

Aleksandra Bujacz

Jak działa Assessment Center? Zagadka trafności ośrodków oceny

W niniejszym artykule rozważana jest kwestia paradoksu trafności *Assessment Center* związanej z niską trafnością teoretyczną tej metody przy jej wysokich możliwościach predykcyjnych. Przedstawione wyniki badań potwierdzające pogląd, iż AC bada bardziej poziom wykonania zadań przez kandydata niż jego kompetencje, rozumiane jako względnie trwałe cechy, oraz ich implikacje dla praktyki selekcji.

Wprowadzenie

Przyjęte standardy traktują ośrodki oceny (*Assessment Center* – AC) jako technikę oceniania zachowania przy pomocy wielu metod [*International Task Force on Assessment Center Guidelines*, 2000]. Program oceny, aby zostać nazwanym AC, musi spełniać określone wymagania. Powinien być poprzedzony analizą pracy i opierać się na klasyfikacji behawioralnej. AC tworzą techniki różne pod względem formy (testy, wywiady, symulacje itp.), ale mierzące określone w analizie pracy kompetencje. Każde AC musi zawierać co najmniej jedną symulację dającą możliwość obserwacji zachowań kandydata, które są traktowane jako wskaźniki ocenianych kompetencji. W tradycyjnym ujęciu AC opiera się więc na mierzeniu konstruktywów psychologicznych, nazywanych kompetencjami, przy pomocy przypisanych im behawioralnych wskaźników.

Assessment Center jest metodą selekcji, której wartość predykcyjną udowodniono wielokrotnie w ciągu ponad trzech dekad jej stosowania [Gaugler i in., 1987]. Badania nad trafnością predykcyjną opierały się na powiązaniu wyników AC ze wskaźnikami kryterium powodzenia w pracy, np. wysokością zarobków. Wyniki badań podłużnych pozwalają stwierdzić, że na podstawie AC można trafnie wnioskować o przyszłym powodzeniu bądź niepowodzeniu w pracy ocenianego kandydata [Arthur i in., 2003, s. 125–154; Jansen i Vinkenburg, 2006, s. 253–266]. Badania nad trafnością AC wykazały jed-

nak, że metodę tę charakteryzuje niska trafność teoretyczna (*construct validity*) [Sackett i Dreher, 1982, s. 401-410]. Ten rodzaj trafności określa, na ile dana metoda odzwierciedla cechę psychologiczną, która ma być przedmiotem pomiaru [Hornowska, 2001, s. 94]. Fakt, że AC charakteryzuje się trafnością predykcyjną, a jednocześnie brakuje mu trafności teoretycznej, nazywany jest paradoksem lub zagadką tej metody [Klimowski i Bickner, 1987, s. 243-260]. Trafność teoretyczna określana jest bowiem jako nadrzędna i podstawowa wobec innych rodzajów trafności [Anastasi i Urbina, 1999 za: Hornowska, 2001, s. 94]. Wiadomo więc, jak stwierdzają Klimowski i Brickner [1987], że AC działa, nadal jednak nie ma pewności, dlaczego działa i co tak naprawdę mierzy.

Wyjaśnianie paradoksu trafności AC

Trafność AC badano metodą macierzy wielu cech wielu metod – MTMM [Cambell i Fiske, 1959, s. 81-105; Brzeziński, 2004, s. 525]. Traktowano więc poszczególne ćwiczenia jako kolejne metody, zaś mierzone w nich kompetencje jako cechy. Dane uzyskane dzięki macierzy MTMM pozwalają sprawdzić, jaka jest trafność zbieżna konstruktów (*convergent validity*), czyli jak silne są korelacje między tymi samymi cechami mierzonymi różnymi metodami. Silne korelacje świadczą o wysokiej trafności zbieżnej, dzięki której wiadomo, że mierzona cecha jest stabilna niezależnie od metody. Macierz MTMM informuje również, jakie są korelacje między różnymi cechami mierzonymi w ramach jednej metody. W tym przypadku oczekiwane są jak najniższe korelacje, które świadczą o wysokiej trafności różnicowej (*discriminant validity*). Ten rodzaj trafności wskazuje na względną niezależność cechy i jej wyodrębnienie spośród innych konstruktów. W przypadku AC podstawową zmienną, jaka podlega badaniu, jest ocena kandydata w ramach danej kompetencji w danym ćwiczeniu (PEDR – *postexercise dimension rating*).

