wszystkie wywiady zostały spisane.

Najważniejsze wyniki

Zadanie badawcze 1. Różnice między recenzentami. Czy można zredukować efekt kolejności?

Realizując zadanie 1. najpierw na 5 różnych zbiorach danych potwierdzono występowanie różnic między recenzentami w zakresie średniej i wariancji wystawianych ocen. Wyniki są jednoznaczne: recenzenci różnią się znacznie stylem oceniania – jedni są mniej a drudzy bardziej surowi, co oznacza, że czynnik losowy związany z wyborem recenzenta wpływa istotnie na przyznanie grantu czy publikację tekstu. Należałoby więc rekomendować, aby konkurujące między sobą projekty, teksty naukowe były oceniane przez ten sam zestaw recenzentów.

W takim schemacie oceny „surowy” recenzent będzie w tym samym stopniu zaniżał oceny wszystkich konkurujących między sobą uczestników. Ale ocena serii przez tych samych recenzentów nie rozwiązuje wszystkich problemów, ponieważ pojawia się stwierdzany w wielu badaniach efekt kolejności - szczególnie w sytuacji, w której recenzent nie może wracać do wcześniej ocenionych obiektów.

Pierwszym krokiem było sprawdzenie na dwóch zbiorach analizowanych danych występowania efektu kolejności. Pierwsza analiza (test hipotezy 1a) została przeprowadzona na zbiorze ocen 673 abstraktów konkurujących o przyznanie grantów pokrywających koszty udziału w konferencji międzynarodowej. Abstrakty oceniane w pierwszej piętnastce serii (każdy recenzent oceniał średnio 61 abstraktów) miały istotnie niższą średnią niż te oceniane później. Hipoteza 1a „Abstrakty konferencyjne oceniane na początku otrzymują niższe oceny niż pozostałe abstrakty” została potwierdzona. W następnym kroku w specjalnie zaplanowanym badaniu eksperymentalnym przetestowano dwie hipotezy i wykazano, że wpływ kolejności na ocenę zależy od jakości obiektu: słabe obiekty zyskują, gdy oceniane na początku, dobre tracą. Wiąże się to z unikaniem skrajnych ocen na początku oceniania serii (potwierdzone zostały hipotezy 1b, 1c i 1d).

Styl oceny i efekt kolejności

Dotychczasowe badania nad efektem kolejności pokazują, że może on wystąpić wszędzie tam, gdzie mamy do czynienia z oceną serii obiektów – niezależnie czy są to oceny wykonanych muzycznych czy projekty grantowe. Stopień znajomości obiektu oceny nie ma tu znaczenia. Warunkiem jest natychmiastowa ocena obiektu jednego po drugim.

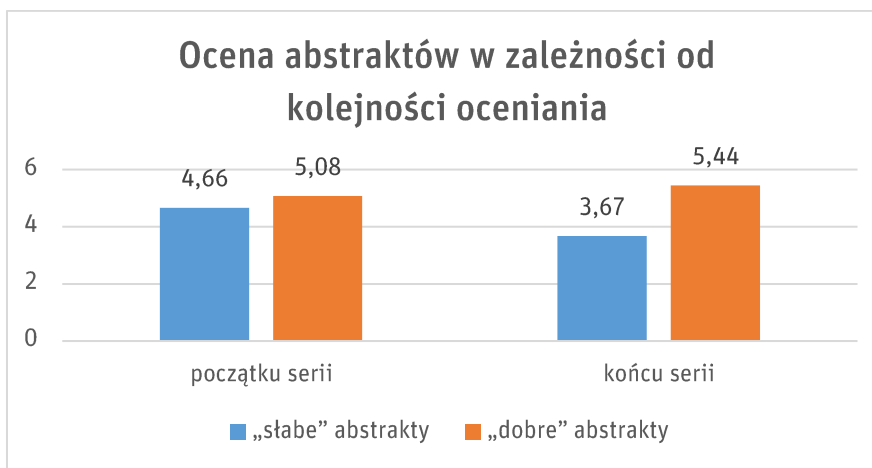
Przeprowadzone analizy zbioru abstraktów konferencyjnych oraz danych pochodzących z badań eksperymentalnych potwierdziły występowanie efektu kolejności. Jego wystąpienie dowiodły niższe średnie oceny projektów ocenianych na

