

wę tę jedynie sygnalizujemy, jako wymagającą głębszej analizy.

Projektując dalsze badania tego typu należy się również zastanowić nad ograniczeniem materiałów do prowadzenia wywiadu o wywiadzie. Duża liczba narzędzi badawczych przesądza o ich małej operatywności. Wydaje się, iż korzystne byłoby wyeliminowanie kart pytań i opracowanie dla każdego pytania kwestionariusza, będącego przedmiotem wywiadu o wywiadzie, graficznych schematów toku wywiadu, w których kolejne poszukiwane informacje byłyby ostatecznie skonkretyzowane. Niecelowe wydaje się również przekazywanie ankietom wzorów pytań do wywiadu o wywiadzie.

Szczególnie staranny powinien być dobór ankietów do tego typu wywiadów. Najlepiej byłoby zaangażować ankietów o dużym doświadczeniu w prowadzeniu wywiadów swobodnych. Ankietery ci winni odznaczać się wysokim poziomem intelektualnym oraz zdolnościami do samokontroli i autokrytycyzmu. Należy uczynić wszystko, co tylko jest możliwe, by wzbudzić w nich intelektualne i emocjonalne zaangażowanie problematyką badań.

Ocena wywiadu o wywiadzie jako metody uzyskiwania informacji o procesie formowania odpowiedzi przez respondenta dotyczy w pierwszym rzędzie jego operatywności, tj. możliwości uzyskania rezultatów zgodnie z przyjętymi celami i założeniami. Ponieważ metoda ta oparta jest na introspekcji, podlega wielu zarzutom, które tutaj tylko sygnalizowaliśmy, przedstawiając nasze przypuszczenia dotyczące faktycznego przedmiotu introspekcyjnych wypowiedzi i możliwości ich interpretacji zgodnie z celami badań. Przytoczona ocena tej metody nie jest więc uzasadniona w pełni. Wydaje się jednak, że wykazanie operatywności jakiegoś sposobu postępowania ma także znaczenie, nawet jeśli nie można — przynajmniej w danej chwili — ocenić go w pełni.

denci o wysokim wyniku w teście słownikowym niż respondenci o niskim wyniku. Współczynnik ten kształtuje się na poziomie $-0,12$, co raczej przemawia przeciw słuszności powyższej hipotezy.

BOHDAN DUDEK
RYSZARD KIETLIŃSKI

ZASTOSOWANIE TESTÓW PSYCHOLOGICZNYCH DO OCENY OSÓB ODPOWIADAJĄCYCH NA PYTANIA W WYWIADZIE KWESTIONARIUSZOWYM

Cztery drogi może wybrać badacz, aby ustalić swój stopień zaufania do wyników uzyskanych w badaniach kwestionariuszowych.

1. Badacz może dążyć do stworzenia optymalnej sytuacji badawczej, tzn. może tak manipulować sytuacją wewnętrzną i zewnętrzną respondenta, aby ten był skłonny i zdolny do udzielania trafnych odpowiedzi¹. Taki sposób postępowania jest rozpowszechniony w socjologii jeśli nie zawsze w praktyce badawczej, to na pewno w świadomości badaczy. Jest to efekt refleksji metodologicznej, która w głównej mierze nastawiona jest na to, aby ustalić, jaka jest optymalna sytuacja badawcza i co trzeba zrobić, aby ją osiągnąć np. drogą zmniejszenia wpływu ankietarskiego, przeprowadzenia pilotażu w celu ustalenia najlepszego sformułowania pytań w kwestionariuszu, odpowiedniego doboru ankietów oraz ich szkolenia².

2. Badacz może przeprowadzić specjalne badania całej grupy lub wylosowanej próbki, aby stwierdzić, czy w procesie odpowiadania na pytania zaszły szkodliwe zjawiska, które spowodowały udzielenie nietrafnych odpowiedzi. Badania te mogą mieć

¹ Wyjaśnienie terminu „trafna odpowiedź” zob. artykuł J. Lutyńskiego, *Cele, przebieg i wstępne wyniki badań weryfikacyjnych we Włocławku w 1973 r.*, w tym tomie.