Wspomniane braki w trafności teoretycznej ujawniają się poprzez wysokie korelacje PEDR w ramach ćwiczenia, a niskie w ramach kompetencji. Zjawisko to nazywane jest efektem ćwiczenia [Schneider i Schmitt, 1992, s. 32-41; Lance i in., 2004, s. 377-385]. Efekt ćwiczenia znalazł potwierdzenie nie tylko w badaniach z wykorzystaniem metodologii macierzy MTMM. Metoda analizy czynnikowej, zarówno w wersji eksploracyjnej, jak i confirmacyjnej, konsekwentnie wykazywała, że na proces AC składa się tyle czynników, ile zadań zawiera proces oceny [Bycio i in., 1987, s. 463-474; Sackett i Harris, 1988, s. 214-229; Lievens i Van Keer, 2001, s. 373-378]. Metaanalizy badań również potwierdziły istnienie tego efektu [Lievens i Conway, 2001, s. 1202-1222], co sugeruje, że jest to zjawisko stałe i względnie niezależne od rodzaju przeprowadzanego AC.

Podstawowym wyjaśnieniem efektu ćwiczenia od momentu jego odkrycia było stwierdzenie, że wynika on z niedostatków metody (*metod bias*) [Sackett i Dreher, 1982]. Uważano go za zagrożenie dla trafności procesu AC i szukano możliwości „na-

prawienia” AC. Efekt ćwiczenia tłumaczono ograniczeniami możliwości poznawczych asesorów, ale by efekt ten wyeliminować, zmniejszano ilość kompetencji podlegających ocenie w danym ćwiczeniu oraz minimalizowano liczbę kandydatów ocenianych przez jednego asesora [Gaugler i Thornton, 1989, s. 611–618; Bycio i in., 1987]. Posługiwano się także specjalnym systemem trenowania asesorów (FOR – *frame-of-reference*) w celu zwiększenia ich umiejętności różnicowania pomiędzy poszczególnymi wymiarami [Woehr i Huffcutt, 1994, s. 189-205]. Jako podstawę do formułowania ocen stosowano listy zachowań – asesor miał za zadanie odznaczać na nich zachowania, które kandydat ujawnił podczas AC, co przyczyniło się do wzrostu zarówno trafności różnicowej, jak i zbieżnej AC [Reilly i in., 1990, s. 71-84.]. Udowodniono, że trafność AC wzrasta także, gdy asesorzy są psychologami [Gaugler i in. 1987]. Manipulacji poddano proces oceny kandydatów, stwierdzając, że kiedy asesorzy podejmują decyzje dopiero po wykonaniu przez ocenianego wszystkich zadań, przyznają kandydatowi ocenę z danej kompetencji. Biorąc pod uwagę wszystkie zachowania, jakie ujawnił w czasie AC (*within-dimension rating*), trafność teoretyczna wzrasta [Robie i in., 2000; Kolk i in., 2002]. Niedawne metaanalizy trafności AC wykazały, że efekt ćwiczenia jest mniej wyraźny w przypadku niektórych kompetencji, np. planowania i organizacji czy umiejętności rozwiązywania problemów [Bowler i Woehr, 2006]. Generalnie jednak próby poprawy trafności, jakkolwiek przyczyniły się do zwiększenia obiektywizmu i standaryzacji AC, nie doprowadziły w żadnym ze znanych autorce przypadków do zlikwidowania efektu ćwiczenia bądź znacznego jego zmniejszenia. Część z nich wiązała się także ze zdecydowanym wzrostem kosztów przeprowadzenia AC, poprzez wymaganie większej liczby asesorów, stając się niemożliwymi do zastosowania w praktyce selekcji [Kolk i in., 2002].

