

Piotr ZIELIŃSKI, Marcin BIERNACKI

**OCENA PRZYDATNOŚCI WSKAŹNIKA
EFEKTYWNOŚCI POZNAWCZEJ W BADANIACH
SELEKCYJNYCH PILOTÓW***

**ASSESSMENT OF COGNITIVE EFFICIENCY INDEX IN
SELECTION OF PILOTS***

Wojskowy Instytut Medycyny Lotniczej, Warszawa
Zakład Psychologii Lotniczej

Military Institute of Aviation Medicine, Warsaw
Department of Aviation Psychology

STRESZCZENIE: Wstęp. W pracy pilota niezbędna jest wysoka efektywność działania, rozumiana jako wykonywanie reakcji szybkich i adekwatnych do sytuacji. Ponieważ testy psychologiczne stosowane w badaniach pilotów nie kontrolują aspektu poprawności i szybkości działania jednocześnie, stworzenie uniwersalnego wskaźnika efektywności może przyczynić się do ułatwienia podejmowania wniosków diagnostycznych. **Cel pracy.** Zadaptowanie współczynnika efektywności poznawczej, znanego z badań nad stylem poznawczym impulsywność-refleksyjność, na potrzeby testów stosowanych w diagnostyce pilotów. **Badani i metoda.** Analizie poddano wyniki 375 kandydatów na pilotów. W badaniu zastosowano test orientacji przestrzennej pilotów (PST), test badający uwagę i koncentrację (COG) oraz, w celu uzyskania wyników kontrolnych, Test Matrycy Ravena w wersji dla Zaawansowanych (TMZ) oraz test rozpiętości pamięci świeżej (CORSI). **Wyniki.** Wyniki analiz wskazują na istotny związek wskaźnika efektywności testu COG z wynikami testów TMZ i CORSI (odpowiednio $r = 0,44$ i $r = 0,36$, $p < 0,05$), a także z przyznaną w wyniku badań selekcyjnych kategorią zdrowia (R^2 Coxa i Snella = 0,19). Podobnie, wskaźnik efektywności testu PST koreluje istotnie z wynikami testów TMZ i CORSI (odpowiednio $r = 0,43$ i $r = 0,33$, $p < 0,05$) oraz z przyznaną w wyniku badań selekcyjnych kategorią zdrowia (R^2 Coxa i Snella = 0,21). **Wnioski.** Uzyskane wyniki wspierają tezę, że pojedynczy, uniwersalny wskaźnik efektywności prowadzi do równie trafnych wniosków diagnostycznych co łączne rozpa-

* Pracę wykonano w ramach realizacji projektu BST 107/2008

trywanie zróżnicowanych wskaźników testowych, będąc równocześnie miarą bardziej jednoznaczną i łatwiejszą w interpretacji

SŁOWA KLUCZOWE: efektywność poznawcza, badania selekcyjne pilotów

SUMMARY: Background. Pilot profession requires high efficiency of actions, i.e. reactions both rapid and adequate in the specific situation. As the psychological tests used in studies of pilots do not control reaction correctness and speed in a uniform way, development of the universal efficacy index may facilitate diagnostic conclusions. **Objectives.** This study aimed at adopting cognitive efficiency index, known from the studies on cognitive reflection-impulsivity studies, to the needs of tests used in pilots' diagnostics. **Participants and methods.** Three hundred seventy five candidates for military pilots were tested with Pilot's Spatial Test (PST) and attention and concentration test (COG). Control results were obtained with Raven's Advanced Progressive Matrices (APM) and Corsi Block-Tapping Test (CORSI). **Results.** The obtained results show significant relationship of COG test efficiency index and the results of RMPT and CORSI ($r = 0,44$ and $r = 0,36$, $p < 0,05$, respectively) and with health category resulting from selection (Cox and Snell R square = 0,19). Similarly, PST efficiency index significantly correlates with the results of APM and CORSI ($r = 0,48$ and $r = 0,33$, $p < 0,05$, respectively), and health category resulting from selection (Cox and Snell R square = 0,21). **Conclusions.** The obtained results confirm the thesis that single, universal index of efficiency lead to equally correct diagnostic conclusions as total analysis of diversified indices of efficiency. At the same time it is more unambiguous and easier to interpret

KEY WORDS: cognitive efficiency, selective examinations of pilots

Wstęp

Praca pilota wymaga ciągłego angażowania złożonych procesów poznawczych w celu monitorowania i oceny aktualnej sytuacji oraz podejmowania decyzji, często w dużym deficycie czasu. Wymaga też szybkiego i precyzyjnego podejmowania działań opartych na tych decyzjach. Z tego względu istotne jest, aby uprawnienia pilota uzyskiwały takie osoby, których sprawność poznawcza i psychomotoryczna jest na tyle rozwinięta, że jej poziom nie stanowi dodatkowego czynnika ryzyka i pozwalał na skuteczne funkcjonowanie w powietrzu. W tym celu kandydaci na pilotów powinni być poddawani szczegółowym badaniom diagnostycznym, podczas których wykwalifikowany psycholog podda ocenie poziom ich funkcjonowania. Przykładem realizacji tego postulatu mogą być np. międzynarodowe przepisy obowiązujące w lotnictwie cywilnym [1] czy badania kwalifikacyjne kandydatów do szkolenia lotniczego w Wyższej Szkole Oficerskiej Sił Powietrznych w Dęblinie, prowadzone przez Wojskowy Instytut Medycyny Lotniczej.

