

Warszawa, dnia 26.03.2019

Recenzja

rozprawy doktorskiej mgr Katarzyny Kingi Kowalczyk

„Rola recenzji w zarządzaniu środkami na badania naukowe”

napisanej pod kierunkiem

prof. dr hab. Grażyny Wieczorkowskiej-Wierzińskiej i dr Grzegorza Króla, 2019, ss. 164

Mgr Katarzyna Kowalczyk podejmuje bardzo ważny i ciekawy problem w swojej pracy doktorskiej, a mianowicie czynników wpływających na sposób oceniania w nauce. Problem ten jest szczególnie istotny, gdyż w sytuacji rozrostu ośrodków badawczych, szybkiego zwiększania się liczby uczelni ocenianych na podstawie osiągnięć niezwykle szybko rosnącej liczby pracowników naukowych, problemem stał się dostęp do finansowania badań, które stają się podstawą publikacji, a te z kolei później stają się obiektem liczenia punktów, które uzyskują poszczególni pracownicy, a w ostatecznym rozrachunku jednostki naukowe. Często mówi się o zjawisku punktozji jako podstawowym kryterium ich ewaluacji. Wady tej metody oceny uczelni i jednostek badawczych zasługują na pogłębioną analizę i są przedmiotem dyskusji w środowisku naukowym, ale to nie jest obiektem zainteresowań autorki. Kwestia oceny dorobku naukowego, poza jego wymiarem ilościowym, jest niezwykle trudna. Proces ten zaczyna się już w fazie początkowej, niejako na starcie, gdy uzyskuje się lub nie możliwość udziału w konferencjach, dostęp do środków finansowych umożliwiających prowadzenie badań bądź nie. Lista uwarunkowań dostępu jest bardzo długa. Częściowo jest efektem relacji mających stosunkowo niewiele wspólnego z jakością proponowanych badań, natomiast jest często uwarunkowana rywalizacją między naukowcami, ośrodkami naukowymi, zamożnością fundacji finansujących, ilością pieniędzy przeznaczanych na naukę w poszczególnych krajach, sposobem ich alokacji kierowanym różnymi względami, często dalekimi od proponowanych tematów i metodologii badań.

Autorka recenzowanej pracy doktorskiej nie interesuje się wymienionymi wyżej uwarunkowaniami. Ogranicza się do roli recenzentów, ich psychologicznych predyspozycji do oceniania/recenzowania abstraktów i aplikacji o granty. Wspomniane ograniczenie ma sens, ale trzeba zdawać sobie sprawę z faktu, że autorka nie bada zarządzania nauką, a jedną z grup przesłanek, które są brane pod uwagę w procesie zarządzania. Mgr Kowalczyk identyfikuje szereg czynników, które mogą wpływać na sposób dokonywanej oceny, stawia jasno sformułowane hipotezy, które weryfikuje w trakcie prowadzonych przez siebie badań.

Pragnę podkreślić, że pracę doktorską w ramach określonych przez mgr Kowalczyk należy uznać za poprawnie zrealizowaną. Wywód jest jasny, praca jest dobrze skonstruowana. Opracowanie statystyczne wielości danych uzyskanych w ramach kolejnych zadań badawczych pokazuje, że weryfikacja sformułowanych hipotez została przeprowadzona zgodnie z regułami użytej metodologii. Autorka formułując hipotezy odwołuje się do literatury polskiej i zagranicznej co powoduje, że podejmowana przez nią problematyka lokuje się w głównym nurcie badań w tym zakresie.

Obecnie przejdę do szczegółowego omówienia poszczególnych części pracy i wątpliwości jakie się nasuwają.

Autorka pisze: "Głównym celem jest analiza ograniczeń procesu recenzowania naukowego i empiryczne przetestowanie dziewięciu hipotez wyprowadzonych z przeglądu literatury. 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) i 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. Na 5 różnych zbiorach danych wyniki badań autorki przyniosły potwierdzenie występowania różnic między recenzentami w zakresie średniej i wariancji wystawianych ocen". Innymi słowy, że jedni recenzenci bardziej surowo oceniają, podczas gdy inni łagodniej. Jak słusznie podkreśla ona, może się zdarzyć, że bardziej surowi recenzenci mogą być wybrani do recenzji określonego tekstu lub wręcz odwrotnie - łagodniejsi bądź zespół oceniający może składać się z jednych i drugich. W ten sposób mgr Kowalczyk potwierdziła poprzez swoje badania znany fakt różnorodności recenzentów, który należy brać pod uwagę próbując ocenić wartość recenzowanych prac.

