

Uniwersytet Warszawski  
Wydział Zarządzania

mgr inż. Bartłomiej Michałowicz

# Ankiety ewaluacyjne w szkolnictwie wyższym: wpływ wyboru ewaluatorów

autoreferat

Praca wykonana pod kierunkiem  
prof. zw. dr hab. Grażyny Wieczorkowskiej-Wierzińskiej  
Wydział Zarządzania, Uniwersytet Warszawski  
Katedra Psychologii i Socjologii Zarządzania

Warszawa, styczeń 2016

## **Spis treści**

- I. Wprowadzenie
- II. Ustalenia terminologiczne
- III. Cel pracy, zadania badawcze, hipotezy
- IV. Opis zbiorów danych
- V. Najważniejsze wyniki analiz
- VI. Implikacje dla zarządzania w szkolnictwie wyższym

## I. Wprowadzenie

Każde podejmowane działanie, niezależnie od dziedziny, powinno być przedmiotem uważnej oceny praktycznej w zakresie skuteczności i ekonomiczności. Badania ewaluacyjne powinny być podstawą do podejmowania decyzji menadżerskich, co tłumaczy wykorzystywanie narzędzi ewaluacyjnych w każdym obszarze życia. Ewaluacja stanowi integralny element zarządzania przez cele i działania oraz promowanego w UE *przejrzystego i efektywnego zarządzania*. Zgodnie z tym podejściem, podstawowym celem każdej ewaluacji jest podnoszenie jakości, spójności i skuteczności oddziaływań.

Przedmiotem zainteresowania w tej rozprawie doktorskiej są badania ewaluacyjne prowadzone w szkolnictwie wyższym. Ich rola w tym obszarze jest niezwykle ważna, ponieważ zarówno badania naukowe, jak i zajęcia akademickie, choć często prowadzone w warunkach rynkowych, nie mogą być oceniane wyłącznie z perspektywy osiągniętego zysku. Ewaluacja ma więc dostarczać kryteriów, metod i środków do oceny efektywności działań, spełniając w szkolnictwie wyższym rolę podobną do mechanizmu rynkowego w sektorze prywatnym.

Choć znaczenie terminu *ewaluacja* rozszerza się znacząco poza samą tylko ocenę działania – traktuje się ją bardziej jako proces, narzędzie interakcji pomiędzy uczestnikami tego procesu, w końcu jako dziedzinę wiedzy – to w rozprawie skoncentruję się tylko na jednym narzędziu ewaluacji, jakim są ankiety wykorzystywane standardowo w zarządzaniu pracownikami w organizacjach sektora edukacyjnego.

Jakość dla usługodawcy mierzy się najczęściej za pomocą ankiet ewaluacyjnych, którym poświęcona jest ta rozprawa. Zbyt często uznaje się, że jeśli uczelnia posługuje się – w sposób bardziej lub mniej regularny – ankietami, w których studenci oceniają wykładowców, to możemy mówić, że taka organizacja posiada system ewaluacji. Tymczasem, samo zbieranie danych ankietowych bez wypracowanych standardów dotyczących interpretacji wyników oraz bez uwzględniania kontekstu, w którym dane były zbierane, trudno uznać za kompletny system ewaluacji.

## II. Ustalenia terminologiczne

Dla ułatwienia lektury rozprawy, poniżej doprecyzowano często używane terminy.

- **Ewaluator** – w literaturze przedmiotu *ewaluatorem* nazywa się osobę przeprowadzającą ewaluację, a nie tylko dostarczającą danych. W rozprawie **zawężono znaczenie tego pojęcia** do osoby odpowiadającej na pytania ankiety ewaluacyjnej (która bada subiektywne opinie na temat jakości usługi edukacyjnej).
- **Ewaluacja seryjna** – sytuacja, w której ta sama osoba ocenia kilka obiektów (np. student ocenia 6 kursów, w których uczestniczył w danym semestrze).
- **Badanie ewaluacyjne wyczerpujące (pełne)** – występuje, gdy ankieta jest kierowana do wszystkich studentów uczestniczących w zajęciach.

- **Procent zwrotu** (*response rate*, tłumaczony też jako stopień realizacji próby, stopa zwrotu) – w ewaluacji seryjnej wyznaczany zarówno na poziomie **jednostkowym**, jak i **grupowym**.
- **Ocena** – dotyczy wykładowcy i jest wystawiana przez studentów w ankiecie ewaluacyjnej (zamiennie nazywana też *ewaluacją*).
- **Stopień** – ocena wystawiana studentom przez wykładowców (wartość liczbową wpisywana do protokołów).