początku (tracą projekty dobre), mniejsza ilość skrajnych ocen wystawianych na początku serii recenzji. Uzyskane wyniki są spójne z teorią kalibracji, zakładającą tworzenie przez oceniającego wewnętrznej skali i częstsze stosowanie skrajnych ocen w toku serii⁶.

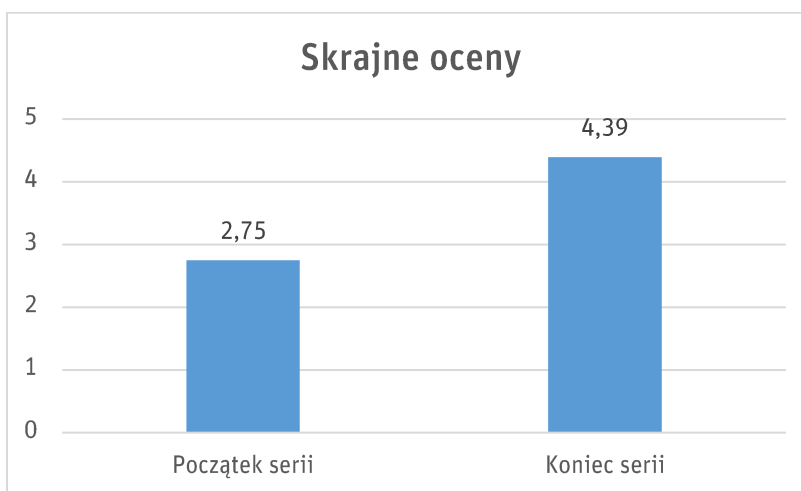
Zmienna zależna: Ocena projektu

Zmienne:	F	p
Wersja (kolejność)	16,26	0,0001
Jakość abstraktu (słabe, dobre)	138,44	0,001
Jakość abstraktu*wersja	7,87	0,01

Tabela 1. Wyniki dwuczynnikowej analizy wariancji z powtarzaniem pomiarem na czynniku JAKOŚĆ abstraktu



Wykres 1. Średnia ocena słabych i dobrych abstraktów ocenianych na początku i na końcu serii



Wykres 2. Średnia liczba skrajnych ocen na początku i na końcu serii.

⁶ Unkelbach, Osthheimer, Fasold and Memmert, 2012

Próba jego minimalizacji przez zastosowanie przerwy nie przyniosła rezultatu. Manipulacja przerwą okazała się być nieskuteczna. W charakterze przerwy zastosowano ocenę logotypów konferencyjnych – być może zasoby uwagi związane z systematem obrazowym aktywizują inne systemy przetwarzania informacji. Zastosowana przerwa mogła być również zbyt krótka, aby odciągnąć uwagę recenzentów i wyczyścić pamięć roboczą. Gdyby zamiast logotypów dać recenzentom inne projekty do oceny, mogłoby to okazać się skuteczniejszym narzędziem. Z drugiej strony mogłoby to również przyczynić się do ich większego przeciążenia, co nie było naszym celem.

Zadanie badawcze 2. Czy można zredukować efekt HALO?

Zadanie badawcze 2. związane było z próbą eliminacji w ocenach recenzentów efektu HALO przejawiającego się w bardzo wysokim skorelowaniu ocen częściowych [pozytywna/negatywna ocena przejawia się w pozytywnych/negatywnych ocenach na wszystkich wymiarach]. W następnym kroku postawiono hipotezę, że na jego siłę może wpływać sposób oceniania serii obiektów, który można przeprowadzić w sposób w OBIEKTOWY (najpierw obiekt 1 na wszystkich wymiarach, potem obiekt 2...) lub WYMIAROWY (najpierw wszystkie obiekty na wymiarze 1., potem wszystkie obiekty na wymiarze 2...). W tym celu przeprowadzono specjalnie zaprojektowane badanie eksperymentalne, które potwierdziło hipotezę (H2) mówiącą, że ocenianie wymiarowe w odróżnieniu od oceniania obiektowego zmniejszy siłę efektu HALO.