² Zob. np. H. H. Hyman, *Interviewing in Social Research*, Univ. Chicago Press 1954.

postać wywiadu o wywiadzie, w którym badacz stara się ustalić, czy badani rozumieli pojęcia wchodzące w skład pytań, czy rozumieli zadanie, jakie było zawarte w pytaniu, czy byli w stanie udzielić odpowiedzi uwzględniając doświadczenie, do jakiego apeluje pytanie, czy nie zaszły jakieś emocjonalne zaburzenia skłaniające badanych do udzielania nietrafnych odpowiedzi itp. W zależności od otrzymanych wyników badacz może korygować lub choćby tylko określić stopień zaufania, jakim może darzyć wyniki uzyskane w badaniach zasadniczych.

3. Badacz może skonfrontować odpowiedzi badanych osób z dostępnymi faktami i ustalić stopień ich rozbieżności. Sposób ten obok wielu zalet posiada wadę polegającą na tym, że można go stosować tylko w ograniczonym zakresie. Niewiele bowiem pytań dotyczy faktów, które badacz mógłby stosunkowo łatwo ustalić.

4. Badacz może zmierzyć odpowiednie cechy respondenta, które mogą mieć wpływ na udzielenie prawidłowej i trafnej odpowiedzi, np. inteligencję (szczególnie werbalną), niektóre cechy osobowości (np. tendencję do pokazywania się w lepszym świetle) itp. Cechy te wpływać mogą na procesy zachodzące w trakcie udzielania odpowiedzi. Jeżeli badacz ustali w sposób empiryczny, że jakaś cecha jest związana z udzielaniem trafnych odpowiedzi na dane pytania, będzie mógł oszacować jakość uzyskanych wyników na podstawie stopnia, w jakim grupa jest nasycona daną cechą.

Niniejszy artykuł poświęcony jest czwartemu sposobowi, który został powyżej skrótowo przedstawiony.

Uznaliśmy, że podstawowymi cechami respondentów, które mogą zaburzyć udzielanie prawidłowych odpowiedzi, a równocześnie stosunkowo łatwo nadają się do mierzenia za pomocą testów, są: zasób słownictwa, tendencja do pokazywania się w lepszym świetle i skłonność do udzielania odpowiedzi na pytania puste³. Można przypuszczać, że respondent o małym zasobie słów

będzie miał trudności ze zrozumieniem pytań oraz adekwatnym sformułowaniem własnych myśli. Druga uwzględniana przez nas cecha powoduje to, że badani starając się pokazać od jak najlepszej strony udzielają (świadomie lub nieświadomie) odpowiedzi akceptowanych społecznie. Dlatego można spodziewać się, że osoby posiadające tę cechę w wysokim stopniu częściej udzielają będą odpowiedzi nietrafnych. Skłonność do udzielania odpowiedzi na pytania puste zgodnie z żądaniem zawartym w pytaniu świadczyć może o tym, że respondent odpowiada bez głębszego zastanawiania się nad treścią pytania, wypowiadając swoje opinie nawet tam, gdzie ich nie posiada.

Do pomiaru wyżej wymienionych cech przygotowaliśmy 3 testy: test znajomości słów (S), test kłamstwa (K) oraz test pytań pustych (P). Wszyscy respondenci, z którymi przeprowadzono wywiad kwestionariuszowy, zostali przebadani tymi testami. Następnie za pomocą wywiadu o wywiadzie i wywiadu weryfikacyjnego ustalono, na które pytania kwestionariusza poszczególni badani udzielali odpowiedzi prawidłowych i trafnych przypadkowo, nietrafnych i nieistotnych⁴. W dalszej części badań przeprowadzono analizę, która miała na celu określenie związków między wynikami w testach a udzielaniem prawidłowych odpowiedzi na pytania kwestionariusza socjologicznego.

1. Przygotowanie testów

Od jakości zastosowanych testów zależało w znacznej mierze powodzenie naszych badań. Dlatego dużo uwagi i pracy poświęciliśmy na ich przygotowanie.

Test słownikowy S oparty był w dużym stopniu na „Teście znajomości słów” opracowanym przez M. Chojnowskiego. Ponieważ zależało nam na tym, aby test nie był zbyt długi (chodziło o krótki czas badania) oraz by zawierał pojęcia występujące

kie pytanie jedyną prawidłową odpowiedzią jest wskazanie błędu zawartego w pytaniu. Por. artykuł P. Daniłowicza, *Refleksje nad stosowaniem pytań pustych w wywiadach kwestionariuszowych*, w tym tomie.

