

Test wiedzy²

Jak rozpoznać rzetelność testów wiedzy

Istotnym zagadnieniem pomiaru wiedzy pracowników jest jego dokładność. Precyzja pomiaru testem nazywana jest rzetelnością testu wiedzy. Rzetelność odpowiada na pytanie, czy dobrze mierzymy to, co mierzymy. W przypadku testów wiedzy odpowiada na pytanie, czy nasza „linijka” jest dobrze wyskalowana.



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

Często można spotkać się z opiniami, że testy wiedzy się nie sprawdzają, gdyż:

- ▲ mierzą tylko wycinek pewnej wiedzy,
- ▲ wyniki testu to ruletka – raz się zdaje, a raz nie, mimo że testuje się ten sam materiał,
- ▲ nie można wykorzystać wyników testów w praktyce.

Z naszej praktyki wynika, że powyższe zarzuty są z reguły zasadne dla testów wiedzy konstruowanych *ad hoc*, bez użycia właściwej metodologii.

Wybrane sposoby szacowania rzetelności

Istnieją różne metody określania rzetelności testów wiedzy. Poniżej zostaną omówione metody w podziale na liczbę pomiarów niezbędnych do obliczenia odpowiednich współczynników rzetelności:

- ▲ Pomiar dwukrotny tym samym testem.
- ▲ Pomiar jednokrotny:
 - metody połówkowe,
 - analiza całego testu – bez podziału na części.

UWAGA CZYTELNICY

W pierwszej części cyklu zostały poruszone zagadnienia dotyczące tego, jak tworzyć bank pytań do testu wiedzy.

Pomiar dwukrotny – stabilność testu

Metoda badania stabilności testu polega na dwukrotnym sprawdzeniu wiedzy pewnej grupy osób przy użyciu tego samego testu. Oczywiście między oboma pomiarami następuje pewna przerwa, np. kilka tygodni. Jeśli nie zachodzą pomiędzy pomiarami żadne istotne zmiany mogące mieć wpływ na testowaną wiedzę (np. pracownicy nie uczestniczą w żadnym szkoleniu) i test daje dokładny pomiar, to należy spodziewać się, że wyniki uzyskane w obu pomiarach nie będą się istotnie różnić. Załóżmy, że w pierwszym pomiarze stwierdzimy, że pracownik A ma najwyższy poziom wiedzy z danego zakresu, pracownik C – najniższy, a poziom wiedzy pracownika B mieści się między pracownikiem A i C. Jeśli test jest rzetelny, to powinniśmy w drugim pomiarze uzyskać podobną hierarchię pracowników ze względu na posiadaną przez nich wiedzę (zobacz wykres 1).

Należy zauważyć, że w tym przypadku interesuje nas uzyskanie podobnej hierarchii wyników, choć wyniki punktowe mogą być nieco inne np. wyższe (efekt uczenia się testu). Jeśli zatem w drugim pomiarze uzyskamy podobne odwzorowanie



wyników (wysoka korelacja wyników) co w pomiarze pierwszym, możemy stwierdzić, że jest rzetelny, czyli dobrze mierzy obszar wiedzy, który podlega testowaniu.

Metody połówkowe – pomiar jednokrotny

Stosowanie metody podwójnego pomiaru nie zawsze jest praktycznie wykonalne. Poza tym pracownicy zwykle nabywają nową wiedzę, chociażby w procesie interakcji z innymi pracownikami czy przełożonymi. Zachodzi także proces zapominania części posiadanych informacji. Dodatkowo osoby badane mogą nabrać pewnej wprawy w rozwiązywanie testu, czy też po prostu zapamiętać niektóre pytania z pierwszego badania i np. poszukać na nie odpowiedzi. Wszystkie te czynniki będą miały wpływ na wyniki kolejnego pomiaru. Nieco inne wyniki drugiego pomiaru mogą sprawiać wrażenie, że jest on nierzetelny, podczas gdy w rzeczywistości zaszły zmiany w poziomie wiedzy pracowników. W metodach sprawdzania rzetelności opartych na jednokrotnym

badaniu wykorzystuje się podział testu na dwie części lub specjalne wskaźniki statystyczne. Pomiar rzetelności przez dzielenie testu na dwie części oznacza, że każda z nich musi spełniać wiele warunków statystycznych. Zakłada się, że obie części testu są tzw. testami równoległymi, co oznacza, że mają one jednakowe średnie wyników oraz jednakowe wariancje (zróżnicowanie wyników). Istnieją jeszcze dwa dodatkowe warunki równoległości testów, dużo trudniejsze do spełnienia, a jednocześnie do sprawdzenia.

