

# Seweryn Rudnicki, Anna Szwed

---

## Ocenianie oceniających : trzy pytania o ewaluację zajęć dydaktycznych

---

Kultura i Polityka : zeszyty naukowe Wyższej Szkoły Europejskiej im. ks. Józefa Tischnera w Krakowie nr 5, 127-139

---

2009

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Seweryn Rudnicki\*  
Anna Szwed\*\*

## OCENIANIE OCENIAJĄCYCH. TRZY PYTANIA O EWALUACJĘ ZAJĘĆ DYDAKTYCZNYCH<sup>1</sup>

### Streszczenie

Artykuł prezentuje problematykę ewaluacji nauczania w szkolnictwie wyższym. Każda z trzech części artykułu stanowi próbę odpowiedzi na konkretne pytanie: czym tak naprawdę jest (lub powinna być) ewaluacja zajęć, jak przeprowadza się ewaluację w uznanych światowych ośrodkach akademickich oraz czy opinie wyrażane przez studentów w ankietach ewaluacyjnych są wiarygodnym źródłem danych. W pierwszej części autorzy wyjaśniają pojęcie ewaluacji, zwracając uwagę na fakt, że obejmuje ona nie tylko surowe dane, ale także element interpretacji, oraz podkreślając jej podwójny: formatywny i sumatywny charakter. W części drugiej przedstawiono „dobre praktyki” funkcjonujące w ramach systemów ewaluacji na wybranych uniwersytetach z kręgu anglosaskiego. Autorzy wspominają tutaj o takich problemach, jak: transparentność systemu, triangulacja zbieranych danych, procedury ewaluacyjne, zaangażowanie studentów itp. Trzecia część dotyczy zagadnień rzetelności i trafności ocen studenckich oraz czynników, które mogą wpływać na tę formę pomiaru.

127

### Słowa kluczowe

Ewaluacja nauczania, ewaluacja formatywna i sumatywna, metodologia ewaluacji nauczania, oceny studenckie, *peer observations*, autoewaluacja.

---

\* **Seweryn Rudnicki** – socjolog i psycholog społeczny, wykładowca w Wyższej Szkole Europejskiej im. ks. Józefa Tischnera i doktorant w Instytucie Socjologii UJ. Interesuje się socjologią internetu, socjologią ciała i metodologią badań społecznych. E-mail: [rudnicki@wse.krakow.pl](mailto:rudnicki@wse.krakow.pl)

\*\* **Anna Szwed** – socjolożka, wykładowczyni w Wyższej Szkole Europejskiej im. ks. Józefa Tischnera w Krakowie, doktorantka na Uniwersytecie Jagiellońskim. Przygotowuje rozprawę na temat obrazu kobiety w nauczaniu Kościoła rzymskokatolickiego oraz w świadomości polskich księży. Zainteresowania naukowe: socjologia kultury, socjologia religii, *gender studies*. E-mail: [aszwed@wse.krakow.pl](mailto:aszwed@wse.krakow.pl)

<sup>1</sup> Prace badawcze, których jednym z rezultatów jest niniejszy tekst, były prowadzone w ramach realizowanego przez Wyższą Szkołę Europejską projektu „Kompetentnie ku przyszłości” finansowanego z Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki, Działanie 4.1.1.



Liczba publikacji naukowych poświęconych ewaluacji dydaktyki pod koniec ubiegłego wieku sięgała dwóch tysięcy pozycji na całym świecie (McKeachie, Kaplan 1996), co zdaje się świadczyć zarówno o stopniu skomplikowania problemu, jak i o jego znaczeniu. O ile nie trzeba nikogo przekonywać o konieczności zapewnienia wysokiej jakości kształcenia w szkolnictwie wyższym, o tyle rzeczowa refleksja nad sposobem prowadzenia ewaluacji dydaktyki wydaje się zasadna. Obserwacja sposobu funkcjonowania uczelni wyższych w naszym kraju wskazuje bowiem na to, że mimo wprowadzania ewaluacji studenckich na coraz szerszą skalę ciągle nie jest powszechne traktowanie studenta jako klienta instytucji edukacyjnej (którego zadowolenie z jakości produktu stanowi jedną z podstawowych racji bytu uczelni) czy jako jednostki współtworzącej unikalną *communitas*, jaką ma być uniwersytet.

Niniejsze opracowanie ma na celu udzielenie odpowiedzi na podstawowe pytania związane z ewaluacją zajęć dydaktycznych na podstawie przede wszystkim literatury przedmiotu, rozwiązań funkcjonujących w krajach anglosaskich, a także doświadczenia autorów w przeprowadzaniu ewaluacji dydaktyki.

