

WŁODZIMIERZ ROSTOCKI

MOŻLIWOŚCI I SPOSOBY OCENY STOPNIA TRUDNOŚCI PYTAŃ KWESTIONARIUSZA

Artykuł niniejszy dotyczy elementarnej jednostki kwestionariusza, jaką jest pytanie. Przedmiotem analizy są pytania zwane ogólnie trudnymi. W analizie wykorzystano wyniki badań empirycznych we Włocławku w 1973 r.

Celem artykułu będzie uzyskanie odpowiedzi na następujące problemy:

1. Co sprawia, że poszczególne pytania są trudne dla respondenta, i jakie są konsekwencje stosowania takich pytań w badaniach socjologicznych?
2. Czy istnieją możliwości oceny stopnia trudności pytań — jeśli tak, to jakie?
3. Czy zwłaszcza można określić stopień trudności pytania wyłącznie na podstawie analizy samego pytania?

Podjęcie problemu pytań trudnych wymagało jednego podstawowego założenia, orzekającego, iż trudność pytania jest cechą stopniowalną, dającą się wyrazić na skali co najmniej porządkowej.

1. Pojęcie i miara stopnia trudności pytania

Przez trudność pytania rozumie się tutaj prawdopodobieństwo wystąpienia zakłóceń w procesach psychicznych respondenta, prowadzących do uformowania prawidłowej odpowiedzi. Zakłócenia mają swą przyczynę w jego ograniczonych możliwościach intelektualnych.

Definicja ta jest podobna do definicji trudności zadania testo-

wego, występującej w psychologii¹. Jej zastosowanie nie jest łatwe. Zakłada ona bowiem, że można ustalić, jakie są procesy psychiczne respondenta, które prowadzą do odpowiedzi przynoszącej prawdziwą informację. Walorem proponowanej definicji jest natomiast to, iż wynika z niej miara stopnia trudności pytań, możliwa do zastosowania w pewnych typach badań. Miara ta będzie procent ujawnionych zakłóceń w prawidłowym przebiegu reakcji wewnętrznej respondentów na pytanie kwestionariusza, mających swe źródło w ich możliwościach intelektualnych. Wykorzystując tę miarę należy zatem dla określenia stopnia trudności pytania obliczyć odsetek osób, u których wystąpiły owe zakłócenia.

W definicji odwołano się do pojęcia odpowiedzi prawidłowej — a więc trafnej, oraz procesów prowadzących do tej odpowiedzi, których wystąpienie zakłada badacz. Posługiwanie się tą definicją implikuje więc wykorzystanie modelu reakcji wewnętrznej respondenta, w którym charakteryzuje się te procesy.

Zastosowanie definicji i związanej z nią miary trudności pytania w badaniach wrocławskich było możliwe dlatego, iż wiele wysiłku włożono w to, aby zbudować odpowiednie modele procesów założonych dla każdego pytania oraz ujawnić zakłócenia, które wystąpiły w przebiegu reakcji wewnętrznej respondentów. W tym celu wykorzystano wywiad o wywiadzie przeprowadzony w odniesieniu do każdego z siedmiu wybranych pytań z zasadniczego kwestionariusza.

Przyjmując wyprowadzoną z definicji miarę stopnia trudności należy zdać sobie sprawę z dość istotnych trudności wiążących się z takim postępowaniem. Są to trudności dwojakiego rodzaju.

Po pierwsze, w procesie rozumienia pytania i formowania odpowiedzi wystąpić mogą zakłócenia różnego typu, a nie tylko takie, które są wywołane małymi możliwościami intelektualnymi

¹ „Fakt, że znikomy procent osobników w danej populacji potrafi wykonać zadanie zadowalająco, bywa niekiedy uznawany za obiektywną miarę trudności. Trudnym problemem jest więc taki problem, którego rozwiązanie dla pojedynczego osobnika jest stosunkowo mało prawdopodobne” (E. D. Berlyne, *Struktura i kierunek myślenia*, Warszawa 1969, s. 349).

respondenta. W analizie bowiem przyczyn zakłóceń nie można pominąć zakłóceń innej natury, zwłaszcza wywołanych przez emocje². Ze względów praktycznych przyjęto tutaj jako miarę stopnia trudności pytań liczbę wszelkich zakłóceń. Uzasadnienie takiego postępowania stanowił fakt, że przy udzielaniu odpowiedzi na pytania z kwestionariusza wrocławskiego inne, zwłaszcza emocjonalne przyczyny występowały bardzo rzadko³. Podstawowe znaczenie miały zakłócenia wywołane niskimi w stosunku do wygórowanych wymagań stawianych przez pytanie możliwościami intelektualnymi respondenta.

Po drugie, stosowana przez nasz zespół metoda ujawniania zakłóceń bynajmniej nie jest metodą niezawodną. Reakcja wewnętrzna respondenta na pytanie kwestionariusza jest najczęściej niesłyszane złożona i dokładne jej zbadanie, jeśli jest w ogóle możliwe, wymagać by musiało znacznie bardziej subtelnych i wyrafinowanych metod niż wykorzystywany wywiad o wywiadzie. Liczba stwierdzonych przez nas zakłóceń dla każdego pytania jest więc zawsze wartością przybliżoną.

