

Janusz F. Górski

O DOKŁADNOŚCI I WIARYGODNOŚCI POMIARU METODĄ LIKERTA

Jakkolwiek w sensie dosłownym tekst ten dotyczy kwestii partykularnej, to jego implikacje zdają się mieć charakter bardziej ogólny. Wiarygodność metod pomiaru wiąże się bowiem ze skutecznością eksplanacji i predykcji, te zaś w nauce nowożytnej decydują o statusie danej dyscypliny w nieformalnej, lecz jednak funkcjonującej w jakiejś mierze hierarchii. Trudno przy tym kwestionować fakt, iż wyższym statusem cieszą się nauki dysponujące rozwiniętymi i skodyfikowanymi teoriami umożliwiającymi dokładne i weryfikowalne prognozowanie¹ lub też – posługując się innym określeniem – zapewniającymi „obniżenie stopnia empiryzmu w rozwiązywaniu problemów”². Cechy te ułatwiają ocenę wartości teorii, zwiększają zgodę poznawczą oraz tempo dezaktualizacji konstruktów starszych, a tym samym przyczyniają się do szybszego rozwoju wiedzy³.

Cechy takie jednak bardzo trudno znaleźć w socjologii, w której codziennością są wątpliwości dotyczące nie tylko rozwiązań szczegółowych, ale również ich podstaw epistemologicznych⁴. Jeśli przy tym „*apetitus societatis*” H. de Groota może bezkolizyjnie koegzystować na przykład z pojęciem anomii, wówczas dyscyplina ta jawi się jako tożsama nie tyle ze swą współczesnością, ile z całą swoją historią – i w tym sensie jako nie podlegającą rozwojowi⁵. Gdyby kwestią wyłącznie semantyczną było pytanie, czy w związku z tym socjologia jest nauką? – to kłopoty autodefinicyjne nie byłyby większe od tych, jakie ma na przykład filozofia w niektórych

¹ Por. S. Cole, *Czy istnieje hierarchia nauk?*, [w:] *Rozwój nauki a społeczny charakter poznania*, Warszawa 1987, s. 54.

² J. B. Conant, *Modern Science and Modern Man.*, Columbia University Press, New York 1952, s. 62.

³ Por. Cole, *op. cit.*, s. 54.

⁴ J. Szacki, *Czy kryzys socjologii?*, [w:] *Czy kryzys socjologii?*, Warszawa 1977, s. 6.

⁵ Por. J. Niżnik, *Socjologia jako nauka nie podlegająca rozwojowi*, [w:] *Rozwój nauki...*, s. 127–141.

swoich obszarach. Rzecz jednak w tym, iż w znacznej mierze jest to również kwestia ambicjonalna, albowiem niemała część badaczy zjawisk społecznych pragnęłaby widzieć swoją dyscypliną jako „naukę normalną”, nie gorszą od przyrodoznawstwa⁶. Te ambicje owocują z jednej strony przerostem autorefleksji i samokrytycyzmu⁷, z drugiej zaś poszukiwaniem recept na „unaukowanie” dyscypliny, przybierającym czasem zabawne formy, o czym świadczy choćby następująca obserwacja recepcji teorii Kuhna wśród przedstawicieli nauk społecznych: „Niejeden badacz nauk społecznych mówił mi, że teraz przynajmniej dowiedział się, jak przekształcić swoją dyscyplinę w «naukę», przez co rozumie oczywiście, iż wie, jak ją udoskonalić. Zdaniem owych badaczy, receptą jest ograniczenie krytycyzmu, zredukowanie wszystkich ogólnych teorii do jednej i stworzenie normalnej nauki, która posiadałaby ową teorię jako swój paradygmat. Studentów należy chronić przed rozważaniami wychodzącymi z odmiennych zasad, a co bardziej niespokojnych kolegów trzeba przystosować do wykonywania poważnej pracy”⁸.

Standardową wszelako procedurą naukowego doskonalenia socjologii – realizowaną przeważnie siłami wewnętrznymi – jest produkcja „przepisów metodologicznych”, osiągająca tak znaczne rozmiary, iż gdyby ilość automatycznie przekształcała się w jakość, zapewniłoby to badaniom społecznym najwyższą naukową perfekcję.

Skala Likerta stanowiąca tutaj główny obiekt zainteresowań jest jedną z ilościowych metod pomiaru postaw i jako taka podlega odnośnym metodologicznym przepisom. Nie wchodząc w rozległe kwestie teoretyczne dotyczące pomiaru i w stosownie obszerną literaturę, skonstatujmy tylko dwie kardynalne cechy poprawnej skali, tj. rzetelność i trafność. W kanonicznym podręczniku metod badań społecznych cechy te definiowane są w sposób następujący: „Skala jest rzetelna, kiedy daje te same rezultaty przy każdym zastosowaniu do tej samej próby”⁹ oraz: „skala jest trafna wtedy, kiedy rzeczywiście mierzy to, co chce mierzyć”¹⁰. Łatwo zauważyć, iż określenia te wywodzące się z ducha operacjonizmu Bridgmana czy głębiej: pragmatyzmu Pierce’a, mają jawnie tautologiczny charakter. Być może nie jest to istotną wadą w odniesieniu do dyscyplin posługujących się wysoce zmatematyzowanymi czy szerzej – skodyfikowanymi teoriami. W ramach jednak badań społecznych oraz charakterystycznych dla nich niedookreślonych i mało precyzyjnych teorii skutkuje to erupcją myśli krytycznej przy minimalnych

⁶ Szacki, *op. cit.*, s. 6.

⁷ *Ibidem*.

⁸ P. K. Feyerabend, *Jak być dobrym empirystą*, Warszawa 1979, s. 201–202.

⁹ W. J. Goode, P. K. Hatt, *Methods in Social Research*, Mc Graw-Hill, New York, Toronto, London 1952, cyt. za: *Metody badań socjologicznych*, red. S. Nowak, Warszawa 1965, s. 261.

¹⁰ *Ibidem*, s. 265.

konstruktywnych rezultatach. Każde bowiem eksplanacyjne czy prognostyczne niepowodzenie da się z łatwością wyjaśnić na przykład: brakiem epistemicznej trafności zastosowanego instrumentu pomiaru¹¹. W rezultacie mamy do czynienia ze swoistym efektem błędnego koła, w którym brak rezultatów poznawczych każdorazowo tłumaczony jest nie tyle niedoskonałością „recept metodologicznych”, ile niedoskonałością stosowania się do nich. Przepisy są zatem właściwe, natomiast ich poprawna realizacja jest jeszcze *in statu nascendi*.