⁴ Wyjaśnienie terminów: „odpowiedź prawidłowa, trafna, trafna przypadkowo, istotna” zob. artykuł J. Lutyńskiego, *op. cit.*

³ Pytanie puste to takie pytanie, które nie spełnia założeń poprawności logicznej lub którego założenia empiryczne nie są spełnione. Na ta-

w pytaniach kwestionariuszy socjologicznych, zrezygnowaliśmy z kilkunastu pytań oryginalnego testu i dołożyliśmy kilka nowych. Test w wersji eksperymentalnej składał się z 40 pytań (*items*). Każde pytanie zawierało na pierwszym miejscu pojęcie, którego znajomość chcieliśmy zbadać, po nim występowały 4 słowa, z których tylko jedno znaczyło to samo lub prawie to samo, co hasło. Zadanie badanych polegało na zakreśleniu właściwej odpowiedzi. Pozostałe słowa pełniły rolę dystraktorów, które utrudniały odgadnięcie rozwiązania. Wynik testowy danej osoby stanowiła liczba poprawnych odpowiedzi.

Test kłamstwa K oparty był na skali kłamstwa z testu Eysencka, który został zaadaptowany do warunków polskich przez M. Choynowskiego⁵, oraz na skali L ze znanego inwentarza osobowości MMPI⁶. Podobnie jak poprzednio wybraliśmy z tych dwóch skal kilkadziesiąt pytań oraz ułożyliśmy nowe, które tematycznie wiązały się z problematyką poruszaną w kwestionariuszu. Pytania w teście kłamstwa dotyczą rzeczy, które zdarzają się prawie każdemu, lecz przyznanie się do nich stawiać może człowiek w niekorzystnym świetle. Np. „Czy chwali się Pan trochę?”, „Czy wyrażał się Pan kiedykolwiek źle o swoich zwierzchnikach?” itp. Osoba, która chce zrobić dobre wrażenie, daje odpowiedź przeczącą. Jeżeli na wiele pytań tego rodzaju odpowie w ten sposób, można uważać, że jej tendencja do pokazania się w lepszym świetle jest wyraźna⁷. Test składał się z 40 pytań. Wynik osoby badanej obliczano przyznając 2 punkty za udzielenia odpowiedzi diagnostycznej (świadczącej o tendencji do pokazania się w lepszym świetle) oraz 1 punkt za zakreślenie „?”, co oznaczało, że badany nie mógł zdecydować się na jednoznaczna odpowiedź. Rezultat stanowiła suma tych punktów.

⁵ M. Choynowski, *Opracowanie polskiej adaptacji „Inwentarza osobowości” H. J. Eysencka „Maudsley Personality Inventory”: Wyniki krajowe, porównania międzynarodowe*, „Biuletyn Psychometryczny”, t. II, Wrocław 1968.

⁶ Zob. np. A. Anastasi, *Psychological Testing*, The Macmillan Company, New York 1961, s. 498 i nast.

⁷ H. J. Eysenck, *Sens i nonsens w psychologii*, Warszawa 1965, s. 239.

Pytania puste układaliśmy uwzględniając dwa kryteria: a) aby nie były zbyt trudne lub zbyt łatwe, b) aby były tematycznie związane z pytaniami kwestionariusza. Za pytania zbyt trudne uznawaliśmy arbitralnie te pytania, na które — jak sądziliśmy — zdecydowana większość ludzi da odpowiedź zgodną z żądaniem zawartym w pytaniu. Natomiast pytania zbyt łatwe to takie, których „pustość” mogła być łatwa do rozpoznania. Ułożyliśmy 9 takich pytań. Wynik stanowiła liczba pytań, na które badany dał się „złapać”.

Ponieważ zasadniczy wywiad kwestionariuszowy miał być przeprowadzany z młodymi robotnikami, jakość testów sprawdziliśmy na podobnej grupie młodzieży. Do badań wybraliśmy 100-osobową grupę uczniów Zasadniczej Szkoły Metalowej w Łodzi, dziewcząt i chłopców w wieku 17–20 lat.

