

Seweryn Rudnicki

Anna Szwed

Ocenianie oceniających. Trzy pytania o ewaluację zajęć dydaktycznych¹

Artykuł prezentuje problematykę ewaluacji nauczania w szkolnictwie wyższym poprzez 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 ich niezbędną interpretację 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 problemach, takich jak: transparentność systemu, triangulacja zbieranych danych, procedury ewaluacyjne, zaangażowanie studentów itp. Trzecia część dotyczy zagadnień rzetelności i trafności ocen studenckich oraz czynników, które mogą potencjalnie wpływać na wyniki dokonywanego przy ich użyciu pomiaru jakości sposobu prowadzenia zajęć.

¹ 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. Tekst ukazał się po raz pierwszy w czasopiśmie naukowym „Kultura i Polityka” (nr 5/2009), wydawanym przez Wyższą Szkołę Europejską.

Wprowadzenie

Liczba publikacji naukowych poświęconych ewaluacji dydaktyki po koniec ubiegłego wieku sięgała 2 tysięcy pozycji na całym świecie (McKeachie, Kaplan 1996), co zdaje się świadczyć zarówno o stopniu skomplikowania problemu, jak i o jego znaczeniu. O ile przekonywanie o konieczności zapewnienia wysokiej jakości kształcenia w szkolnictwie wyższym jest truizmem, o tyle rzeczowa refleksja nad sposobem prowadzenia ewaluacji dydaktyki wydaje się zasadna. Potoczna obserwacja sposobu funkcjonowania uczelni wyższych w naszym kraju wskazuje bowiem, że mimo wprowadzania ewaluacji studenckich na coraz szerszą skalę, ciągle nie jest powszechne traktowanie studenta jako klienta instytucji edukacyjnej (którego zadowolenie z wysokiej 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 zestaw podstawowych pytań związanych z prowadzeniem ewaluacji zajęć dydaktycznych w oparciu przede wszystkim o literaturę przedmiotu, rozwiązania funkcjonujące w krajach anglosaskich, a także doświadczenie 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 są wciąż tematami stosunkowo nowymi choćby w porównaniu z krajami anglosaskimi, w których należą one do akademickiej rutyny, ewaluację najczęściej utożsamia się z semestralnymi ocenami studencki-

mi 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, takich jak: ankiety studenckie, hospitacje, *peer observations*, autoewaluacja). Interpretacja, a więc nadanie znaczenia (także w oparciu o kontekst oceniania²), jest zatem równie istotnym elementem ewaluacji, co same dane.

Drugie ważne zastrzeżenie terminologiczne związane jest z celem ewaluacji. W literaturze przedmiotu najczęściej stosuje się niezwykle praktyczne rozróżnienie na tzw. ewaluację formatywną i ewaluację sumatywną (Felder, Brent 2004). W przypadku pierwszej z nich nacisk kładziony jest na rozwijanie kompetencji dydaktycznych prowadzącego. 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 raczej koncentrowały się na stwarzających możliwość porównywania danych ilościowych. W przypadku ewaluacji formatywnej szczególnego znaczenia nabierać będą natomiast wszelkie dane o charakterze jakościowym (np. uzyskane za pomocą pytań o charakterze otwartym).

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

Na czym polega system ewaluacji zajęć dydaktycznych?

W odpowiedzi na to pytanie proponujemy przegląd „dobrych praktyk”³ stosowanych w wybranych uniwersytetach z kręgu anglosaskiego⁴. 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.). Świadczyć o tym może 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 dla 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 polepszanie 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 dla celów operacyjnych (ocena pracownika, wynagrodzenie, nagrody itp.). Zarówno w tych, jak i w innych analizowanych przypadkach, zwraca uwagę silny nacisk na podwójną funkcję ewaluacji – formatywną i sumatywną.

3 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 wypracowane dokumenty są wynikiem nabytych przez konkretną uczelnię doświadczeń).