Tekst niniejszy generalnie abstrahuje od epistemologicznych sporów toczących się w socjologii. Pisany jest on z punktu widzenia opcji kwantytatywnej, jednakże wnioski – jak się okaże – nie są dla tej opcji szczególnie korzystne. W dalszym ciągu bowiem zaprezentowane zostaną wyniki badań weryfikacyjnych dokładności pomiaru metodą Likerta, których efekty określić można jako umiarkowanie pesymistyczne. W typowej interpretacji, wyniki te należałoby tłumaczyć niedoskonałym zastosowaniem recept metodologicznych. Ponieważ jednak wykorzystane instrumenty pomiaru formalnie spełniały kryteria metodologicznej poprawności, nie można więc wykluczyć, iż przyczyną ogólnie mało zadowalających rezultatów są strukturalne wady samej metody pomiaru, względnie założeń teorii stanowiącej podstawę prognozy bądź też koincydencji obu czynników, co niestety w przypadku badań społecznych wydaje się najbardziej prawdopodobną hipotezą.

Wyniki, które zostaną dalej przedstawione, uzyskano w trakcie realizacji szerszego projektu badawczego dotyczącego możliwości wykorzystania metod symulacji komputerowej w prognozie socjologicznej. Ujmując rzecz w skrócie, testowano predykcyjną przydatność programów komplikujących proste sytuacje teoretyczne za pośrednictwem ich randomizacji, czyli wprowadzania elementu przypadkowości – generowania chaosu. Zasadnicze pytanie brzmiało: przy którym stopniu randomizacji analizowanego układu zmiennych, przewidywania teoretyczne pokryją się z danymi rzeczywistymi? Odpowiedź na to pytanie informowała o możliwej dokładności prognozy. O ile optymalne dopasowanie przewidywań udało się uzyskać na niskich poziomach randomizacji badanego układu, a więc przy bardzo ograniczonym działaniu zmiennej losowej (na przykład w danej sytuacji może wystąpić jedynie zdarzenie A) – istniałaby możliwość dokładnego prognozowania. Jeśli natomiast dopasowanie to wystąpiłoby na poziomach wysokich (na przykład w danej sytuacji może wystąpić zdarzenie: A, B, C, D, E; albo: w danej sytuacji może wystąpić zdarzenie A lub dowolne inne zdarzenie – pełny indeterminizm) – wtedy dokładność prognozowania byłaby bardzo ograniczona lub też przewidywanie w ogóle nie byłoby możliwe.

¹¹ Por. np. M. Marody, *Sens teoretyczny a sens empiryczny pojęcia postawy*, Warszawa 1976, s. 53 i n.

Jak łatwo zauważyć, do realizacji tego projektu – obok programu symulacyjnego, generującego różne stopnie/poziomy chaos – niezbędne były przynajmniej trzy elementy: 1) teoria stanowiąca podstawę przewidywań, 2) dane rzeczywiste dla konfrontacji z przewidywaniami teorii, 3) instrumenty pomiaru czynników określających daną sytuację oraz jej ewentualne/prognozowane zmiany.

W kwestii pierwszej (teoria) oparto się na ugruntowanym założeniu, iż pozytywnemu stosunkowi do pracy powinna towarzyszyć wyższa produktywność, zaś negatywnemu – niższa. Założenie to, charakterystyczne m. in. dla funkcjonalizmu, wiąże się z kilkoma zastrzeżeniami. Dotyczą one rodzaju pracy, składu grup roboczych, internalizacji wartości itp.¹², w związku z czym przyjęto, iż badana rzeczywistość powinna w miarę możliwości odpowiadać wymogom teorii. Stosowne kryteria dotyczyły przede wszystkim doboru próby do badań i były następujące:

- uznano, iż próba powinna obejmować zespoły z możliwie dużej ilości zakładów pracy o różnym profilu działalności;
- wybrane zespoły robocze powinny być zróżnicowane pod względem swej produktywności (a więc dobre, średnie i złe);
- wykonywana praca powinna mieć charakter zespołowy, oparty na współdziałaniu pracowników;
- w okresie badania techniczno-organizacyjne warunki pracy zespołu powinny być optymalne i ustabilizowane, w celu wykluczenia innych – poza psychospołecznymi – czynników wpływających na efektywność działania (produktywność);
- efekty produkcyjne powinny być możliwe do wyliczenia (oszacowania).

W rezultacie do badań zakwalifikowano 44 zespoły robocze (w sumie 300 respondentów) pracujące w 7 przedsiębiorstwach o różnym profilu produkcji. Same zaś badania¹³ polegały na ustaleniu – za pomocą dwóch niezależnych metod – w którym miejscu na 100-punktowej skali produktywności sytuują się rzeczywiste wyniki pracy zespołu (podstawową jednostką analizy był zespół, nie osoba) oraz na określeniu przeciętnego stosunku do pracy (postaw) występującego w danym zespole. Temu ostatniemu zadaniu służyć miała bateria skal Likerta mierząca istotne, z punktu widzenia teorii, aspekty stosunku do pracy poszczególnych członków zespołu, a więc określająca subiektywny wymiar sytuacji (pracy). Użyto do tego następujących skal (zob. aneks):

- 1) stosunek do pracy ogólnie – C,
- 2) stosunek wobec wydajności pracy – D,

¹² Por. np. D. Katz, R. L. Kahn, *Spoleczna psychologia organizacji*, Warszawa 1979, s. 568–600.

¹³ Badania przeprowadzone zostały w 1988 r.

- 3) stosunek wobec jakości pracy – E,
- 4) stosunek wobec dyscypliny pracy – F,
- 5) stosunek do zakładu pracy (identyfikacja) – G,
- 6) stosunek do treści pracy (samorealizacja) – H.

Zgodnie z przyjętymi założeniami teorii wysokie wyniki uzyskiwane na skalach Likerta powinny korelować się dodatnio z rzeczywistą produktywnością, tj. odpowiadać im powinny wysokie przedziały punktacji skali produktywności. Świadczyłoby to o istotnej zbieżności deklarowanych przekonań i realnego zachowania, a zarazem o trafności prognostycznej wykorzystanych założeń i metod pomiaru. W sytuacji idealnej, występującej wszelako jedynie w naukach fizycznych, zbieżność ta powinna mieć charakter proporcjonalności prostej, co graficznie przedstawia rys. 1.