Przeprowadzone badanie eksperymentalne potwierdza występowanie efektu halo. O jego wystąpieniu świadczy brak różnicowania wymiarów częściowych przez recenzentów. Ogólne wrażenie wywołane przez projekt wpływa na ocenę pozostałych kryteriów. Manifestuje się to poprzez jednoczynnikową strukturę ocen częściowych, przez to traci się wielowymiarowy obraz danego obiektu.

W przeprowadzonym eksperymencie próba zminimalizowania efektu przez zastosowanie wymiarowego w zastępstwie obiektowego sposobu oceny przyniosła pozytywny efekt. Umożliwiła rozbicie jednoczynnikowej struktury, a także zmniejszenie procentu wyjaśnianej wariancji. Zastosowanie takiego sposobu pozwala również wyeliminować jeden z potencjalnych powodów wystąpienia efektu halo – opóźnienia pomiędzy obserwacją a oceną, które zwiększają prawdopodobieństwo polegania na globalnym wrażeniu⁷.

⁷ Murphy & Balzer, 1986

% wyjaśnianej wariancji przez pierwszy czynnik		
	Standardowy obiektowy proces oceny	Wymiarowy proces oceny
Projekt1	66,99	45,95
Projekt 2	66,90	55,75

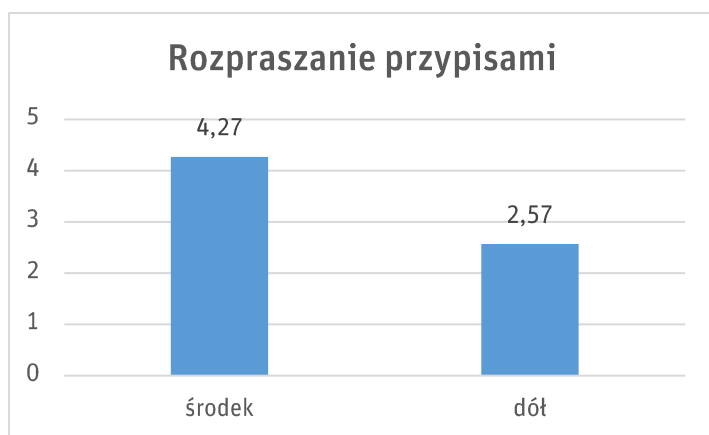
Tabela 2. Procent wyjaśnianej wariancji przez pierwszy czynnik.

Zadanie badawcze 3. Wpływ przeciążenia poznawczego

Trzecie zadanie badawcze dotyczyło wpływu przeciążenia poznawczego, któremu podlegają wszyscy pracownicy nauki borykający się z niekontrolowanym przyrostem informacji.

Przedmiotem badania eksperymentalnego w zadaniu 3 było sprawdzenie konsekwencji stosowania dwóch odmiennych standardów cytowań: w środku tekstu (np.: APA, Harvard, MLA) vs na dole strony (np. Chicago/Turabian, Oxford).

Założono, że przypisy umieszczane w nawiasach w środku tekstu, często dzielące zdanie na dwie lub więcej części, niepotrzebnie przeciążają umysł czytającego utrudniając syntezę sensu zdania. W tym zadaniu w specjalnie zaplanowanych dwóch eksperymentach przetestowano trzy hipotezy: (H3a) Przypisy środkowe powodują większe rozproszenie uwagi „recenzentów” niż przypisy dolne; (H3b) Na ocenę projektu wpływa stan psychoenergetyczny „recenzenta”: podwyższa ją silniejsza motywacja, obniża: stopień zmęczenia i stopień rozproszenia; (H3c) Osoby oceniające projekty opisane w standardzie przypisów środkowych deklarują większe rozproszenie, niż ci, którzy oceniali projekty opisane w standardzie przypisów dolnych. Efekt jest tym silniejszy, im silniejsza jest u „recenzenta” tendencja do sekwencyjnego przetwarzania informacji. Wszystkie hipotezy uzyskały poparcie empiryczne, choć warto dodać, że należałoby to badanie zreplikować na prawdziwych recenzentach, a nie tylko badanych wczuwających się w ich rolę.