Taką sprawność działania – której wysoki poziom jest pożądanym u pilotów z racji specyfiki ich zawodu – na potrzeby niniejszego tekstu można zdefiniować jako zdolność do wykonywania reakcji szybkich i adekwatnych do sytuacji. Poziom tak rozumianej sprawności w zadaniach złożonych jest wynikiem m. in. szybkości i do-

kładności procesów związanych z percepcją, oceną i podejmowaniem decyzji, a także umiejętności kontroli reakcji (m. in. powstrzymywania reakcji odruchowych, gdy są one nieadekwatne) i poziomu pracy efektorów (czyli m. in. szybkości i koordynacji motorycznej). Wszystkie te aspekty funkcjonowania powinny być więc uwzględnione w procesie diagnostycznym i rozpatrywane zarówno z perspektywy czasowej (szybkość), jak i pod względem poprawności reakcji (adekwatność).

Wśród narzędzi diagnostycznych stosowanych w psychologii wyróżniamy zazwyczaj testy mocy, testy szybkości oraz – jako grupę pośrednią, łączącą cechy dwóch pozostałych – testy z ograniczeniem czasu [2]. W każdej z tych grup trochę inaczej kształtuje się problem pomiaru sprawności działania. Testy mocy nie mają ograniczenia czasowego, lub jest ono na tyle liberalne, aby badany miał szansę podejść do każdego z zadań testowych. Z drugiej strony, zadania testowe mają zróżnicowany, stopniowany poziom trudności, a liczba poprawnie rozwiązanych zadań jest wyznacznikiem maksymalnego poziomu wykonania diagnozowanej osoby. W testach mocy brany jest więc pod uwagę aspekt poprawności reakcji, bez uwzględniania szybkości analizy i podejmowania decyzji. Istnieje w nich dodatkowa możliwość pomiaru czasu pracy jako osobnej zmiennej, choć trafność takiego pomiaru w znacznym stopniu zależy od instrukcji testowej – jeśli podkreśla ona brak presji czasu, zróżnicowane tempo działania nie będzie wyrazem faktycznych możliwości, a jedynie preferencji osób badanych do pracy szybkiej lub wolnej.

W odróżnieniu od testów mocy, klasyczne testy szybkości składają się z zadań o w miarę wyrównanym, niewielkim poziomie trudności, nakładając równocześnie na badanego surowe ograniczenie czasowe. Sprawia to, że wynik, czyli liczba poprawnie rozwiązanych zadań, odzwierciedla jedynie tempo pracy diagnozowanej osoby. Wypowiadanie się o poprawności działania jest możliwe np. przez analizę liczby błędów (tj. liczby zadań, które zostały podjęte, ale reakcja była nieprawidłowa), choć sensowność takiego wskaźnika zależy od konstrukcji testu (specyficzna jest też jego interpretacja, która może odwoływać się np. do cech temperamentalnych i świadczyć o tendencji do dezorganizacji zachowania przy dużej presji czasu).

Pośredni wariant, czyli testy z ograniczeniem czasu, składają się z zadań o zróżnicowanym poziomie trudności, zmuszając równocześnie badanego do pracy w poczuciu deficytu czasu. Wynik (liczba poprawnie rozwiązanych zadań) zależy więc zarówno od tempa pracy, jak i poziomu mierzonych zdolności (zadania zbyt trudne nawet przy dużym tempie pracy nie będą poprawnie rozwiązane). Testy z ograniczeniem czasu są więc najbliższe idei pomiaru sprawności działania, której potrzeba została zasygnalizowana na początku artykułu. Wadą tych narzędzi jest jednak problem z interpretacją wyników obniżonych – nie dają one jednoznacznej odpowiedzi, czy przyczyną takiego wyniku jest obniżony poziom zdolności, czy np. jedynie wolne tempo działania.