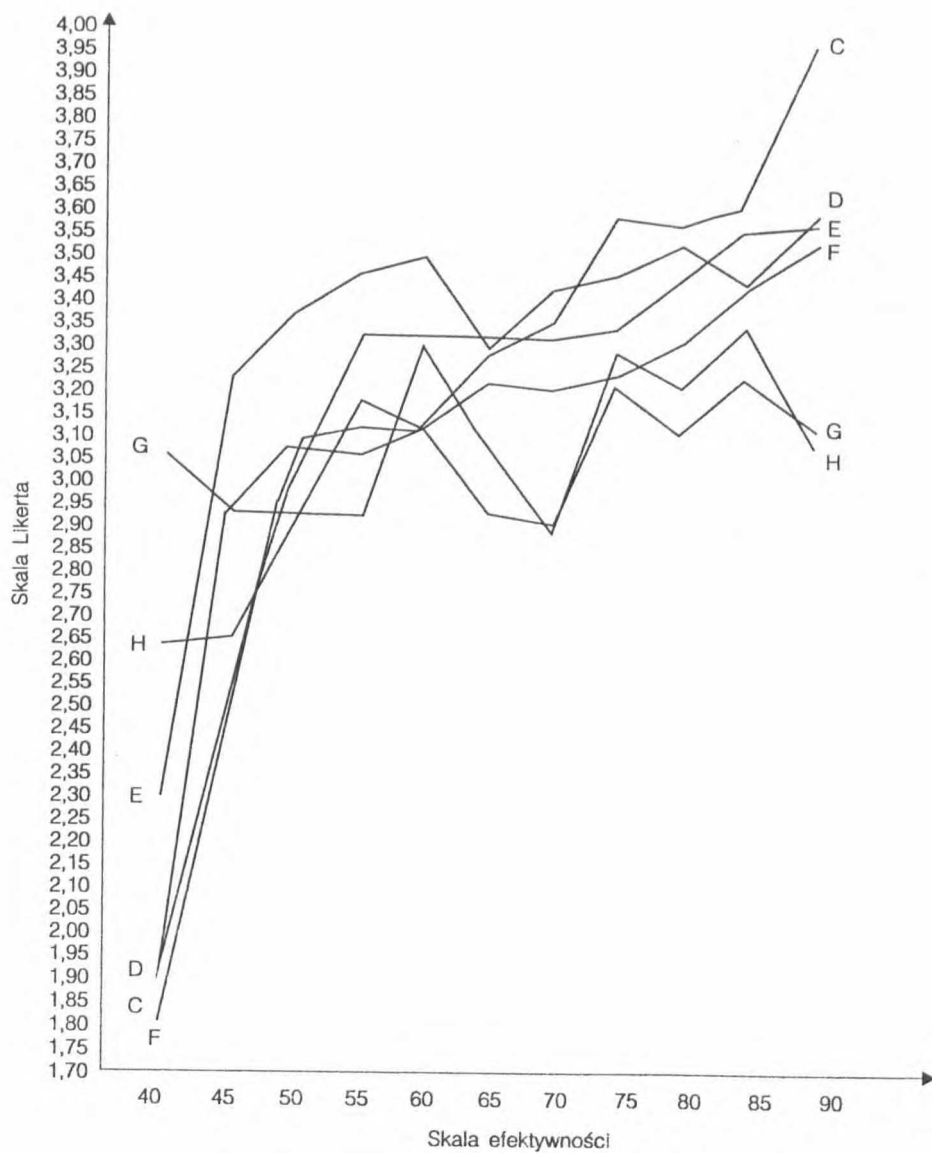
Trudno oczywiście spodziewać się, że zjawiska społeczne da się opisać równaniem stopnia pierwszego, niemniej analiza wyników pokazała miała, w jakim stopniu zastosowane procedury zbliżają się do tego (rys. 1) predykcyjnego ideału.

Graficzne odzwierciedlenie uzyskanych danych empirycznych stanowi rys. 2. Prezentuje on krzywe zbieżności wyników skali Likerta ze skalą produktywności, dla wszystkich wykorzystanych w badaniach skal. Jak widać, obraz ten niewiele ma wspólnego z ideałem, jest on jednak dokładnym odwzorowaniem uzyskanych danych liczbowych (zob. aneks, tab. 1). Regularność przebiegu krzywych poszczególnych skal jest zróżnicowana – choć generalnie niezbyt duża. Najlepiej pod tym względem prezentuje się skala C, najgorzej skale G i H, których krzywe w ogóle nie wyrażają jakichkolwiek teoretycznie prawidłowych tendencji. Próba dopasowania tych wyników do przewidywań teorii powiodła się dopiero na 15 poziomie randomizacji – w skali, w której poziom 20 oznacza pełny indeterminizm (chaos). Oznacza to, iż prognoza oparta na sumie uzyskanych wyników, miałaby charakter losowy (a więc nie byłaby stricte prognozą, lecz zgadywaniem), gdyż prawdopodobieństwo jej trafności w najlepszym wypadku wynosiłoby jak 1:4.

Wypada jednak zauważyć, że wniosek ten dotyczy sytuacji, w której podstawą prognozy byłyby dane z wszystkich skal, rys. 2 wskazuje natomiast, iż przebieg poszczególnych krzywych był mocno zróżnicowany. Tak więc o ile obraz rysowany przez skale G i H całkowicie rozmija się z przewidywaniami teorii, o tyle wyniki skal C i F są z nią w znacznej



Rys. 1. Przykład zbieżności idealnej



Rys. 2. Rozkład wyników 6 skal Likerta (dane podstawowe)

mierze zbieżne. W związku z tym w kolejnych etapach analizy zastosowano zabiegi korekcyjne umożliwiające bardziej precyzyjny ogląd wyników.

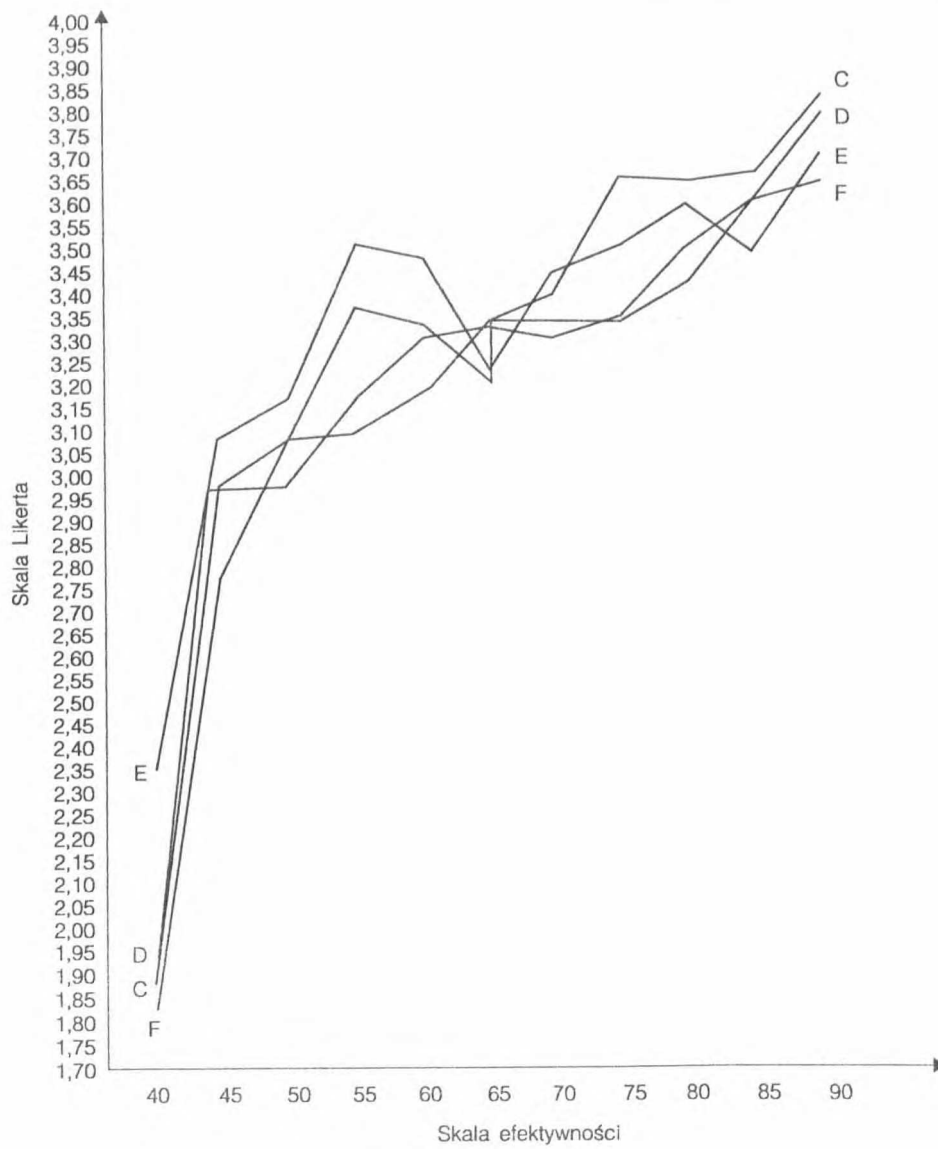
Aby wyeliminować problemy interpretacyjne związane ze wspomnianym zróżnicowanym przebiegiem krzywych wynikowych poszczególnych skal przeprowadzono ich normalizację. Zabieg ten, jak wiadomo, polega na przekształceniu wyników „surowych” w tzw. skalę standardową w celu poprawienia ich symetrii, tj. rozkładu pod krzywą normalną. Pozwolił on ustalić, iż ten sam wynik liczbowy (dane zob. aneks, tab. 2) oznacza co innego (inne natężenie postaw) na skalach G i H niż na pozostałych, potwierdził trudności interpretacyjne dotyczące wyników oscylujących wokół środkowej wartości punktacji i dowiódł, iż błędem jest traktowanie środka tej skali jako jej naturalnego zera.