### **III. Cel pracy, zadania badawcze, hipotezy**

Przedmiotem rozprawy doktorskiej jest analiza przydatności ankiet ewaluacyjnych wykorzystywanych do oceny pracowników akademickich. Na podstawie przeanalizowanych w części teoretycznej doniesień z literatury, sformułowano następujące zadania badawcze:

#### **III.2. Określenie procentu zwrotu i czynników na niego wpływających**

Standardowo określa się procent zwrotu dla poszczególnych zajęć. W przeprowadzonych analizach procent zwrotu jest traktowany także jako charakterystyka studenta (połączone zostały ankiety wypełniane przez tego samego studenta w trzech badaniach prowadzonych w trzech różnych semestrach). Celem analiz była odpowiedź na pytanie, ile zajęć ocenił ten sam student. Skoro tylko nieliczni studenci wypełniają ankiety, to ważne jest pytanie, czy ewaluatorzy różnią się od „nieewaluatorów”. Jest to więc pytanie o **efekt autoselekcji**. Jedynym wymiarem, na którym można było dokonać porównania, są stopnie uzyskiwane przez studentów.

#### **III.3. Określenie struktury czynnikowej ankiet ewaluacyjnych**

Przed przygotowaniem narzędzi ewaluacji jakości kształcenia w formie ankiety konieczne jest zadanie trzech pytań: (1) Jak długa powinna być ankieta? (2) Jakie pytania powinna zawierać? oraz (3) Czy analizowane będą odpowiedzi na pojedyncze pytania, czy też raczej wskaźniki syntetyczne w formie skal lub indeksów? Zadanie badawcze 2. polegało na sprawdzeniu struktury czynnikowej odpowiedzi na pytania analizowanych danych.

#### **III.4. Ocena metodologicznej poprawności porównania wykładowców za pomocą wyników ankiet ewaluacyjnych**

Celem zadania badawczego 3. było sprawdzenie, czy wyniki ankiet ewaluacyjnych pozwalają na ocenę pracy wykładowców. W przedstawionych przykładach pokazano, że nagradzani są wykładowcy otrzymujący najwyższe średnie. W tym zadaniu badawczym sprawdzano, czy ignorowanie błędów standardowych z wyliczonych średnich jest poprawne metodologicznie.

#### **III.5. Określenie wpływu wystawianych przez wykładowców stopni na ewaluację zajęć. Przewidywania ekspertów**

W tym zadaniu badawczym szacowano związek między przewidywanymi stopniami a ocenami wystawianymi wykładowcy. Sprawdzano także opinie wykładowców na ten temat stwierdzenia „dobre wystawione stopnie przekładają się na dobrą ocenę zajęć”. Sprawdzano

także, czy opisywane w literaturze zjawisko inflacji stopni można zaobserwować także w Polsce.

### III.6. Hipotezy

W ramach omówionych powyżej problemów badawczych sformułowano i weryfikowano następujące hipotezy:

- H1: Niski procent zwrotu wiąże się efektem autoselekcji polegającym na tym, że ankiety wypełniają studenci lepsi i gorsi.
- H2: Ewaluatorzy oceniają zajęcia „jednowymiarowo”, czego przejawem jest efekt aureoli.
- H3: Porównywanie wykładowców za pomocą wyników ankiet ewaluacyjnych przeprowadzane jest w sposób poprawny metodologicznie.
- H4: Studenci otrzymujący lepsze stopnie wyżej oceniają zajęcia.
- H5: Chęć otrzymania wyższych ocen zajęć powoduje obniżenie przez wykładowców wymagań przejawiające się w inflacji stopni.

## IV. Opis zbiorów danych

### IV.1. Badania audytoryjnych ankiet ewaluacyjnych

Analizowane dane pochodziły z badania ewaluacyjnego przeprowadzonego w formie papierowej (audytoryjnej) w latach 2010 i 2011 (4 semestry). Z wszystkich możliwych 1502 grup zajęciowych otrzymano dane od 1395 grup, uzyskując wyniki dla ponad 33 000 ankiet.

Tabela 1. Liczba ankiet w kolejnych semestrach (ankieta audytoryjna). Opracowanie własne.

Semestr	Liczba grup	Liczba osób	Liczba ankiet
Zima 2009/2010	338	19189	9106
Lato 2009/2010	323	18617	7546
Zima 2010/2011	341	18203	8678
Lato 2010/2011	393	18928	7929
Suma	1395	74937	33259

### IV.2. Badania internetowych ankiet ewaluacyjnych 0., 1., 2., 3.

We wstępnym badaniu przeanalizowano wyniki ankiet z 4 kursów internetowych (N=178), dla których dopasowano oceny kursu ze stopniami uzyskanymi przez ewaluatorów.

W Badaniach 1., 2., i 3., prowadzonych w kolejnych trzech semestrach, zebrano wyniki z **ponad 100 tysięcy internetowych ankiet** ewaluacyjnych. Unikatowość tego zbioru danych pozwala na **połączenie** ankiet ewaluacyjnych dotyczących różnych przedmiotów wypełnianych przez tę

samą osobę z jej stopniami. Nieznane były jednak inne cechy ewaluatora, takie jak np. wiek, płeć, rok studiów.