### **Czy dane mówią same za siebie?**

Pojęcie ewaluacji dydaktyki wymaga szczególnej uwagi w związku z pojawiającymi się często nieścisłościami w jego rozumieniu. Na gruncie polskiego szkolnictwa wyższego, gdzie ewaluacja nauczania i problematyka zapewniania odpowiedniej jakości kształcenia wciąż są tematami stosunkowo nowymi w porównaniu np. z krajami anglosaskimi, w których należą one do akademickiej rutyny, ewaluację najczęściej utożsamia się z semestralnymi ocenami studenckimi dokonywanymi za pomocą ankiet. W rezultacie uznaje się, że jeśli uczelnia posługuje się, w sposób bardziej lub mniej regularny, ankietami (najczęściej mającymi dodatkowo charakter wyłącznie lub głównie ilościowy), w których studenci oceniają przedmiot i wykładowcę, to możemy mówić, że taka jednostka posiada system ewaluacji. Nic bardziej mylnego. Oceny studentów zbierane za pomocą ankiet są tylko jednym ze źródeł danych wykorzystywanych w procesie ewaluacji, który nie sprowadza się do zaprezentowania surowych danych uzyskanych z jednego źródła, lecz zakłada interpretację ocen będących, w wersji optymalnej, wynikiem tzw. triangulacji (wykorzystania różnych narzędzi zbierania danych, np. ankiety studenckie, hospitacje, *peer observations*, auto-

ewaluacja). Interpretacja, a więc nadanie znaczenia (także w związku z kontekstem oceniania<sup>2</sup>), jest zatem równie istotnym elementem ewaluacji, co same dane.

Drugie ważne zastrzeżenie terminologiczne wiąże się z celem ewaluacji. W literaturze przedmiotu najczęściej stosuje się niezwykle praktyczne rozróżnienie na tzw. ewaluację formatywną i ewaluację sumatywną (Felder, Brent 2004). W wypadku pierwszej z nich kładzie się nacisk na rozwijanie kompetencji dydaktycznych prowadzącego zajęcia. Na podstawie zgromadzonych danych (m.in. z ankiet studenckich, obserwacji przez zaprzyjaźnionych wykładowców, hospitacji) prowadzący, wspomagany przez doradcę edukacyjnego lub inną kompetentną osobę, podejmuje refleksję dotyczącą mocnych i słabych stron swojego stylu nauczania, możliwych obszarów poprawy jego jakości i kierunku rozwoju. W ewaluacji formatywnej punktem wyjścia i celem, na którym się koncentrujemy, jest zatem sam proces dydaktyczny. Ewaluacja sumatywna ma natomiast na celu dostarczenie danych, które mogą zostać wykorzystane w rozstrzygnięciach o charakterze personalnym, takich jak: decyzja o dalszym prowadzeniu określonego przedmiotu lub jego zmianie, rozstrzygnięcia dotyczące zatrudnienia i jego formy, awansu, wynagrodzenia, nagród itp.

Choć w obu typach ewaluacji możliwe jest wykorzystanie podobnych narzędzi zbierania danych, można jednak z dużym prawdopodobieństwem zakładać, że osoby dokonujące ewaluacji dla celów podsumowujących będą się raczej skupiały na stwarzających możliwość porównywania danych ilościowych. W wypadku ewaluacji formatywnej szczególnego znaczenia będą natomiast nabierać wszelkie dane o charakterze jakościowym (np. uzyskane za pomocą pytań o charakterze otwartym).

129

### **Na czym polega system ewaluacji zajęć dydaktycznych?**

W odpowiedzi na to pytanie proponujemy przegląd „dobrych praktyk”<sup>3</sup> spotykanych na wybranych uniwersytetach z kręgu anglosaskie-

<sup>2</sup> Tym kontekstem może być np. typ przedmiotu, którego uczy prowadzący/a (kurs obligatoryjny, kurs fakultatywny), doświadczenie zawodowe prowadzącego/ej (czy jest to jego/jej pierwszy semestr nauczania) itp.

<sup>3</sup> Wykorzystane tu informacje zostały zaczerpnięte z ogólnodostępnych źródeł w postaci materiałów zamieszczonych na stronach internetowych uczelni (ich lista znajduje się w Bibliografii). Jest to zatem raczej analiza typów idealnych rozwiązań ewaluacyjnych niż ich realnego funkcjonowania (choć można w sposób uzasadniony zakładać, że przywołane dokumenty powstały w wyniku doświadczeń konkretnej uczelni).

go<sup>4</sup>. Tym, co łączy wszystkie brane pod uwagę uczelnie, jest z pewnością duży stopień transparentności systemu ewaluacji (nie oznacza on oczywiście upublicznienia danych o charakterze poufnym, osobowym itp.). Może o tym świadczyć sam fakt zamieszczenia szczegółowych informacji o funkcjonującym na uczelni systemie ewaluacji zajęć dydaktycznych na ogólnodostępnej stronie internetowej. Z pewnością oprócz roli informacyjnej zabieg ten pełni także funkcję uwiarygodniającą, wskazującą na dbałość o jakość nauczania, a pośrednio również i promocyjną.