W kolejnym etapie analizy wyeliminowano z krzywych wynikowych poszczególnych skal wartości skrajne, o najwyższych odchyleniach standardowych, traktując je jako prawdopodobne artefakty. Obniżyło to stopień dyspersji danych, a tym samym umożliwiło uzyskanie wyraźniejszego obrazu tendencji przebiegu krzywych. Rezultaty tych zabiegów prezentują rys. 3, 4, 5, zawierające – w kolejności – krzywe wynikowe skal C, D, E, F (rys. 3), następnie skal G i H (rys. 4) o najmniej prawidłowych przebiegach, wreszcie skal C i F (rys. 5), prezentujących się najlepiej (dane liczbowe – zob. aneks, tab. 3).

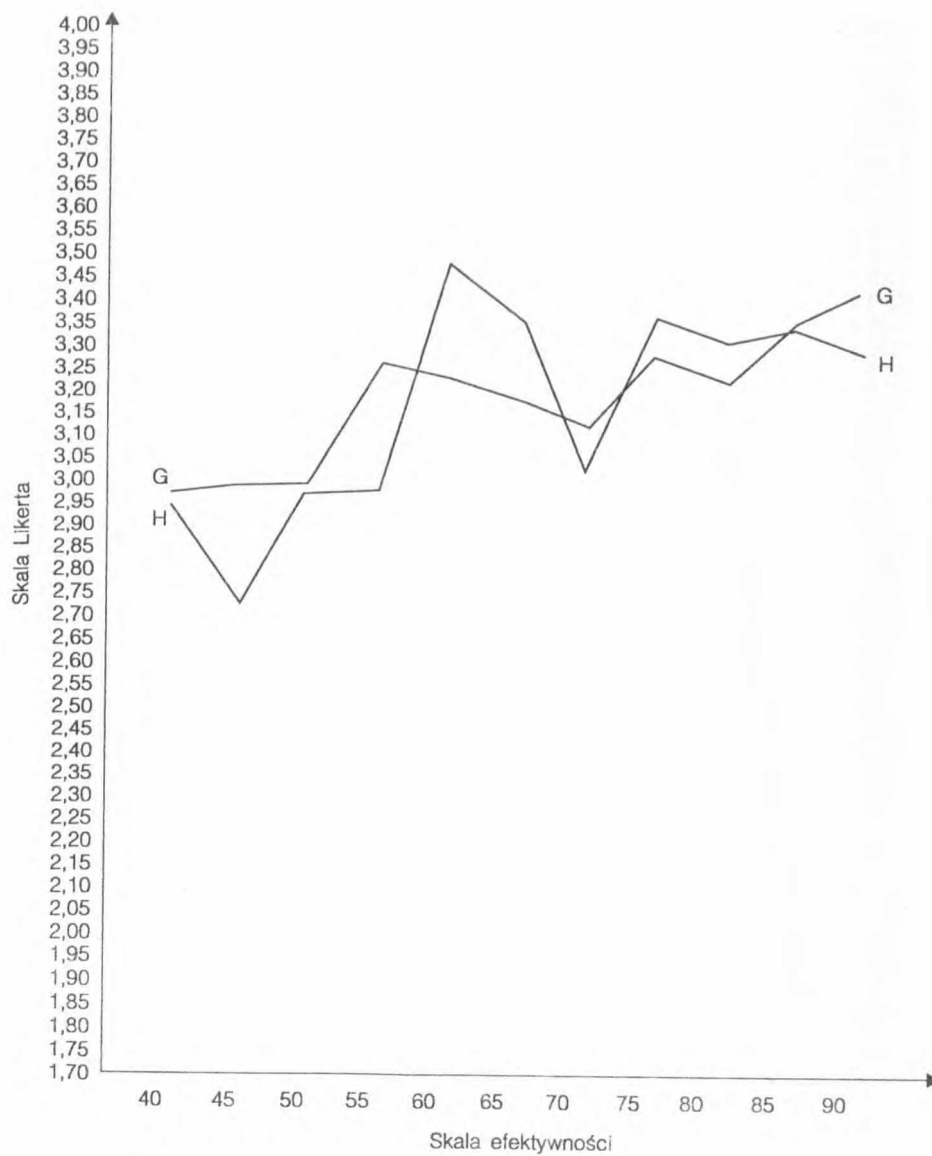
Po eliminacji wartości skrajnych przeprowadzono szereg testów, mających na celu określenie prognostycznej dokładności poszczególnych skal. Nie wchodząc w szczegóły metody wyjaśnimy tylko, że analizowano „rozrzut”, jaki wyniki pomiaru postaw wykazują w stosunku do danych rzeczywistych (produktywności). Układem odniesienia, a więc swoistą „tarczą”, była podziałka skali produktywności. Ponieważ wartości zmieniały się na niej skokowo, co 5 punktów, tym samym takie było kryterium górnej granicy dokładności pomiarowej testowanych skal Likerta. Wielkość owego „rozrzutu” wahała się w zależności od skali, w przypadku skrajnym – skala G – pokrywając obszar 40 stopni skali produktywności. Najdokładniejsza okazała się skala C, najmniej dokładne: skale G i H – ponieważ również przebieg ich krzywych nie wykazywał żadnej sensownej tendencji, zostały one usunięte z dalszej analizy.