Równoległe do prób wyjaśniania efektu ćwiczenia błędami w metodologii AC rozwijał się drugi nurt tłumaczenia tego zjawiska. Wychodząc z założenia, że efekt ćwiczenia nie jest błędem, a właśnie tym tajemniczym czynnikiem, który sprawia, że AC jest trafnym i skutecznym narzędziem, poszukiwano wyjaśnień dla jego znaczenia. Jednym z nich jest twierdzenie, że efekt ten jest problematyczny jedynie, gdy przyjmujemy tradycyjne rozumienie AC jako procesu oceny względnie stałych charakterystyk zachowań [Lance i in., 2000]. Odrzucenie tego założenia i dopuszczenie możliwości, że to sam kandydat i zmienność jego zachowań w zależności od sytuacji prowadzą do powstania efektu ćwiczenia, rzuca zupełnie nowe światło na rzekome braki w trafności AC. Sugeruje się więc, że efekt ćwiczenia powodowany naturalną zmiennością zachowań ludzi w zależności od sytuacji, w której się znajdują, nie stanowi problemu czy zagrożenia dla trafności AC [Neidig i Neidig, 1984, s. 182–186]. Jednak czy akceptując (i ignorując) międzysytuacyjną zmienność zachowań kandydata, nie pomija się właśnie tego, co jest w AC najcenniejsze?

Hipotezy

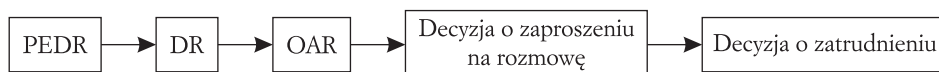
Omawiane badanie było próbą odpowiedzi na to pytanie poprzez sprawdzenie, na ile od sytuacji (ćwiczenia, zadania) zależą wyniki uzyskiwane przez kandydata w AC. Postawiono hipotezę o zależności otrzymywanych ocen od typu zadania, jak i od kompetencji, jaka jest w tym zadaniu oceniana. Potwierdzenie hipotezy mówiącej o łącznym wpływie tych dwóch zmiennych na wynik otrzymywany przez kandydata świadczyłoby o sytuacyjnej zależności oceny danej kompetencji przez asesora. Do analizy wybrano dwie metody, test i ćwiczenie grupowe, z których tylko jedna zakładała wpływ asesora. Wykorzystano także dwie powszechnie stosowane w AC kompetencje – pracę w grupie oraz rozwiązywanie problemów. Wymiary te reprezentują dwie podstawowe grupy kompetencji, które są oceniane w AC [Shore i in., 1990]. Jednocześnie postanowiono sprawdzić, czy oceny z poszczególnych ćwiczeń i kompetencji, a także ocena z całego procesu (OAR – *overall assessment rating*) mają wpływ na decyzję o zatrudnieniu kandydata. Decyzja ta, podejmowana przez kierownika działu odpowiedniego dla stanowiska, o które ubiega się kandydat, traktowana była jako kryterium powodzenia kandydata w pracy.

Badanie

Badanie prowadzone było od czerwca do grudnia 2008 roku przez firmę konsultingową zajmującą się doradztwem w zakresie rekrutacji pracowników. W badaniu uczestniczyło 440 osób narodowości polskiej, ocenionych w ramach programu rekrutacyjnego na stanowiska produkcyjne w nowoczesnej fabryce należącej do międzynarodowego koncernu. Ośrodek oceny składał się z testu sytuacyjnego badającego kwalifikacje na stanowiska produkcyjne, ćwiczenia grupowego (grupy od 3 do 6 osób) oraz wywiadu behawioralnego. W sumie proces ten mierzył 9 kompetencji, z czego 5 podlegało badaniu dwukrotnie różnymi metodami. Asesorzy byli psychologami. Dokonywali oceny kompetencji w danym ćwiczeniu (PEDR) niezależnie od siebie, zazwyczaj zaraz po ćwiczeniu. Test oceniany był według klucza odpowiedzi. Każdy kandydat był oceniany przez jednego asesora, w każdym zadaniu innego. Współczynnik liczby ocenianych kandydatów na jednego asesora wahał się od 1:1 (rozmowa) do 1:3 (ćwiczenie grupowe). Dla kompetencji ocenianych dwukrotnie dane były integrowane, tworząc jedną ocenę z danej kompetencji (DR – *dimension rating*). Następnie zgodnie z ustalonym standardem (kluczem) kandydaci otrzymywali ocenę końcową (OAR – *overall assessment rating*). Znając OAR oraz DR kandydata, kierownicy fabryki podejmowali decyzję o zaproszeniu kandydata na rozmowę z przyszłym przełożonym, na podstawie której tenże przełożony decydował o ewentualnym zatrudnieniu kandydata.