² Zakłócenia o charakterze emocjonalnym mogą być wywołane drażliwością pytania. Przez pytanie drażliwe rozumie się tutaj pytanie, które ze względu na swoją treść wywołuje u respondenta emocje takie, jak wstyd, strach, obawa przed zmniejszeniem prestiżu w oczach ankietera czy chęć jego podwyższenia. Wśród naszych pytań za drażliwe można uznać tylko pytanie 21. Pewne pytania dla części badanych mogą być jednocześnie trudne i drażliwe. Poza tym osoby nie umiejące odpowiedzieć na pytanie mogą czuć się taką sytuacją zdenerwowane i speszzone, co nie pozostaje rzeczą jasną bez wpływu na stopień ich zaangażowania w dalszej części wywiadu.

³ Wśród zakłóceń, których źródła dało się ustalić, tylko przy odpowiadaniu na pytanie 17 wystąpiły zakłócenia z przyczyn emocjonalnych w bardzo małej liczbie przypadków (4 przypadki). Pytanie 21, z kolei najbardziej narażone, naszym zdaniem, na drażliwość, okazało się najłatwiejsze ze wszystkich. Oczywiście nie można być pewnym, że przy odpowiadaniu na inne pytania zakłócenia emocjonalne nie wystąpiły. Można jednak z dużym prawdopodobieństwem przyjąć, że jeśli tak było, to liczba zakłóceń tego rodzaju była zbliżona we wszystkich pytaniach. Nie zmienia to ich kolejności, która jest dla nas ważna ze względu na wykorzystywane statystyczne miary.

Aby wyjaśnić bliżej charakter ujawnianych zakłóceń, posłużymy się kilkoma przykładami z przeprowadzonych wywiadów o wywiadzie.

Ankieter w wywiadzie o wywiadzie pyta o to, co respondent rozumie przez słowo „konflikt”. Respondent odpowiada:

„Konflikt... Pomoc młodym czy pomoc młodszych starszym no... itd. No konflikt... Powiedzmy, jakieś wspólne zrozumienie, pomoc wzajemna... Konflikt... Jest to nic innego jak wspólne dążenie do czegoś” (wywiad nr 65, pytanie 16).

Respondent odpowiada na pytanie ankietera, dlaczego udzielił takiej odpowiedzi, a nie innej:

„Odpowiedziałem tak, bo nie rozumiem tego pytania” (wywiad nr 56, pytanie 22).

Ankieter pyta o to, co respondent rozumie przez termin „pochodzenie społeczne”. Respondent odpowiada:

„Pochodzenie społeczne to jak ktoś chodzi do pracy społecznej, pracuje społecznie” (wywiad nr 52, pytanie 24).

Odpowiedź na pytanie, dlaczego respondentka udzieliła takiej odpowiedzi, a nie innej:

„Nie zrozumiałam dobrze pytania... Nie zrozumiałam tego: mniejsze, większe, równe [...] Było za długie” (wywiad nr 7, pytanie 14).

Ankieter pyta respondenta o to, jak rozumie on słowo „solidarność”. Respondent odpowiada:

„Znaczy się solidarność... solidarność w stosunku do pracy [...] Jeżeli mają pracować przykładowo, na jakimś odcinku pracy, na pracy zależy tak czy inaczej. Taki majster czy brygadzysta zostawi ich i powie: No słuchajcie, to i to trzeba zrobić. Jeżeli jest przy nich, pracują, a jeżeli nie, to idą sobie gdzieś, czy na piwo, czy idą spać” (wywiad nr 19, pytanie 19).

Ankieter pyta o to, co respondent rozumie przez zwrot „polityka kadrowa”. Respondent odpowiada:

„Polityka — to mnie się wydaje, że to jest tak, że kadry mają swoje zdanie i pracownicy swoje” (wywiad nr 2, pytanie 22).

We wszystkich tych przypadkach uznano, że proces formowania odpowiedzi na pytanie był zakłócony. Wybrane przykłady do-

tyczą rozumienia pytań, gdyż zakłóceń w tym zakresie było najwięcej.

Na zakończenie rozważań w tej części przykład używania zaproponowanej miary. Współczynnik trudności pytania dla danej zbiorowości, wynoszący 0,8, oznacza, iż u 80% osobników, pochodzących z tej zbiorowości, w ich reakcji wewnętrznej na pytanie kwestionariusza ujawnione zostały zakłócenia. Pytanie z trudnością określoną na poziomie 0,80 jest pytaniem trudniejszym od innego, którego trudność wynosi dajmy na to, 0,30 (u 30% osób stwierdzono zakłócenia).

2. Konsekwencje i przyczyny trudności pytań

Wykorzystując empiryczną miarę stopnia trudności dokonać można uporządkowania pytań od najtrudniejszego do najłatwiejszego. Wyniki przedstawia tabela 1.

Tabela 1. Pytania kwestionariusza włocławskiego, uporządkowane według stopnia trudności

Numer pytania	Liczba osób, u których stwierdzono zakłócenia	% osób, u których stwierdzono zakłócenia	Ranga
16	94	77,0	1
22	92	75,4	2
14	87	71,3	3
24	57	46,7	4
19	56	45,9	5
17	51	41,8	6
21	6	4,9	7

Z tabeli wyraźnie wynika, że łatwe było tylko pytanie 21, pozostałe 6 pytań należy zaliczyć do pytań bardzo trudnych (17, 19, 24) lub krańcowo trudnych (pozostałe pytania) dla badanej zbiorowości.

Tak duży procent ujawnionych zakłóceń musiał mieć określone konsekwencje dla liczby prawidłowych odpowiedzi, uzyskiwanych

na te pytania. Uporządkowanie (szereg rangowy) ustalone według rosnącej liczby prawidłowych odpowiedzi przedstawia tabela 2.