4 Uczelnie, do których rozwiązań odwołujemy się, 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).

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 dokonujące oceny bardziej ogólnej – całości programu studiów, warunków studiowania itp. (np. Univeristy of Auckland – „University Review”, „Graduating Year Review”, „Student Experience and Satisfaction Survey” ; „Oxford University – Student Course Experience Questionnaire”). Ankiety studenckie uznawane są 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 uczeni, wiele placówek wykorzystuje także narzędzia standaryzowane o zweryfikowanej trafności i rzetelności, które dodatkowo dają możliwość porównywania wyników ewaluacji między uczelniami⁵.

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 określana jest wymagana częstotliwość tej formy oceny dla każdego z uniwersyteckich stanowisk, np. w przypadku asystentów – raz w roku). Innym proponowanym narzędziem ewaluacji jest samoocena prowadzącego zajęcia. W wybranych przypadkach stanowić ona może 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: prowadzone kursy, indywidualna praca ze studentem, praca ze specjalnymi grupami studentów, doradztwo, doskonalenie dydaktyczne prowadzącego wprowadzane innowacje; ewaluacje – i m.in. wyniki ocen studenckich, *peer review*, uzyskane nagrody itp.

5 Przykładowo Instructional Development and Effectiveness Assesment (IDEA), Student Instructional Report, Student Evaluation of Educational Quality (SEQ), Instructor and Course Evaluation System, Student Instructional Rating System, Instructional Assessment System (Hobson, Talbot 2001).

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. readers czy wizualne pomoce naukowe). Innym źródłem cennych opinii na temat dydaktyki (choć w mniejszym stopniu dotyczących 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 na jakąś formę decentralizacji. Dokonują jej, delegując opracowanie procedury na poziom wydziału 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 studentów oceniających kurs (np. Stanford University deklaruje, że studenci wypełniają elektroniczne ankiety na zewnętrznym, niezależnym serwerze). W drugim przypadku 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 itd. Niemal wszystkie z analizowanych uczelni organizują oceny studenckie w końcowej części semestru (lub innej jednostki rozliczeniowej), przed egzaminami.

Te z uczelni, które w sposób szczegółowy 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 także uwzględnienie roli studentów w procesie ewaluacji, zawarte w większości analizowanych dokumentów. W najbardziej oczywistym wymiarze wiąże się ono z oparciem ewaluacji dydaktyki na ocenach studentów. Inne wybrane aspekty tego zaangażowania to w zależności od uczelni: rola organizacji studenckich jako podmiotu zainteresowanego wynikami prowadzonej 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, związany jest 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 kadre i studentów pytań dotyczących ewaluacji – FAQ (SU), czy też 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/ą? Czy najlepszych ocen nie dostają zwłaszcza te osoby, u których łatwo dostać 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. powtórne 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 uzy-

skiwane 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 (ze względu na mnogość prac 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 także 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 średnio 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 również wysoka stopa zwrotów, bowiem osoby, które uczestniczyły w zajęciach, a nie oceniły ich, mogą różnić się znacząco 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 bardzo 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 jednak 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 Foxa – 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, s. 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 przeprowadzonymi 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 uczyli 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 narosło także wiele błędnych przekonań dotyczących czynników, od których ich wyniki są uzależnione i które mogą zniekształcać rezultaty oceny, czyniąc ją nietrafną (Simpson, Siguaw 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)⁶:

- 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 obowiązkowe;
- 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 podwyższanie ocen (a właściwie za dawanie do zrozumienia, że oceny będą wysokie).
- 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;

⁶ 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 przypadkach mamy jednak zapewne do czynienia z sytuacjami, w których np. wymagający prowadzący są oceniani nisko albo pleć osoby prowadzącej obniża lub zawyża oceny.

- brak zapewnionej anonimowości oceniających – opinie wyrażane nieanonimowo są zazwyczaj wyższe;
- obecność osoby prowadzącej zajęcia w czasie oceny – ten pozornie nieważny szczegół w istocie ma znaczenie, ponieważ wypełniane w obecności wykładowców ankiety prowadzą 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 entuzjazm, 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 warsztat⁷.

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 poprawić się może wiedza czy warsztat osoby prowadzącej), dorobku naukowego, wieku, płci i osobowości studentów, momentu dokonywania ewaluacji (pod warunkiem, że ewaluacja jest przeprowadzana w 2. połowie semestru), rozmiaru grupy zajęciowej (choć część badań pokazuje, że mniejsze grupy oceniają prowadzących wyżej).