Poszczególne kompetencje (PEDR) oceniano według skali 5-stopniowej, zaś OAR przypisywano w skali 4-stopniowej. Wobec obu tych zmiennych zaistniał problem określenia, czy powinny być traktowane jako porządkowe czy ilościowe. Poziom pomiaru w skalach szacunkowych budzi bowiem kontrowersje – popularne podręczniki dopuszczają uznawanie takiej skali za ilościową [Bedyńska i Brzezińska, 2007], zaś opracowania metodologów zdecydowanie nakazują traktowanie jej jak skalę porządkową [Brzezińska i Brzeziński, 2004]. W tym badaniu oceny PEDR z testu sytuacyjnego przypisywane były ze względu na liczbę poprawnie wskazanych odpowiedzi. W ćwiczeniu grupowym zaś zastosowano listy behawioralne, na których asesory zaznaczali, ile danego rodzaju zachowań przejawia kandydat. Przypisywanie ocen od 1 do 5 zależało od sztywnego kryterium liczby poprawnych zachowań, odpowiadającej każdemu z tych poziomów. Taka metoda oceniania kandydatów daje podstawy do uznania ocen z poszczególnych kompetencji za wskaźniki natężenia cechy, co jest kryterium uznania danych za interwałowe [Brzeziński, 2004]. Analizując wyniki badania, uznano więc, że oceny PEDR mogą zostać potraktowane jako dane przedziałowe. Całościowa ocena z AC (OAR) powstawała poprzez „przyłożenie” ocen z poszczególnych kompetencji do stworzonego na potrzeby danego stanowiska klucza. Rolą tego klucza było określenie, na ile kandydat o pewnym profilu kompetencyjnym sprawdzi się w przyszłej pracy. Z tego względu oceny OAR zostały potraktowane jako dane jakościowe.

Rysunek 1. Przebieg badania



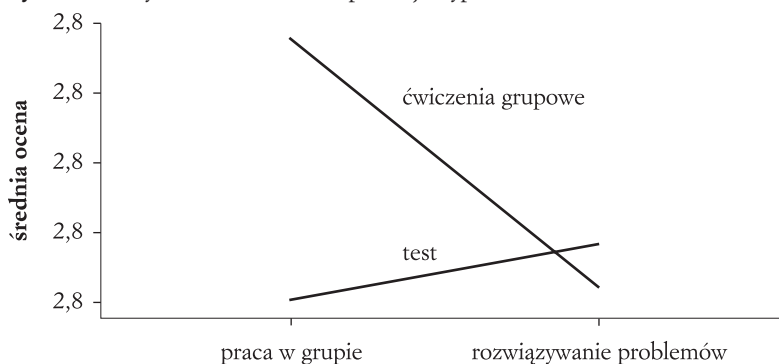
Źródło: opracowanie własne

Wyniki

Przede wszystkim postanowiono sprawdzić, czy w badaniu również ujawni się efekt ćwiczenia. W tym celu policzono korelacje między ocenami kompetencji z ćwiczenia (PEDR) w ramach danej kompetencji (*within-dimension*) oraz w ramach poszczególnych ćwiczeń (*within-exercise*). Korelacje w ramach poszczególnych ćwiczeń okazały się wyższe ($M r = 0,4$) niż w ramach poszczególnych kompetencji ($M r = 0,2$). Wszystkie korelacje były istotne na poziomie $p < 0,05$, co wskazuje na ujawnienie się efektu ćwiczenia. Jednocześnie wykonano eksploracyjną analizę czynnikową struktury ocen uzyskanych po AC, poprzez którą uzyskano trzy wyraźne czynniki związane z trzema zadaniami składającymi się na proces oceny kandydata. Czynniki te wyjaśniały łącznie 54% wariacji wyników.