Tabela 2. Liczba ankiet w kolejnych badaniach (ankieta internetowa). Opracowanie własne.

Semestr	Liczba grup	Liczba stopni	Liczba ankiet	Liczba studentów	Liczba wykładowców
Badanie 1. Zima 2011/2012	3586	87303	58663	11598	851
Badanie 2. Lato 2011/2012	3546	88256	35231	11070	928
Badanie 3. Zima 2012/2013	4038	78969	32142	11550	903

### **IV.3. Badanie inflacji stopni**

W badaniu zebrano dane dotyczące stopni ukończenia studiów na dwóch polskich uczelniach. Analizowano dane pozyskane w latach 2009–2014 (Uniwersytet SWPS) i 2010–2015 (UW) i zebrano odpowiednio 4020 oraz 9857 obserwacji.

### **IV.4. Badanie opinii na temat ankiet ewaluacyjnych**

Na pytania dotyczące stosunku do ankiet ewaluacyjnych odpowiedziało 115 studentów i doktorantów Uniwersytetu Warszawskiego biorących udział w kursach metodologii badań.

### **IV.5. Sonda ekspercka**

Sondaż przeprowadzony wśród pracowników dydaktycznych i doktorantów Wydziału Zarządzania UW i Wydziału Psychologii UW. Ankietowani odpowiadali na jedno pytanie zamknięte dotyczącej ich opinii na temat tego, czy stopień, który otrzymał student, ma wpływ na polecenie innym danego przedmiotu. Przeanalizowano dane od 58 nauczycieli akademickich – pracowników naukowych Wydziału Zarządzania, których podzielono na tych z mniejszym (co najwyżej 12 lat) i większym doświadczeniem dydaktycznym.

## V. Najważniejsze wyniki analiz

### V.1. Czynniki wpływające na procent zwrotu

W zadaniu badawczym 1. analizowano procent zwrotu w internetowych ankietach ewaluacyjnych w trzech kolejnych semestrach (Badanie 1., 2., 3.) oraz porównano procenty zwrotu w ankietach audytoryjnych i internetowych zajęć obowiązkowych (ćwiczenia, warsztaty) z nieobowiązkowymi (wykłady).

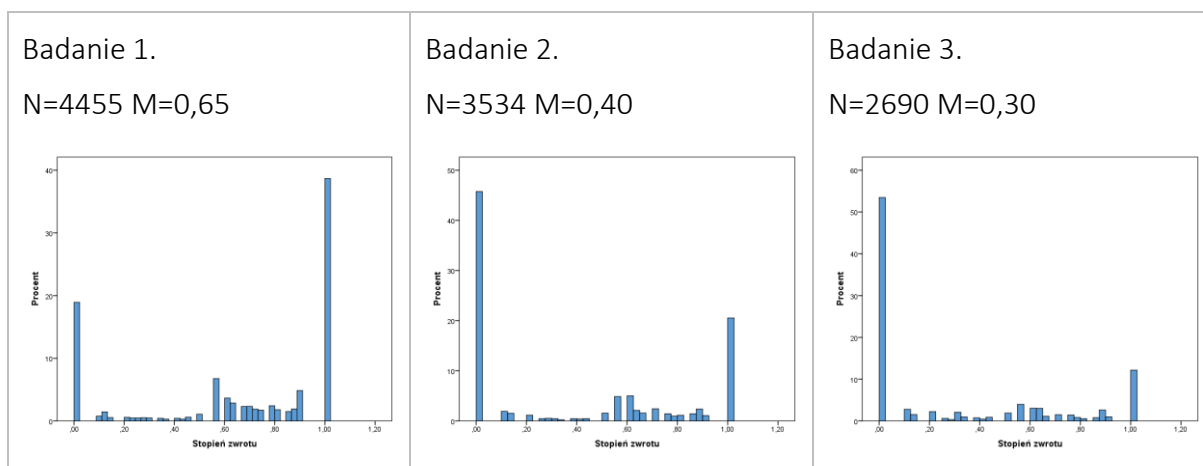
Porównania procentu zwrotu w **internetowych ankietach ewaluacyjnych** w trzech kolejnych semestrach (Badanie 1., 2., 3.) wykazało, że średni procent zwrotu zależał od badania. Zgodnie z przewidywaniami, najwyższy procent zwrotu (65%) uzyskano w Badaniu 1., ponieważ różniło się ono sposobem przeprowadzenia od dwóch pozostałych – **student dostawał monity dotyczące wypełnienia ankiet** – sugerujące, że jeśli nie oceni co najmniej 50% zajęć, w których uczestniczył, to nie będzie mógł zapisać się na zajęcia w kolejnym semestrze. Taka procedura okazała się skuteczna, ponieważ średni procent zwrotu wyniósł 65%. W następnych badaniach, gdy nie stosowano monitów, średni procent zwrotu spadł do 40 i 30%.