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, jakkolwiek 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 polecaną 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

7 Istnieją również badania pokazujące, że postrzegana przez studentów osobowość prowadzącego jest czynnikiem, którym można wyjaśnić na 73% wariacji ocen zajęć (Clayson, Sheffet 2006). Co więcej, istotny wpływ tego czynnika da się stwierdzić już po 5 minutach od rozpoczęcia pierwszych zajęć.

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 słów refleksji natury ogólnej. 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 nie mniej 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 dziedzinie), a jej kryteria nie zawsze są czytelne i intersubiektywne. Braków w tej sferze polskiego szkolnictwa wyższego, 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 podwyższania jakości kształcenia może być zarówno czynnikiem wprowadzania zmian związanych z przeorientowaniem uczelni w kierunku nastawienia na studenta, jak i budowania przewagi konkurencyjnej względem innych placówek.

Bibliografia:

Ali, Donna L., Sell, Yvonne (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 [15.02.2009].

Cashin, William 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, John 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, Dennis E. (1999), *Students' Evaluation of Teaching Effectiveness: Some Implications of Stability*, „Journal of Marketing Education”, 21(1), s. 68–75.

Clayson, Dennis E., Sheffet, Mary Jane (2006), *Personality and the Student Evaluation of Teaching*, „Journal of Marketing Education”, 28(2), s. 149–160.

Eizsler, Charles F. (2002), *College Students' Evaluations of Teaching and Grade Inflation*, „Research in Higher Education”, 43(4), s. 483–501.

Felder, Richard M. (1992), *What Do They Know, Anyway?*, „Chemical Engineering Education”, 26(3), s. 134–135.

48 Felder, Richard, Brent, Rebecca (2004), *How to evaluate teaching?*, „Chemical Engineering Education”, 38(3), s. 200–202.

Greenwald, Anthony 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, Steven E. (2007), *Student Evaluations of Teaching Effectiveness and the Leniency Hypothesis: A Literature Review*, „Educational Research Quarterly”, 30(3), s. 56–69.

Gursoy, Dogan, Umbreit, Terry W. (2005), *Exploring Students' Evaluations of Teaching Effectiveness: What Factors Are Important?*, „Journal of Hospitality and Tourism Research”, 29(1), s. 91–109.

Hobson, Suzanne M., Talbot, Donna M. (2001), *Understanding Student Evaluations: What All Faculty Should Know*, „College Teaching”, 49(1), s. 26–31.

Johnson, Trav (2002), *Online Students Rating: Will Students Respond?*, Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Overall, Jesse U. IV, Marsh, Herbert W. (1982), *Students' Evaluations of Teaching: An Update*, „AAHE Bulletin” 1982, s. 9–12.

Paswan, Audhesh K., Young, Joyce A. (2002), *Student Evaluation of Instructor: A Nomological Investigation Using Structural Equation Modeling*, „Journal of Marketing Education”, 24(3), s. 193–202.

Simpson, Penny M., Siguaw, Judy A. (2000), *Student Evaluations of Teaching: An Exploratory Study of the Faculty Response*, „Journal of Marketing Education”, 22(3), s. 199–213.

Stapleton, Richard John, Murkison, Gene (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.

Strony internetowe uczelni:

Ohio State University, http://oaa.osu.edu/eval_teaching/index.html, <http://www.ureg.ohio-state.edu/ourweb/scansurvey/index.html> [13.07.2009].

Oxford University, <http://ceq.oucs.ox.ac.uk/intro.cfm>, <http://www.learning.ox.ac.uk/>, <http://ceq.oucs.ox.ac.uk/interpretOSCEQ.cfm> [13.07.2009].

Stanford University, http://registrar.stanford.edu/faculty/course_evaluations/index.htm [13.07.2009].

University of Auckland, 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> [13.07.2009].

Washington State University, <http://www.wsu.edu/provost/teaching.htm> [13.07.2009].

University of British Columbia, http://www.vpacademic.ubc.ca/policy_on_student_evaluation_of_teaching.pdf [13.07.2009].