Po zebraniu wyników przeprowadziliśmy statystyczną analizę pytań (*item analysis*), określając moc dyskryminacyjną i trudność pytań w testach S i K⁸. Na podstawie przeprowadzonej analizy zaakceptowaliśmy 28 pytań do testu S i 19 do testu K. Tak przygotowanymi testami przebadaliśmy wszystkie osoby, z którymi przeprowadzono wywiad kwestionariuszowy. W oparciu o uzyskane wyniki ponownie przeprowadzono statystyczną analizę pytań. Wykazała ona, że wybór pytań był poprawny, gdyż tylko 1 pytanie w teście S i 2 pytania w teście K okazały się nieodpowiednie ze względu na niską moc dyskryminacyjną. Testy i rezultaty drugiej statystycznej analizy pytań zamieszczamy w Aneksie C.

Analiza pytań pustych wykazała, że pytania, które wydawały nam się niezbyt trudne, prowokowały większość badanych do udzielania odpowiedzi zgodnych z żądaniem zawartym w pytaniu. Mimo uzyskania niekorzystnych rozkładów odpowiedzi w badaniach wstępnych, zdecydowaliśmy się na zastosowanie w badaniach zasadniczych 5 pytań, które stosunkowo najczęściej rozpoznawane były jako puste. Umieszczone one zostały w kwestionariuszu między innymi pytaniami zasadniczymi i funkcyjno-

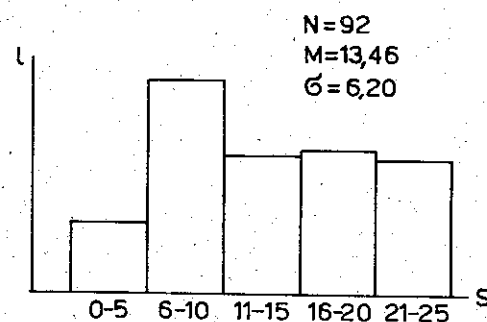
⁸ A. Anastasi, *op. cit.*, s. 168–172, 432–436.

wały jak zwyczajne pytania kwestionariusza. Nie przeprowadzono tylko na ich temat wywiadu o wywiadzie, co uczyniono w stosunku do zasadniczych pytań kwestionariusza. Treść pytań pustych oraz wyniki uzyskane w badaniach zasadniczych przedstawiamy również w Aneksie C.

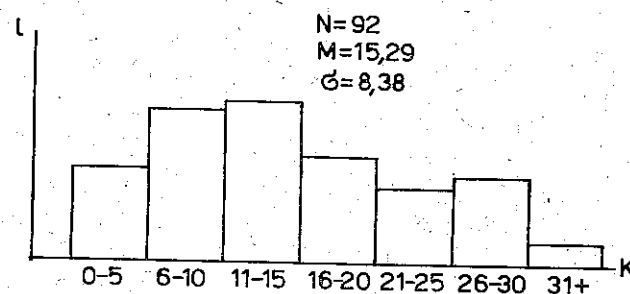
2. Charakterystyka grupy ze względu na mierzone zmienne

Grupa składała się ze 122 osób. Badani, o których nie uzyskano kompletnych informacji, byli wyłączani przy niektórych analizach. W związku z tym w różnych tabelach pojawia się różna liczba osób.

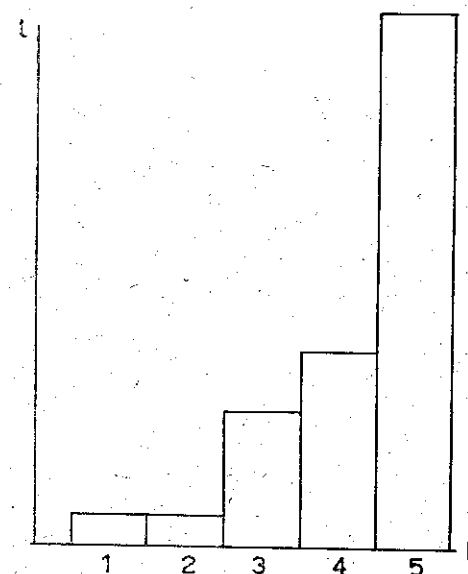
Poniżej przedstawiamy rozkłady liczebności trzech zmiennych niezależnych i jednej zależnej.