W celu zbadania, czy rodzaj wykonywanego ćwiczenia oraz oceniana kompetencja wpływają na wyniki w ćwiczeniu (PEDR), została przeprowadzona dwuczynnikowa analiza wariancji z powtarzanym pomiarem w następującym schemacie: (2) rodzaj zadania (ćwiczenie grupowe, test) \times (2) kompetencja („praca w grupie”, „rozwiązywanie problemów”). Zmienną zależną była ocena kompetencji w ćwiczeniu (PEDR) mierzona na skali od 1 (niewielka liczba zachowań świadczących o danej kompetencji) do 5 (bardzo duża liczba zachowań świadczących o danej kompetencji). Analiza wariancji została przeprowadzona w modelu jednozmiennowym (*univariate repeated measures* – ANOVA). Analiza wykazała istotny efekt interakcji rodzaju ćwiczenia i typu kompetencji $F(1,440) = 26$; $p < 0,001$; $\eta^2 = 0,056$. Oznacza to, że różne kompetencje były różnie oceniane w zależności od typu zadania. Aby zrozumieć, na czym polegał ten efekt, przeprowadzono dwie jednoczynnikowe analizy wariancji z powtarzanym pomiarem oddzielnie dla każdego efektu głównego w modelu jednozmiennowym. Analiza dla kompetencji „rozwiązywanie problemów” wykazała nieistotny wpływ rodzaju zadania na oceny w tej kompetencji $F(1,440) = 0,98$; ni. Z kolei efekt dla kompetencji „praca w grupie” okazał się istotny $F(1,440) = 25,8$; $p < 0,001$; $\eta^2 = 0,055$. Można więc stwierdzić, że w ćwiczeniu grupowym kompetencja „praca w grupie” oceniana była istotnie statystycznie wyżej niż w teście.

Rysunek 2. Wykres zależności kompetencji i typu zadania



Źródło: opracowanie własne.

W celu sprawdzenia, czy decyzja o zatrudnieniu jest zależna od otrzymanych przez kandydata wyników, przeprowadzono test niezależności dwóch zmiennych *chi*-kwadrat. Analizowano zmienne odpowiadające średniemu wynikowi uzyskanemu w poszczególnych zdaniach, które zostały dychotomizowane, oraz OAR i zmienną kryterialną przyjmującą dwie wartości (kandydat został zatrudniony; kandydat nie został zatrudniony). Wzięto pod uwagę jedynie dane dotyczące kandydatów, którzy zostali zaproszeni na rozmowę

z kierownikiem. Porównania wyników testu, rozmowy oraz OAR nie wykazały zależności tych zmiennych z oceną kierowników. Jedynie średni wynik z ćwiczenia grupowego wykazał związek z kryterium $\chi^2(1, N = 159) = 2,2; p < 0,09$. Związek ten przedstawiał się podobnie dla kompetencji „pracy w grupie” $\chi^2(3, N = 159) = 6,9; p < 0,075$, jak i dla kompetencji „rozwiązywania problemów” $\chi^2(3, N = 159) = 6,8; p < 0,077$. Należy pamiętać, że przy obliczaniu przedstawionych tu wyników nie wzięta została pod uwagę grupa kandydatów, którzy otrzymali wysokie wyniki w ćwiczeniu grupowym, a jednak nie zostali zaproszeni na rozmowę z kierownikiem. Średnia ocena z ćwiczenia grupowego, ze względu na użycie klucza OAR, determinowała bowiem jedynie 14% wyniku OAR, zaś aż 50% OAR określał test. Te dane pozwalają przypuszczać, że otrzymana zależność między średnim wynikiem ćwiczenia grupowego a zatrudnieniem kandydata jest godna uwagi, choć poziom istotności tego związku wymaga dodatkowych potwierdzeń.