W badaniach kwalifikacyjnych, w których oceniane są predyspozycje do szkolenia lotniczego, wykorzystywanych jest wiele testów psychologicznych pozwalających na ocenę zarówno podstawowych zdolności intelektualnych (np. poziom inteligencji płynnej), jak i ocenę poziomu wykonania zadań specyficznych, angażujących równocześnie różne funkcje poznawcze (np. rotacja wyobraźniowa samolotu przez określanie jego położenia względem horyzontu). Ostateczna decyzja na temat odpowiednich kwalifikacji (lub ich braku) u badanego oparta jest nie na pojedynczym wyniku testowym, ale na wiązce wyników, w której ważny jest zarówno poziom, jak i wzajemne

relacje wskaźników testowych. Aby decyzje oparte na analizie zbioru wyników były trafne, każdy ze stosowanych wskaźników testowych musi w sposób maksymalnie jednoznaczny charakteryzować poszczególne aspekty funkcjonowania badanego (oczywiście, każdy ze stosowanych testów musi też cechować się zadowalającym poziomem rzetelności pomiaru). Jak wspomniano wcześniej, zasadne jest, aby te aspekty były rozważane zarówno z perspektywy szybkości, jak i poprawności działania.

Wiele z testów stosowanych w procesie diagnostycznym to typowi reprezentanci kategorii testów mocy i testów szybkości. W takich narzędziach analiza sprawności działania wymaga uwzględnienia – poza ogólnym wynikiem testu – także wskaźników dodatkowych. Niejednokrotnie wskaźniki takie są nieznormalizowane, w niektórych testach duża ich liczba znacznie komplikuje proces interpretacji. Przydatne i wskazane byłoby więc stworzenie uniwersalnego wskaźnika, który pozwalałby określić sprawność działania w różnego rodzaju testach mierzących poziom istotnych dla pilota funkcji poznawczych i psychomotorycznych.

Inspiracją mogą być prace dotyczące opisanego przez J. Kagana stylu poznawczego impulsywność – refleksyjność, zwanego czasem tempem poznawczym [wg 3]. Styl ten opisuje preferencje poznawcze do działania, w którym priorytetem jest albo szybkość (przy znacznym ignorowaniu możliwości popełnienia błędu), albo dokładność (której wysoki poziom zapewniany jest kosztem szybkości), nie jest natomiast wyznacznikiem wyższego lub niższego poziomu badanych zdolności. Badania nad tempem poznawczym w większości prowadzone były z zastosowaniem testu MFF autorstwa J. Kagana [wg 4] i m. in. ze względu na specyfikę tego testu dotyczyły głównie dzieci (w młodym wieku tempo poznawcze podlega największym zmianom rozwojowym). Wynikami testu MFF są, odrębnie mierzone, czas pracy i liczba popełnionych błędów, zaś klasyfikacja osób do grupy refleksyjnych, impulsywnych lub do typu pośredniego odbywa się na podstawie wzajemnej relacji tych dwóch wskaźników. Opracowano szereg metod ułatwiających taką klasyfikację, z których najciekawsza wydaje się być opisana przez Salkinda i Wrighta [5]. Zaproponowali oni wyliczanie pojedynczego wskaźnika impulsywności według wzoru:

$$I = z_E - z_T \quad (1)$$

gdzie z_E oznacza standaryzowany wynik z liczby popełnionych błędów, a z_T oznacza standaryzowany wynik z czasu wykonania zadania.

Dodatnia wartość wskaźnika I oznacza osoby impulsywne (duża liczba błędów i szybkie tempo pracy), zaś ujemna wartość I oznacza osoby refleksyjne (mała liczba błędów i wolne tempo pracy). Ponieważ I jest opisany na skali ciągłej, pozwala traktować refleksyjność-impulsywność jako wymiar, a nie jedynie jako kategorie klasyfikacyjne. Niejako wtórnym wynikiem jest możliwość prostego przekształcenia tego wzoru na wskaźnik efektywności, czyli:

$$E = z_E + z_T \quad (2)$$

gdzie oznaczenie z_E i z_T są takie same jak w poprzednim wzorze.

Ujemna wartość wskaźnika E oznacza dużą efektywność działania (niewielka liczba błędów i szybkie tempo pracy). Dodatnia wartość oznacza niewielki poziom efektywności. Wskaźnik E pozwala więc w pojedynczym wyniku przedstawić poziom sprawności działania na skali niezależnej od stosowanego testu, zaś w razie potrzeby pozwala na analizę jego części składowych w celu ustalenia, w jakim stopniu do efektywności przyczyniło się tempo, a w jakim dokładność reakcji. Konstrukcja tego wskaźnika pozwala też na jego stosowanie także w testach, w których – ze względu na konstrukcję czy materiał testowy – analiza stylu poznawczego impulsywność-refleksyjność jest bezzasadna.

