



Test wiedzy³

Benchmarki w testach wiedzy, czyli jak wyznaczyć próg zaliczenia testu

Właściwie skonstruowany test wiedzy pozwala na trafniejsze wyłowienie właściwych kandydatów w trakcie procesu selekcji. Używając wyników z testów wiedzy, możliwe jest uszeregowanie ludzi według wzrastającej kolejności uzyskanych wyników. W praktyce jednak bardzo często pojawia się problem podjęcia decyzji o tym, jaki poziom wiedzy jest akceptowalny. Innymi słowy, chodzi o to, jaka wartość jest wystarczająca do tego, żeby uznać, że dana osoba posiada wystarczający zasób wiedzy.



Dariusz Ambroziak

Psycholog organizacji (SWPS) z dyplomem MBA (Akademia L. Koźmińskiego). Partner zarządzający w Instytucie Analiz im. Karola Gaussa Sp. z o.o. Posiada wieloletnie doświadczenie na rynku usług doradztwa organizacyjnego.
d.ambroziak@iakg.pl



Krzysztof Fronczyk

Doktor psychologii (UW) ze specjalnością psychometria, starszy konsultant w Instytucie Analiz im. Karola Gaussa Sp. z o.o. Posiada wieloletnie doświadczenie w zakresie projektowania narzędzi badawczych.
k.fronczyk@iakg.pl

Takie decyzje podejmowane są np. przy sprawdzaniu, czy dane szkolenie odniosło oczekiwany rezultat (pracownicy zdobyli zakładany poziom wiedzy), czy też przy przyznawaniu różnego typu certyfikatów oraz we wspomnianym procesie selekcyjnym. Kluczowe jest zatem umiejętne określenie minimalnej wartości, przy której można uznać, że dana osoba osiągnęła zakładany poziom wiedzy. Jest to problem określenia tak zwanych punktów odcięcia zwanych inaczej benchmarkami. Istnieją różne sposoby wyznaczania punktów odcięcia, np.:

- ▲ Określenie kryteriów zdawalności jest niezbędne przed pierwszym badaniem testowym.
- ▲ Określenie kryteriów zdawalności po zebraniu wyników z badania testowego.

o minimalnym, ale wystarczającym w danej sytuacji poziomie wiedzy. Powyższa grupa jest specyficzną grupą reprezentatywną lub grupą odniesienia. Od każdego eksperta wymaga się oszacowania, jaka część takiej grupy osób udzieli odpowiedzi poprawnej na każde z pytań testu. Można więc uznać, że eksperci szacują trudność każdego z pytań. Im pytanie łatwiejsze w opinii ekspertów, tym więcej wyobrażonych osób rozwiązujących test udzieli odpowiedzi poprawnej. Eksperci wyrażają swoje oceny w proporcjach osób udzielających prawidłowych odpowiedzi. W dalszej części tekstu zobaczymy wykorzystanie metody Angoffa w metodzie Ebela.

Metoda Ebela

Inną metodę wyznaczania punktów odcięcia opracował R. Ebel (1972). W jego metodzie każde pytanie oceniane jest przez ekspertów dwukrotnie. Pierwsza ocena polega na oszacowaniu trudności pytań. Dla uproszczenia przyjmuje się tu trzy kategorie trudności: duża, średnia, niska. Druga ocena dotyczy związku każdego z pytań z mierzonym zakresem wiedzy. W tym wypadku każdemu pytaniu eksperci przyporządkowują jedną z kategorii:

Metoda Angoffa

Jedną z pierwszych metod przeznaczonych do wyznaczania benchmarków, gdy jeszcze nikt nie został przetestowany, opracował W.H. Angoff (1971). Metoda ta jest relatywnie prosta. Polega na ocenie przez grupę niezależnych ekspertów każdego z pytań testu wiedzy. Ocena ta wymaga od eksperta wyobrażenia sobie grupy 100 osób rozwiązujących test

Tabela 1

Wykorzystanie pierwszej fazy metody Ebela do określenia punktu odcięcia w przykładowym teście

Związek z mierzonym zakresem wiedzy	Trudność		
	niska	średnia	duża
Zasadniczy	Pytania: 3, 7, 8, 13 Szacunek prawidłowych odpowiedzi: 95%	Pytania: 11, 15, 22 Szacunek prawidłowych odpowiedzi: 85%	Pytania: 2, 4 Szacunek prawidłowych odpowiedzi: 80%
Ważny	Pytania: 21, 6, 9 Szacunek prawidłowych odpowiedzi: 90%	Pytania: 10, 14, 20 Szacunek prawidłowych odpowiedzi: 75%	Pytania: 16, 25 Szacunek prawidłowych odpowiedzi: 60%
Akceptowalny	Pytania: 5 Szacunek prawidłowych odpowiedzi: 80%	Pytania: 12, 18 Szacunek prawidłowych odpowiedzi: 55%	Pytania: 19, 23 Szacunek prawidłowych odpowiedzi: 35%
Dyskusyjny	Pytania: 1 Szacunek prawidłowych odpowiedzi: 50%	Pytania: brak Szacunek: nie ma potrzeby dokonywania oszacowań	Pytania: 17, 24 Szacunek prawidłowych odpowiedzi: 20%