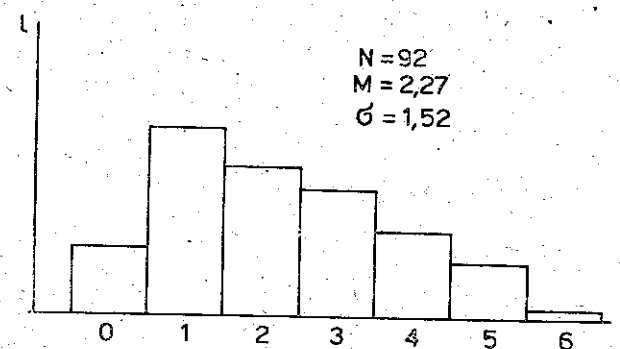
Rys. 1. Rozkład liczebności wyników testu S



Rys. 2. Rozkład liczebności wyników w teście K



Rys. 3. Rozkład liczebności wyników testu P



Rys. 4. Rozkład liczebności ze względu na liczbę udzielanych prawidłowych odpowiedzi na pytania kwestionariusza

3. Analiza wyników

Zasadniczą hipotezę naszych badań można wyrazić następująco: respondenci charakteryzujący się wysokimi wynikami w teście słownikowym (*S*), niskimi wynikami w teście kłamstwa (*K*) i pytań pustych (*P*) częściej udzielają prawidłowych odpowiedzi na pytania kwestionariusza socjologicznego niż respondenci o przeciwnych wynikach w testach.

Weryfikacja tej hipotezy sprowadza się do znalezienia związków między wynikami w testach a zmienną kryterialną *C*, którą była liczba prawidłowych odpowiedzi udzielanych przez respondenta na pytania kwestionariusza. Ponieważ badani udzielali odpowiedzi na 7 pytań, mogli uzyskać ocenę od 0 (żadna odpowiedź nie była prawidłowa) do 7 (wszystkie odpowiedzi prawidłowe). Rozkład ocen jest wyraźnie skośny dodatnio ze średnią 2,27 i odchyleniem standardowym 1,52 (zob. rys. 4).

W celu ustalenia siły związku między testami *S* i *K* a kryterium *C* obliczono korelację Pearsona. Wynosiła ona:

$$r_{SC} = +0,65$$

$$r_{KC} = -0,40$$

Nieco inaczej obliczono związek wyników testu pytań pustych *P* z kryterium *C*. Ponieważ rozkład w teście *P* był wybitnie skośny ujemnie, zastosowano dychotomizację wyników, przyjmując arbitralnie, że jedną kategorię tworzą wyniki 0-3 (osoby w mniejszym stopniu dające się „złapać” na pytanie puste), a drugą 4-5 (osoby udzielające istotnych odpowiedzi na pytania puste). W związku z tym zastosowano współczynnik korelacji biseryjnej⁹, który wynosił:

$$r_{bi_{PG}} = -0,53$$

W oparciu o otrzymane współczynniki korelacji można ułożyć równania regresji¹⁰, które na podstawie wyniku testowego

⁹ J. P. Guilford, *Podstawowe metody statystyczne w psychologii i pedagogice*, Warszawa 1960, s. 334.

¹⁰ *Op. cit.*, s. 408 i nast.

pozwołą przewidywać liczbę (lub procent) prawidłowych odpowiedzi udzielanych przez respondenta na pytania kwestionariusza. Równania te pozwolą więc wartościować respondentów na podstawie wyników, jakie uzyskują w testach.

Gdybyśmy nie dysponowali informacjami o wynikach testowych respondenta, moglibyśmy z największym prawdopodobieństwem przypuszczać, że uzyska on średnią liczbę prawidłowych odpowiedzi na pytania. Ponieważ istnieje związek między wynikami w testach a liczbą udzielanych prawidłowych odpowiedzi, to znając wyniki testowe możemy przewidywać, w jakim stopniu powiększy się lub zmniejszy w stosunku do średniej liczba prawidłowych odpowiedzi udzielanych przez respondenta.

Skonstruowano w tym celu 2 równania regresji, które pozwalają na szacowanie, o ile procent zmieni się w stosunku do średniej (=100%) liczby trafnych odpowiedzi. Wynik dodatni wskazuje na procent poprawy trafności, natomiast ujemny o jej pogorszeniu¹¹.

$$C' = -93 + 6,85 T_S \quad (1)$$

$$C' = +48 - 3,15 T_K \quad (2)$$

Na podstawie równań (1) i (2) można przewidywać zmianę liczby prawidłowych odpowiedzi z dwóch testów oddzielnie. Istnieje możliwość przewidywania prawidłowości w oparciu o dwa wyniki testów łącznie. Służy do tego równanie regresji wielokrotnej, opartej na korelacji wielokrotnej¹² ($R_{C,SK} = +0,65$).