Ostatnim krokiem podjętym dla poprawienia regularności rozkładu wyników była agregacja danych pomiarowych pochodzących ze skal: C, D, E, F (dane liczbowe – zob. aneks). Uzyskano w ten sposób jedną krzywą – przedstawia ją rys. 6 – której parametry wyraźnie się poprawiły. Posiadając dwa maksima i dwa minima, stanowi ona parabolę trzeciego stopnia dającą się sprowadzić do postaci liniowej¹⁴.

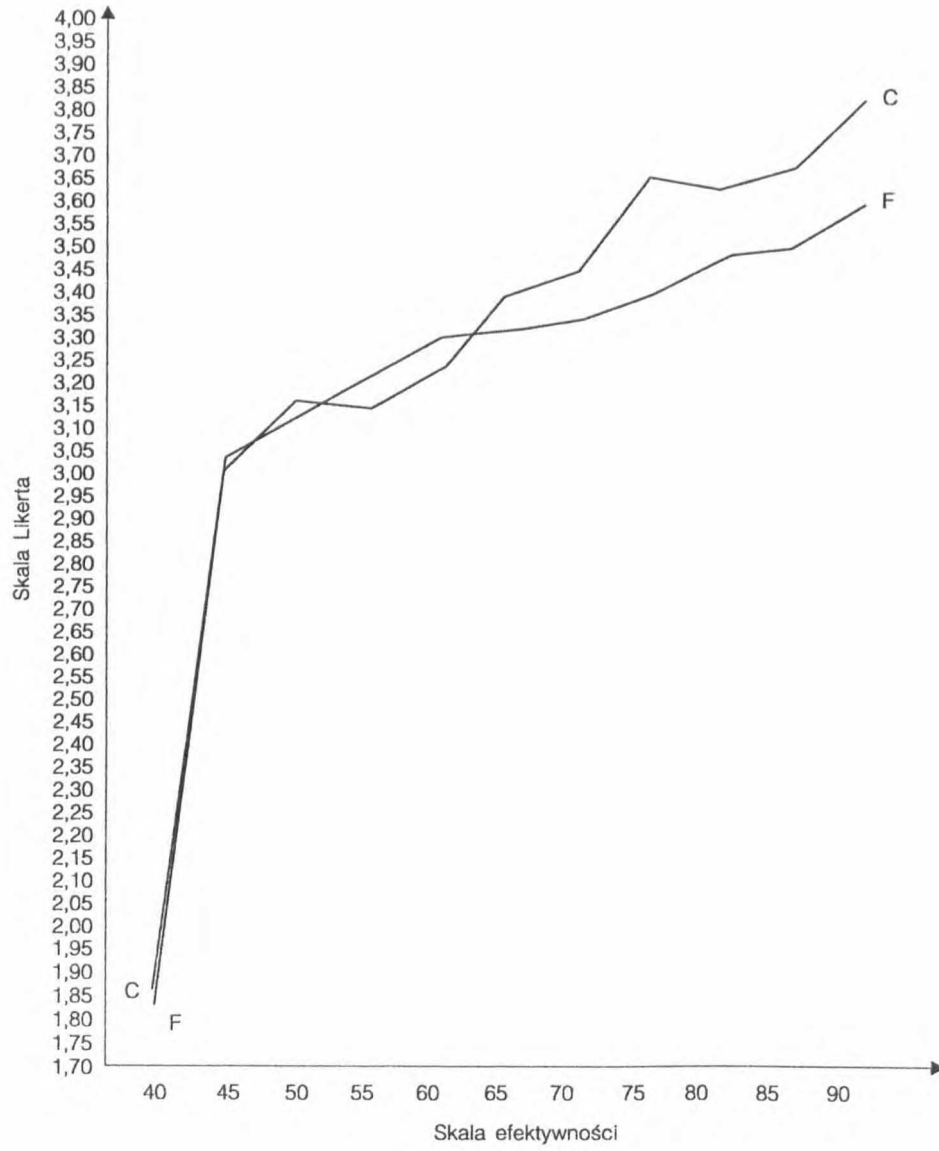
¹⁴ Por. J. Brzeziński, *Elementy metodologii badań psychologicznych*, Warszawa 1985, s. 46.



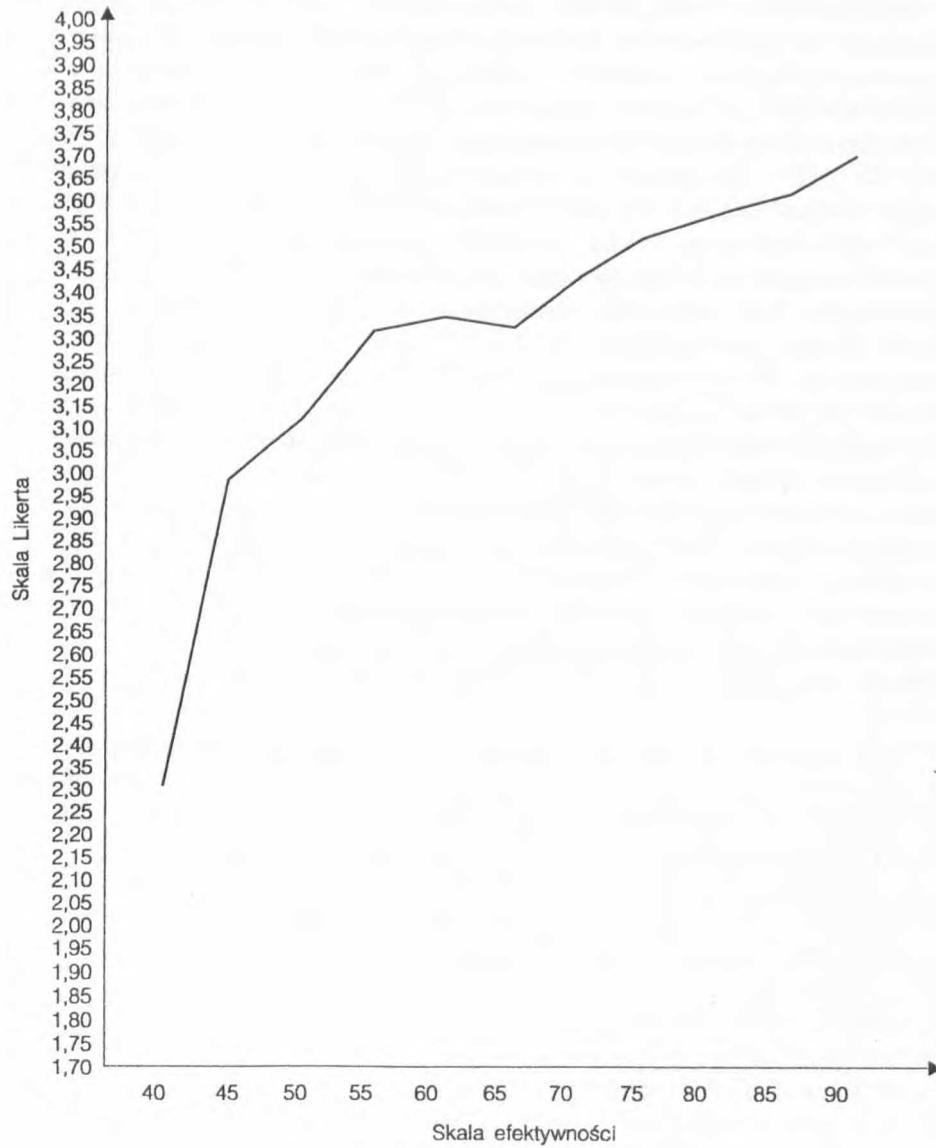
Rys. 3. Krzywe wynikowe dla skal: C. D. E. F. po eliminacji wartości ekstremalnych