Wykres 3. Stopień rozproszenia ze względu na umiejscowienie przypisów.

Analiza regresji

Zmienna zależna: Ocena projektu

Zmienne:

Zwersja	0,06
Zrozpraszanie	-0,068
Wersja*rozpraszanie	0,084*
Zaangażowanie	0,111**
Zmęczenie badaniem	-0,282***
Płeć	0,063
Wiek	0,023

Stała: 0,298 $t(666)=-619$ $p=ni$; Tabela przedstawia współczynniki Beta:

*** $p < 0.001$ (test 2-stronny); * $p < 0.05$ (test 2-stronny)

Tabela 3. Wyniki analizy regresji (zmienna zależna: ocena projektu) ($F(7,666) = 15,379$; $p < 0.001$).

Analiza regresji

Zmienna zależna: Rozpraszanie

Predyktory:

Zwersja (1=środkowy, 2=dolny)	-0,244*
Zstandard przypisów (1=środkowy, 2=dolny)	0,228
symultaniczność	0,211
Wersja*symultaniczność	0,284*
Płeć (1=mężczyźni, 2=kobiety)	0,283*

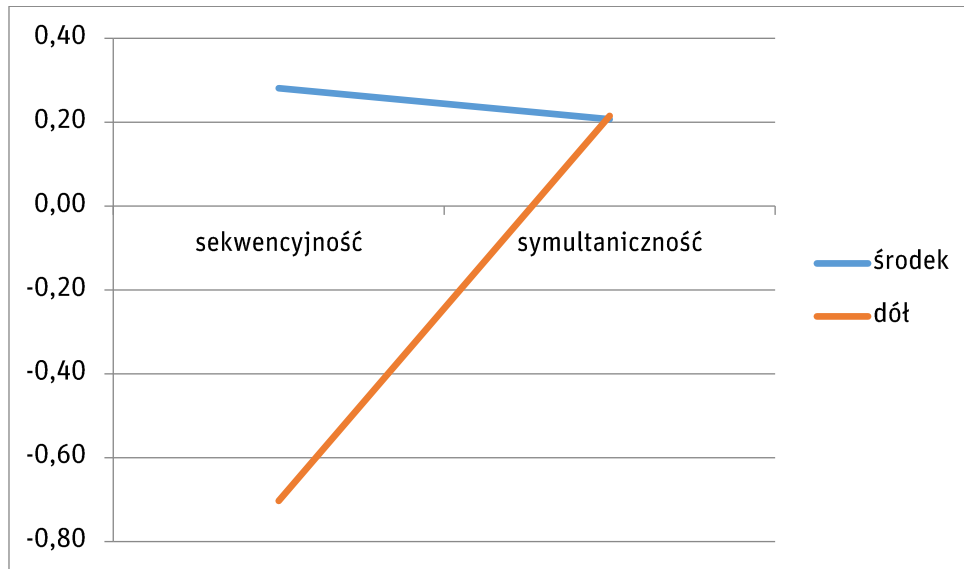
Stała: 1,180, $t(53)=2,936$, $p=0,05$; Tabela przedstawia współczynniki Beta:

*** $p < 0.001$ (test 2-stronny); * $p < 0.05$ (test 2-stronny)

Tabela 4. Wyniki analizy regresji (zmienna zależna poziom rozproszenia) ($F(5,53)=4,343$; $p < 0.001$).