▲ równość korelacji pytań między sobą w każdej części,

▲ równość korelacji każdej z części z innym, zewnętrznym względem testu pomiarem, np. efektywnością pracy.

Pojawia się pytanie o to, czy w ogóle warto stosować taką metodę oceny rzetelności, skoro oznacza ona konieczność spełnienia tak wielu warunków. W rzeczywistości jednak wystarczą pewne przybliżenia co do realizacji tych warunków, żeby można było dobrze oszacować rzetelność testu.

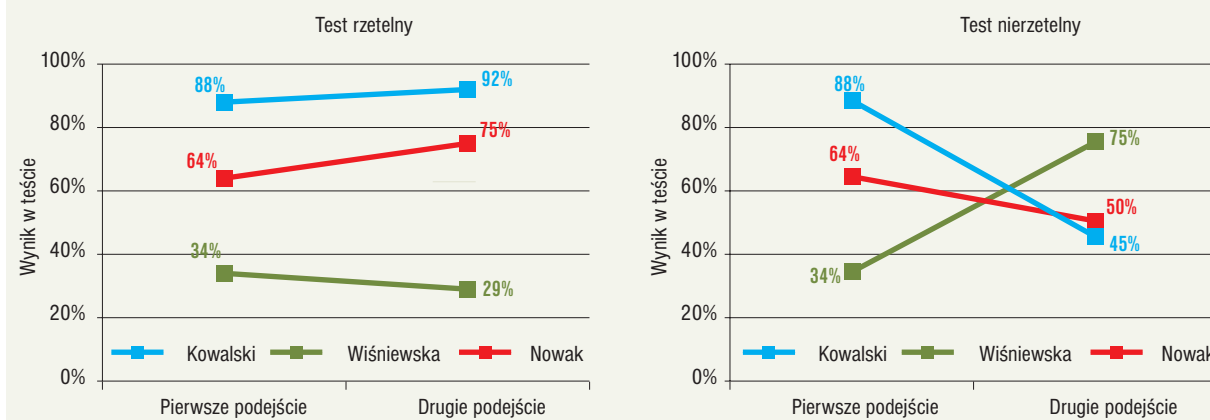
W jaki sposób to osiągnąć? Po pierwsze, podziału testu na dwie części nie można dokonywać w sposób bezrefleksyjny i czysto mechaniczny, np. pierwsze dziesięć pytań stanowi jedną połowę, a drugie dziesięć kolejnych pytań – drugą połowę. Takie postępowanie jest niezalecane, gdyż pytania w testach wiedzy układane są często we wzrastającej kolejności trudności. Podział testu w sposób opisany powyżej oznaczałby, że pierwsza część testu będzie dużo łatwiejsza niż druga, co oznacza, że na pewno średnie wyników nie będą równe.

WAŻNE

Czasami pytania w testach wiedzy łączone są w bloki tematyczne. Trzeba jednak pamiętać, że wiedza z jednego zakresu tematycznego może być opanowana przez pracowników w innym zakresie, niż wiedza z innego zakresu tematycznego. Innymi słowy – bloki tematyczne również różnią się swoją trudnością. W takiej sytuacji wspomniany podział również jest niewskazany, gdyż może zdarzyć się tak, że każda z części testu wiedzy będzie zawierała pytania tylko z jednego zakresu tematycznego. Oznacza to, że obie części będą znacznie różniły się pod względem treści, a co za tym idzie – trudności. ▲

WYKRES 1

Hierarchia wyników uzyskanych w teście wiedzy. Test wiedzy rzetelny v. test wiedzy nierzetelny



ŹRÓDŁO: Instytut Analiz im. Karola Gaussa Sp. z o.o., październik 2011 r.

WYKRES 2

Korelacja między wynikami uzyskanymi w dwóch połówkach testu



ŹRÓDŁO: Instytut Analiz im. Karola Gaussa Sp. z o.o.