Równanie to ma następującą postać:

$$C' = -81,4 + 6,5 T_S - 0,5 T_K \quad (3)$$

Do równania (3) można wstawiać dowolne wyniki uzyskane przez respondenta w testach *S* i *K*. Tabela 1 ilustruje, jak zmie-

¹¹ Np. badany nr 38 uzyskał w teście *S* wynik równy 12. Po podstawieniu tego wyniku do równania (1) w miejsce T_S i wykonaniu potrzebnych obliczeń uzyskuje się $C' = -10,8$. Oznacza to, że należy oczekiwać, że osoba nr 38 udzieli (lub udzieliła) 10,8% mniej prawidłowych odpowiedzi niż średnia. Podobnie postępować należy z innymi równaniami regresji.

¹² Korelacja i regresja wielokrotna zob. J. P. Guilford, *op. cit.*, s. 435.

nia się przewidywana liczba prawidłowych odpowiedzi, wyrażona w procentach, jeżeli wyniki w testach przybierają wartości średnie lub odchylają się od niej w obie strony o jedno odchylenie standardowe.

Tabela 1. Przewidywana zmiana prawidłowości odpowiedzi w procentach przy różnej kombinacji wyników w testach *S* i *K* (wyniki zaokrąglone)

<i>S</i> \ <i>K</i>	<i>M</i> - 1σ	<i>M</i>	<i>M</i> + 1σ
<i>M</i> - 1σ	-37	-40	-44
<i>M</i>	+4	0	-4
<i>M</i> + 1σ	+44	+40	+37

Z tabeli wynika np., że osoba, która uzyskała w obu testach wyniki niższe od średniej o jedno odchylenie standardowe, udzieli o 37% mniej prawidłowych odpowiedzi niż przeciętna w grupie. Inna osoba, która np. uzyskuje średni wynik w teście *K*, a w teście *S* wynik o jedno odchylenie standardowe wyższy od średniej, udzieli 40% więcej prawidłowych odpowiedzi niż przeciętnie.

Drugi sposób zweryfikowania postawionej hipotezy polega na sprawdzeniu istotności różnic w liczbie udzielanych prawidłowych odpowiedzi na pytania kwestionariuszowe przez osoby uzyskujące wysokie i niskie wyniki w testach. Punkt podziału wyników w testach *S* i *K* był zbliżony do mediany. Osoby o wysokich wynikach w teście *S* uzyskiwały średnio 3,24 prawidłowych odpowiedzi w kwestionariuszu, a przy niskich wynikach — 1,34.

Przy podziale badanych na uzyskujących wysokie i niskie wyniki w teście *K* uzyskuje się średnią liczbę prawidłowych odpowiedzi odpowiednio: 1,90 i 2,76.

Podobną analizę wykonaliśmy dla testu *P*. Osoby, które częściej dawały się „złapać” na pytania puste, udzielały średnio 1,69 prawidłowych odpowiedzi na pytania kwestionariusza, natomiast osoby z drugiej kategorii — 3,09.

Wszystkie wymienione wyżej różnice są wysoce istotne statystycznie.

Dla sprawdzenia hipotezy, że prawidłowe odpowiedzi współwystępują z wysokimi wynikami w teście *S* oraz niskimi w teście *K*, wykonaliśmy osobne analizy dla każdego z siedmiu pytań kwestionariusza. W tym celu obliczono średnie wyniki testów dla osób dających na dane pytania odpowiedzi prawidłowe oraz dla tych osób, u których wystąpiły różnego typu zaburzenia w procesie udzielania odpowiedzi. W każdym pytaniu kwestionariusza najwyższe wyniki w teście *S* wystąpiły u osób dających odpowiedzi prawidłowe. Te same osoby uzyskiwały równocześnie najniższe wyniki w teście *K*.

Analizę tę wykonano również dla wszystkich pytań łącznie, tzn. obliczono średnie wyniki w testach u osób, które dawały różne typy odpowiedzi na wszystkie pytania kwestionariusza. Dla obliczenia średniej traktowano więc oddzielnie odpowiedzi każdej osoby na poszczególne pytania („osobo-pytanie”). Wyróżniono następujące typy odpowiedzi: prawidłowe, trafne przypadkowo i nietrafne, nieistotne. Wyniki osobno dla obu testów przedstawia tabela 2.