Procent zwrotu w ankietach audytoryjnych zależy od tego, czy zajęcia są obowiązkowe, czy też nie. Przykładowo stopień zwrotu ankiet oceniających nieobowiązkowe wykłady (29%) jest o wiele mniejszy, niż stopień zwrotu dotyczących obowiązkowych ćwiczeń (77%). Różnice w procencie zwrotu między wykładami a ćwiczeniami znikają w ankietach internetowych, ale niestety procent zwrotu bez dodatkowych monitów zwykle nie przekracza 30%.

#### Procent zwrotu jako charakterystyka ewaluatora

Standardowo określa się procent zwrotu dla poszczególnych zajęć. W przeprowadzonych analizach **procent zwrotu** analizowany był także jako **charakterystyka studenta** i jest to – wg naszej wiedzy – pierwsze takie badanie na świecie.

Forma, w jakiej zbierano dane, pozwalała na monitorowanie aktywności studenta w czasie każdego z trzech badań internetowych. W tym celu połączone **zostały ankiety wypełniane przez tego samego studenta** wraz z informacją o liczbie zajęć, w których uczestniczył (i które mógł potencjalnie ocenić). Celem analiz była odpowiedź na pytanie, ile zajęć ocenił ten sam student. Najliczniejsza grupa w Badaniu 1. wypełnia **wszystkie** ankiety, w pozostałych badaniach najliczniejsze są grupy nie wypełniające **żadnej**.



Rysunek 1. Rozkład procentów zwrotu w trzech badaniach. Opracowanie własne.

#### Podsumowując:

- W Badaniu 1. (N=4455) **38,7%** (N=1723) wypełniło wszystkie, **18,9%** (N=843) nie wypełniło żadnej ankiety.
- W Badaniu 2. (N=3534) **20,5%** (N=726) wypełniło wszystkie, **45,8%** (N=1617) nie wypełniło żadnej ankiety.
- W Badaniu 3. (N=2690) **12,2%** (N=327) wypełniło wszystkie, **53,5%** (N=1438) nie wypełniło żadnej ankiety.

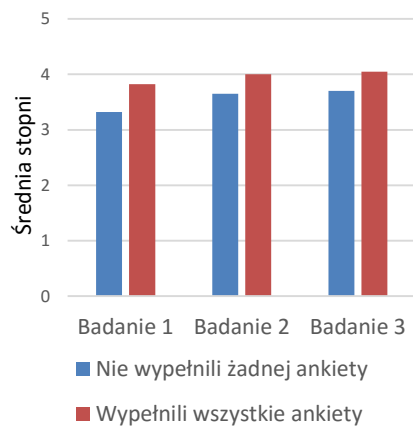
#### Test efektu autoselekcji

Mimo zakładanego charakteru wyczerpującego typu badania ewaluacyjnego (wszyscy studenci są zaproszeni do wzięcia w nim udziału), uzyskiwane dane są efektem procesu autoselekcji ewaluatorów. Skoro tylko nieliczni studenci wypełniają ankiety, to ważne jest pytanie, czy ewaluatorzy różnią się od nieewaluatorów. Jedynym wymiarem, na którym można było dokonać porównania, są stopnie. Przedmiotem testowania była hipoteza mówiąca, że dobrowolne anonimowe ankiety ewaluacyjne częściej wypełniają studenci lepsi (uzyskujący lepsze stopnie) i studenci gorsi (mający problem z zaliczeniem).

Na dwóch zbiorach danych, pozwalających na różne sposoby analizy, okazało się, że ankiety ewaluacyjne częściej wypełniają lepsi studenci niż gorsi.



**Poziom analizy: porównanie międzypersonowe dla danych zagregowanych dla każdego studenta. Porównanie średniej stopni studentów, którzy wypełnili wszystkie ankiety ewaluacyjne z tymi, którzy nie wypełnili żadnej**



Badanie 1.:

$M=3,32$  (N=843) <  $M=3,82$  (N=1723), (dla 3 kategorii  $F(2,4452)=14584,225$ ,  $p=0,000$ , Tukey  $p=0,000$ )

Badanie 2.:

$M=3,65$  (N=1617) <  $M=4,00$  (N=726), (dla 3 kategorii  $F(2,3531)=16094,355$ ,  $p=0,000$ , Tukey  $p=0,000$ )

Badanie 3.:

$M=3,70$  (N=1438) <  $M=4,05$  (N=327) (dla 3 kategorii  $F(2,2687)= 8159,570$ ,  $p=0,000$ , Tukey  $p=0,000$ ).

Rysunek 2. Porównanie średniej stopni osób wypełniających wszystkie ankiety ewaluacyjne ze stopniami osób, które nie wypełniły żadnej (dla 3 badań). Opracowanie własne.

Średnia ze stopni tych, którzy nie wypełnili **żadnej** ankiety w każdym z badań jest **niższa** niż tych, którzy wypełnili **wszystkie**.