W celu sprawdzenia przydatności wskaźnika E w narzędziach stosowanych w diagnostyce pilotów, do analizy wybrano wyniki dwóch testów wchodzących w skład Wiedeńskiego Systemu Testów: PST i COG. Ich szczegółowa charakterystyka zostanie przedstawiona w dalszej części artykułu. Każdy z tych testów pozwala na pomiar zarówno tempa, jak i poprawności reakcji, żaden z nich nie ma jednak pojedynczego wskaźnika obrazującego sprawność działania badanych. Ocena wskaźnika E oparta będzie na analizie korelacji tego parametru oraz podstawowych wskaźników testów z wynikami testów mierzących podstawowe aspekty funkcjonowania poznawczego (inteligencję płynną i pojemność pamięci świeżej), a także na analizie mocy predykcyjnej tego wskaźnika jako zmiennej przewidującej pozytywną kwalifikację do szkolenia lotniczego.

Badani i metoda

Badani. W badaniu analizie poddano wyniki 375 osób (w tym 55 kobiet), średni wiek 22 lata, będących kandydatami do szkolenia lotniczego w Wyższej Szkole Oficerskiej Sił Powietrznych w Dęblinie. Każdy z badanych w wyniku procesu diagnostycznego (kwalifikacja oparta na rozbudowanej baterii narzędzi psychologicznych, których wyniki podlegają szczegółowej ocenie) miał na podstawie badań psychologicznych przydzieloną kategorię zdrowia, opisującą zdolność do pełnienia służby w powietrzu lub do służby w ramach naziemnego zabezpieczenia lotów. Na potrzeby niniejszego badania zmienna ta została zdychotomizowana, dzięki czemu badani zostali przydzieleni do dwóch grup: osób pozytywnie zakwalifikowanych do podjęcia szkolenia na pilota wojskowego (przydzielona w ramach badań psychologicznych kategoria IA, IB lub IC) oraz osób niezakwalifikowanych (przydzielona kategoria II lub III albo niezdolny do podjęcia nauki w WSOSP). Wśród 375 badanych, pozytywnie zakwalifikowanych było 295 osób.

Narzędzia. Analiza koncentrowała się na wynikach dwóch testów wchodzących w skład Wiedeńskiego Systemu Testów. Był to test orientacji przestrzennej pilotów (PST [6]) oraz ogólny test wykonawczy badający uwagę i koncentrację, tzw. Cognitron (COG [7]). Test PST jest klasycznym testem mocy, w którym osoba badana wykonuje 13 zadań o zróżnicowanym poziomie trudności, zaś ograniczenie czasowe jest bardzo liberalne (dwie i pół minuty na zadanie, przy przeciętnym czasie rozwiązywania zadania wynoszącym mniej niż minutę). Końcowym wynikiem, świadczącym o poziomie zdolności przestrzennych i umiejętności rotacji wyobraźniowej, jest liczba poprawnie rozwiązanych zadań. Jest to jedyny znormalizowany wskaźnik testowy. Poza nim, w teście kontrolowany jest też czas wykonania, lecz ze względu na sposób konstrukcji testu nie podlega on interpretacji. Na

potrzeby wyliczenia wskaźnika E oba parametry zostały wystandaryzowane. Oryginalny sposób wyliczenia wskaźnika E daje w rezultacie skalę, w której wyższa efektywność jest opisywana przez wartości ujemne – jest to zrozumiałe w kontekście pracy Salkinda i Wrighta, która koncentrowała się na wymiarze refleksyjności-impulsywności (zaś wskaźnik E był jedynie wtórną modyfikacją wskaźnika I), jednak na potrzeby niniejszych analiz przyjęto bardziej intuicyjną skalę, w której wysokie wartości E świadczą o wysokiej efektywności. Dodatkowo – ponieważ w teście PST raportowana jest liczba poprawnych odpowiedzi, a liczba błędów (potrzebna do wyliczenia oryginalnego wskaźnika E) jest jej prostym dopełnieniem ze względu na stałą liczbę zadań, po odpowiednich modyfikacjach do wyliczenia wskaźnika E zastosowano nowy wzór:

$$E_{\text{PST}} = z_C - z_T \quad (3)$$

gdzie z_C oznacza standaryzowany wynik z liczby poprawnych odpowiedzi, z_T oznacza standaryzowany wynik z czasu wykonania zadania, zaś sam wskaźnik E_{PST} przyjmuje dodatnie wartości dla wysokiej efektywności (działanie szybkie i dokładne) i ujemne dla efektywności niskiej (działanie długie i niedokładne).