Dyskusja

Wykazano, że efekt ćwiczenia ujawnił się także w przypadku testu sytuacyjnego, gdzie nie występował ewentualny błąd związany z nierzetelną oceną asesora. Ta informacja pozwala na stwierdzenie, że efekt ćwiczenia nie może być przypisany jedynie występowaniu efektu halo. Błąd w procesie oceniania nazwany efektem halo, definiowany jako wpływ globalnej oceny kandydata na ocenę w ramach poszczególnych wymiarów, związany jest bowiem z osobą asesora [Solomonson i Lance, 1997]. Można przypuszczać więc, że spójność ocen w ramach poszczególnych ćwiczeń nie wynika z błędów w procesie oceniania, a raczej odzwierciedla pewną konsekwentnie ujawniającą się w AC charakterystykę psychologiczną. Rzeczywiście seria badań mających na celu sprawdzenie, jaka zmienna kryje się za wynikającymi z efektu ćwiczenia czynnikami, potwierdza tę hipotezę [Lance i in., 2000]. Czynniki te okazały się istotnie korelować ze wskaźnikami zdolności poznawczych, jak również z poziomem wiedzy profesjonalnej i rezultatami osiąganymi w pracy (*job performance*). Przyjmując założenie, że błąd nie wykazuje tak systematycznych korelacji, można uznać, że efekt ćwiczenia odzwierciedla raczej poziom wykonania czy radzenia sobie kandydata z każdym kolejnym ćwiczeniem, niż powodowany metodą błąd. Stwierdzając, że efekt ćwiczenia nie jest błędem, ale istotną zmienną odzwierciedlającą międzysytuacyjną zmienność zachowań kandydata, natrafiamy na kolejną zagadkę AC. W jaki sposób oceny asesorów, trenowanych w ocenianiu kompetencji i przekonanych, że te właśnie wymiary oceniają, okazują się odzwierciedlać inną zmienną? Odpowiedzi na to pytanie szukano, analizując proces decyzyjny asesorów [Lance, 2000; Lance i in., 1994]. Postawiano hipotezę o występowaniu tzw. dominującej kompetencji (*SD – salient dimension*). Według tego schematu każde ćwiczenie powoduje bardziej wyraźne ujawnianie się jednej kompetencji, w którą są niejako wpisywane umiejętności wykonania

zadania przez danego kandydata. Otrzymane wyniki zdają się potwierdzać tę hipotezę, ujawniając widoczny efekt interakcji kompetencji i ćwiczenia grupowego właśnie dla wymiaru pracy w grupie. Najciekawszy z punktu widzenia praktycznego wykorzystania AC wydaje się być fakt, że to właśnie ćwiczenie grupowe, w którym stwierdzono występowanie wariacji wynikającej z interakcji kompetencji i metody, okazało się najlepszym predyktorem powodzenia kandydata w pracy. Wyniki te zdają się potwierdzać stwierdzenie sformułowane przez Russella i Domma [1995], że asesory tak naprawdę nie wystawiają ocen w zakresie kompetencji, ale raczej wydają trafne predykcje oceny umiejętności radzenia sobie kandydata z zadaniami w trakcie AC.

Konkluzja

Przeszło trzydziestoletnia historia badań nad trafnością AC, obfitując w zagadki, prowadzi do przewrotnych wniosków. Okazuje się bowiem, że zamiast mierzyć stałe charakterystyki zachowań kandydata, AC mierzy raczej zależne od sytuacji wskaźniki wykonania czy radzenia sobie z zadaniami [Sackett i Harris, 1988]. Wyzwania, przed którymi staje kandydat w trakcie procesu AC, stają się dla niego bardziej próbkami pracy (*work samples*) niż bodźcami powodującymi ujawnienie się stałych cech zachowań [Lance i in., 2000]. W tym kontekście dobór ćwiczeń do AC nabiera nowego znaczenia, wyprzedzając ważnością dobór ocenianych kompetencji. Jednocześnie oceny asesorów stają się odzwierciedleniem bardziej poziomu wykonania zadania przez kandydata niż jego kompetencji [Russell i Domm, 1995, s. 25-47].