Opracowane na podstawie R. Ebel (1972) Essential of educational measurement

zasadnicze, ważne, akceptowalne, dyskusyjne. Wszystkie pytania zostają więc pokategoryzowane według dwóch kryteriów. Łączne ich uwzględnienie daje 12 różnych kategorii (3 poziomy trudności x 4 poziomy związków pytań z mierzonym zakresem wiedzy). Przykładowe przyporządkowanie każdego z pytań przedstawiono w tabeli 1, dla testu zawierającego 25 pytań.

W kolejnym kroku eksperci oceniają, na jaki odsetek pytań z danej kategorii osoby posiadające minimalny, ale akceptowalny poziom wiedzy są w stanie udzielić prawidłowej odpowiedzi (wykonanie tylko tego kroku oznacza zastosowanie tylko metody Angoffa). W związku z tym, że metoda R. Ebela jest nieco bardziej złożona, również oceny dokonywane przez ekspertów są dokonywane w nieco bardziej skomplikowany sposób. Livingston i Zieky (1982) proponują, aby ocen tych dokonywać w następujący sposób:

1. Sędziowie dokonują wstępnej klasyfikacji każdego z pytań do jednej z 12 kategorii. Przyporządkowania dokonywane są niezależnie przez każdego z sędziów.
2. W toku dyskusji eksperci ustalają wspólnie przyporządkowanie każdego pytania do kategorii.

3. Sędziowie dokonują wstępnego oszacowania odsetka pytań z każdej kategorii, na jakie zostanie podana prawidłowa odpowiedź przez osoby o minimalnym akceptowanym poziomie wiedzy. Oszacowania dokonywane są niezależnie przez każdego z sędziów.

4. W toku dyskusji eksperci ustalają wspólnie wartości wyżej wspomnianych odsetków.

Mając oszacowania procentów pytań z każdej kategorii, na jakie zostanie podana prawidłowa odpowiedź, można przystąpić do określenia punktu odcięcia. W tym celu należy:

1. Pomnożyć oszacowany odsetek pytań z danej kategorii, na które podawana jest prawidłowa odpowiedź przez liczbę pytań z tej kategorii. W ten sposób obliczony zostanie oczekiwany wynik danej kategorii pytań uzyskiwany przez minimalnie akceptowane osoby pod względem swojej wiedzy.
2. Powtórzyć operację z pkt 1 dla wszystkich kategorii.
3. Zsumować oczekiwane wyniki każdej kategorii pytań. Uzyskana suma jest punktem odcięcia.

W tabeli 2 przedstawiono przykład takich obliczeń, które przeprowadzono dla danych z tabeli 1.

Podsumowując, wartość benchmarku wynosi 17,6 pkt. Przekroczenie wartości granicznej oznacza spełnienie oczekiwań co do minimalnego poziomu wiedzy.

Powyżej opisane dwie metody wymagają wykorzystania ekspertów z danej dziedziny wiedzy. Nie zawsze jednak tacy eksperci są dostępni. Co więcej, ocena ekspercka może być, mimo wybrania najlepszych specjalistów, obciążona pewną dozą subiektywizmu czy własnych specyficznych doświadczeń. Istotne zmniejszenie błędów szacowania jest możliwe wtedy, gdy mamy do dyspozycji faktyczne wyniki testowe.

Metoda kryterium zewnętrznego

Gdy wyniki testowe zostały już zebrane, można wykorzystać metodę kryterium zewnętrznego. Jak zatem możemy zdefiniować kryterium zewnętrzne? Wyniki testów wiedzy powinny być skorelowane z różnymi innymi pomiarami, co do których można założyć, że są związane z określonym poziomem wiedzy. W przypadku wiedzy produktowej można przypuszczać, że im lepszą wiedzę



o sprzedawanych produktach posiada sprzedawca, tym lepiej je sprzedaje. Wynika stąd, że wyniki testów wiedzy produktowej powinny korelować z wielkością sprzedaży lub z innym miernikiem skuteczności handlowej, np. trafnością zadawania pytań podczas rozmowy handlowej itp.