W analizowanych dokumentach opisujących „polityki ewaluacyjne” uczelni ważnym elementem jest określenie celów, jakim ma służyć ewaluacja nauczania. Przykładowo, Stanford University zwraca uwagę na trzy zasadnicze zadania: dostarczenie informacji zwrotnej wykładowcom, dostarczenie studentom informacji pomocnej w wyborze kursów, dostarczenie ogólnej informacji o jakości kursów do celów administracyjnych (polityka kadrowa, wynagrodzenia itp.). W podobnym duchu ewaluacja rozumiana jest w polityce University of British Columbia, gdzie za jej podstawowe cele uznaje się: dostarczenie danych, które pozwolą na poprawę jakości nauczania studentów, dostarczenie danych o ogólnej jakości nauczania na danym kierunku, wydziale itp., dostarczenie danych wykładowcom (także po to, by mogli dalej rozwijać swoje umiejętności), dostarczenie danych uniwersytetowi do celów operacyjnych (ocena pracownika, wynagrodzenie, nagrody itp.). Zarówno w tych, jak i w innych analizowanych wypadkach zwraca uwagę silny nacisk na podwójną funkcję ewaluacji – formatywną i sumatywną.

Wszystkie analizowane uczelnie zwracają także szczególną uwagę na konieczność triangulacji metodologicznej w zbieraniu danych, czyli na wykorzystanie różnych sposobów uzyskiwania informacji na temat prowadzonych zajęć dydaktycznych. Najbardziej popularną, podstawową i obecną we wszystkich jednostkach formą są oczywiście ankiety studenckie (*student ratings*), wśród których można wyróżnić zarówno takie, które oceniają prowadzenie pojedynczego przedmiotu, jak i takie, które dokonują oceny bardziej ogólnej – całości programu studiów, warunków studiowania itp. (np. University of Auckland – University Review, Graduating Year Review, Student Experience and Satisfaction Survey; Oxford University – Student Course Experience Questionnaire).

<sup>4</sup> Uczelnie, do których rozwiązań się odwołujemy, to (w nawiasach podano skróty stosowane w dalszej części artykułu): University of British Columbia (BC), Ohio State University (OSU), Stanford University (SU), University of Auckland (UA). W ograniczonym zakresie korzystamy także z propozycji Oxford University (OxU) oraz Washington State University (WSU).

Ankiety studenckie są uznawane za najlepszą pojedynczą metodę zbierania danych wykorzystywanych w procesie ewaluacji dydaktyki; są też metodą najczęściej stosowaną i mającą największe znaczenie (zob. Paswan, Young 2002). Oprócz ankiet opracowywanych przez ekspertów na potrzeby konkretnych uczelni wiele placówek wykorzystuje także narzędzia standaryzowane o zweryfikowanej trafności i rzetelności, które dodatkowo dają możliwość porównywania wyników ewaluacji między uczelniami<sup>5</sup>.

Większość analizowanych systemów ewaluacji uwzględnia także jakąś formę recenzji koleżeńskiej, najczęściej w postaci tzw. *peer observations*, w trakcie których wykładowcy obserwują i oceniają swoje zajęcia (często określa się wymaganą częstotliwość tej formy oceny dla każdego stanowiska uniwersyteckiego, np. asystenci – raz w roku). Innym proponowanym narzędziem ewaluacji jest samoocena prowadzącego zajęcia. W wybranych wypadkach może ona stanowić część tzw. *teaching portfolio*. Obszerny i szczegółowy opis tej formy zbierania danych zamieszcza na swoich stronach Washington State University. Przygotowywane przez samego prowadzącego portfolio, choć oryginalne i twórcze, powinno zgodnie z zaleceniami uczelni zawierać między innymi następujące elementy: cele stawiane sobie przez prowadzącego w zakresie dydaktyki; obowiązki – w tym m.in.: prowadzone kursy, indywidualna praca ze studentem, praca ze specjalnymi grupami studentów, doradztwo, doskonalenie dydaktyczne prowadzącego, wprowadzane innowacje; ewaluacje – w tym m.in. wyniki ocen studenckich, *peer review*, uzyskane nagrody itp. Ostatnia część portfolio nosi nazwę „rezultaty” i obejmuje m.in. informacje dotyczące sukcesów studentów, w których osiągnięciu wykładowca miał swój udział, opublikowane artykuły na temat nauczania oraz przygotowane przez prowadzącego materiały dydaktyczne (podręczniki, tzw. *readery*, wizualne pomoce naukowe etc.). Innym źródłem cennych opinii na temat dydaktyki (choć w mniejszym stopniu poszczególnych zajęć), a także innych aspektów funkcjonowania uczelni mogą być wyniki badania absolwentów.