Rys. 4. Krzywe dla skal: G. H. po eliminacji wartości ekstremalnych



Rys. 5. Krzywe dla skal: C. F. po eliminacji wartości ekstremalnych



Rys. 6. Krzywe dla zagregowanych wyników skal: C. D. E. F.

Dopasowanie w ten sposób ujednoliconych danych do założeń teorii powiodło się na czwartym poziomie randomizacji układu, co oznacza, iż prawdopodobieństwo trafności opartej na nich prognozy wynosi mniej więcej 0,8. W porównaniu z pierwszą próbą jest to ewidentna poprawa – przyczyniły się do niej przeprowadzone zabiegi statystyczne oraz usunięcie skali G i H – czy jednak zadowolająca, to rzecz inna. Pewnej poprawie uległa również dokładność zagregowanych wyników. Ich maksymalny „rozrzut” nie przekracza obecnie 15 stopni skali produktywności. Jest to co prawda znacznie mniej niż 40 stopni, ciągle jednak dużo. Tak więc uogólniając przytoczone dane należałoby stwierdzić, iż dokonując pomiaru za pomocą skali Likerta trzeba liczyć się z 15–20-procentowym marginesem jego niepewności. W przypadku korzystania z jednej tylko skali, rozsądniej byłoby go jeszcze zwiększyć.

Podsumowując należy przestrzec przed formułowaniem prognoz na podstawie wyników pomiaru dokonanego pojedynczą skalą Likerta. Z zebranych doświadczeń wynika, że prawdopodobieństwo minimalnie zadowolającego działania skali jako narzędzia prognozy wynosi jak 1:6. Ryzyko to można zmniejszyć stosując monotematyczną baterię kilku skal i agregując ich wyniki – wymaga to jednak skomplikowanych zabiegów statystycznych. Najpewniejsza jest wszakże weryfikacja skali, czy oszacowanie błędu, z jakim pracuje, co jednak rzadko jest możliwe w badaniach standardowych.

Tak oto przedstawiają się uzyskane dane empiryczne i wynikające z nich podstawowe wnioski. Konkluzja wydaje się zatem oczywista: prognozowanie zachowań rzeczywistych na podstawie pomiaru natężenia postawy dokonywanego omawianą metodą jest zawodne, a bez dodatkowej korekty statystycznej praktycznie niemożliwe. Przestrzeganie standardowych procedur analizy rzetelności skali okazuje się niewystarczające, wszystkie bowiem stwierdzenia testowanych skal charakteryzowały się zarówno wysoką mocą rozdzielczą, jak i zbieżnością z wynikiem ogólnym pomiaru (poziomy istotności: 0,001 i 0,0005), mimo to jednak uzyskane krzywe w znacznej mierze pozbawione były oczekiwanego kształtu i regularności. Czy więc zawiodła trafność? Trudno to ocenić nawet *a posteriori*, zważywszy choćby fakt, iż krzywe wynikowe skal G i H biegły niejako w poprzek wyjściowo przyjętych założeń – badana populacja nie potwierdziła tezy, że satysfakcja i zadowolenie z pracy ma jakikolwiek związek z jej efektami¹⁵. Tak więc uzyskane rezultaty najprawdopodobniej dowodzą koincydencji obu rodzajów błędów, tj. błędów teorii i błędów pomiaru, co dla metodologa oznaczać może tyle, iż popełniono błędy i one właśnie – nie zaś sam „przepis” – są

¹⁵ Por. Katz, Kahn, *op. cit.*, s. 579. Por. też. R. Likert, *Developing Patterns of Management*, New York 1961.

przyczyną niepowodzeń. Jednakże dla badacza kwestii merytorycznych oznacza to po prostu, że nie ma on narzędzi (w szerokim rozumieniu), których dokładności i wiarygodności mógłby zaufać.

Skala Likerta nie jest oczywiście reklamowana jako idealna metoda pomiaru. Znane są powszechnie trudności w interpretowaniu wyników środkowych itp. Wszelako zrelacjonowane badania podjęte zostały nie po to, aby potwierdzić jej doskonałość, lecz po to, aby sprawdzić jej prognostyczną użyteczność. Otóż fakt, iż w tym zakresie może ona okazać się bezużyteczna, wydaje się mieć znaczenie więcej niż partykularne. Nie ma bowiem powodów, aby fakt ten ograniczać się miał jedynie do tej metody. Przeciwnie, sądzić można, iż podobnie bezużyteczne są inne techniki oparte na zbliżonych procedurach jak również funkcjonujące według odmiennych założeń. Rozumowanie takie prowadzi przy tym nieuchronnie do niezbyt oryginalnego wniosku o niedoskonałości warsztatu badawczego kwantytatywnej socjologii. Jednakże w tej banalnej tezie – byle traktowanej serio – upatrywać można nadziei na poprawę prognostycznej sprawności socjologii. Czyż bowiem przedstawione wyniki poddają się wyłącznie negatywnej interpretacji? Rozważmy: okazało się, że niektóre skale są dla predykcji zupełnie nieprzydatne. Ale okazało się także, że inne bardziej się do tego celu nadają. Ponadto uzyskano wiedzę o prognostycznej sprawności poszczególnych skal i o zakresie błędu, z jakim pracują, co może być z powodzeniem wykorzystane w przyszłych badaniach z użyciem tych samych skal. Wreszcie ustalone zostało, iż najlepszą prognozę – wyraźniej już zbliżoną do ideału – uzyskać można metodą zespoloną, dokonując pomiaru postaw przy użyciu skal C, D, E i F, a następnie agregując wyniki – i to jest właśnie efekt pozytywny, konkretne narzędzie prognozy produktywności, pracujące z wysokim, jak na charakter analizowanych zjawisk, prawdopodobieństwem trafności.