**Poziom analizy: porównanie wewnątrzosobowe. Ewaluatorzy wybierają do oceny przedmioty, z których dostali lepsze stopnie**

Skoro ponad 30% studentów w każdym z badań ocenia tylko część zajęć, w których uczestniczyli, to warto zadać pytanie, czy wybierają do oceny zajęcia, z których dostali lepsze stopnie, czy też może wolą oceniać te, z których stopnie były gorsze.

Porównanie (wewnątrz osób) średnimi stopni z przedmiotów ocenionych i nieocenionych można było przeprowadzić dla:

- 1662 czyli 42,4% studentów z pierwszego badania,
- 1123 czyli 33,7% studentów z drugiego badania,
- 804 czyli 34,4% studentów z trzeciego badania.

Wyniki pokazują tendencję do **wybijerania** do oceny **przedmiotów**, z których studenci otrzymują **lepsze stopnie**, ponieważ średni stopień z przedmiotów ocenionych ( $M_1$ ) był istotnie wyższy (test-t dla prób zależnych) niż średni stopień przedmiotów nieocenionych ( $M_2$ ).

- Badanie 1.  $M_1=3,85 > M_2=3,53$ ,  $t(1662) = 17,21$ ;
- Badanie 2.  $M_1=4,08 > M_2=3,84$ ,  $t(1123) = 11,27$ ;
- Badanie 3.  $M_1=4,08 > M_2=3,91$ ,  $t(804) = 6,44$ .

Porównanie międzypersonowe – między grupą pierwszą (która wypełniła wszystkie ankiety) i trzecią (która nie wypełniła żadnej ankiety) prowadzi do analogicznych wniosków.

Taką samą zależność zaobserwowano w przypadku podziału obserwacji na semestry. W każdym z nich średnia stopni z ocenionych zajęć była wyższa od średniej stopni z nieocenionych zajęć.

Podsumowując: nie udało się potwierdzić hipotezy H1: Niski procent zwrotu wiąże się efektem autoselekcji polegającym na tym, że ankiety wypełniają studenci „lepsi” i „gorsi”. Gdy sprawdzano procent zwrotu w podziale na otrzymane stopnie zawsze otrzymywano zależności liniowe „im niższy stopień tym niższy poziom zwrotu”.

Brak potwierdzenia statystycznej zależności z hipotezy 1 nie oznacza, że zły stopień nie motywuje studentów do wypełnienia ankiety ewaluacyjnej. W badaniu opinii 115 studentów i doktorantów na temat ankiet ewaluacyjnych aż **63% z nich przyznało, że zawsze lub często wypełnia ankietę, gdy otrzyma złą ocenę i chce w ten sposób ukarać wykładowcę.**

## V.2. Struktura czynnikowa ankiet ewaluacyjnych

Analiza czynnikowa głównych składowych odpowiedzi na pytania zawarte w ponad 94 tys. ankiet (N= 94130) wykazała 1 czynnik wyjaśniający 66,3% wariacji zmiennych wejściowych.

Taka globalna analiza czynnikowa obarczona jest błędem metodologicznym tj. **nie został spełniony warunek niezależności obserwacji**. Jedna osoba wypełnia więcej niż jedną ankietę, więc jej oceny nie są niezależne. Jest to błąd popełniany w wielu analizach, np. w publikacji w *The Economist*, gdzie przeprowadzono analizę ocen wykładowców dokonanych przez anonimowe osoby na portalu <http://www.ratemyprofessors.com>. Badacze wnioskowali, że studenci doceniają wykładowców łagodnych i mówiących ciekawie, a nisko cenią tych, którzy są surowi i nudni. **Poprawną analizę można przeprowadzić tylko na poziomie pojedynczego kursu** (jeden student miał możliwość ocenienia jednych zajęć tylko raz). Dla rygorystów metodologicznych, aby wyniki analizy czynnikowej można było uznać za sensowne, liczba obserwacji powinna stanowić przynajmniej dziesięciokrotność badanych zmiennych. Dlatego do dalszych analiz **wybrano przedmioty, które zostały ocenione przez przynajmniej 100 osób**. W sumie było to 50 kursów z Badań 1. i 2. (każdy prowadzony był przez jednego wykładowcę) i dla każdego z nich wykonano analizę czynnikową dla ograniczonego zestawu pytań (pominięto pytanie o dostępność wykładowcy i o przydatność zajęć w toku studiów).

**Wyniki 50 analiz czynnikowych** pokazały, że w **48** przypadkach wyodrębniono **jeden czynnik** wyjaśniający od 54,6% do 80,7% zmienności pytań ankiety. Tylko w dwóch przypadkach wyodrębniono dwa czynniki.

Podsumowując: **odpowiedzi na pytania ankiet ewaluacyjnych są wysoko skorelowane** – można mówić o **efekcie aureoli** – np. jeśli wykładowca jest postrzegany jako przyjazny studentom, to jest postrzegany także jako prezentujący materiał w sposób jasny i czytelny.