W publikowanych dokumentach uczelnie na różnych stopniach szczegółowości określają procedury związane z ewaluacją. Ze względu na duże rozmiary instytucji często decydują się one na jakąś formę decentralizacji, np. delegując opracowanie procedury na poziom wy-

---

<sup>5</sup> Np. Instructional Development and Effectiveness Assessment (IDEA), Student Instructional Report, Student Evaluation of Educational Quality (SEEQ), Instructor and Course Evaluation System, Student Instructional Rating System, Instructional Assessment System (Hobson, Talbot 2001).

działu czy departamentu lub dopuszczając do dywersyfikacji narzędzi ze względu na specyfikę jednostek uczelnianych, potrzeby poszczególnych wykładowców czy organizacji studenckich. Ważnym elementem każdego z analizowanych systemów ewaluacji są kwestie poufności w procesie zbierania danych i publikowania informacji. Pierwsza z nich wiąże się z zapewnieniem anonimowości studentom oceniającym kurs (np. Stanford University deklaruje, że studenci wypełniają elektroniczne ankiety na zewnętrznym, niezależnym serwerze). W drugim wypadku uczelnie często zastrzegają sobie prawo do nieewaluowania lub niebrania pod uwagę wyników ewaluacji kursów prowadzonych po raz pierwszy (lub będących pierwszymi kursami prowadzącymi), kursów ze zbyt małą liczbą oceniających studentów itp. Niemal wszystkie z analizowanych uczelni organizują oceny studenckie w końcowej części semestru (lub innej jednostki rozliczeniowej), przed egzaminami.

Te uczelnie, które szczegółowo określają procedurę ewaluacji dydaktyki, często umieszczają w niej także obowiązujący w ramach instytucji obieg informacji związanych z oceną prowadzących i zajęć, w tym kto i do jakich informacji ma dostęp. Warto podkreślić jest także uwzględnienie roli studentów w procesie ewaluacji w większości analizowanych dokumentów. W najbardziej oczywistym wymiarze wiąże się ono z oparciem ewaluacji dydaktyki na ocenach studentów. Inne aspekty tego zaangażowania to w zależności od uczelni m.in. rola organizacji studenckich jako podmiotu zainteresowanego wynikami ewaluacji, możliwość wprowadzenia przez taką organizację własnych pytań do formularza ankiety, a także przeprowadzenia oceny zajęć na wniosek studentów (UA).

Ostatni element, o którym warto wspomnieć w kontekście systemów ewaluacji funkcjonujących w analizowanych uczelniach, wiąże się z formatywnym wymiarem ewaluacji. Mowa tutaj o rozmaitych mechanizmach wsparcia, jakie udzielane jest wykładowcom zarówno w samym procesie dydaktycznym, jak i w kontekście prowadzonej ewaluacji i jej wyników. Niemal wszystkie uczelnie dysponują specjalnie wydzielonymi jednostkami zajmującymi się jakością nauczania, które stanowią nie tylko podmiot w mniejszym lub większym stopniu czuwający nad realizacją polityki ewaluacyjnej, ale także udzielający porad i wsparcia dydaktykom. Godne uwagi są także proste w swej zasadzie, a niezwykle użyteczne przedsięwzięcia w postaci przygotowania zbioru najczęściej zadawanych przez kadrę i studentów pytań dotyczących ewaluacji – FAQ (SU) czy dokumentu wyjaśniającego przyjęte kryteria oceny oraz wspomagającego proces interpretacji uzyskanych danych (OSU).

### Czy oceny studentów są wiarygodne?

W podtekście tego pytania kryje się brak zaufania do poprawności ocen przyznawanych prowadzącym przez studentów i przekonanie o ich możliwym zniekształceniu przez różne czynniki. Czy studenci mają wystarczające kompetencje, by prawidłowo oceniać sposób prowadzenia zajęć? Czy ich ocena nie jest swoistym „konkurem” na najbardziej lubianego prowadzącego/a? Czy najlepszych ocen nie dostają te osoby, u których łatwo otrzymać dobrą ocenę? Czy nie lepiej byłoby, gdyby prowadzących oceniali inni przedstawiciele kadry?

Warto przypomnieć, że ocena zajęć dokonywana przez studentów jest w istocie specyficznym rodzajem pomiaru, który zgodnie z podstawami metodologii powinien być rzetelny (a więc m.in. powtarzalne – zmierzenie tego samego obiektu tym samym narzędziem powinno przynieść ten sam wynik) i trafny (a więc narzędzie powinno mierzyć to, do czego zostało zaprojektowane, np. ocena sposobu prowadzenia zajęć może dotyczyć umiejętności wyjaśniania trudnych zagadnień przez wykładowcę, ale jej wynik nie powinien zależeć od motywacji studentów do nauki, trudno bowiem karać prowadzącego za gorsze oceny uzyskiwane w grupach wyraźnie mniej chętnych do pracy). Idealne spełnienie tych warunków pozostaje oczywiście poza zasięgiem jakiegokolwiek pomiaru, jednak ankiety studenckie są uznawane za rzetelne i dość trafne narzędzia.