Test COG służy do pomiaru koncentracji uwagi oraz pomiaru sprawności analizy wzrokowej u badanych. Polega on na porównywaniu prezentowanej figury geometrycznej z zestawem wzorcowym i udzielaniu odpowiedzi, czy figura ta znajduje się w tym wzorcowym zestawie, czy też nie. Konstrukcja testu zbliżona jest do klasycznych testów szybkości – z dużą liczbą zadań (200 w stosowanej w omawianym badaniu wersji testu) o względnie równoważnym poziomie trudności, umożliwiającym przeciętnemu badanemu udzielenie poprawnych odpowiedzi we wszystkich itemach testowych. Złożoność zadań wydłuża jednak czas ich rozwiązywania, i właśnie tempo pracy, a konkretnie średni czas poprawnego oceniania figur niepasujących do wzorca (jako zadanie bardziej złożone poznawczo niż znajdowanie zgodności), jest głównym wskaźnikiem testowym. W odróżnieniu od klasycznych testów szybkości nie ma w stosowanej wersji COG ograniczenia czasu, jednak jest on głównym przedmiotem pomiaru, a instrukcja testowa podkreśla znaczenie dużego tempa pracy. Równocześnie, badani zwiększając tempo swej pracy, ryzykują więcej błędów wynikających z niedostatecznej oceny prezentowanych figur. Wskaźniki uzupełniające testu to średni czas poprawnych ocen figur zgodnych z wzorcem, a także liczba poprawnych ocen figur zgodnych i liczba poprawnych ocen figur niezgodnych. Wszystkie cztery niezależne wskaźniki są znormalizowane. Dodatkowo kontrolowany jest całkowity czas wykonania testu, nie podlega on jednak normalizacji. W rezultacie diagnosta musi oceniać nie tylko poziom poszczególnych wyników, ale też ich wzajemne relacje, co znacząco utrudnia porównywanie wyników pochodzących od kilku różnych osób.

W niniejszej analizie z wskaźników testu COG stworzono pojedynczy wskaźnik efektywności E_{COG} . W pierwszym kroku zsumowano liczbę poprawnych odpowiedzi, następnie tę sumę oraz czas wykonania testu wystandaryzowano. Warto podkreślić, że czas wykonania testu jest bardzo wysoki (w badanej grupie r -Pearsona wynosi 0,98) skorelowany z głównym wskaźnikiem testowym, tj. przeciętnym czasem poprawnego odrzucania figur niezgodnych. Uzyskane wartości podstawiono następnie do wzoru:

$$E_{\text{COG}} = z_{\text{C}} - z_{\text{T}} \quad (4)$$

gdzie, podobnie jak przy teście PST, z_{C} oznacza standaryzowany wynik z liczby poprawnych odpowiedzi, z_{T} oznacza standaryzowany wynik z czasu wykonania testu, zaś sam wskaźnik E_{COG} przyjmuje dodatnie wartości dla wysokiej efektywności (działanie szybkie i dokładne) i ujemne dla efektywności niskiej (działanie długie i niedokładne). Jako miary kontrolne, służące ocenie trafności opisanych wyżej wskaźników E , przyjęto wyniki Testu Matryc Ravena w wersji dla Zaawansowanych (TMZ [8]), ogólnie traktowane jako dobra miara inteligencji płynnej, oraz wyniki testu CORSI [9] służącego do pomiaru rozpiętości wizualno-przestrzennej pamięci świeżej, która odgrywa istotną rolę w procesie analizy poznawczej w trakcie wykonywania testów COG i PST [10].

Wyniki

Statystyki opisowe

Statystyki opisowe wszystkich analizowanych zmiennych ilościowych znajdują się w tab. 1.

Tab. 1. Statystyki opisowe. $N = 375$
Tab. 1. Descriptive statistics ($N = 375$)

Zmienna	Średnia	Odchylenie standardowe	Skośność	Kurtoza
COG przeciętny czas poprawnych ocen figur niezgodnych [COG ₁]	2,17	0,44	0,76	0,63
COG przeciętny czas poprawnych ocen figur zgodnych [COG ₂]	2,04	0,4	0,64	0,23
COG liczba poprawnych ocen figur niezgodnych [COG ₃]	115,45	3,15	-0,98	0,84
COG liczba poprawnych ocen figur zgodnych [COG ₄]	75,57	3,32	-1,43	2,76
COG czas wykonania testu (w sekundach)	424,16	80,91	0,65	0,27
COG Efektywność [E_{COG}]	0	1,09	-0,41	-0,04
PST liczba poprawnych odpowiedzi [PST ₁]	7,02	2,66	-0,4	-0,3
PST czas wykonania testu (w sekundach)	382,42	167,92	1,35	3,24
PST Efektywność [E_{PST}]	0	1,5	-0,68	0,6
TMZ	23,9	4,53	-0,57	0,83
CORSI	6,8	1,13	0,14	-0,13

Rozkłady wszystkich zmiennych wykazują niewielkie, ale istotne statystycznie (w teście Shapiro-Wilka $p < 0,05$) odstępstwa od rozkładu normalnego. Biorąc jednak pod uwagę wielkość próby, nie stanowi to znaczącej przeszkody do stosowania, w toku dalszej analizy metod parametrycznych.