Metody podziału testu

Zalecanym sposobem postępowania jest dobór losowy pytań do obu części testu. Zakłada się, że taki dobór spowoduje, że pytania o różnej trudności będą reprezentowane w jednakowy sposób w obu połowach testu. Inny sposób postępowania to dobór pytań do obu części testu przez ekspertów w ten sposób, aby obie części były równoważne pod względem treściowym, tj. w obu połówkach wszystkie zagadnienia objęte testem były w jednakowy sposób reprezentowane. Podobnie jak w przypadku metody powtarzanego pomiaru, miarą rzetelności testu jest współczynnik korelacji, tyle że w tym przypadku jest to korelacja pomiędzy oboma połowami testu. Chodzi tu zatem o stopień zgodności uporządkowań pracowników dokonanych za pomocą obu połówek testu.

Podział testu na dwie równoważne pod względem treści połowy nie zawsze jest możliwy. Zdarza się, że niektóre zagadnienia reprezentowane są czasem w teście zaledwie przez jedno pytanie. Również podział losowy nie zawsze musi dać w pełni równoważne

części testu. W takich sytuacjach wykorzystuje się inne wskaźniki.

Jednym z najbardziej popularnych wskaźników rzetelności jest tzw. współczynnik alfa Cronbacha, czyli wartość wskazująca stopień, w jakim poszczególne pytania testu są wzajemnie skorelowane. Innymi słowy, jak bardzo odpowiedzi na pytania są ze sobą powiązane, czy osoby zdające test, jak również niezdające – odpowiadają w sposób systematyczny czy losowy na pytanie testowe.

Jak interpretować wskaźniki rzetelności?

Im wyższy wskaźnik rzetelności, tym lepiej. Wskazuje on bowiem na to, że test dokładniej mierzy badany zakres wiedzy. Jednak wskaźnik ten zależy od kilku charakterystyk testu oraz badanej grupy. Na wykresie 2 zamieszczono przykładowe współczynniki rzetelności dla testu rzetelnego i testu nierzetelnego wraz z graficzną prezentacją rozproszenia wyników. Ilustracja dotyczy analizy współczynnika rzetelności dla metody połówkowej. Zatem im mniejsze rozproszenie wyników, tym bardziej

„spójnie” odpowiadały osoby testowane w pierwszej i drugiej części testu. Natomiast im większe rozproszenie wyników, tym mniej „spójnie” odpowiadały osoby testowane w pierwszej i drugiej części testu.

Rozproszenie wyników (wariancja)

Im większa wariancja, czyli im większe zróżnicowanie pracowników pod względem badanego zakresu wiedzy, tym większy wskaźnik rzetelności. Mówiąc bardziej przystępnie, im pracownicy są bardziej podobni do siebie pod względem posiadanej wiedzy, tym wskaźnik rzetelności jest niższy. Jeśli natomiast mamy pracowników bardzo różniących się ze względu na swą wiedzę, tzn. mamy pracowników o bardzo dużej wiedzy, jak również pracowników, którzy nie wiedzą prawie nic, to łatwiej jest osiągnąć wyższy wskaźnik rzetelności. Jest to dość ważne.

Pracownicy a kandydaci

W większości przypadków, gdy badanie prowadzone jest wśród



pracowników, można przypuszczać, że ich wiedza jest w jakiś sposób do siebie podobna. Z reguły są to ludzie, którzy przeszli przez sito selekcyjne, pracują w jednym zespole i nawzajem od siebie, nawet bez specjalnych szkoleń, zdobywają wiedzę związaną z wykonywaną pracą. Jeśli natomiast badamy kandydatów do pracy, to możemy oczekiwać, że w tym wypadku ich zróżnicowanie pod względem posiadanej wiedzy będzie większe, gdyż wśród takiej grupy znajdują się prawdopodobnie zarówno absolwenci, jak i osoby z dłuższym stażem pracy. Zróżnicowane doświadczenia kandydatów będą warunkować ich poziom wiedzy. Stąd należy oczekiwać wyższej rzetelności w przypadku kandydatów niż w przypadku pracowników.

Klucz odpowiedzi

Kolejny czynnik, jaki wpływa na rzetelność testu, to jakość opracowanego klucza odpowiedzi. Poprzez jakość klucza odpowiedzi rozumiemy to, czy w poprawny sposób określono, co jest odpowiedzią poprawną oraz czy przypisano odpowiedni sposób punktowania. Nie zawsze jest to takie jasne. Na przykład w pewnych bardzo specyficznych zagadnieniach nawet eksperci mogą nie być w pełni zgodni co do określenia w pełni prawidłowych odpowiedzi. Tak może zdarzyć się np. w testach dotyczących niektórych zagadnień związanych z technikami sprzedaży, gdyż prawidłowa odpowiedź zależy od konkretnego kontekstu sytuacyjnego. Stąd też popularność sytuacyjnych testów wiedzy i kompetencji dopasowanych do specyfiki danego biznesu oraz organizacji.