Rzetelność ankiet potwierdzono w wielu badaniach, w których zazwyczaj stosowano wystandaryzowane i używane powszechnie na uczelniach narzędzia (z powodu mnogości badań w dziedzinie ewaluacji podawane poniżej wyniki zazwyczaj opierają się na metaanalizach wielu badań). Większość poprawnie skonstruowanych narzędzi legitymuje się wewnętrzną spójnością na co najmniej akceptowalnym poziomie (współczynniki spójności na poziomie między 0,7 a 0,9 [Cashin 1995; Ali, Sell 1998]). Wykazano również, że podstawowym czynnikiem decydującym o ocenie jest osoba prowadzącego, a nie wykładany przedmiot, co dodatkowo wpływa na stabilność ocen (Cashin 1995). O wysokiej rzetelności świadczy duża zgodność oceny danych zajęć wśród uczestniczących w nich studentów, a także podobieństwo oceny tych samych zajęć przez obecnych studentów i absolwentów minimum rok po ukończeniu studiów (mediana korelacji na poziomie 0,83 [Overall, Marsh 1980 za: Overall, Marsh 1982]). Nie do utrzymania byłby zatem argument, że studenci nie są w stanie adekwatnie ocenić zajęć lub że mając większe doświadczenie zawodowe i życiowe, będą oceniać je inaczej. Wreszcie okazuje się, że studenci są śred-



nio znacznie bardziej zgodni w ocenie danych zajęć niż przedstawiciele kadry dydaktycznej, którzy również je obserwowali (wewnętrzna korelacja ocen wśród przedstawicieli kadry była na poziomie 0,26 [Hobson, Talbot 2001]). Co więcej, oceny innych dydaktyków nie korelowały wystarczająco wysoko z innymi kryteriami jakości dydaktyki, chyba że oceniający zostali wcześniej przeszkoleni pod kątem zasad dokonywania oceny.

Warto zaznaczyć, że warunkiem poprawności ewaluacji jest wykorzystanie danych zebranych w odpowiedni sposób. Szczególnie istotne jest tutaj traktowanie z ostrożnością wyników ankietyzacji w małych grupach zajęciowych (poniżej 15 osób [Cashin 1995]), ponieważ mała liczba badanych jest zagrożeniem dla rzetelności wyników. Warunkiem koniecznym jest też wysoka stopa zwrotów, bowiem osoby, które uczestniczyły w zajęciach, a nie oceniły ich, mogą się różnić od tych, które wyraziły swoje opinie (przykładowo może się okazać, że ankiety są najchętniej wypełniane przez najbardziej niezadowolonych lub najbardziej zadowolonych studentów, co w oczywisty sposób zniekształca wyniki). Czynnikiem stopy zwrotów jest istotny zwłaszcza w wypadku ankiet internetowych (Johnson 2002), często wybieranych z racji niskich kosztów zbierania i obróbki danych. Popularne sposoby „zachęcenia” do wypełniania ankiet (np. zablokowanie możliwości podglądu ocen egzaminacyjnych, brak możliwości zaliczenia semestru itp.) mogą budzić wątpliwości co do jakości danych uzyskiwanych w ten sposób (ankiety mogą być wypełniane pośpiesznie lub nieuważnie).

Większe wątpliwości wiążą się z problemem trafności ocen studenckich. Zakłada się, że jeżeli metody oceny są trafne, prowadzący, na których zajęciach studenci uczą się najwięcej i najefektywniej, powinni być oceniani najwyżej. Sprzeczny z tym przekonaniem jest tzw. efekt doktora Focha – udający prowadzącego aktor, który wygłosił dla studentów efektowny, ale pozbawiony merytorycznej treści wykład, został bardzo wysoko oceniony (Naftalin, Ware, Donnelly 1973 za: Overall, Marsh 1982: 11). Jak się okazało, eksperyment miał istotne metodologiczne słabości, ale pewne wątpliwości pozostały, tym bardziej że inne badania pokazały zależność ocen od ekspresyjności, pasji i dynamiki prowadzącego.

Najczęściej stosowanym sposobem wykazania trafności ocen studenckich jest porównanie ich wyników z ocenami przeprowadzanymi w inny sposób. Metaanalizy pokazują, że istnieje umiarkowanie silny związek oceny zajęć przez studentów ze stopniem wynoszonej przez nich wiedzy potwierdzonej przez zewnętrzne egzaminy (d'Apollonia, Abrami 1997 za: Ali, Sell 1998) i wspólne egzaminy grup, które uczy-

li różni prowadzący (Cashin 1995; Felder 1992). Choć związek ten nie jest bardzo silny, typowo uzyskiwane wartości współczynników korelacji (ok. 0,3-0,5) są zwyczajowo uznawane w naukach społecznych czy w psychologii za pokazujące ważne związki. Dodatkowo warto zauważyć, że ostateczny wynik procesu kształcenia nie zależy przecież wyłącznie od umiejętności prowadzącego, ale także od takich czynników, jak motywacja studentów, czas poświęcony na naukę czy kapitał kulturowy wyniesiony z domu rodzinnego. Większość badań potwierdza również trafność ocen studenckich udokumentowaną korelacjami z wynikami innych metod oceny prowadzących. Stwierdzono podobieństwo na średnim poziomie między wynikami ankiet studenckich a samooceną prowadzących, średnie lub wysokie podobieństwo ocen studentów i absolwentów oraz różny poziom podobieństwa opinii studentów i innych prowadzących, którzy z kolei są zazwyczaj mniej zgodni ze sobą niż studenci (Ali, Sell 1998; Cashin 1995; Felder 1992; Gursoy, Umbreit 2005; Hobson, Talbot 2001).