Analiza statystyczna

Pierwszym krokiem oceny wskaźników efektywności dla testu COG i PST była ocena związku tych wskaźników ze standardowymi wskaźnikami testowymi oraz z wynikami narzędzi do pomiaru podstawowych aspektów funkcjonowania poznawczego. W tab. 2 prezentowana jest macierz korelacji wskaźników służących do interpretacji wyników testów COG i PST wraz z wyliczonym na potrzeby niniejszej analizy wskaźnikami E_{COG} i E_{PST} oraz wynikami testów TMZ i CORSI.

Tab. 2. Korelacje wskaźników analizowanych testów z nowo utworzonymi wskaźnikami efektywności $N = 375$

Tab. 2. Correlation of analyzed tests indices with newly formed indices ($N = 375$)

	COG₂	COG₃	COG₄	PST₁	TMZ	CORSI	E_{COG}	E_{PST}
COG₁	0,86 *	0,31 *	0,33 *	-0,12 *	-0,21 *	-0,21 *	-0,54 *	-0,31 *
COG₂		0,41 *	0,28 *	-0,03	-0,12 *	-0,18 *	-0,48 *	-0,24 *
COG₃			0,34 *	0,24 *	0,27 *	0,11 *	0,41 *	0,14 *
COG₄				0,15 *	0,23 *	0,19 *	0,48 *	0,05
PST₁					0,46 *	0,28 *	0,31 *	0,75 *
TMZ						0,36 *	0,44 *	0,43 *
CORSI							0,36 *	0,33 *
E_{COG}								0,38 *

Uwaga! Kodowe oznaczenia wskaźników testowych identyczne jak w tab. 1. * $p < 0,05$

Dodatkowo wyliczono też, za pomocą regresji wielokrotnej, współczynniki R opisujące związek wyników testów TMZ i CORSI ze wszystkimi czterema standardowymi wskaźnikami testu COG równocześnie. W przypadku związku wskaźników COG z wynikami TMZ, współczynnik R wynosił 0,46 ($R^2 = 0,21$), zaś w przypadku związku z wynikami testu CORSI współczynnik R wynosił 0,36 ($R^2 = 0,13$)

Kolejny krok polegał na ocenie, w jakim stopniu pierwotne wskaźniki testowe oraz wskaźniki efektywności pozwalają przewidzieć pozytywną lub negatywną kwalifikację badanej osoby do szkolenia lotniczego. Ponieważ kwalifikacja miała charakter zmiennej zdychotomizowanej, w celu przeprowadzenia porównań zastosowano regresję logistyczną, w której zmienną przewidywaną była kwalifikacja do szkolenia, zaś predyktorami były poszczególne wskaźniki testowe. Ponieważ wszystkie modele regresji były budowane z wykorzystaniem tego samego zbioru danych, z tą samą zmienną zależną, a celem był ocena mocy predykcyjnej modelu, jako podstawę do porównań między modelami przyjęto współczynnik R^2 Coxa i Snella (jest to współczynnik opisujący poprawę dopasowania modelu w stosunku do modelu zerowego, a jego maksymalna wartość jest mniejsza niż 1, przez co nie może być interpretowany jak klasyczny współczynnik R^2 znany z regresji liniowej).

Dla testu COG na wstępie skonstruowano cztery modele, po jednym dla każdego z standardowych wskaźników testowych. W przypadku zmiennych opisujących tempo pracy, tj. przeciętny czas poprawnych ocen figur niezgodnych (COG_1) i przeciętny czas poprawnych ocen figur zgodnych (COG_2), nie odnotowano istotnej poprawy względem modeli zerowych, a R^2 Coxa i Snella przyjmowało wartości poniżej 0,01. W przypadku modelu z liczbą poprawnych ocen figur niezgodnych (COG_3) jako predyktorem, model wskazywał istotnie lepsze dopasowanie od modelu zerowego, a R^2 Coxa i Snella wynosiło 0,09. Podobnie, w modelu z liczbą poprawnych ocen figur zgodnych (COG_4) jako predyktorem, model wskazywał istotnie lepsze dopasowanie od modelu zerowego, a R^2 Coxa i Snella również wynosiło 0,09. Następnym był model uwzględniający wszystkie cztery podstawowe wskaźniki testowe równocześnie. Okazał się on istotnie lepiej dopasowany do danych niż model zerowy, z R^2 Coxa i Snella wynoszącym 0,21. Istotne ($p < 0,001$) były wszystkie predyktory z wyjątkiem przeciętnego czasu poprawnych ocen figur zgodnych ($p = 0,327$). Analogiczny model, w którym klasyfikacja była przewidywana na podstawie wskaźnika E_{COG} , był istotnie lepiej dopasowany od modelu zerowego, a R^2 Coxa i Snella wynosiło 0,17.