Można także przypuszczać, że wyniki testów wiedzy powinny korelować z rzeczywistym zachowaniem pracownika podczas czynności, w których wykorzystywana jest wiedza z testowanego zakresu. Ocena zachowania podczas pracy jest dokonywana przez odpowiednio przygotowanych sędziów (asesorów). Czasem wykorzystuje się do tego celu przełożonych, ale nie jest to metoda polecana, ze względu na ewentualne uprzedzenia i ogólną wiedzę przełożonych o swoich pracownikach, która może wpływać na spostrzeganie wykonywania konkretnych czynności.

Warto o tym pamiętać, że wymienione mierniki (wielkość sprzedaży, ocena zachowania pracownika podczas pracy) to przykłady zewnętrznych kryteriów. Są to więc miary niedotyczące bezpośrednio zakresu posiadanej wiedzy, lecz tego, jak ta wiedza wykorzystywana jest w praktyce.

Założmy, że wyniki pewnego testu wiedzy korelują z ważnym dla danej firmy wskaźnikiem efektywności pracy, jakim jest procent wykonanego planu sprzedaży. Sytuacja ta jest przedstawiona w sposób schematyczny na schemacie 1. Każdy punkt na wykresie reprezentuje konkretnego pracownika. Położenie każdego punktu odpowiada wynikowi w teście (oś pozioma) oraz poziomowi efektywności pracy (oś pionowa). W tej sytuacji na podstawie testu wiedzy wnioskujemy o efektywności pracy. Jak więc ustalić punkt odcięcia na bazie pozyskanych danych?

W celach decyzyjnych konieczne jest ustalenie, jaki wynik testu jest satysfakcjonujący, a jaki nie. W tym celu

Tabela 2

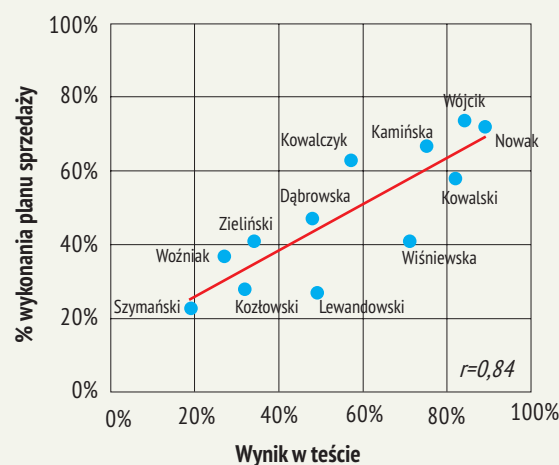
Wykorzystanie drugiej fazy metody Ebela do określenia punktu odcięcia w przykładowym teście

Kategoria	Oszacowanie odsetka pytań o prawidłowych odpowiedziach	Liczba pytań w kategorii	Oczekiwane wyniki dla pytań z kategorii
zasadnicze			
niska	95	4	$0,95 \times 4 = 3,8$
średnia	85	3	$0,85 \times 3 = 2,55$
duża	80	2	$0,8 \times 2 = 1,6$
ważne			
niska	90	3	$0,9 \times 3 = 2,7$
średnia	75	3	$0,75 \times 3 = 2,25$
duża	60	2	$0,6 \times 2 = 1,2$
akceptowalne			
niska	80	1	$0,8 \times 1 = 0,8$
średnia	55	2	$0,55 \times 2 = 1,1$
duża	35	2	$0,35 \times 2 = 0,7$
dyskusyjne			
niska	50	1	$0,5 \times 1 = 0,5$
średnia		0	
duża	20	2	$0,2 \times 2 = 0,4$
Suma			17,6

Opracowane na podstawie Ebel (1972) Essential of educational measurement

SCHEMAT 1

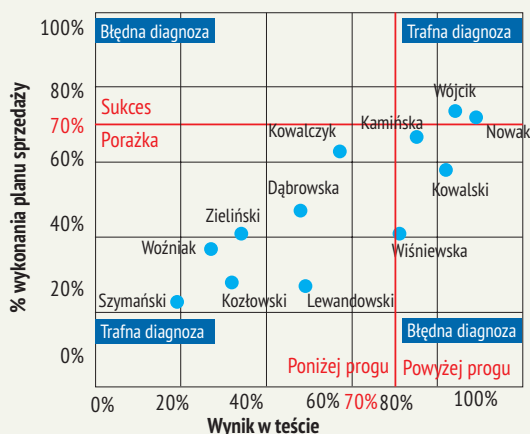
Korelacja między wynikami w teście a wynikami sprzedaży



Opracowanie własne Autorów

SCHEMAT 2

Obszary trafnej i błędnej diagnozy określone punktem odcięcia oraz poziomem efektywności zawodowej