Jak się wydaje, wokół ankiet studenckich nagromadziło się także wiele błędnych przekonań dotyczących czynników, od których są uzależnione ich wyniki i które mogą zniekształcać rezultaty oceny, czyniąc ją nietrafną (Simpson, Sigauw 2000). Warto zatem zdawać sobie sprawę z tego, jakie czynniki rzeczywiście wpływają na wysokość ocen studenckich – oto ich lista oparta na metaanalizach wielu badań (Cashin 1995; Felder 1992; Stapleton, Murkison 2001)<sup>6</sup>:

- motywacja studentów – wyższe oceny przyznają studenci, którzy są zainteresowani przedmiotem lub mocno zmotywowani do nauki;
- dobrowolność udziału w kursie – wyżej oceniane są zajęcia wybierane niż obligatoryjne;
- spodziewana ocena – ten czynnik był szczególnie starannie badany (zob. Centra 2003; Eiszler 2002; Greenwald 1996; Gump 2007), a uzyskiwane wyniki nie są zupełnie jednoznaczne, jednak istotna część badań pokazuje niestety niewielki (współczynniki korelacji 0,1-0,3) dodatni związek między oceną za kurs, jakiej spodziewają się studenci, a ich opinią na temat prowadzącego, istnieje zatem pewna możliwość „kupowania” przychylności studentów za podwyższenie ocen (a właściwie za dawanie do zrozumienia, że oceny będą wysokie);

<sup>6</sup> Warto jednak zaznaczyć, że zdecydowaną większość z nich przeprowadzono w Stanach Zjednoczonych, a więc w innym kontekście kulturowym. Co więcej, z reguły mają one charakter korelacyjny, potwierdzają zatem pewne ogólne i niejako uśrednione prawidłowości – w poszczególnych wypadkach mamy jednak zapewne do czynienia z sytuacjami, w których np. wymagający prowadzący są oceniani nisko albo pleć osoby prowadzącej obniża lub zawyża oceny.

- ekspresyjny styl prezentacji – pokazuje to wspomniany wyżej „efekt doktora Foxa”, ale akurat ta cecha wydaje się wspierać proces uczenia się, zatem nie przekreśla trafności ocen;
- poziom zaawansowania kursu – wyżej oceniane są kursy zaawansowane niż wprowadzające, jednak związek ten jest słaby;
- dyscyplina akademicka – zgodnie z oczekiwaniami wyżej oceniane są kursy z dziedzin humanistycznych, nieco niżej z nauk społecznych, a najniżej przedmioty matematyczno-przyrodnicze;
- wymagany nakład pracy – przeciwnie do oczekiwań wyżej (choć związek ten jest słaby) oceniane są zajęcia wymagające od studentów znacznego nakładu pracy, jakkolwiek część badań pokazuje związek krzywoliniowy, tzn. przekroczenie pewnego poziomu wymagań skutkuje pogorszeniem wyników ankiet;
- brak zapewnionej anonimowości oceniających – opinie wyrażane nieanonimowo są zazwyczaj lepsze;
- obecność osoby prowadzącej zajęcia w czasie oceny – ten pozornie nieważny szczegół w istocie ma znaczenie, ponieważ wypełnianie w obecności wykładowców ankiety prowadzi do wyższej oceny ich pracy;
- ujawnienie celu ocen – jeżeli studenci zdają sobie sprawę, że wyniki ewaluacji będą używane do decyzji personalnych, oceniają wyżej, niż gdy opinie mają służyć jedynie do poprawy warsztatu osoby prowadzącej;
- niektóre cechy osobowości osoby prowadzącej – szczególnie entuzjizm, energiczność i pozytywna samoocena; związek ten jest o tyle istotny, że osobowość nie ulega zasadniczym zmianom w ciągu życia, trudno więc w tym zakresie doskonalić swój warsztat<sup>7</sup>.