Liczba pytań w teście

Kolejnym czynnikiem wpływającym na rzetelność testu jest liczba pytań. Im więcej pytań, tym bardziej rzetelny test. Dzieje się tak z kilku

powodów. Po pierwsze, więcej pytań oznacza bardziej zróżnicowane odpowiedzi, czyli większą wariację wyników, która przyczynia się do zwiększenia wskaźnika rzetelności. Po drugie, dłuższy test daje bardziej wszechstronny obraz wiedzy pracownika. Przykładowo 5-pytaniowy test wiedzy o produkcie może dać bardzo przypadkowe odpowiedzi, dlatego że ze względu na małą liczbę pytań dotyczy on tylko wybranych, być może specyficznych, aspektów produktu. Ponadto w tak krótkim teście jest większe prawdopodobieństwo, uzyskania wysokiego wyniku w sposób przypadkowy lub przez zgadywanie. Ten sam test wydłużony do 20 pytań da już znacznie lepszą orientację w ocenie wiedzy pracownika.

Rozważając rzetelność testu, należy wziąć zawsze pod uwagę jego długość. W sytuacji porównywania testu 5-pytaniowego i testu 20-pytaniowego o tej samej rzetelności warto wybrać test krótszy. Skoro jednak udało się osiągnąć taką samą rzetelność jak w przypadku testu dłuższego, to świadczy to na korzyść testu krótszego. Patrząc z punktu ekonomii pomiaru, taniej i efektywniej jest stosować krótsze testy, jeśli posiadają zadowalającą rzetelność.

Jeśli zatem dłuższy test jest bardziej rzetelny niż krótszy, to wydaje się, że dołożenie dodatkowych pytań powinno podnieść rzetelność testu. Jednak nie jest to zasada bezwyjątkowa. Test po dodaniu dodatkowych pytań zwiększy swą rzetelność jedynie wtedy, gdy nowe pytania są skorelowane z już istniejącymi pytaniami oraz są dobrej jakości (np. jednoznaczne, dopasowane treściowo).

Poziom trudności testu

Z innych czynników wpływających na rzetelność testu warto jeszcze wymienić jego poziom trudności. Generalnie, testy bardzo trudne albo bardzo łatwe

charakteryzują się niską rzetelnością. Testy o przeciętnej trudności mają zwykle rzetelność największą. Dzieje się tak, gdyż w teście łatwym dużo osób osiąga bardzo wysokie wyniki, a w teście trudnym – dużo osób osiąga bardzo niskie wyniki. W obu tych przypadkach zakres zróżnicowania wyników jest niewielki, gdyż wszyscy osiągają podobne wyniki. Takie ograniczone zróżnicowanie badanych owocuje niską rzetelnością testu wiedzy. Dodatkowym czynnikiem obniżającym rzetelność bardzo trudnych testów jest zgadywanie. Część osób zdających test, nie znając prawidłowych odpowiedzi, będzie próbowała je odgadnąć. Przyczynia się to do powstawania dodatkowego szumu informacyjnego i spadku rzetelności.

Podsumowanie

Wszystkie dotychczas wspomniane metody szacowania rzetelności testu dotyczą sprawdzenia, czy test bądź jego część (a w przypadku alfy Cronbacha pojedyncze pytania) dają podobne uszeregowania pracowników pod względem posiadanej wiedzy. Jednak w pomiarze wiedzy nie chodzi jedynie o uszeregowanie ludzi (pod względem poziomu wiedzy), lecz także o porównanie ich wiedzy z zewnętrznymi kryteriami (np. z wymaganiami niezbędnymi, aby być skutecznym pracownikiem czy też uzyskać dany certyfikat). W tym wypadku w powtórnych badaniach chodzi nie o podobne uporządkowanie osób wypełniających test, lecz o to, czy na jego podstawie zostanie sformułowana ta sama decyzja (np. stwierdzenie posiadania określonego minimum wiedzy i przyznanie lub nieprzyznanie certyfikatu, zaświadczenia itp.).

UWAGA CZYTELNICY

Trzecia część cyklu będzie poświęcona metodzie tzw. wyznaczania benchmarków (punktów odniesienia) dla omawianych testów wiedzy.