

Katarzyna Kinga Kowalczyk

Streszczenie pracy doktorskiej

Rola recenzji w zarządzaniu środkami na badania naukowe

Praca wykonana pod kierunkiem
Prof. Grażyny Wieczorkowskiej-Wierzbińskiej
i doktora Grzegorza Króla
Wydział Zarządzania
Uniwersytetu Warszawskiego

Recenzje naukowe (peer-review) stanowią fundament oceny wartości pracy naukowej. Są podstawą dla akceptacji bądź odrzucenia artykułów naukowych i projektów grantowych. W sposób pośredni lub bezpośredni wywierają wpływ nie tylko na decyzje o podziale środków na badania, ale mają podstawowe znaczenie dla pozycji badacza w świecie naukowym, są narzędziem, który jest używane przez zarządzających środkami na badania naukowe.

Rozprawa ma charakter empiryczny. Głównym jej celem jest analiza ograniczeń procesu recenzowania naukowego i empiryczne przetestowanie dziewięciu hipotez wyprowadzonych z przeglądu literatury.

Praca składa się z 3 części. W części pierwszej, teoretycznej, przedstawiono: poznawczy model procesu oceniania i przeanalizowano czynniki mogące zaburzać przebieg tego procesu takie, jak indywidualny styl oceniania recenzenta (poziom surowości, niechęć do różnicowania ocen dla różnych obiektów), efekt halo (brak różnicowania ocen na cząstkowych wymiarach oceny), efekt kolejności pojawiający się przy ocenie serii obiektów oraz wpływ przeciążenia informacyjnego.

Część empiryczna pracy składa się z opisu badań jakościowych (pogłębione wywiady z 35 doświadczonymi recenzentami) i ilościowych. Na dane ilościowe składają się analizy danych zastanych (oceny 673 abstraktów konkurujących o granty konferencyjne) i przeprowadzone przez Autorkę 4 badania eksperymentalne, w których uczestniczyło w sumie 912 osób. 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.

Logika przedstawionego w rozprawie wywodu została wsparta realizacją 4 zadań badawczych. Realizując zadanie 1. najpierw na 5 różnych zbiorach danych potwierdzono występowanie różnic między recenzentami w zakresie średniej i wariacji 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. Fakt, że zazwyczaj pracę recenzują co najmniej dwie osoby nie zabezpiecza przed „wylosowaniem” dwóch surowych lub dwóch łagodnych recenzentów. 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 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).

Kolejnym wyzwaniem badawczym było sprawdzenie czy wprowadzenie PRZERW może wyeliminować efekt kolejności w ocenianiu serii obiektów.

Próba eliminacji tego efektu przez wprowadzenie w jednej z grup przerwy (w czasie której wcielający się w rolę recenzentów uczestnicy badań wykonywali dodatkowe zadanie polegające na ocenie estetycznej logotypów) zakończyła się porażką, więc hipoteza 1e nie zyskała poparcia empirycznego.

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]. Trzeba pamiętać, że 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 np. 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, który został też zreplikowany w rozprawie doktorskiej. W następnym kroku postawiono hipotezę, że na jego siłę może wpływać sposób oceniania serii obiektów, którą można przeprowadzić w sposób 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.

Trzecie zadanie badawcze dotyczyło wpływu przeciążenia poznawczego, któremu podlegają wszyscy pracownicy nauki borykający się z niekontrolowanym przyrostem informacji określanym w literaturze różnymi terminami m.in.: przeciążenie informacyjne information overload¹)zalew informacji (flood of information)², smog informacyjny (data smog)³ czy overflow⁴. Jednym z głównych źródeł przeciążenia są publikacje naukowe, których liczba rośnie w zastraszającym tempie – rocznie publikuje się około 15000 międzynarodowych czasopism naukowych⁵. Zwiększeniu uległa nie tylko liczba publikacji i istniejących czasopism, ale także wzrosła objętość i złożoność publikacji, w tym liczba cytowań⁶. Odnotowano wzrost liczby cytowanych publikacji we wszystkich dyscyplinach naukowych⁷, w niektórych czasopismach aż ośmiokrotnie⁸. Analiza przeprowadzona na prawie tysiącu artykułach z siedmiu czasopism⁹ pokazała, że coraz więcej autorów używa dwóch lub więcej odwołań, aby uzasadnić ten sam argument, zwiększając w ten sposób całkowitą objętość publikacji¹⁰. Wzrost ten nie dziwi, albowiem zwiększenie liczby przypisów w artykule jest podawane jako jedna ze strategii zwiększających szanse publikacji¹¹.

¹ Toffler, 1970

² Eppler, Mengis, 2004

³ Shenk, 1997

⁴ Czarniawska i Löfgren, 2012

⁵ Liniowa ekstrapolacja analiz Thompsona z 2005 roku, za: Wieczorkowska, Wierzbiński, Król, 2015

⁶ Reis i Stiller, 1992; Gasparyan i inni, 2015

⁷ Gasparyan i inni, 2015

⁸ Adair i Vohra, 2003

⁹ Adair i Vohra, 2003

¹⁰ Adair i Vohra, 2003

¹¹ Ebrahim i inni, 2013

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 dwie 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. Obie hipotezy uzyskały poparcie empiryczne, choć warto dodać, że należałoby to badanie zreplikować na prawdziwych recenzentach, a nie tylko badanych wczuwających się w ich rolę.

W rozprawie zamieszczono także efekty czwartego zadania badawczego, które polegało na przeprowadzeniu pogłębionych wywiadów 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.

Wywiady pokazały m.in., przyzwyczajenie do standardu cytowań. Większość z recenzentów, w dyscyplinach z obowiązującym środkowym standardem, nie uwierzy w wyniki badań eksperymentalnych sugerujących, że prowadzi on do niepotrzebnego przeciążenia. Nie można wykluczyć, że ich umysły zautomatyzowały ignorowanie zawartości środkowych nawiasów.

Rozprawa się kończy się trzema rekomendacjami dla zarządzających regulacjami dotyczącymi procesu recenzji.

Ze względu na powszechnie występujące różnice w stopniu surowości recenzentów należy albo: (1) zacząć 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, albo (2) 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.

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.

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