Opracowanie własne Autorów

należy wykorzystać akceptowalny poziom efektywności pracy. Załóżmy np., że w wyżej opisanym przykładzie zadowalający jest 80-proc. poziom realizacji planu sprzedaży. Wszyscy, którzy realizują plan na takim poziomie, mogą być uznani za dobrych pracowników, a ci, którzy go nie realizują, za złych pracowników. Natomiast w przypadku testu wiedzy wstępnie określono punkt zdawalności na 75 proc. Próg ten zostanie w dalszej części analizy zweryfikowany w zależności od przeznaczenia testu. Wstępną kategoryzację osób zdających zamieszczono na schemacie 2. Na rysunku 2 uwzględniono jeszcze jeden dodatkowy podział. Jest to podział ze względu na wielkość wyniku testowego. Wszyscy wypełniający test zostali podzieleni na osoby o wysokich wynikach i osoby o niskich wynikach ze względu na punkt odcięcia. Biorąc pod uwagę te dwa podziały: ze względu na poziom efektywności pracy oraz ze względu na wynik w teście otrzymujemy podział badanych na cztery kategorie:

A. **Pierwsza** z nich to osoby, które osiągnęły wysoki wynik w teście i odnoszą sukcesy w pracy. Są to

osoby trafnie zdiagnozowane pod względem poziomu swej wiedzy.

B. **Druga** grupa to osoby, które wprawdzie osiągnęły wysoki wynik w teście, ale odnoszą porażki w pracy. Są to osoby błędnie zdiagnozowane pod względem poziomu swej wiedzy. Pozostają jeszcze dwie grupy osób, które mają niskie wyniki w teście. Są wśród nich tacy, którzy odnoszą sukcesy w pracy.

C. **Trzecia** grupa to błędne diagnozy poziomu wiedzy, bo według testu osoby te mają mały zasób wiedzy, mimo to sprawdzają się w pracy.

D. **Czwarta** ostatnia grupa to osoby, które odnoszą porażki w pracy i jednocześnie mają niskie wyniki w teście wiedzy. Są to więc trafne diagnozy poziomu wiedzy.

Jak zatem zweryfikować, czy punkt odcięcia został prawidłowo określony? Analizując schemat 2, można zauważyć, że jest możliwe takie ustawienie punktu odcięcia pod względem wyników testu, aby zminimalizować obszar błędnych diagnoz. Błędne przyjęcie kogoś do pracy wiąże się zwykle z kosztami, zatem istotą sprawy jest maksymalizacja obszaru trafnych akceptacji. Decyzje co do precyzyjnego

ustawienia punktów odcięcia są zależne od konkretnych celów biznesowych, jakie ma spełniać dany test. Posiadając dane dotyczące wyników testu, wybranego kryterium efektywności pracy oraz akceptowalnego poziomu tego kryterium, można tak ustalić punkt odcięcia pod względem wyniku testu, aby jak najbardziej zoptymalizować decyzje dotyczące wiedzy pracowników. Zaletą opisaną metody jest to, że uwzględnia ona pewien margines błędów w klasyfikacji pracowników pod względem posiadanej wiedzy. Oznacza to, że wiemy, jakim konkretnie błędem są obciążone decyzje selekcyjne. Natomiast w celu precyzyjnego wyznaczenia punktów odcięcia w opisaną metodę wykorzystuje się bardziej złożone metody analizy statystycznej. Analizy tego typu zwykle wykonują eksperci z dziedziny psychometrii i statystyki. Rezultatem analiz jest precyzyjne określenie punktów odcięcia w zależności od przyjętych kryteriów wynikających z aktualnych i przyszłych potrzeb danej organizacji. Możliwe jest także łączenie różnych metod ustalania punktów odcięcia. Można np. przy użyciu metody Angoffa ustalić punkt odcięcia, a następnie przetestować jego zasadność w odniesieniu do jakiegoś zewnętrznego kryterium itp. Zatem mamy pełną swobodę działania, o ile zostaną zachowane reguły metody oraz narzędzie testowe będzie sprawdzać się w praktyce biznesowej, bo to jest istotą sprawy. ▀

BIBLIOGRAFIA

1. W. H. Angoff, *Scales, norms, and equivalent scores*, w: R. L. Thorndike (red.), *Educational measurement*, American Council on Education, Washington 1971, s. 508–600.
2. R. L. Ebel, *Essential of Educational measurement*, Prentice-Hall, Inc., New Jersey 1972.
3. S. A. Livingston, M. J. Zieky, *Passing Scores: A manual for setting standards of performance on educational and occupational tests*, Educational Testing Service, Livingston 1982.