Tabela 2. Pytania kwestionariusza wrocławskiego, uporządkowane według rosnącej liczby trafnych odpowiedzi

Numery pytania	Liczba osób, których odpowiedzi zostały uznane za prawidłowe	% odpowiedzi uznanych za prawidłowe	Ranga
16	9	7,4	1
14	11	9,0	2
22	14	11,5	3
17	41	33,6	4
19	43	35,3	5
24	46	37,7	6
21	95	77,5	7

Korelacja rang Spearmana między szeregiem rangowym przedstawionym w tabeli 1 i szeregiem rangowym z tabeli 2 wynosi 0,83. Wartość tego współczynnika przy $n=7$ pozycji rangowych jest istotna na poziomie $p<0,05$ ⁴.

Wynik ten wymaga kilku zdań wyjaśnienia. Interpretując otrzymaną zależność można stwierdzić, że w przypadku zastosowania pytań trudnych do badań przeprowadzanych w populacji młodych robotników trudność była główną przyczyną małej liczby prawidłowych odpowiedzi. Pytania trudne okazały się więc pytaniami bardzo mało efektywnymi w stosunku do tej grupy respondentów.

Stwierdzenie to nie stanowi żadnego odkrycia, ponieważ prawidłowość odpowiedzi oceniono na podstawie zakłóceń. Objaśnienia wymaga więc raczej to, iż korelacja nie jest jeszcze wyższa. Wiąże się to z faktem, iż w pozycjach rangowych kilku pytań w

⁴ Istotność statystyczną wartości współczynnika korelacji rang Spearmana określano na podstawie tablicy L zamieszczonej w podręczniku J. R. Guilforda, *Podstawowe metody statystyczne w psychologii i pedagogice*, Warszawa 1964, s. 550; por. także G. Clauss, M. Ebner, *Podstawy statystyki dla psychologów, pedagogów i socjologów*, Warszawa 1972, s. 287.

dwu przedstawionych powyżej hierarchiach, według empirycznej miary stopnia trudności oraz według liczby prawidłowych odpowiedzi wystąpiły przesunięcia. Wywołane zostały one różnicami w efektywności wywiadu o wywiadzie w wykrywaniu zakłóceń w odniesieniu do różnych pytań. W niektórych pytaniach (np. 17) procent nie otrzymanych informacji o różnych rodzajach zakłóceń był większy niż w pytaniach pozostałych. Przypadków tych nie uwzględniano przy zliczaniu zakłóceń, natomiast odpowiedzi na te pytania nie mogły być uznane za prawidłowe.

Poszukując wyjaśnienia trudności pytań przyjęto, iż wywołują je przyczyny dwóch rodzajów, które zawsze występują łącznie. Są nimi, z jednej strony, cechy samego pytania, a z drugiej, cechy osób odpowiadających na to pytanie⁵. Ponieważ szczegółowa analiza cech pytań przeprowadzona będzie w dalszej części niniejszej pracy, przeto w tej części ograniczono się do próby wykazania powiązań między stopniem trudności pytania a jedną z ważnych tutaj cech respondenta, jaką stanowi umiejętność rozumienia słów, mierzona wynikiem odpowiedniego testu⁶.

Zależność tę badano w uwikłaniu, biorąc za podstawę następujące rozumowanie. Jeśli dane pytanie jest tym trudniejsze dla respondenta, im mniej bogaty jest jego zasób słów, to respondenci dysponujący małym zasobem słów powinni rzadziej udzielać prawidłowych odpowiedzi na pytania trudne.

Dla zweryfikowania tej hipotezy podzielono osoby badane na takie, które uzyskały wysoki oraz niski wynik w teście słownikowym. Punkt podziału zbliżony był do mediany. Wśród 7 pytań w kwestionariuszu za trudne uznano wszystkie z wyjątkiem py-

⁵ Podobne stanowisko zajmuje się w psychologii w kwestii trudności zadań testowych. „Ogólnie biorąc trudność zadania D jest funkcją jego złożoności K i odpowiedniej przeciętnej sprawności funkcjonalnej S , osoby rozwiązującej to zadanie lub populacji osób rozwiązujących to zadanie” — pisze Z. Żimny, *Pomiary dynamiki funkcji intelektualnych*, [w:] *Problemy psychologii matematycznej*, red. J. Koziebecki, Warszawa 1973, s. 242.

⁶ B. Dudek, R. Kietliński, *Zastosowanie testów psychologicznych do oceny osób odpowiadających na pytania w wywiadzie kwestionariuszowym*, w niniejszym tomie.

tania 21. Osoby o wynikach wysokich w teście rozumienia udzielały średnio 2,23 prawidłowych odpowiedzi na pytania trudne, zaś osoby o niskim wyniku — 0,31. Różnica między średnimi jest istotna statystycznie na poziomie $p < 0,001$.

Następna analiza dotycząca tej samej kwestii oparta jest na podobnym rozumowaniu. Jeśli dane pytanie jest tym trudniejsze dla respondenta, im mniej bogaty jest jego zasób słów, to powinna wystąpić pozytywna zależność między stopniem trudności pytania dla całej grupy a przeciętnym wynikiem w teście rozumienia słów u osób, które odpowiedziały prawidłowo (a więc trafnie) na to pytanie. Innymi słowami, im trudniejsze pytanie, tym większy powinien być zasób słów respondenta, aby mógł on dać sobie z nim radę.

Aby sprawdzić tę hipotezę obliczono średnie arytmetyczne wyników w teście słownikowym tych osób, które odpowiedziały prawidłowo na dane pytanie kwestionariusza. Otrzymano w ten sposób kolejne uporządkowanie pytań.