Wykazano także interakcyjny wpływ zmiennej WERSJA i symultaniczności (beta = 0,248). Współczynnik kierunkowy wynosił 1,827 ($t(53)=2,937$; $p=0,005$).

Interpretacja interakcyjnego wpływu przedstawiona jest na poniższym rysunku



Wykres 4. Interakcyjny wpływ wersji i symultaniczności na poziom rozproszenia „recenzenta”. Na osi OX znajduje się poziom symultaniczności, na osi OY – ocena rozproszenia (im wyższa tym wyższe rozproszenie przez nazwiska).

Rekomendacje dla procesu zarządzania

Proces recenzowania należy do grona tych, które decydują o jakości nauki i tego, jak jest ona odbierana na zewnątrz. Obserwator spoza środowiska może go nie dostrzegać lub nawet nie być świadomym jego istnienia w przeciwieństwie do np. dydaktyki lub badań. Ale to właśnie sprawny i rzetelny przebieg procesu recenzji ma wpływ na to, jak naukę odbiera społeczeństwo, bowiem decyduje o jej wiarygodności.

Jakość dokonywanych recenzji znajduje przełożenie nie tylko na to, jak środowisko naukowe postrzegane jest przez społeczeństwo, a w konsekwencji zarówno polityków, jak i przedsiębiorców, ale ma wielowymiarowy wpływ na stosunki wewnętrzne w świecie naukowym. Dotyka zarówno aspektów osobistych, jak i zawodowych. Wpływa na prestiż osoby i jej pozycję w społeczności naukowej za pośrednictwem siły oddziaływania jaką mają publikacje dokonywane w prestiżowych periodykach oraz możliwości przeprowadzania badań, dzięki zdolności pozyskania odpowiednich funduszy. Odgrywa też niebagatelną rolę w tempie rozwoju naukowej kariery.

Recenzenci nie mają poczucia, że ich praca jest doceniana adekwatnie do wkładanego przez nich wysiłku. W niniejszej pracy zostało także pokazane, że sam proces dokonywania recenzji jest daleki od doskonałości i nie jest wolny od systemowych zniekształceń. Można wręcz zaryzykować i postawić tezę, że system wolny jest od elementów i narzędzi wspomagających zarządzanie jego jakością.

Na podstawie badań i analiz przeprowadzonych w niniejszej pracy można sformułować rekomendacje mające na celu polepszenie jakości procesu recenzowania m.in. dzięki eliminacji pozamerytorycznych czynników.

Ze względu na powszechnie występujące różnice w stopniu surowości recenzentów należy:

1. Rozpocząć od procesu kalibracji poprzez zlecenie recenzentom oceny projektów/publikacji o ustalonej wcześniej wartości i wykluczyć z procesu recenzowania tych, których oceny znacząco odbiegają (in plus lub in minus) od ustalonej wcześniej wartości. Kalibracja powoduje, że wzorce porównawcze są możliwie do siebie zbliżone.
2. Alternatywnie zlecać ocenę wszystkich konkurujących ze sobą projektów/publikacji do danego numeru czasopisma temu samemu np. trójosobowemu zespołowi recenzentów wymuszając na nich ocenianie wymiarowe, a nie obiektowe i płacąc za wykonaną pracę, co znacznie skróciłoby czas recenzowania. Przy ocenie poszczególnych wymiarów cząstkowych należy rotować kolejność ocenianych obiektów, aby uniknąć efektu kolejności. Czynnikiem sprzyjającym jego wystąpieniu jest również niemożliwość samodzielnego decydowania o kolejności oceny przez eksperta lub zmiany raz postawionej oceny.
3. Trzeba też pamiętać, że mnożenie kryteriów cząstkowych niepotrzebnie przeciąża recenzentów, którzy zaczynają kierować się zasadą zgodności ewaluatywnej (dobre lub złe oceny na wszystkich wymiarach). Choć nie zostało to zbadane w tym programie można przewidywać, że warto minimalizować liczbę kryteriów cząstkowych.