Kontrowersyjna wydaje mi się jej sugestia, że surowi recenzenci powinni być eliminowani z procesu oceniania abstraktów/aplikacji o granty, gdyż – jak pisze - zaburzą proces oceny powodując obniżenie średniej oceny. Autorka zakłada, że komisje ostatecznie selekcjonujące aplikacje o granty czy abstrakty czynią to mechanicznie opierając się li tylko na średniej ocen. Zapewne bywa różnie. I takiej sytuacji nie można wykluczyć. W znanej mi rzeczywistości oceny aplikacji o granty w Horizon 2020 przez panele recenzentów w Brukseli, proces oceny przebiega inaczej niż sugeruje autorka. Co dzieje się z bardzo surowymi ocenami, gdy odbiegają one od ocen reszty recenzentów/panelistów? Po wnikliwym zapoznaniu się z oceną "średnią", ale też ocenami cząstkowymi, nieraz podejmuje się decyzję o wyeliminowaniu tej recenzji poprzestając na ocenach pozostałych recenzentów, których z zasady jest kilku. Warto jednak zdawać sobie sprawę, że "surowy recenzent" może zwrócić uwagę na jakiś aspekt pominięty z różnych przyczyn przez innych recenzentów. Przed podjęciem decyzji o pominięciu jego recenzji, członkowie panelu zapoznają się z jego opinią. Moje doświadczenia z udziału w komisji przydzielającej granty w ramach Ministerstwa Szkolnictwa Wyższego i Nauki też pokazywały, że choć działały inaczej niż panel w Brukseli to nie działało się to mechanicznie. Innymi słowy, autorka pracy przecenia i nadmiernie uogólnia mechaniczność działania gremiów kwalifikujących aplikacje o finansowanie badań. Szkoda więc, że nie poświęciła

więcej uwagi temu problemowi, badaniach jakościowych, gdy rozmawiała z osobami, które faktycznie uczestniczą w tym procesie.

Innym zadaniem, które autorka sobie postawiła, 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).

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.

Mój komentarz: ustalenia autorki są ważne i dobrze udokumentowane, skądinąd zbieżne z obserwacjami uzyskanymi w czasie egzaminów, gdy podstawą oceny są wypowiedzi egzaminowanych, a nie testy w przypadku których z góry jest ustalone, które odpowiedzi są prawidłowe.

Kolejne zadanie badawcze (nr 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].

Jak pisze mgr Kowalczyk (s.83) „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, który 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 zmniejsza siłę efektu HALO”. Komentarz: Autorka słusznie stwierdza, że mnożenie kryteriów/wymiarów oceny może nie służyć właściwej ocenie projektu czy abstraktu, gdyż kryteria mogą zachodzić na siebie, w pewien sposób replikować się wzajemnie. Równocześnie powstaje problem, że stosując wymiarowe a nie obiektowe ocenianie, zatracą się obraz całości poszczególnych ocenianych obiektów. Zatem prawdą jest to co ustaliła autorka metodami statystycznymi, że zwiększa się niezależność ocen na poszczególnych wymiarach, ale w efekcie traci się obraz obiektów jako pewnych całości. Powstaje pytanie: co jest ważniejsze? Co chce się premiować? Myślę, że

wartość obiektu. Pojawia się tu jeszcze inny problem, o którym mgr Kowalczyk pisze: unikanie ocen krańcowych i trzymanie się środka skali przez badanych. Zjawisko to jest dobrze znane w naukach społecznych, gdy badane są opinie respondentów na różne tematy. Ma charakter bardziej ogólny, nie jest specyficzne dla oceny projektów.

Autorka stwierdziła, że przerwa w ocenianiu projektów, nie wpłynęła na jego sposób. Być może decydował o tym sposób jej "zagospodarowania", który polegał na ocenie logotypów, czyli też ocenie. Być może należałoby uzyskany rezultat zweryfikować w przyszłych badaniach "wypełniając" przerwę inną czynnością.