W przypadku testu PST, stworzone zostały dwa modele – jeden uwzględniający standardowy wskaźnik testu, tj. liczbę poprawnie rozwiązanych zadań, i drugi ze wskaźnikiem E_{PST} w roli predyktora. Model ze standardowym wskaźnikiem testu PST okazał się istotnie lepiej dopasowany do danych niż model zerowy, a R^2 Coxa i Snella wynosiło 0,31. Podobnie, istotnie lepiej dopasowany niż model zerowy był model ze wskaźnikiem efektywności. R^2 Coxa i Snella było jednak znacznie niższe i wynosiło 0,21.

Omówienie

Na wstępie należy zwrócić uwagę na istotną statystycznie korelację o umiarkowanym natężeniu między wskaźnikami E_{PST} i E_{COG} , znacznie większą niż korelacje wyników testu PST z każdym z pojedynczych wyników testu COG. Sugeruje to, iż efektywność poznawcza opisywana przez wskaźnik E jest trafnym, uniwersalnym konstruktem, częściowo niezależnym od typu wykonywanego zadania. Poza tym, szczegółowe wnioski, jakie można wyciągnąć na podstawie analizy wyników, są nieco odmienne w przypadku testu COG i testu PST.

W teście COG obliczony według podanego wcześniej wzoru współczynnik efektywności E_{COG} okazuje się trafniejszą miarą funkcjonowania poznawczego oraz lepszym predyktorem pozytywnej kwalifikacji do szkolenia niż każdy z czterech wskaźników testowych rozpatrywanych pojedynczo. Równocześnie modele regresji wielokrotnej (w ramach klasycznego modelu liniowego oraz regresji logistycznej) pokazują, że współczynnik E_{COG} prowadzi do praktycznie identycznych przewidywań jak standardowe wskaźniki testu COG rozpatrywane łącznie. W tej sytuacji jego przewaga nad standardowymi współczynnikami jest znacząca, gdyż opisanie funkcjonowania badanych za pomocą pojedynczego, jednoznacznie interpretowalnego wskaźnika ułatwia zarówno proces orzekania, jak i porównania międzyosobowe. Można więc powiedzieć, że w teście szybkości, którego głównym

celem jest pomiar efektywności pracy, zastosowanie pojedynczego wskaźnika efektywności nie zubaża diagnozy, a równocześnie ułatwia interpretację wyników.

W przypadku testu PST wyniki kształtują się nieco inaczej. Ponieważ PST jest testem mocy, standardowo nie uwzględnia on czasowego aspektu wykonywania zadań poznawczych. Rozszerzenie wyniku PST o element szybkości, jak we współczynniku E_{PST} , nieznacznie obniża korelację z poziomem inteligencji płynnej (wynik testu TMZ), zwiększając równocześnie siłę związku z rozpiętością pamięci świeżej (wynik testu CORSI), warunkującą większą dynamikę procesów poznawczych. Zmiany współczynników korelacji są jednak w obu przypadkach nieistotne statystycznie (odpowiednio $p = 0,61$ w przypadku korelacji z TMZ i $p = 0,452$ w przypadku korelacji z CORSI, co zostało wyliczone w wyniku porównania współczynników r po transformacji z-Fishera). Równocześnie widoczna jest niższa moc predykcyjna modelu z E_{PST} w przewidywaniu klasyfikacji do szkolenia lotniczego.

W tym miejscu należy jednak zaznaczyć, że procedura zestawiająca wskaźniki testów COG i PST z przydzieloną kategorią zdrowia i kwalifikacją do szkolenia lotniczego jest obciążona efektem tzw. skażenia kryterium, gdyż kwalifikacja badanych była oparta na wynikach baterii testów, w skład której wchodziły m. in. testy COG i PST. Jest to poniekąd wada prezentowanej analizy, choć – ze względu na to, że proces diagnostyczny nie poddaje się prostej algorytmizacji – uzyskanie innego materiału badawczego w chwili obecnej wydaje się niemożliwe. Równocześnie skażenie kryterium tłumaczy w pewnym stopniu uzyskane wyniki – w przypadku testu COG w kwalifikacji brane są pod uwagę wszystkie parametry wykorzystane przy tworzeniu współczynnika E_{COG} (jak wskazano wcześniej, korelacja czasu wykonania testu z przeciętnym czasem poprawnej oceny figur niezgodnych jest niemal idealna), stąd predykcja na podstawie takich wskaźników jest niemal identyczna. Równocześnie, w standardowej ocenie testu PST, którą dokonuje się w procesie diagnozy i kwalifikacji, szybkość wykonywania testu nie jest uwzględniana. Tłumaczy to pogorszenie trafności predykcji w przypadku współczynnika E_{PST} i może stanowić argument, aby w przyszłości w procesie kwalifikacji uwzględnić także aspekt czasu, gdyż wskaźnik E_{PST} wykazuje podobną trafność teoretyczną co standardowy wynik testu PST (co widać dzięki korelacjom z wynikami TMZ i CORSI), mając równocześnie większą trafność treściową (uwzględnia bowiem aspekt czasowy, niezbędny do oceny efektywności).