Tabela 3. Hierarchia pytań według średniej arytmetycznej osiągnięć w teście słownikowym osób, które udzieliły odpowiedzi uznanej za prawidłową

Numer pytania	Średni wynik w teście	Ranga
16	21,22	1
22	20,64	2
19	18,86	3
24	18,48	4
14	17,82	5
17	15,20	6
21	14,46	7

Otrzymane uporządkowanie pytań poddano analizie w celu ustalenia stopnia zbieżności z hierarchią pytań według miary empirycznej (por. tabela 1). Wartość współczynnika korelacji Spearmana dla obu szeregów rangowych wyniosła 0,86, co przy $n=7$ pozycjach rangowych jest istotne na poziomie $p < 0,05$.

Otrzymane wyniki dowodzą, że liczba prawidłowych odpowiedzi na pytania trudne zależy w dużym stopniu od natężenia cechy, którą stanowi dysponowanie większym lub mniejszym zasobem słów. Osoby o większym zasobie słów, rozumiejące więcej pojęć, miały znacznie większe szanse udzielenia prawidłowej odpowiedzi niż osoby o bardzo ograniczonym zasobie słów, ponieważ pytania te były dla nich mniej trudne. Relatywnie duża sprawność werbalna, obejmująca czynny słownik danej jednostki, oraz umiejętność korzystania z niego w określonych sytuacjach spowodowały, że liczba zakłóceń w prawidłowym przebiegu reakcji wewnętrznej respondenta na pytania była mała.

Operowanie małym zasobem słów nie jest oczywiście cechą wrodzoną danego respondenta i wiąże się z kolei z całym szeregiem czynników. Wśród nich warto zwrócić uwagę na poziom wykształcenia. W badanej zbiorowości znalazły się osoby z wykształceniem podstawowym, zasadniczym zawodowym i średnim. Dla grup osób wyróżnionych według kryterium wykształcenia obliczono średnie wyniki w teście słownikowym. Otrzymane wartości średnich były następujące: 10,29 — dla osób z wykształceniem podstawowym, 13,30 — dla osób z wykształceniem zasadniczym zawodowym i 19,81 — dla podzbiorowości z wykształceniem średnim⁷. Różnice między otrzymanymi wynikami są statystycznie bardzo istotne ($p < 0,01$ i $p < 0,001$).

Zależność między poziomem wykształcenia a wynikiem w teście starano się ustalić również w inny sposób. Obliczono mianowicie h — współczynnik dla skal porządkowych Goodmana-Kruskala⁸. Wartość tego współczynnika okazała się bardzo duża. Wyniosła ona 0,85⁹.

⁷ Liczebność grup wykształcenia przedstawiała się następująco: 49 osób z wykształceniem podstawowym, 42 osoby z zasadniczym zawodowym i 31 osób z wykształceniem średnim.

⁸ A. Siciński, *Współczynniki zależności zmiennych niemierzalnych*, [w:] *Metody statystyczne w socjologii*, red. K. Szaniawski, Warszawa 1968, s. 243 - 248.

⁹ W celu obliczenia tego współczynnika podzielono respondentów na dwie kategorie według wyniku w teście (wysoki, niski, podział wg mediany) i trzy podane wyżej kategorie wykształcenia.

Jak z tego widać, zasób słów zależy od wykształcenia, które pośrednio oddziałuje więc na to, jakie i ile pytań sprawia trudność respondentowi. Pośrednio o występowaniu tej oczywistej zresztą zależności świadczy także występowanie wyraźnego związku między wykształceniem a liczbą prawidłowych odpowiedzi¹⁰.

Rola zasobu słów jako cechy respondenta wiąże się ściśle ze stopniem trudności językowej jako cechy pytania. Do sprawy tej przejdziemy w dalszym ciągu artykułu.

3. Metody oceny stopnia trudności pytań

W socjologicznym procesie badawczym problem przewidywania stopnia trudności pytań dla określonej grupy respondentów ma bardzo duże znaczenie. Przeprowadzona przed właściwymi badaniami umiejętna ocena stopnia trudności pytań pozwala uniknąć później zarówno odpowiedzi nieistotnych, jak i nietrafnych, a więc braku informacji i informacji błędnych. Badaczowi przeto powinno zależeć na tym, aby ocenę stopnia trudności przeprowadzić jak najwcześniej, jak najdokładniej i oczywiście jak najprościej.

a) Liczba odpowiedzi nieistotnych na pytanie jako miara jego stopnia trudności

Często w pilotażach socjologicznych wykorzystuje się jako prostą miarę stopnia trudności liczbę odpowiedzi nieistotnych, np. odpowiedzi „nie wiem”, „nie rozumiem”¹¹. Miarę tę zastosowano również w analizowanych badaniach. Wyniki przedstawia tabela 4.

¹⁰ Dla każdej z trzech grup wykształcenia obliczono średnie arytmetyczne udzielanych prawidłowych odpowiedzi na 6 pytań trudnych. Osoby z wykształceniem podstawowym udzielały średnio 0,7 prawidłowej odpowiedzi, z zasadniczym zawodowym 1 prawidłową odpowiedź, zaś osoby z wykształceniem średnim 2,55 prawidłowych odpowiedzi. Bardzo istotne statystycznie były różnice w średnich prawidłowych odpowiedzi między respondentami z wykształceniem średnim i respondentami z wykształceniem zasadniczym zawodowym oraz podstawowym — (w obydwu wypadkach $p < 0,001$). Różnice w średnich między osobami z wykształceniem podstawowym i zasadniczym zawodowym były nieistotne statystycznie.