Asesory przyznają więc oceny w ramach poszczególnych kompetencji przede wszystkim z powodu wymagań narzuconej im procedury, zaś wynik ogólny AC (OAR) opierają raczej na zachowaniach kandydata niż, jak się zakłada, na kompetencjach [Jackson i in., 2007]. Pojawia się pytanie, czy wyniki te detronizują powszechny kompetencyjny model AC? Podążając za argumentacją Lance'a, można stwierdzić, że zależy od tego, jak mocno jesteśmy przywiązani do tradycyjnego rozumienia kompetencji [Lance i in., 2000]. Jeżeli bowiem będziemy twierdzić, że kompetencje odzwierciedlają stałe cechy i umiejętności kandydata, to należałoby w świetle przedstawionych wyników badań uznać, że wyniki AC nie odzwierciedlają trafnie założonego konstruktów teoretycznego. Jeśli zaś uznamy, że umiejętność radzenia sobie w poszczególnych, ważnych z punktu widzenia roli zawodowej zadaniach jest podstawą predykcijnej trafności AC, a kompetencje potraktujemy jako konstrukty określające poszczególne fragmenty tego wykonania, uprawnieni będziemy do uznania AC za trafną metodę przewidywania sukcesu zawodowego. Pojęcie kompetencji stanie się więc w nowym rozumieniu swego rodzaju szkłem powiększającym, które będzie miało za zadanie uwypuklić poszczególne aspekty wykonania zadania przez kandydata.

Literatura

- Arthur W. Jr., Day E.A., McNelly T.L., Edens P.S. (2003), *A meta-analysis of the criterion-related validity of assessment-center dimensions*, „Personnel Psychology”, No. 56.
- Bowler M.C., Woehr D.J. (2006), *A Meta-Analytic Evaluation of the Impact of Dimension and Exercise Factors on Assessment Center Ratings*, „Journal of Applied Psychology”, No. 91.
- Bedyńska S., Brzezicka A., red. (2007), *Statystyczny drogowskaz*, Wydawnictwo SWPS Akademicka, Warszawa.
- Brzezińska A., Brzeziński J. (2004), *Skale szacunkowe w badaniach diagnostycznych [w:] Metodologia badań psychologicznych. Wybór tekstów*, Wydawnictwo Naukowe PWN, Warszawa.
- Brzeziński J. (2004), *Metodologia badań psychologicznych*, PWN, Warszawa.
- Bycio P., Alvares K.M., Hahn, J. (1987), *Situational specificity in assessment center ratings: A confirmatory factor analysis*, „Journal of Applied Psychology”, No. 72.
- Campbell D.T., Fiske D.W. (1959), *Convergent and discriminant validation by multitrait-multimethod matrix*, „Psychological Bulletin”, No. 56.
- Gaugler B.B., Rosenthal D.B., Thornton G.C.III, Bentson C. (1987), *Meta-analysis of assessment center validity*, „Journal of Applied Psychology”, No. 72.
- Gaugler B.B., Thornton G.C.III (1989), *Number of assessment center dimensions as a determinant of assessor generalizability of the assessment center ratings*, „Journal of Applied Psychology”, No. 74.
- Hornowska E. (2001), *Testy psychologiczne. Teoria i praktyka*, Scholar, Warszawa.
- International Task Force on Assessment Center Guidelines (2000), *Guidelines and Ethical Considerations for Assessment Center Operations*, „Public Personnel Management”, No. 29.
- Jackson D.J.R., Barney A.R., Stillman J.A., Kirkley W. (2007), *When Traits Are Behaviors: The Relationship Between Behavioral Responses and Trait-Based Overall Assessment Center Ratings*, „Human Performance”, No. 20.
- Jansen P.G.W., Vinkenburg C.J. (2006), *Predicting management career success from assessment center data: A longitudinal study*, „Journal of Vocational Behavior”, No. 68.
- Klimowski R., Brickner M. (1987), *Why do assessment centers work? The puzzle of assessment center validity*, „Personnel Psychology”, No. 40.
- Kolk N.J., Born M.P., Van der Flier H. (2002), *Impact of Common Rater Variance on Construct Validity of Assessment Center Dimension Judgments*, „Human Performance”, No. 15.
- Lance C.E., Foster M.R., Nemeth Y.M., Gentry W.A., Drollinger S. (2007), *Extending the Nomological Network of Assessment Center Construct Validity: Prediction of Cross-Situationally Consistent and Specific Aspects of Assessment Center Performance*, „Human Performance”, No. 20.
- Lance C.E., Gatewood R.D., Newbolt W.H., Foster M.S., French N.R., Smith D.E. (2000), *Assessment center exercise factors represent cross-situational specificity, not method bias*, „Human Performance”, No. 13.
- Lance C.E., Lambert T.A., Gewin A.G., Lievens F., Conway J.M. (2004), *Revised Estimates of Dimension and Exercise Variance Components in Assessment Center Postexercise Dimension Ratings*, „Journal of Applied Psychology”, No. 89.
- Lance C.E., LaPointe J.A., Stewart A.M. (1994), *A test of the context dependency of three causal models of halo rater error*, „Journal of Applied Psychology”, No. 79.
- Lievens F., Conway J.M. (2001), *Dimension and Exercise Variance in Assessment Center Scores: A Large-Scale Evaluation of Multitrait-Multimethod Studies*, „Journal of Applied Psychology”, No. 86.
- Lievens F., Van Keer E. (2001), *The construct validity of a Belgian assessment centre: A comparison of different models*, „Journal of Occupational and Organizational Psychology”, No. 74.