Uogólniając zaryzykujemy tezę, iż pomiar i opartą na nim prognozę dałoby się w socjologii znacznie udoskonalić nie tylko drogą doskonalenia metod, ale również poprzez doskonalenie wiedzy o ich praktycznych możliwościach. Przeszkodą są tu jednak zakorzenione nawyki i stereotypy. Przyjmijmy za oczywiste, iż większość socjologów nie miałaby nic przeciwko temu, żeby ich dyscyplina spełniała funkcję predykcyjną równie skutecznie jak nauki przyrodnicze, problem jednak w tym, że podchodzą do tego odwrotnie. W przyrodznawstwie priorytetem jest odkrycie i weryfikacja, metoda zaś – nierzadko – produkcją uboczną, której poprawność sprawdza się *ex post*. W kwantytatywnej socjologii, dotkniętej manierą neopozytywizmu, postuluje się poznanie podporządkowane metodologicznej poprawności, w niej bowiem upatrywana jest recepta na naukę – priorytetem zatem staje się metoda. Różnica jest więc taka jak między obrazem a wyprodukowanym na jego podstawie puzzlem. Fakt, że kopia nie dorównuje oryginałowi, przypisuje się na ogół błędom w układaniu, czyli stosowaniu metody.

To jednak nie wszystko, albowiem dochodzi tu jeszcze immanentny dla badań społecznych czynnik, który można by określić mianem efemeryczności metody: metodę należy stosować zgodnie z zaleceniami metodologii i szanujący się badacz traktuje te sprawy poważnie. Jednakże kwestionariusz ankiety, skala czy pytanie są w badaniach społecznych najczęściej narzędziami jednorazowego użytku – wykorzystane w jednym badaniu, w innym zastępowane są przez następne, nie podlegając, oprócz standardowych analiz, żadnej pogłębionej refleksji. O skuteczności predykcijnej stosowanej przez siebie metody – jeśli go to w ogóle interesuje – badacz dowiaduje się z metodologii. Wiedza ta jednak, siłą rzeczy, często ma się nijak do konkretnego (jednorazowego) narzędzia, które stworzył i którym zamierza się posłużyć. W ten sposób trafna i precyzyjna prognoza wykluczona zostaje niejako z góry, nie można bowiem poznać możliwości konkretnego, indywidualnego przedmiotu wyłącznie z opisu ogólnej ich klasy.

Tak więc, nie przesądzając o kierunku dalszej ewolucji dyscyplin społecznych, wydaje się że to, co mogłoby przyczynić się do poprawy ich predykcijnej skuteczności, leży w zasięgu ręki każdego badacza. Jest to sprawdzenie praktycznych możliwości przygotowanego narzędzia – wiedza ta przyda się do następnych badań. Trzeba tylko przestać opierać się wyłącznie na metodologii.

ANEKS

Tabela 1

Wyniki na skalach Likerta a efektywność działania grupy (dane podstawowe)

Stopień efektywności	C	D	E	F	G	H
90	4.0268 ^a	3.6821 ^a	3.6632 ^a	3.5847 ^a	3.2174	3.1521 ^a
85	3.7060 ^a	3.6573	3.5320 ^a	3.4933	3.3827 ^a	3.4227 ^a
80	3.6387	3.5262 ^a	3.6495	3.4150 ^a	3.2103 ^a	3.2070 ^a
75	3.7000	3.4521 ^a	3.5356	3.3361 ^a	3.3211	3.3989 ^a
70	3.4521 ^a	3.3484	3.4923	3.2846	2.9906 ^a	2.8807 ^a
65	3.3874	3.3623	3.3958 ^a	3.3181	3.0683	3.1355 ^a
60	3.2080 ^a	3.3770	3.6020 ^a	3.1980 ^a	3.2110	3.3830 ^a
55	3.1333 ^a	3.4367	3.5667	3.2233	3.2933 ^a	3.0300 ^a
50	3.1655	3.1110 ^a	3.4788	3.1657	3.0985	3.0318
45	3.0161	2.8211 ^a	3.3467 ^a	3.0256 ^a	3.0822	2.7200 ^a
40	1.8700 ^a	1.9200 ^a	2.3650 ^a	1.8350 ^a	3.1600	2.7000

^a Istotna statystycznie różnica pomiędzy dwoma wynikami w kolumnie.

Tabela 2

Normalizacja. Granice przedziałów skal Likerta według podziału na steny skali standardowej (Steny 5 i 6 – wartości przeciętne)

Steny	Przedziały					
	C	D	E	F	G	H
1.	1,60–2,00	1,07–2,00	1,07–2,47	1,00–2,13	1,00–2,14	1,23–1,82
2.	2,07–2,46	2,13–2,64	2,53–2,87	2,20–2,66	2,18–2,32	1,91–2,19
3.	2,47–2,82	2,67–2,87	2,92–3,07	2,67–2,87	2,36–2,68	2,23–2,50
4.	2,85–3,15	2,93–3,07	3,13–3,29	2,92–3,13	2,72–2,91	2,52–2,80
5.	3,20–3,40	3,13–3,33	3,33–3,50	3,14–3,31	2,95–3,13	2,82–3,00
6.	3,47–3,78	3,40–3,57	3,57–3,69	3,33–3,50	3,14–3,36	3,05–3,38
7.	3,80–4,07	3,58–3,82	3,73–3,93	3,53–3,73	3,38–3,59	3,41–3,62
8.	4,13–4,33	3,87–4,06	4,00–4,20	3,75–3,93	3,60–3,82	3,63–3,89
9.	4,36–4,54	4,07–4,33	4,23–4,47	4,00–4,14	3,86–4,01	3,91–4,14
10.	4,67–4,93	4,40–4,93	4,67–4,73	4,20–4,60	4,05–4,64	4,18–4,68

Tabela 3

Wyniki sześciu skal według stopnia efektywności po eliminacji wartości ekstremalnych

Stopień efektywności	C	D	E	F	G	H
90	3.8864 ^a	3.8406 ^a	3.6632 ^a	3.5847 ^a	3.4509 ^a	3.2653
85	3.7060 ^a	3.6573 ^a	3.5320	3.4933	3.3827	3.4227 ^a
80	3.6887	3.5262	3.6495	3.4794	3.2584	3.3488
75	3.7000	3.3728 ^a	3.5356	3.3361 ^a	3.3211	3.3989
70	3.4471 ^a	3.3817	3.4452 ^a	3.2846	3.1441 ^a	3.0020 ^a
65	3.3874	3.2443 ^a	3.2140 ^a	3.3181	3.2244	3.3883 ^a
60	3.2350 ^a	3.2800	3.4720 ^a	3.2900	3.2500	3.5086
55	3.1333	3.4367	3.5667	3.2233 ^a	3.2933	3.0300 ^a
50	3.1055 ^a	3.1110	3.2050 ^a	3.0233	3.0985	2.9917
45	3.0161	2.8211 ^a	3.1650	3.0256	3.0822	2.7571 ^a
40	1.8700 ^a	1.9200 ^a	2.3650 ^a	1.8350 ^a	2.9461	2.9645 ^a

^a Istotna statystycznie różnica pomiędzy dwoma wynikami w kolumnie.