¹¹ Wspominają o tym między innymi W. J. Goode i P. K. Hatt w swoim podręczniku *Methods in Social Research*, New York 1952, s. 159.

Tabela 4. Hierarchia pytań według częstości odpowiedzi nieistotnych

Numery pytań	Liczba odpowiedzi nieistotnych	% odpowiedzi nieistotnych	Ranga
14	48	39,3	1
16	46	37,7	2
19	17	13,9	3
17	16	13,1	4
24	13	10,7	5
22	4	3,28	6
21	1	0,82	7

Współczynnik korelacji rang Spaermanna obliczony dla hierarchii pytań według liczby odpowiedzi nieistotnych i hierarchii według empirycznej trudności pytań (por. tabela 1) wyniósł 0,46, co nie jest istotne na poziomie $p < 0,05$. Uzyskana wartość współczynnika korelacji rang dowodzi, iż prognozowanie stopnia trudności pytań jedynie na podstawie liczebności odpowiedzi nieistotnych obarczone jest w przypadku niektórych pytań (np. 22) bardzo dużym błędem.

b) Metoda sędziów

Poszukując nowych i łatwych do stosowania w codziennej praktyce badawczej socjologa sposobów oceny stopnia trudności pytań posłużyliśmy się między innymi opiniami sędziów. 18 pracowników naukowych Instytutu Socjologii UŁ, posiadających doświadczenie badawcze, wystąpiło w roli arbitrów, oceniając stopień trudności siedmiu przedstawionych w pierwszej części pracy pytań kwestionariusza włocławskiego¹².

¹² Każdy sędzia przed oceną pytań znał definicję trudności pytania, przyjętą przez badaczy, definicję trafności i istotności odpowiedzi oraz dysponował szeregiem danych z kart pytań. Dane te były następujące: 1. informacja poszukiwana w pytaniu, 2. alternatywy prezentowane przez pytanie, 3. rozumienie terminów przyjęte przez badacza, 4. założenia empiryczne pytania. Dysponując takim zestawem informacji arbiter przystępował do oceny.

W badaniach zastosowano metodę porównywania parami. Każdy z arbitrów dokonywał więc 21 porównań dwu pytań, jednocześnie dokonując ich oceny. Na podstawie ocen utworzono hierarchię pytań według ich trudności dla każdego sędziego.

Wstępna analiza otrzymanych wyników dowiodła, że oceny sędziów dotyczące stopnia trudności pytań wykazują zgodność statystycznie istotną, czyli, że arbitrzy kierują się w swoich ocenach jakimś kryterium wspólnym. Wartość współczynnika W Kendalla mierzącego stopień zgodności między 18 hierarchiami sędziów wyniosła $W=0,5219$, co jest istotne na poziomie $p<0,001$ ¹³.

Następnie wyniki ocen 18 sędziów uporządkowano, budując kolejną hierarchię stopnia trudności pytań. Ponieważ wyniki oceny każdego sędziego rangowano od najtrudniejszego pytania (ranga 1) do najłatwiejszego (ranga 7), przeto budując globalny szereg rangowy posłużono się kryterium sumy rang uzyskanych przez pytanie w ocenie 18 arbitrów. Pytanie najtrudniejsze miało zatem najmniejszą sumę rang, pytanie najłatwiejsze — największą. Wyniki ocen sędziów prezentuje tabela 5.

Tabela 5. Hierarchia trudności pytań w ocenach 18 sędziów

Numer pytania	Suma rang uzyskanych przez pytanie	Ranga globalna
24	37	1
22	50,5	2
16	56,5	3
19	68	4
14	73	5
17	102,5	6
21	115,5	7

Wartość współczynnika Spearmana, obliczona dla szeregów rangowych według ocen sędziów i według empirycznej miary stop-

¹³ Wykorzystując w pracy współczynnik W Kendalla opierałem się na artykule M. Manturzewskiej, *Zgodność ocen wykonawstwa pianistycznego*, „Biuletyn Psychometryczny”, tom I, Wrocław 1966, s. 111-115. Istotność W obliczano wzorem $x^2=k(n-1)W$, gdzie k to liczba sędziów, tamże s. 113.

nia trudności (tabela 1), wynosi 0,68. Jest to wynik dość wysoki, jednak statystycznie nieistotny. Przy skrajnie małej liczbie pozycji rangowych ($n=7$), aby wynik był istotny na poziomie $p<0,05$, R musi osiągnąć wartość 0,714¹⁴. Zbieżność między dwoma analizowanymi szeregami rangowymi nie jest więc istotna na poziomie $p<0,05$.

Posługując się metodą sędziów osiągnięto tylko połowiczny sukces. Wyjaśnienia braku pełnego sukcesu można szukać w tym, że jak się wydaje, sędziowie oceniali trudność pytania jedynie w zakresie jednego wymiaru, jakim jest trudność językowa, pomijając inne, które zostaną opisane w dalszej części artykułu. Potencjalnych przyczyn trudności pytań dla młodych robotników sędziowie szukali przede wszystkim w ich warstwie słownej, co w przypadku niektórych pytań zawiodło¹⁵. Ogólna jednak ocena proponowanego sposobu określania stopnia trudności pytań, mimo szeregu zastrzeżeń, jakie wywołuje, jest, jak się wydaje, pozytywna. Można przypuszczać, że przy większej liczbie sędziów i większej liczbie pytań odznaczających się bardziej wyraźnym zróżnicowaniem stopnia trudności metoda ta da znacznie lepsze efekty. Trzeba jeszcze zaznaczyć, że jest to metoda dość pracochłonna.