Przeprowadzone analizy potwierdzają hipotezę drugą H2: **Ewaluatorzy oceniają zajęcia „jednowymiarowo”, czego przejawem jest efekt aureoli**. Można więc budować wskaźniki uśredniając odpowiedzi. Trzeba jednak pamiętać o warunku trafności fasadowej – choć odpowiedź na pytanie o stopień zaangażowania studenta w naukę przedmiotu może

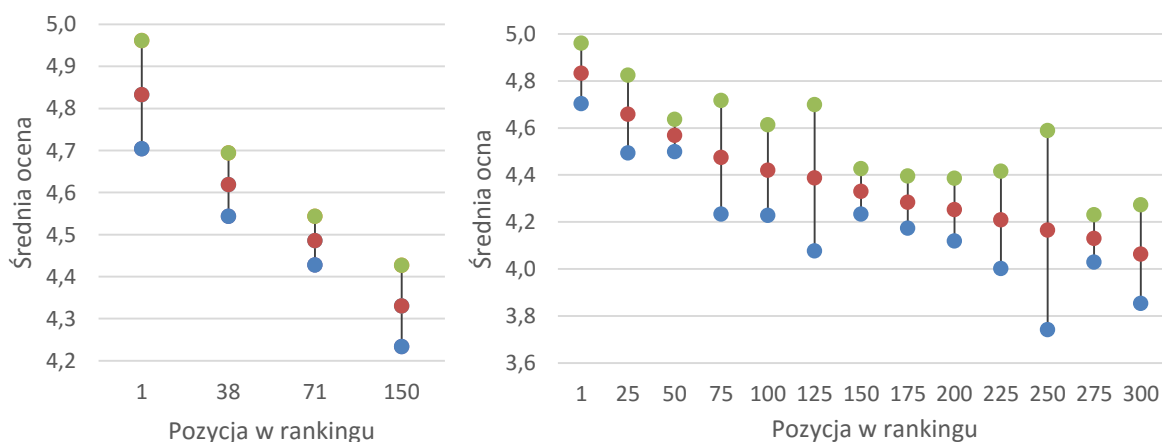
korelować z odpowiedziami na inne pytania, to nie powinna ona wchodzić w skład wskaźnika oceny wykładowcy, ponieważ poziom zaangażowania studenta w zajęcia może zależeć np. od jego planów zawodowych, na które wykładowca nie ma wpływu.

### V.3. Ocena metodologicznej poprawności porównania wykładowców za pomocą wyników ankiet ewaluacyjnych

Celem zadania badawczego 3. jest sprawdzenie, **czy wyniki ankiet ewaluacyjnych pozwalają na ocenę pracy wykładowców**. Po pierwsze, można zapytać o **rzetelność ocen studentów** rozumianą jako stałość odpowiedzi w czasie. W badaniach ewaluacyjnych zwykle nikt nie pyta o ocenę tego samego wykładowcy wielokrotnie w tym samym semestrze, natomiast w Badaniach 1. i 2. znalazły się pytania o oszacowanie swojej średniej stopni. **Studenci wypełniali co najmniej 6 ankiet po kolei, za każdym razem odpowiadając na to samo pytanie o średnią z ostatniego roku studiów** – można więc oczekiwać, że ich odpowiedzi będą takie same. Co najmniej 20% ewaluatorów (**28% z N=515** w badaniu pierwszym, **22% z N=562** w badaniu drugim) **zmienia zdanie na temat średniej wypełniając kolejne ankiety**. Dodatkowa analiza czasu wypełniania ankiet w badaniu 3 pokazuje, że 70% studentów wypełnia ankiety w jednym podejściu. Nawet jeśli założyć, że student wypełnia ankiety z przerwami, trudno bronić tezy, że w tym czasie zmieniła się jego średnia ocen z ostatniego semestru.

W ocenie wykładowców najczęściej używa się **rankingu średnich arytmetycznych**, ignorując informacje o tym, na podstawie jakich danych ta średnia została wyliczona (liczba ankiet i wariacje ich ocen).

W Badaniu 1. policzono średnią ze wszystkich otrzymanych ocen dla **723 wykładowców**, których oceniło **co najmniej 15 studentów** i na tej podstawie zbudowano **listę rankingową**. Można byłoby na tej podstawie przyznać nagrody dydaktyczne dwudziestu najwyższej notowanym wykładowcom (reguła wykorzystywana w tej uczelni), ale gdy przyjrzeć się przedziałom ufności dla tych wyników, zauważyć można, że aż **44 wykładowców nie różni się istotnie od lidera rankingu**.



Rysunek 3. Listy rankingowe wykładowców – wybrane przedziały ufności ocen pracy wykładowcy. Opracowanie własne.