136

Zaskakujące jest jednak, jak wiele czynników w istocie nie ma wpływu na wysokość ocen studenckich. Większość badań nie pokazuje wpływu (Clayson 1999): wieku, płci i rasy osoby prowadzącej, doświadczenia w nauczaniu (oceny prowadzących są stabilne w czasie z wyjątkiem pierwszego okresu pracy, dokładniejsze badania pokazują jednak, że nie zmienia się komponent związany z oceną osobowości, a może się poprawić wiedza czy warsztat osoby prowadzącej), dorobku naukowego, wieku, płci i osobowości studentów, momentu dokonywania ewaluacji (pod warunkiem, że ewaluacja jest przeprowadzana w dru-

<sup>7</sup> Są również badania pokazujące, że sposób postrzegania przez studentów osobowości prowadzącego jest czynnikiem, którym można wyjaśnić 73% wariacji ocen zajęć (Clayson, Sheffet 2006). Co więcej, istotny wpływ tego czynnika da się stwierdzić już po pięciu minutach od rozpoczęcia pierwszych zajęć.

giej połowie semestru), rozmiaru grupy zajęciowej (choć część badań pokazuje, że mniejsze grupy oceniają prowadzących wyżej).

Na koniec warto przypomnieć, że ankiety badają opinie studentów, są więc w procesie ewaluacji danymi wejściowymi, a nie jej wynikiem. Oceny studentów pozostają najbardziej rzetelnym i trafnym sposobem zbierania danych ewaluacyjnych, chociaż ostateczna ewaluacja nie jest zadaniem studentów, lecz osób zarządzających instytucją edukacyjną. Podatność na pewne zniekształcenia nie przekreśla ich wartości (tym bardziej że trudno je zastąpić), lecz wymaga świadomości i sposobów przeciwdziałania, z których najbardziej polecane jest wykorzystanie danych z dodatkowych źródeł. Prosty sposób jest wykorzystanie w ewaluacji nie tylko odpowiedzi na pytania ilościowe (np. mierzone na skali 1-5), ale także na jakościowe pytania otwarte (np. o słabe i mocne strony kursu czy osoby prowadzącej). Naturalnie konieczne jest w tym wypadku uniemożliwienie identyfikacji osoby udzielającej odpowiedzi. Informacje uzyskiwane w ten sposób są znacznie bogatsze i zawierają o wiele więcej konkretnych wskazówek dla prowadzących niż niewiele mówiące wartości liczbowe. Ich wartość formacyjna jest zatem o wiele większa.



Na koniec kilka ogólnych refleksji. Kultury ewaluacyjnej, którą zdefiniować można jako formalne i nieformalne zakorzenienie (również na poziomie zwyczajów, praktyk i sposobów myślenia) zasad wzajemnego i opartego na jasnych kryteriach oceniania oraz niemniej ważnej otwartości na wyniki tej oceny, nie buduje się ani łatwo, ani szybko. Mimo że ocena jest stałym elementem życia uniwersyteckiego, zwykle dokonywana jest ona hierarchicznie (np. ocena studenta przez egzaminatora czy wnioskodawcy grantu przez komisję składającą się z uznanych ekspertów w danej dziedzinie), a jej kryteria nie zawsze są czytelne i intersubiektywne. Braków polskiego szkolnictwa wyższego w tej sferze, o których może najlepiej świadczy niewielka liczba krajowych publikacji i konferencji na ten temat, nie usuną formalne wymagania Państwowej Komisji Akredytacyjnej. Praktycznym czynnikiem wymuszającym rozwój może być natomiast presja demograficzna i rosnąca konkurencja, także na globalnym rynku edukacyjnym. W tej sytuacji wykorzystanie ewaluacji jako narzędzia poprawy jakości kształcenia może być zarówno czynnikiem wprowadzania zmian związanych z przeorientowaniem uczelni w kierunku nastawienia na studenta, jak i budowania przewagi konkurencyjnej nad innymi placówkami.