4. Propozycja wielowymiarowej analizy stopnia trudności pytań

W dotychczasowych rozważaniach wskazywano już na to, że traktowanie cechy trudności pytania jako cechy jednowymiarowej nie prowadzi do efektywnych sposobów jej oceny. Bardziej interesujące wydaje się traktowanie trudności pytania jako cechy wielowymiarowej. Umożliwia to zastosowanie innego niż poprzednie sposobu oceny pytania. Sposób ten w omawianym przypadku dał lepsze wyniki od pozostałych.

¹⁴ J. R. Guilford, *op. cit.*, s. 550, tablica L.

¹⁵ W ocenach sędziów przeważało kryterium trudności językowej, jednak w ramach tego kryterium każdy arbiter stosował szereg kryteriów szczegółowych, prowadzących do różnic w ocenach.

Założenia i cechy prowadzonej według niego analizy są następujące:

1. Wymiary trudności pytania ustalone zostały przy wykorzystaniu modelu reakcji wewnętrznej respondenta na pytanie kwestionariusza i odpowiadają przyjętym w tym modelu rodzajom procesów założonych. Są one następujące:

I. Odbiór pytania.

II. Zrozumienie pytania w zakresie stawianego przez nie zadania.

III. Zrozumienie pytania w zakresie zrozumienia sensu alternatyw.

IV. Formowanie stanowiska i odpowiedzi.

Spośród sześciu grup procesów wyróżnionych w modelu nie uwzględniono tu dwóch: akceptacji pytania oraz uznania uformowanego stanowiska za prawdziwe. Jest to zrozumiałe, ponieważ zajęcie ich nie wymaga intelektualnego wysiłku. Natomiast rozumienie pytania rozbito na dwa wymiary ze względu na znaczenie tego procesu¹⁶.

Trudność pytania należy zatem rozpatrywać w zakresie tych procesów, w czterech odpowiadających im wymiarach.

2. W ramach wymiarów konieczne wydawało się wyróżnienie podwymiarów. Oznaczono je w zamieszczonej poniżej liście dużymi literami alfabetu. Wymiary oznaczono cyframi rzymskimi.

3. Zasady punktacji są proste. Jest to punktacja zerojedynkowa. Gdy pytanie odznacza się jakąś cechą wymienioną w wykazie wymiarów i podwymiarów trudności, otrzymuje ono 1 punkt, gdy nie jest wymienione — uzyskuje 0 punktów.

4. Propozycja analizy stopnia trudności pytania odnosi się do pytań wprost, żądających wyboru jednej spośród niewielu alternatyw, na które odpowiedzi są jednocześnie odpowiedziami na pytanie badacza albo stanowią bezpośredni wyraz przekonań respondenta, których dotyczy poszukiwana informacja.

5. Szereg decyzji podejmowanych w tej analizie, odnoszących

¹⁶ Por. B. Tuchańska, *Analiza modelu wewnętrznej reakcji respondenta na pytanie w wywiadzie kwestionariuszowym*, w niniejszym tomie.

się przede wszystkim do punktacji, to decyzje arbitralne, lecz oparte na tzw. zdrowym rozsądku i wykorzystujące doświadczenia ankieterskie autora.

Przejdźmy teraz do przedstawienia listy wymiarów i podwymiarów. Charakteryzują one zarazem strukturę cechy „trudności pytania”.

I Wymiar. Odbiór słuchowy pytania.

A. Trudność wynikająca z długości pytania mierzonej liczbą słów występujących w pytaniu.

Pytania dłuższe niż 20 słów otrzymują 1 punkt.

B. Trudność wywołana skomplikowaną strukturą gramatyczną pytania, polegającą na występowaniu w pytaniu zdań złożonych, w tym zdań wtrąconych.

Gdy w pytaniu występowały zdania złożone, otrzymywało ono 1 punkt.

II Wymiar. Rozumienie zadania stawianego przed respondentem w pytaniu¹⁷.

A. Trudności wynikające z występowania w pytaniu dużej liczby alternatyw, spośród których należy wybrać jedną będącą odpowiedzią na to pytanie.

Przyjęto, że pytania posiadające 4 i więcej alternatyw otrzymują 1 punkt¹⁸.

B. Trudności związane z tym, iż pytanie jest właściwie zbitką dwóch pytań i jego odpowiedzi są odpowiedziami na różne pytania, należące do dwóch poziomów.

¹⁷ W modelu rozumienie zadania określono następująco: „Respondent zidentyfikował prawidłowo alternatywy i zrozumiał, że ma dokonać wyboru jednej z nich. Prawidłowa identyfikacja oznacza, że respondent umiał wskazać właściwe alternatywy jako odrębne, nadając im odpowiednie nazwy (*Ogólny model reakcji wewnętrznej respondenta na pytanie kwestionariusza*, tekst powielony, s. 6).

¹⁸ Dla dokonania wyboru jednej alternatywy metodą porównań trzeba $\frac{n(n-1)}{2}$ porównań między wszystkimi alternatywami prezentowanymi

przez pytanie. Zatem dla dwóch alternatyw konieczne jest dokonanie 1 porównania, dla trzech — 3 porównań, dla czterech — 6 porównań, dla pięciu alternatyw — 10 porównań itp.

Przykład stanowi tutaj pytanie 16 naszego kwestionariusza¹⁹.

Jeśli pytanie stanowi zbitkę dwóch pytań, otrzymuje 1 punkt.