"Trzecie zadanie badawcze – jak pisze autorka- 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 overload*) zalew informacji (*flood of information*) smog informacyjny (*data smog*) czy overflow. Jednym z głównych źródeł przeciążenia są publikacje naukowe, których liczba rośnie w zastraszającym tempie – rocznie publikuje się około 15000 międzynarodowych czasopism naukowych. Zwiększeniu uległa nie tylko liczba publikacji i istniejących czasopism, ale także wzrosła objętość i złożoność publikacji, w tym liczba cytowań."

Mgr Kowalczyk udowodniła w tym przypadku, że sposób umieszczania informacji o wykorzystanej literaturze przy pisaniu własnego artykułu: "środkowy" bądź "dolny" mają wpływ na rozproszenie uwagi recenzentów. Występuje ono rzadziej w przypadku "dolnego zapisu", ale też, jak słusznie podkreśla, znaczącą rolę odgrywa przyzwyczajenie czytającego do tekstu. Uwaga ta wydaje się bardzo ważna, ponieważ oceniający musi też zaopiniować czy wykorzystana literatura jest tą właściwą, a nie niebudzącymi zaufania przyczynkami, nie zawsze właściwie ocenionymi przy kwalifikacji ich do druku w różnych pismach czy wydawnictwach.

Zwrócenie uwagi na rodzaj zapisu jako czynnika wpływającego na ocenę świadczy dobrze o autorce i jej poszukiwaniach w tym zakresie. Równocześnie należy zdawać sobie sprawę z faktem, że obecnie żyjemy w świecie, w którym prawie każde wydawnictwo ma specyficzne dla siebie oczekiwania. Fakt ten utrudnia pracę recenzentom, którzy każdorazowo muszą dostosowywać się do innego zapisu wykorzystywanej literatury.

Budzi wątpliwości i... sprzeciw sformułowanie autorki: "W idealnym świecie prowadzona przez uczelnie działalność badawcza – rozwojowa powinna wywierać bezpośredni wpływ na gospodarkę" Takie stwierdzenie jest uważane za nadmierne i niekategoryczne, gdyż eliminuje badania podstawowe, które odgrywają bardzo istotną rolę w świecie nauki, a w konsekwencji są często asumptem do kolejnych badań w różnych sferach życia, nie tylko w gospodarce.

Badana przez autorkę sfera działalności naukowców jest bardzo złożona, co zresztą podkreśla w swojej pracy odwołując się do bogatej literatury psychologicznej. Podkreśla również trudności w dostępie do rzeczywistych recenzentów. Zastąpienie ich przez studentów oraz osoby po prostu z wyższym wykształceniem (z próby Ariadny) może budzić szereg zastrzeżeń. Osoby te oceniają aplikacje dotyczące problemów im całkowicie nieznanymi co powoduje oczywiste problemy. Poza tym, nie są szkolone w dokonywaniu ewaluacji jako takiej, gdyż kształcenie u nas na ogół ma na celu doprowadzenie do opanowania przez studentów materiału z określonej dziedziny, a nie krytycznego myślenia.

Warto tu dodać, że część badań została zrealizowana w ramach projektu naukowego NCN nr „NCN Preludium nr UMO-2016/21/N/HS4/00528 nt.: „Recenzowanie w czasach zalewu

(overflow), konsekwencje dla zarządzania środkami na badania naukowe”, którego mgr Kowalczyk jest kierownikiem.

Podsumowując, pragnę podkreślić, że praca mgr Kowalczyk stanowi interesujący przyczynek do naszej wiedzy dotyczącej czynników psychologicznych wpływających na dokonywane oceny aplikacji w nauce, a tym samym potencjalnego wpływu wyników tych ocen na losy naukowców. Praca jest dobrze skonstruowana, poprawna stylistycznie co stanowi jej dodatkowe zalety. Niewątpliwie zachęca do kontynuowania badań w tej dziedzinie.

Uważam, że niniejsza praca spełnia wymogi ustawy z dnia 24 marca 2003 roku o stopniach naukowych I tytule naukowym oraz stopniach I tytule w zakresie sztuki. Niniejszym wnioskuję do Rady Wydziału Zarządzania Uniwersytetu Warszawskiego o dopuszczenie do publicznej obrony. Praca mieści się w dziedzinie nauk ekonomicznych w dyscyplinie nauki o zarządzaniu.