W kolejnej analizie porównano oceny dwóch grup studenckich mających zajęcia z dwóch przedmiotów prowadzonych przez tych samych wykładowców. Przeprowadzona analiza wykazała istotny efekt grupy i brak istotnego efektu wykładowcy/przedmiotu. Wniosek – **ten sam wykładowca uczący tego samego przedmiotu uzyskuje odmienne oceny od różnych grup studentów**, co pokazuje, że ocena jakości procesu dydaktycznego nie zależy tylko umiejętności prowadzącego, ale także od uczącej się grupy.

W świetle przedstawionych analiz wykazujących należy odrzucić hipotezę trzecią: H3: Porównywanie wykładowców za pomocą wyników ankiet ewaluacyjnych przeprowadzane jest w sposób poprawny metodologicznie.

#### **V.4. Wpływ przewidywanych stopni na ewaluację wykładowcy**

W zadaniu badawczym 4. sprawdzono związek między uzyskanym przez ewaluatora stopniem (zarówno wystawionym w protokole jak i przewidywanym przez studenta) a oceną nauczyciela akademickiego (na skali 1 do 5).

Przeanalizowano wyniki ponad 22 tysięcy ankiet (N=22 405), które zostały połączone z informacją o stopniach (wystawianych na skali 2–6).

Analizy przeprowadzono w sumie dla **332 kursów** (213 w Badaniu 1. oraz 119 w Badaniu 2.), dla których liczba wypełnionych ankiet była większa niż 30.

Okazało się, że **przewidywany stopień** był lepszym predyktorem polecenia niż stopień otrzymany, co tłumaczyć można faktem, że ankiety ewaluacyjne wypełniane są przed sesją egzaminacyjną.

**Analiza rozkładu współczynników korelacji** dla 332 kursów pokazuje, że wyniki są różne – istotnie dodatni współczynnik korelacji między przewidywanym stopniem a poleceniem kursu wystąpił w przypadku 121 kursów, co stanowi 36% zbioru.

W dotychczasowych analizach pokazano, że **wypełniający ankiety studenci charakteryzują się wyższymi stopniami** od tych, którzy tych ankiet nie wypełniają (zależność była replikowana w trzech badaniach, zarówno w porównaniach między osobami, jak i wewnątrz osób). Oznacza to, że współczynniki korelacji między stopniami a poleceniem zajęć są liczone na „obciętej” próbie, w której **niedoreprezentowani są studenci z gorszymi stopniami**.

Hipotezę 4 mówiącą, że studenci otrzymujący lepsze stopnie lepiej oceniają zajęcia można uznać za częściowo potwierdzoną, ponieważ wykazano:

- jej prawdziwość dla 36% kursów;
- związek między procentem zwrotu a istotnością współczynnika korelacji.

Trzeba jednak podkreślić, że związek ten nie jest silny i trzeba go traktować jako tylko tendencję.

Związek ewaluacji z wystawianymi stopniami: przewidywania ekspertów

Ważne jest nie tylko to, czy rzeczywiście istnieje związek pomiędzy wystawianymi stopniami, ale także przekonania wykładowców dotyczące występowania owej zależności. W psychologii dobrze opisane jest zjawisko spełniających się przepowiedni, pokazujące, że przewidywania mogą mieć tendencję do potwierdzania. Konsekwencją przewidywania, że istnieje związek między wystawianym stopniem a ewaluacją, może być tendencja do wystawiania lepszych stopni.

O wyrażenie opinii dotyczącej istnienia zależności między wystawianymi stopniami a wynikami ewaluacji poproszono 162 wykładowców Wydziału Zarządzania UW oraz 33 członków Rady Naukowej Instytutu Studiów Społecznych UW o odpowiedź na tylko jedno pytanie:

*[...] Czy stopień na skali od 2 do 5, jaki student otrzymał, ma wpływ na to, czy poleci te zajęcia innym (skala 1–5)? (a) TAK, im lepszy stopień tym większa szansa na polecenie zajęć; (b) TAK, im gorszy stopień tym większa szansa na polecenie zajęć; (c) NIE, nie ma związku liniowego/krzywoliniowego.*

Odpowiedź zwrotną uzyskano od 59 wykładowców. W sumie wśród interpretowalnych odpowiedzi związek dodatni przewidywało 46% (27 osób), a brak związku 31 osób. Kiedy wykładowców biorących udział w sondzie eksperckiej podzielono na grupę z większym (powyżej 12 lat; N=31) i mniejszym (reszta; N=21) doświadczeniem dydaktycznym okazało się, że **przekonanie o wpływie stopni na oceny jest podzielana dużo częściej przez młodszych (76%) niż przez starszych (25%)**.

Test hipotezy 5. Upowszechnienie ankiet ewaluacyjnych powoduje inflację wystawianych stopni

Skoro wykładowcy mogą podejrzewać, że obowiązuje tu reguła wzajemności, to (jeśli chcą otrzymywać dobre oceny ewaluacyjne od studentów) mogą mieć tendencję do dawania im lepszych stopni. W związku z tym, że z każdym rokiem coraz więcej uczelni wprowadza wyniki ankiet do ocen pracowniczych, powinno być możliwe zaobserwowanie inflacji średnich stopni w kolejnych latach. Dlatego w następnym kroku przetestowano hipotezę 5., zakładającą występowanie inflacji wystawianych przez wykładowców stopni na 2 uczelniach.