Do trudności związanych z rozumieniem zadania pytania należałoby zaliczyć jeszcze trudność polegającą na tym, iż zadanie przedstawiono w sposób skomplikowany z punktu widzenia logiki. Dzieje się tak wtedy, gdy na przykład w pytaniu występuje podwójne przeczenie. Trudności tego rodzaju nie występowały w naszych pytaniach.

III Wymiar. Rozumienie alternatyw.

A. Trudności będące następstwem występowania w alternatywach nazw i zwrotów relatywnie rzadko używanych w badanej zbiorowości. Chodzi tutaj o terminy takie, jak na przykład: solidarność, pochodzenie społeczne, poziom kulturalny itp. Za występowanie jednego takiego słowa pytanie otrzymuje 1 punkt.

B. Trudności płynące z występowania w pytaniu terminów lub zwrotów, które mogą mieć dla respondenta znaczenie odrębne od przyjętego przez badacza.

Jeśli w pytaniu występuje jedno takie słowo, otrzymuje ono 1 punkt.

C. Trudności będące następstwem używania w pytaniu pojęć abstrakcyjnych, które wywołują kłopoty w identyfikowaniu przez respondenta różnorodnych zjawisk i ich układów, będących desygnatami tego pojęcia²⁰.

¹⁹ Pytanie to zanalizowała B. Tuchańska w artykule *Analiza modelu wewnętrznej reakcji respondenta na pytanie w wywiadzie kwestionariuszowym*, w niniejszym tomie.

²⁰ O występowaniu terminów abstrakcyjnych w kwestionariuszach pisze w sposób następujący Z. Gostkowski: „Często też wersja kwestionariusza przygotowana przez badacza-teoretyka sformułowana jest w kategoriach zbyt abstrakcyjnych, zbyt obcych sposobowi myślenia i rozumowania prostych ludzi, którzy mają odpowiadać na pytania wywiadu czy ankiety. Z reguły badacze-teoretycy, projektujący kwestionariusze, zafascynowani swoją teoretyczną wizją, przeceniają poziom intelektualny i zdolność ro-

Za występowanie jednego takiego terminu pytanie otrzymuje 1 punkt.

IV Wymiar. Formowanie stanowiska i odpowiedzi.

A. Trudności w przypominaniu zdarzeń sprzed dłuższego okresu czasu.

B. Trudności w wyobrażaniu sobie sytuacji hipotetycznych, co jest niezbędne dla uformowania stanowiska i odpowiedzi.

C. Trudności w porównaniu i uogólnianiu większej liczby obserwacji jednostkowych²¹.

D. Trudności w dokonywaniu obliczeń wymaganych przez pytanie.

Za każdą posiadaną cechę wymienioną pod literami A-D pytanie otrzymuje 1 punkt.

Przeprowadzimy obecnie analizę stopnia trudności pytania 19 z kwestionariusza włocławskiego. Pytanie to brzmiało: „Czy Pana zdaniem robotnicy w Celulozie są solidarni, czy nie?”

I Wymiar:

A — 0 punktów. Pytanie jest krótkie — 10 słów.

B — 0 punktów. Budowa gramatyczna pytania jest prosta.

II Wymiar:

A — 0 punktów. W pytaniu są tylko 2 możliwe odpowiedzi.

B — 0 punktów, albowiem możliwe odpowiedzi należą do jednego poziomu.

III Wymiar:

A — 1 punkt: solidarność nie jest terminem używanym w codziennym języku badanych.

B — 1 punkt: możliwe jest rozumienie solidarności jako solidności w ramach całej badanej grupy²².

zumienia pytań przez respondentów” (Z. Gostkowski, *O humanizację badań surveyowych*, „Studia Socjologiczne”, 1974, nr 1).

²¹ Por. rozważania S. Szostkiewicza na temat pytań analitycznych i syntetycznych, zawarte w artykule *Pytania analityczne i syntetyczne w kwestionariuszach do badań społecznych*, „Biuletyn dla Współpracowników”, 1973, nr 1.

²² Wykorzystano tutaj doświadczenia z pilotażu przeprowadzonego w IFiS PAN. Jak się okazało, termin „solidarność” bywa rozumiany jako so-

- C — 1 punkt: solidarność jest pojęciem abstrakcyjnym.
- IV Wymiar:
- A — 0 punktów. Odpowiedź na pytanie nie wymaga przypominania zdarzenia sprzed długiego okresu czasu.
- B — 0 punktów: pytanie nie wymaga wyobrażania sobie sytuacji hipotetycznych.
- C — 1 punkt: dojście do uformowanego stanowiska wymaga uogólnienia dużej liczby jednostkowych obserwacji.
- D — 0 punktów: formułowanie odpowiedzi na pytanie nie wymaga obliczeń.

Łącznie pytanie 19 otrzymało 4 punkty²³.

Taką analizę przeprowadzono dla każdego z siedmiu pytań kwestionariusza. Jej wyniki przedstawia tabela 6.

Tablica 6. Stopień trudności pytań kwestionariusza według wyników analizy wielowymiarowej

Numer pytania	I wymiar		II wymiar		III wymiar			IV wymiar				Liczba punktów
	A	B	A	B	A	B	C	A	B	C	D	
14	1	1	0	0	1	1	1	0	0	1	0	6
16	1	1	0	1	1	1	0	0	0	1	0	6
17	1	0	0	1	0	0	0	0	0	1	0	3
19	0	0	0	0	1	1	1	0	0	1	0	4
21	0	0	0	0	1	0	0	0	0	0	0	1
22	0	0	0	0	1	1	1	0	0	1	0	4
24	0	0	0	0	1	0	1	0	0	1	0	3

Na podstawie ocen punktowych zbudowano kolejną hierarchię stopnia trudności pytań. Przedstawia ją tabela 7.

liłość. Wspomina o tym również K. Lutyńska w artykule *Pilotaż „pogłębiony”. Koncepcja, realizacja i analiza materiałów pilotażowych*, w niniejszym tomie.