Podsumowanie

Przedstawione analizy zachęcają do stosowania wskaźnika efektywności w procesie diagnostycznym. Wskaźnik taki może być skonstruowany dla każdego testu, w którym kontrolowana jest zarówno dokładność pracy, jak i jej tempo. Wnioski, jakie płyną z prezentowanych w niniejszym artykule analiz można podsumować następująco:

- a) wskaźnik E okazuje się miarą względnie uniwersalną, przynajmniej w obrębie testów poznawczych – wskazuje na to umiarkowana, istotna korelacja efektywności wykonania dwóch testów różniących się zarówno pod względem treści (test uwagi i test zdolności przestrzennych), jak i formy (test szybkości i test mocy).

- b) równocześnie wskaźnik ten jest łatwo interpretowany w kategoriach odchylenia od przeciętnego poziomu efektywności (czyli od zera), a dodatkowe przekształcenie polegające na podzieleniu wskaźnika E przez 2 sprawi, że będzie on wyrażony w jednostkach analogicznych do odchylenia standardowego.
- c) powyższe cechy sprawiają, że stosowanie wskaźnika E może znacząco ułatwić decyzje diagnostyczne tam, gdzie są one dokonywane na podstawie zbioru wyników, opisujących zarówno szybkość, jak i dokładność działania. Dzięki wskaźnikowi E , zamiast złożonych porównań międzyprofilowych można bowiem zastosować pojedynczy, nadrzędny wymiar efektywności, który pozwala na proste uszeregowanie badanych w celu np. wyselekcjonowania osób, które spełniają przyjęte wcześniej kryteria. Może to stanowić duże ułatwienie dla psychologów-diagnostów, pracujących zarówno w lotnictwie wojskowym, jak i cywilnym.

Na koniec trzeba jednak zwrócić uwagę na pewne problemy ze stosowaniem w praktyce opisanych w niniejszym artykule wskaźników efektywności. Po pierwsze, wskaźniki te otrzymujemy na drodze przekształceń arytmetycznych, co jest procesem bardziej skomplikowanym i pracochłonnym niż odniesienie uzyskanego wyniku surowego do tabeli norm. Po drugie, taka procedura, w której transformacja wyniku opiera się na parametrach rozkładu (średnie i odchylenia standardowe wykorzystywane w procesie standaryzacji) sprawia, że współczynniki E są bardziej podatne na odstępstwa rozkładu wyników od rozkładu normalnego, przez co w skrajnych przypadkach mogą prowadzić do nietrafnych interpretacji – w odróżnieniu np. od często stosowanych norm centylowych, które mają charakter skali porządkowej i są niewrażliwe na kształt rozkładu wyników surowych. Z tego względu bezpieczne i wygodne wydaje się stosowanie współczynników E dopiero *post factum* na zebranych uprzednio zbiorze wyników, jako współczynniki do oceny wewnątrzgrupowej (po uprzednim sprawdzeniu parametrów statystycznych uzyskanych zmiennych).

Piśmiennictwo

1. Joint Aviation Authorities: JAA Manual of Civil Aviation Medicine. Global Engineering Documents, Englewood, Colorado 2009.
2. Anastasi A., Urbina S.: Testy psychologiczne. Pracownia Testów Psychologicznych Pol. Tow. Psychol., Warszawa 1999.
3. Strelau J.: Psychologia różnic indywidualnych. Wyd. „Scholar”, Warszawa 2006.
4. Matczak A.: Test Porównywania Znanych Kształtów (MFF) J. Kagana. Pracownia Testów Psychologicznych Pol. Tow. Psychol., Warszawa 1992.
5. Salkind N.J., Wright J.: The development of reflection-impulsivity and cognitive efficiency. *Human Development* 1977, 20, 377-387.
6. Grössenbrunner P., Neuwirth W.: Pilot's Spatial Test. Release 22.00. Schuhfried GmbH, Mödling 2002.

7. Wagner M., Karner T.: Cognitrone. Version 37.00. Schuhfried GmbH, Mödling 2006.
8. Jaworowska A., Szustrowa T.: Podręcznik do Testu Matryc Ravena. Wersja dla Zaawansowanych. Pracownia Testów Psychologicznych Pol. Tow. Psychol., Warszawa 1991.
9. Schellig D.: Corsi Block-Tapping test. Version 22.00. Schuhfried GmbH, Mödling 2007.
10. Zieliński P., Biernacki M.: Benefits of working memory in mental rotation task: spatial memory span connected with efficiency, not correctness in Pilot's Spatial Test. Proceedings of the 29th Conference of the European Association for Aviation Psychology, Budapeszt 2010, 367-371.

Nadesłano: 11.07.2011 r.

Zaakceptowano do publikacji: 26.07.2011 r.