Analizy zostały przeprowadzone osobno na trzech zbiorach danych zawierających informacje o osiągnięciach studentów kończących:

- studia licencjackie na kierunku *Zarządzanie* (N=1059) w latach **2010–2015**;
- studia licencjackie na kierunku *Finanse i rachunkowość* (N=1026) w latach **2011–2015**;
- jednolite studia magisterskie na kierunku *Psychologia* (N=2928) w latach **2009–2014**.

Analiza dostępnych danych nie potwierdziła hipotezy przewidującej wzrost wystawianych stopni. Może to być spowodowane tym, że współczynniki korelacji mogą być zaniżone, jeśli jedna ze zmiennych ma ograniczony zakres zmienności. Przedział o długości 4 lub 6 lat może być zbyt krótki, aby móc zaobserwować trendy. Niestety nie dysponowano danymi obejmującymi szerszy zakres czasowy.

## VI. Implikacje dla zarządzania w szkolnictwie wyższym

W rozprawie doktorskiej przeanalizowano **ponad 130 tysięcy ankiet ewaluacyjnych** z dwóch polskich uczelni. Nie jest to próba reprezentatywna dla wszystkich uczelni polskich. Jednak fakt, że stwierdzone zależności są zbieżne na obu analizowanych uczelniach oraz że analizy były powtarzane osobno na danych zbieranych w różnych latach zwiększa zaufanie do możliwości generalizowania opisanych zależności.

Wprowadzenie ankiet ewaluacyjnych na uczelniach wyższych **zmieniło wzajemną pozycję wykładowców i studentów**. Oceniają się oni wzajemnie. Różnica między nimi polega na tym, że tylko wykładowcy wystawiają stopnie w sposób jawny – co więcej, muszą je szczegółowo uzasadniać, studenci zaś oceniają wykładowców w sposób anonimowy i bez brania jakiegokolwiek odpowiedzialności za wyrażane opinie. Ta sytuacja może być źródłem frustracji i owocować spadkiem motywacji do pracy. Wraz ze wzrostem liczby studentów liczba nauczycieli akademickich na świecie wzrosła o prawie 50% w ciągu zaledwie ośmiu lat (z ok. 6,5 mln w 1999, do 9,5 mln w 2007). Przewiduje się, że liczba studentów i nauczycieli akademickich będzie rosła jeszcze przez przynajmniej dwadzieścia lat. Współczesne uniwersytety są wspólnotami, które konkurują między sobą dążąc do sukcesu biznesowego, tak jak firmy w warunkach wolnorynkowych. Jako **wskaźniki efektywności pracy dydaktycznej** coraz częściej przyjmuje się wyniki ankiet ewaluacyjnych, ale jak zostało to wykazane, jest to wskaźnik wątpliwy ze względu na chociażby niski procent zwrotu.

W dobrowolnych badaniach ewaluacyjnych – tak jak w badaniach sondażowych – procent zwrotu jest coraz niższy. **Procent zwrotu w internetowych ankietach ewaluacyjnych nie przekracza 30%**, jeśli nie stosuje się dodatkowych zabiegów, np. monitów. Przeprowadzone badanie wykazało, że **ponad 50% studentów nie wypełnia żadnej ankiety**. Nie wiemy więc, jak oceniają zajęcia. Niski procent zwrotu może nie być dyskwalifikujący, gdy ewaluacja ma pełnić funkcję formatywną, przy założeniu, że informacja pochodzi tylko od osób, które chcą pomóc w podnoszeniu jakości dydaktyki. Nie można jednak na podstawie tak małego procentu zwrotu stosować ewaluacji do jej funkcji sumatywnej: oceny pracy wykładowców.

Zastrzeżenie budzi fakt, że na uczelniach, na których pilnuje się **standardów stosowania metod naukowych**, prowadzi się **badania ankietowe lekceważące te standardy**. Z powodu nadmiernej troski o anonimowość respondenta ankiety są anonimizowane na etapie zbierania danych, a nie na etapie udostępniania ich do analizy. Dbałość o anonimowość, skądinąd godna pochwały, nie pozwala **sprawdzać pod względem zgodności logicznej** otrzymanych danych.

W unikalnym zbiorze danych, jakim dysponowałem gdy możliwe było połączenie wyników ankiet wypełnianych przez tego samego studenta okazało się, że **co najmniej 20% wśród tych, którzy wypełniają ankiety nie robi tego rzetelnie**: wypełniając co najmniej 6 ankiet po kolei, zmienia odpowiedzi w pytaniu o średnią z ostatniego roku studiów.