## Bibliografia

- Ali, D.L., Sell, Y. (1998), *Issues Regarding the Reliability, Validity and Utility of Student Ratings of Instruction: A Survey of Research Findings*, Calgary: University of Calgary APC Implementation Task Force on Student Ratings of Instruction.
- Cashin, W. E. (1995), *Student Ratings of Teaching: The Research Revisited*, „IDEA Paper”, 32, Manhattan, KS: Center for Faculty Evaluation and Development, Division of Continuing Education, Kansas State University.
- Centra, J. A. (2003), *Will Teachers Receive Higher Student Evaluations By Giving Higher Grades and Less Course Work?*, „Research in Higher Education”, 44 (5), s. 495-518.
- Clayson, D. E. (1999), *Students' Evaluation of Teaching Effectiveness: Some Implications of Stability*, „Journal of Marketing Education”, 21 (1), s. 68-75.
- Clayson, D. E., Sheffet, M. J. (2006), *Personality and the Student Evaluation of Teaching*, „Journal of Marketing Education”, 28 (2), s. 149-160.
- Eizsler, C. F. (2002), *College Students' Evaluations of Teaching and Grade Inflation*, „Research in Higher Education”, 43 (4), s. 483-501.
- Felder, R. (1992), *What Do They Know, Anyway?*, „Chemical Engineering Education”, 26 (3), s. 134-135.
- Felder, Richard, Brent, Rebecca (2004), *How to Evaluate Teaching*, „Chemical Engineering Education”, 38 (3), s. 200-202.
- Greenwald, A. G. (1996), *Applying Social Psychology to Reveal a Major (But Correctable) Flaw in Student Evaluations of Teaching*, Papers presented at the Annual Meeting of the American Psychological Association, New York.
- Gump, S. E. (2007), *Student Evaluations of Teaching Effectiveness and the Leniency Hypothesis: A Literature Review*, „Educational Research Quarterly”, 30 (3), s. 56-69.
- Gursoy, D., Umbreit, W. T. (2005), *Exploring Students' Evaluations of Teaching Effectiveness: What Factors Are Important?*, „Journal of Hospitality and Tourism Research”, 29 (1), s. 91-109.
- Hobson, S. M., Talbot, D. M. (2001), *Understanding Student Evaluations: What All Faculty Should Know*, „College Teaching”, 49 (1), s. 26-31.
- Johnson, T. (2002), *Online Students Rating: Will Students Respond?*, Presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Overall, J. U. IV, Marsh, H. W. (1982), *Students' Evaluations of Teaching: An Update*, „AAHE Bulletin”, s. 9-12.
- Paswan, A. K., Young, J. A. (2002), *Student Evaluation of Instructor: A Nomological Investigation Using Structural Equation Modeling*, „Journal of Marketing Education”, 24 (3), s. 193-202.
- Simpson, P. M., Siguaw, J. A. (2000), *Student Evaluations of Teaching: An Exploratory Study of the Faculty Response*, „Journal of Marketing Education”, 22 (3), s. 199-213.
- Stapleton, R. J., Murkison, G. (2001), *Optimizing the Fairness of Student Evaluations: A Study of Correlations Between Instructor Excellence, Study Production, Learning Production, and Expected Grades*, „Journal of Management Education”, 25 (3), s. 269-291.

138

## Strony internetowe uczelni

University of British Columbia, [http://www.vpacademic.ubc.ca/policy\\_on\\_student\\_evaluation\\_of\\_teaching.pdf](http://www.vpacademic.ubc.ca/policy_on_student_evaluation_of_teaching.pdf) [13 lipca 2009].

Ohio State University, [http://oaa.osu.edu/eval\\_teaching/index.html](http://oaa.osu.edu/eval_teaching/index.html), <http://www.ureg.ohio-state.edu/ourweb/scansurvey/index.html> [13 lipca 2009].

Stanford University, [http://registrar.stanford.edu/faculty/course\\_evaluations/index.htm](http://registrar.stanford.edu/faculty/course_evaluations/index.htm) [13 lipca 2009].

University of Auckland, [http://cad.auckland.ac.nz/content/files/apg/evaluation\\_of\\_teaching.pdf](http://cad.auckland.ac.nz/content/files/apg/evaluation_of_teaching.pdf), <http://www.education.auckland.ac.nz/uoa/fms/default/education/about/faculty/committee/docs/Review%20and%20Evaluation%20of%20Teaching%20TLQ%20Policy.pdf> [13 lipca 2009].

Oxford University, <http://ceq.oucs.ox.ac.uk/intro.cfm>, <http://www.learning.ox.ac.uk/>, <http://ceq.oucs.ox.ac.uk/interpretOSCEQ.cfm> [13 lipca 2009].

Washington State University, <http://www.wsu.edu/provost/teaching.htm> [13 lipca 2009].



**Seweryn Rudnicki** is a sociologist and a social psychologist, a teacher at the Tischner European University and a doctoral student at the Jagiellonian University. His academic interests are: sociology of the Internet, sociology of the body and methodology of the social sciences. E-mail: [srudnicki@wse.krakow.pl](mailto:srudnicki@wse.krakow.pl)

**Anna Szwed** is a sociologist, a lecturer at Tischner European University in Kraków, and a PhD candidate at the Jagiellonian University. She is preparing her PhD dissertation on the Roman Catholic Church's teachings on women, and the perception of women among the Roman Catholic priests in Poland. Her scientific interests include: the sociology of culture, sociology of religion, and gender studies. E-mail: [aszwed@wse.krakow.pl](mailto:aszwed@wse.krakow.pl)

## Abstract

### *Evaluating Evaluators. Three Questions about Evaluation*

The paper discusses the issue of evaluating teaching in tertiary education. It consists of three parts, each being an answer to one question: what the evaluation of teaching really is (or should be), how it is conducted at well-known universities throughout the world, and whether students' evaluations of teaching are indeed a valid and important source of data. In the first part, the authors explain the very idea of evaluation, which includes not only raw data, but also interpretation, underlining its formative and summative character. In the second, various "good practices" from Anglo-Saxon universities are presented. Such issues as system transparency, data triangulation, evaluation procedures, student involvement etc. are mentioned. In the third part, the authors focus on the reliability and validity of student ratings, and on factors that may affect this kind of measurement.

## Keywords

Evaluation of teaching, formative and summative evaluation, methodology of evaluation of instruction, student ratings, peer observations, self-evaluation.