Tabela 2. Średnie wyniki w testach *S* i *K* dla osób dających różne rodzaje odpowiedzi na pytania kwestionariusza

Odpowiedzi	Test <i>S</i>	Test <i>K</i>
Prawidłowe	15,58	12,73
Trafne przypadkowo i nietrafne	12,19	17,11
Nieistotne	10,70	15,92

Różnice między średnimi osób dających odpowiedzi prawidłowe i nietrafne oraz trafne przypadkowo są wysoce istotne ($p < 0,001$) dla obydwu testów. Natomiast różnica między nietrafnymi wraz z trafnymi przypadkowo a nieistotnymi jest znacząca statystycznie tylko dla testu *S* ($p < 0,05$).

Badane testy mogą służyć do wartościowania respondentów

pod względem udzielania przez nich prawidłowych odpowiedzi. Postanowiliśmy zatem wybrać na podstawie wyników testowych dwie kontrastowe grupy: grupa „dobrych”, która uzyskiwała w teście słownikowym *S* wyniki powyżej przeciętnej, w teście kłamstwa *K* poniżej przeciętnej, a w teście pytań pustych *P* wyniki 0-3. (Wszystkie te warunki musiały wystąpić łącznie). Grupa „złych” charakteryzowała się przeciwnymi wynikami w testach. Tabela 3 przedstawia rozkład liczebności prawidłowych odpowiedzi na pytania kwestionariusza w grupie „dobrych” i „złych”.

Tabela 3. Rozkład liczebności prawidłowych odpowiedzi osób uzyskujących dobre i złe wyniki w testach

Grupa	Liczba prawidłowych odpowiedzi								Razem
	0	1	2	3	4	5	6	7	
„Dobra”	—	2	1	5	4	3	1	—	16
„Zła”	9	14	8	1	—	—	—	—	32

Celem ustalenia siły związku między przynależnością do określonej grupy a liczbą prawidłowych odpowiedzi obliczono współczynnik korelacji biseryjnej. Wynosi on:

$$r_{bi} = +0,96$$

4. Dyskusja i wnioski

Dyskusję rozpoczniemy od zagadnień związanych z konstrukcją testów. Można sądzić na podstawie analizy, że testy *S* i *K* zostały skonstruowane prawidłowo. Wskazuje na to fakt, że powtórna statystyczna analiza pytań wykluczała tylko jedno pytanie w teście *S* i dwa pytania w teście *K*. W teście *P* pytania okazały się zbyt trudne, o czym świadczy silnie ujemnie skośny rozkład wyników. Interkorelacja między testami jest umiarkowana ($r_{SK} = -0,55$, $r_{SP} = -0,56$, $r_{KP} = +0,37$), w związku z czym uzasadnione jest traktowanie zbioru tych trzech testów jako baterii testowej.

Zasadnicza hipoteza naszych badań została potwierdzona. Osoby uzyskujące wysokie wyniki w teście słownikowym, niskie wyniki w teście kłamstwa oraz niskie wyniki w teście pytań pustych udzielają istotnie większej liczby prawidłowych odpowiedzi niż osoby cechujące się przeciwnymi wynikami testowymi.

Związki między wynikami pojedynczych testów a liczbą udzielanych prawidłowych odpowiedzi są zaskakująco wysokie. Korelacja testu z jakimś zewnętrznym kryterium rzadko kiedy osiąga uzyskane przez nas wartości (0,40 - 0,65)¹³. Istnieje ponadto prawdopodobieństwo, że silnie skośny rozkład wyników kryterium zaniżył jeszcze uzyskane współczynniki korelacji.

Wysokość tych współczynników mogłaby wzbudzać wątpliwości, lecz nie możemy dopatrzeć się jakiejś tendencyjności w dokonywaniu pomiarów i w analizie, które spowodowałyby sztuczne zawyżenie korelacji. Arbitrzy dokonujący oceny prawidłowości odpowiedzi na pytania kwestionariusza nie znali wyników testowych respondentów. Wyniki testów obliczane były w sposób obiektywny, także i w tym przypadku nie mogła wystąpić tendencyjna interpretacja. Dlatego z dużym zaufaniem można wyciągnąć wniosek, że testy uchwyciły właściwe i ważne cechy respondentów ze względu na udzielanie przez nich prawidłowych odpowiedzi.