²³ Przypisanie pytaniom pewnej liczby punktów nie jest równoznaczne z umiejscawianiem ich na skali interwałowej. Budowana skala ma więc charakter porządkowy.

Tabela 7. Hierarchia trudności pytań według ocen punktowych

Numer pytania	Liczba punktów (ocena)	Ranga
14	6	1,5
16	6	1,5
19	4	3,5
22	4	3,5
17	3	5,5
24	3	5,5
21	1	7

Jaki jest stosunek tej hierarchii do innych, budowanych poprzednio? Obliczono korelację rang między hierarchią według wyników analizy wielowymiarowej i hierarchią trudności, ustaloną empirycznie na podstawie liczby zakłóceń. Współczynnik korelacji *R* Spearmana osiągnął wartość 0,84, która jest istotna na poziomie $p < 0,05$ i zbliża się do $p < 0,01$. Korelacja rang między uporządkowaniem według wyników analizy i kolejnością pytań według liczby prawidłowych odpowiedzi wynosi 0,91, co jest istotne na poziomie $p < 0,01$.

Otrzymane wartości współczynnika *R* uzasadniają przekonanie o przydatności stosowanej analizy do oceny stopnia trudności pytań. Między uporządkowaniem pytań według empirycznej miary trudności i uporządkowaniem według ostatnio zastosowanej miary istnieje duża zbieżność. Nie popełnimy dużego błędu przewidując miejsce w hierarchii trudności pytań, ustalonej ze względu na kryterium empiryczne na podstawie analizy, którą teraz przeprowadziliśmy.

5. Uwagi końcowe

Z trzech przedstawionych w drugiej części niniejszego artykułu sposobów oceny stopnia trudności pytań najbardziej efektywny okazał się sposób ostatni — analiza samego pytania prze-

prowadzona przy uwzględnieniu czterech wymiarów jego trudności. Trzeba tu zaznaczyć, że zasady punktacji przyjęte w tej metodzie w głównych zarysach przygotowano na początku analizy, przed obliczeniem korelacji. Po dokonaniu pierwszych obliczeń zasady te zostały tylko nieznacznie zmienione. Zmiany zwiększyły nieco wysokość współczynnika korelacji rang.

W podjętej próbie dokonano oceny stopnia trudności pytań bez odwoływania się do wyników specjalnie przeprowadzonych badań. Omawiany sposób zakłada jednak wykorzystanie jakichś doświadczeń, zwłaszcza w zakresie wymiaru III A i B. Doświadczenia te mogą, a nawet powinny być zastąpione przez badania. Najlepiej byłoby, aby były to badania przeprowadzone nie *ad hoc*, na przykład w pilotażu danego pytania, lecz wcześniej, tak aby ich wyniki mogły być wykorzystane przy układaniu pytań.

Omawiając dodatnie cechy proponowanej metody nie można jednak pominąć jej dość istotnych ograniczeń. Okazała się ona efektywna w odniesieniu do bardzo małej liczby specjalnie skonstruowanych pytań, co nie oznacza, że będzie użyteczna w przypadku pytań innych. To oczywiście można sprawdzić. Zapewne jednak w ewentualnych dalszych analizach ulegnie zmianie także przyjęty tutaj system punktacji, który zresztą wymaga dyskusji i w swojej obecnej postaci.

JERZY KONIAREK

WERYFIKACJA ZEWNĘTRZNA INFORMACJI UZYSKANYCH W WYWIADZIE KWESTIONARIUSZOWYM

Analizy weryfikacyjne bywają najczęściej przeprowadzane przy okazji badań, których zasadnicze cele są inne niż metodologiczne, wtedy gdy badacz prowadzący te badania chce mieć rozeznanie co do stopnia wiarygodności informacji uzyskanych we właściwych badaniach. Rzadsze są sytuacje, gdy badania od początku są planowane i realizowane jako badania o wyłącznie metodologicznych celach. Artykuł, który jest jednym z kilku zamieszczonych w tym tomie, przedstawia częściowe wyniki takich właśnie metodologicznych badań. Omawia on rezultaty weryfikacji różnych danych uzyskanych w wywiadzie kwestionariuszowym przeprowadzonej dwoma odmiennymi sposobami. Wspólnie dla obu sposobów było to, że porównywano informację zawartą w odpowiedzi respondenta na pytanie kwestionariusza z tą samą informacją uzyskaną inną drogą. Tak więc oba sposoby weryfikacji, przedstawione w artykule, należą do tego typu weryfikacji, którą J. Lutyński proponuje nazwać zewnętrzną¹.

Część pierwsza artykułu poświęcona będzie omówieniu metody i wyników weryfikowania siedmiu pytań zasadniczych kwestionariusza, z których sześć było pytaniami o opinię. Weryfikacja w tym przypadku przeprowadzona była za pomocą wywiadu pogłębianego. W części drugiej przedstawione zostaną wyniki we-

¹ J. Lutyński, *Analizy weryfikacyjne w badaniach z zastosowaniem wywiadu kwestionariuszowego, ich rodzaje i możliwości*, w tym tomie.