

W jaki sposób, spośród kilku testów przeznaczonych do badania określonej zmiennej (np. inteligencji, odporności na stres, temperamentu), wybrać najlepszy?

Aleksandra Jaworowska

W Polsce nie istnieje żaden system certyfikowania czy oceniania testów. Recenzje testów dostępne w czasopismach psychologicznych można policzyć na palcach jednej ręki. Oznacza to, że psycholog, który chce uzyskać odpowiedź na postawione w tytule pytanie (lub pytania podobne, np. czy dany test nadaje się do określonego celu), musi umieć samodzielnie ocenić test i jego przydatność do planowanych zastosowań. Jak to zrobić?

Rozważając przydatność testu, warto wziąć pod uwagę następujące kwestie i odpowiedzieć sobie na związane z nimi pytania:

1. Cel badania – do czego potrzebny jest nam wynik testu? Na jakie pytania chcemy odpowiedzieć na podstawie wyników badania testowego? Innego testu inteligencji będziemy poszukiwać, gdy chodzi nam o wykluczenie upośledzenia umysłowego jako przyczyny trudności ucznia w czytaniu, a innego – gdy celem badania jest pomoc rodzicom w określeniu mocnych i słabych stron ich dziecka.
2. Przedmiot pomiaru – co chcemy mierzyć? Wydaje się oczywiste, że psycholog to wie (np. chce badać samoocenę, odporność na stres, inteligencję), ale z reguły tak ogólne określenie przedmiotu pomiaru nie wystarcza. Innych testów będziemy poszukiwać, gdy chcemy mierzyć inteligencję płynną, a innych – gdy chodzi nam o pomiar inteligencji skryzalizowanej. W zależności od celu badania istotny może być pomiar jednej bądź drugiej.
3. Populacja – kogo będziemy badać? Dzieci czy dorosłych? Dorosłych z wyższym wykształceniem, czy też z wykształceniem co najwyżej średnim?

Poszukiwania odpowiedniego testu rozpoczynamy dopiero wtedy, gdy znamy odpowiedź na postawione wyżej pytania. Ale jak zorientować się, czy dany test odpowiada naszym potrzebom? I czy jest dobrym testem? Aby móc to ocenić, niezbędna jest lektura podręcznika testowego. W podręczniku szukamy informacji na temat:

1. Założeń teoretycznych leżących u podstaw danego testu – mówią nam one, jak autor testu rozumie zmienną będącą przedmiotem pomiaru. Nazwa zmiennej jeszcze nam tego nie mówi, bowiem to samo określenie może być używane na oznaczenie różnych konstruktów. Na przykład inteligencja emocjonalna może być rozumiana jako cecha osobowości (i wtedy pomiar dotyczy dyspozycji motywacyjnych) lub jako zdolność (pomiar dotyczy dyspozycji instrumentalnych). Koncepcja teoretyczna wyznacza ramy dla interpretacji wyniku. Jej znajomość jest niezbędna do oceny, czy test pozwoli zmierzyć, to co chcemy, oraz do tego, by poprawnie zinterpretować wyniki.
2. Rzetelności, czyli dokładności, z jaką będziemy mierzyć interesującą nas zmienną. Jeśli test ma niską rzetelność, czyli mierzy mało dokładnie, jedynie z dużym przybliżeniem pozwala coś powiedzieć o nasileniu danej zmiennej, to ewentualnie można go wykorzystać do porównywania grup (np. chcąc stwierdzić czy poziom zdolności matematycznych jest taki sam w klasach A i B), natomiast na pewno nie będzie się nadawał do wykorzystania w diagnozie indywidualnej. Konsekwencją niskiej rzetelności testu jest bowiem niska rzetelność indywidualnego wyniku sprawiająca, że przedział ufności, czyli przedział, w jakim – z określonym prawdopodobieństwem – mieści się prawdziwy wynik jednostki, jest bardzo szeroki.

3. Trafności testu – korelacje danego testu z innymi testami mierzącymi podobne konstrukty, a także rezultaty analiz czynnikowych dostarczą nam bliższych informacji na temat mierzonej zmiennej i jej charakterystyki. Ale z praktycznego punktu widzenia najważniejsze będą informacje dotyczące trafności kryterialnej (prognostycznej, diagnostycznej) mówiące o związku wyników testu z jakimś zewnętrznym kryterium (zachowaniem lub zjawiskiem, które chcemy diagnozować lub prognozować). I w tym miejscu warto przypomnieć, że nie ma testów ogólnie trafnych, czyli takich, które są trafne dla różnych zastosowań, sytuacji lub grup osób. Trafność zawsze dotyczy konkretnego zastosowania. Oceniając przydatność konkretnego testu, szukamy więc w podręczniku informacji potwierdzających jego użyteczność dla celu, w jakim chcemy go zastosować. Konkretny test może być trafnym narzędziem oceny np. pamięci fonologicznej, ale jeśli nie będzie różnicował dzieci z dysleksją i dzieci nie mających trudności w opanowaniu czytania, to do diagnozowania dysleksji nie będzie się nadawał.
4. Norm – normy dostarczają nam informacji o rozkładzie wyników testowych w grupie odniesienia, i służą do oceny miejsca, jakie zajmuje nasz badany na tle tejże grupy. Powinny więc być opracowane na podstawie wyników badania grupy, której charakterystyka demograficzna odpowiada charakterystyce osób, jakie mamy zamiar badać. Jeśli będziemy badać testem osobowości tylko (lub głównie) mężczyzn, a normy zostały opracowane na podstawie wyników uzyskanych w badaniu studentek, to takie normy będą mało użyteczne, ponieważ pozwolą nam ocenić jedynie to, jak ci mężczyźni wypadają na tle kobiet (a nie na tle innych mężczyzn). W przypadku testów inteligencji ważne jest też, by normy były aktualne, tzn. pochodziły z badań zrealizowanych nie więcej niż przed kilku laty. Tzw. efekt Flynna opisujący zjawisko starzenia się norm w testach inteligencji, wskazuje na systematyczny wzrost wyników testowych wraz z upływem lat. Normy stare będą więc nietrafnie odzwierciedlać miejsce badanego na tle jego grupy odniesienia.

Tekst opracowany na podstawie: A. Jaworowska (1996). Co to jest test psychologiczny? W: A. Ciechanowicz, A. Matczak (red.): Testy, prawo, praktyka. Warszawa: Pracownia Testów Psychologicznych PTP.