Poza tym wysokość tych współczynników wskazuje na rzetelność ocen prawidłowości odpowiedzi ustalonych przez koderów-arbitrów. Gdyby ich oceny miały niską rzetelność, uzyskanie tak wysokich współczynników korelacji byłoby niemożliwe¹⁴.

Opracowane w oparciu o współczynniki korelacji równania regresji mają charakter orientacyjny i nie można w tej chwili ocenić, czy mogą być zastosowane w innych badaniach. Stanowią one ilustrację możliwości zastosowania tego typu zabiegów w badaniach socjologicznych. Dalsze badania mogą dopiero potwierdzić ogólność tych równań.

¹³ A. Anastasi, *op. cit.*

¹⁴ J. P. Guilford, *op. cit.*, s. 489.

Równanie regresji wielokrotnej (3) może wskazywać na względną ważność cech ujmowanych przez testy *S* i *K* w procesie udzielania trafnych odpowiedzi. Ponieważ rozkłady wyników (średnie i odchylenia standardowe) w obu testach są do siebie zbliżone, wagi występujące w równaniu można interpretować jako wskaźniki ich ważności. Waga występująca przy teście *S* jest wielokrotnie wyższa niż waga testu *K*. Świadczy to prawdopodobnie o tym, że procesy intelektualne, związane z operowaniem pojęciami, są ważniejsze przy udzielaniu prawidłowych odpowiedzi na dane pytania niż procesy typu emocjonalnego, związane z tendencją do pokazywania się w lepszym świetle. Ilustruje to dodatkowo tabela 1, w której widać, że lepszy lub gorszy wynik w teście *K* ma mniejszy wpływ na przewidywanie prawidłowej odpowiedzi niż wyniki w teście *S* (mniejsza zmienność w wierszach niż w kolumnach).

Celem dalszej analizy było sprawdzenie, czy ustalone związki między wynikami testowymi a prawidłowością odpowiedzi odnoszą się do wszystkich pytań kwestionariusza osobno, czy też mają tylko charakter statystyczny. Okazało się, że w każdym pytaniu osoby udzielające prawidłowych odpowiedzi uzyskiwały średnio najwyższe wyniki w teście *S* i najniższe w teście *K* w porównaniu z osobami, które dawały inne rodzaje odpowiedzi.

Po stwierdzeniu, że w każdym pytaniu występuje taka sama prawidłowość, obliczyliśmy średnie wyniki w testach *S* i *K* dla osób udzielających odpowiedzi: prawidłowych, trafnych przypadkowo i nietrafnych, nieistotnych. Najniższe wyniki w teście *S* uzyskują osoby udzielające nieistotnych odpowiedzi. Natomiast najgorsze wyniki w teście *K* uzyskują osoby dające wyniki nietrafne oraz trafne przypadkowo (zob. tabela 2). Wprawdzie statystycznie nieznacząco lepsze wyniki w teście *K* uzyskują osoby udzielające odpowiedzi nieistotnych, ale gdyby mimo to interpretować te różnice, można postawić hipotezę o mechanizmie udzielania odpowiedzi nietrafnych (wraz z trafnymi przypadkowo) i nieistotnych. Udzielanie odpowiedzi nieistotnych związane jest częściej z niedostatkiem procesów poznawczych (intelektual-

nych) respondenta, natomiast udzielanie odpowiedzi nietrafnych i trafnych przypadkowo wynika z procesów typu emocjonalnego.

Z całej dotychczasowej analizy wynika, że testy są dobrym narzędziem do wartościowania respondentów pod względem udzielania przez nich prawidłowych odpowiedzi. Dodatkowe sprawdzenie potwierdziło słuszność tego wniosku. 81% osób uzyskujących dobre wyniki w 3 testach udziela większej niż przeciętna liczby prawidłowych odpowiedzi. Natomiast aż 97% osób uzyskujących złe wyniki w testach nie przekracza średniej liczby prawidłowych odpowiedzi (tabela 3).

Uzyskane pozytywne rezultaty pozwalają sądzić, że celowe jest kontynuowanie analiz testowych w badaniach nad funkcjonowaniem pytań kwestionariusza w danej populacji respondentów. Szczególnie obiecujący wydaje się kierunek analizy, który pozwoli ustalić, jaki poziom wyników w testach jest konieczny, aby można było oczekiwać dostatecznie wysokiej proporcji prawidłowych odpowiedzi na pytania o danym stopniu trudności lub emocjonalnej drażliwości.