- Neidig R.D., Neidig P.J. (1984), *Multiple assessment center exercises and job relatedness*, „Journal of Applied Psychology”, No. 69.
- Reilly R.R., Henry S., Smither J.W. (1990), *An examination of the effects of using behavior checklists on the construct validity of Assessment Center dimensions*, „Personnel Psychology”, No. 43.
- Robie C., Osburn H.G., Morris M.A., Etchegaray J.M., Adams K.A. (2000), *Effects of the Rating Process on the Construct Validity of Assessment Center Dimension Evaluations*, „Human Performance”, No. 13.
- Russell C.J., Domm D.R. (1995), *Two field tests of an explanation of assessment centre validity*, „Journal of Occupational and Organizational Psychology”, No. 68.
- Sackett P.R., Dreher G.F. (1982), *Constructs and assessment center dimensions: Some troubling empirical findings*, „Journal of Applied Psychology”, No. 67.
- Sackett P.R., Harris M. (1988), *A further examination of the constructs underlying assessment center ratings*, „Journal of Business and Psychology”, No. 3.
- Schneider J.R., Schmitt N. (1992), *An exercise design approach to understanding assessment center dimension and exercise constructs*, „Journal of Applied Psychology”, No. 77.
- Shore T.H., Thornton G.C.III, Shore L.M. (1990), *Construct validity of two categories of assessment center dimension ratings*, „Personnel Psychology”, No. 43.
- Solomonson A.L., Lance C.E. (1997), *Examination of the Relationship Between True Halo and Halo Error in Performance Ratings*, „Journal of Applied Psychology”, No. 82.
- Woehr D.J., Huffcutt A.I. (1994), *Rater training for performance appraisal: A meta-analytic review*, „Journal of Occupational and Organizational Psychology”, No. 67.

Aleksandra Bujacz – studentka V roku psychologii w Instytucie Psychologii Uniwersytetu im. Adama Mickiewicza w Poznaniu, specjalność: psychologia pracy i organizacji. Zainteresowania naukowe oscylują w zakresie badania potencjału pracowników, zwłaszcza w aspekcie badania twórczego rozwiązywania problemów.