Tabela 4

Zagregowane wyniki skal C, D, E, F według stopni produktywności

Stopień efektywności	Wyniki zagregowane	Poziom istotności różnic
90	3,6233	----
85	3,5323	----
80	3,4818	----
75	3,4441	----
70	3,680	----
65	3,3048	----
60	3,3332	----
55	3,2806	----
50	3,0878	----
45	2,9792	----
40	2,3083	----

SKALA C

1. Należy troszczyć się o to, żeby dużo zarabiać, niewiele pracując.
2. Bez pracy życie byłoby nudne.
3. W swoim życiu człowiek powinien na pierwszym miejscu stawiać pracę.
4. Do pracy powinno się posyłać ludzi tylko za karę.

5. Tylko frajerzy powinni pracować.
6. Pracę należy traktować tylko jako uciążliwy obowiązek.
7. Prawdziwą satysfakcję może dawać tylko praca.
8. Człowiek powinien traktować pracę jako swój najwyższy przywilej.
9. W życiu człowieka tylko praca jest coś warta.
10. Życie byłoby znacznie przyjemniejsze, gdyby nie trzeba było pracować.
11. Człowiek nie został stworzony do pracy.
12. Praca jest tylko marnotrawieniem sił życiowych człowieka.
13. Praca jest tylko złem koniecznym.
14. Praca powinna być głównym celem w życiu człowieka.
15. Człowiek powinien żyć, przede wszystkim po to, żeby pracować.

Poszczególne stwierdzenia skali C, jak również skal pozostałych analizowanych były za pomocą testu Welcha – sprawdzanie mocy rozdzielczej – oraz testu chi-kwadrat – zbieżność stwierdzenia z wynikiem ogólnym skali.

Wszystkie wykorzystane stwierdzenia – we wszystkich skalach – charakteryzują się poziomem istotności 0,001 – dla testu Welcha i 0,0005 dla testu chi-kwadrat.

SKALA D

1. Nie ma sensu być wydajnym pracownikiem.
2. Głównym celem każdego pracownika powinna być wydajna praca.
3. W pracy należy kierować się zasadą: im więcej tym lepiej.
4. Tylko głupcy pracują wydajnie.
5. Kto nie pracuje wydajnie, powinien być z pracy wyrzucony.
6. Wydajnie pracują tylko ci, którzy chcą się podlizać przełożonym.
7. Pracowników mało wydajnych należy zmuszać do szybszej pracy.
8. Nie należy pracować bardziej wydajnie niż inni pracownicy.
9. Dobry pracownik powinien mało mówić, a dużo robić.
10. Wysoko wynagradzana powinna być tylko wydajna praca.
11. Nawet w najtrudniejszych warunkach powinno się pracować wydajnie.
12. Nie należy pracować z przesadną wydajnością.
13. Wydajna pracy przynosi więcej szkody niż pożytku.
14. Należy szukać takiej pracy, żeby mało robić, a dużo zarobić.
15. Praca z nadmierną wydajnością jest sprzeczna z interesami innych pracowników.

SKALA E

1. W pracy są sprawy daleko ważniejsze niż kwestia jakości.
2. Brakoróbstwo powinno być surowo karane.

3. Nie ma takiej pracy, której nie należałoby wykonywać solidnie.
4. Najważniejszym obowiązkiem pracownika powinna być dbałość o dobrą jakość wykonywanej pracy.
5. Człowiek nie powinien starać się pracować lepiej od innych.
6. Pracownik nie powinien przejmować się jakością własnej pracy.
7. Nie należy karać za złą jakość pracy.
8. Nie ma takiego powodu, którym można byłoby usprawiedliwić źle wykonaną pracę.
9. Nie należy wymagać dobrej pracy od ludzi otrzymujących niskie zarobki.
10. Trzeba szukać takiej pracy, w której jakość się nie liczy.
11. Największym powodem do dumy powinna być dla człowieka dobrze wykonana praca.
12. Za brakoróbstwo należy z pracy zwalniać.
13. Pracownik powinien otrzymywać wynagrodzenie tylko za dobrze wykonaną pracę.
14. Nawet produkując buble można być dobrym pracownikiem.
15. Partactwo w pracy powinno być surowo karane.

SKALA F

1. Nie należy zawracać sobie głowy przepisami regulaminu pracy.
2. Dyscyplina pracy niczemu dobremu nie służy.
3. Dyscyplina jest dobra w wojsku, ale nie w zakładzie pracy.
4. Dyscyplina pracy jest dobra tylko do szykanowania ludzi.
5. Regulamin pracy jest po to, żeby go łamać.
6. Najważniejszą sprawą w pracy powinno być przestrzeganie dyscypliny.
7. Dobry pracownik powinien bezwzględnie podporządkowywać się poleceniom przełożonych.
8. Nie warto być zdyscyplinowanym pracownikiem.
9. Przestrzeganie dyscypliny na ogół ułatwia pracę.
10. Dobry pracownik powinien być zdyscyplinowany.
11. Wszelkie polecenia przełożonych należy wykonywać bez dyskusji.
12. Dyscyplina pracy powinna być wdrażana nawet siłą.
13. Żeby być dobrym pracownikiem, niekoniecznie trzeba być zdyscyplinowanym.
14. Pracownik powinien postępować w pracy tak, jak mu się podoba.
15. Nawet drobne naruszenie dyscypliny pracy powinno być surowo karane.

SKALA G

Uważam, że:

1. W tym zakładzie warto pracować.
2. Należy szanować swój zakład pracy.
3. Trzeba czuć się współodpowiedzialnym za całość interesów firmy.
4. Interesy zakładu pracy są często sprzeczne z interesami pracownika.
5. Nie opłaca się być lojalnym wobec własnego zakładu.
6. Jest wiele innych przedsiębiorstw, w których byłoby lepiej pracować.
7. Należy dbać o dobre imię firmy.
8. Warto jak najczęściej zmieniać zakłady pracy.
9. W tym zakładzie ludziom często dzieje się krzywda.
10. W tym zakładzie opłaca się dobrze pracować.
11. Jest wiele rzeczy w tym zakładzie, z którymi rozsądny człowiek nie może się zgodzić.
12. Takie zakłady jak nasz są bardzo potrzebne w społeczeństwie.
13. Interesy i cele zakładu powinno się traktować jak własne.
14. Ogólnie biorąc praca w tym zakładzie oparta jest na rozsądnych zasadach.
15. Nie należy działać wbrew interesom firmy, nawet gdyby to było dla nas korzystne.
16. Większość przepisów, na których opiera się praca w tym zakładzie, należałoby zmienić.
17. Lepiej troszczyć się o własne sprawy niż przejmować się losami przedsiębiorstwa.
18. Ogólnie biorąc ten zakład działa zgodnie z interesami pracowników.
19. Większość wymogów stawianych przed pracownikami naszego zakładu uznać można za słuszne.
20. Należałoby gruntownie zmienić zasady pracy w naszym przedsiębiorstwie.
21. Od zatrudnionych u nas pracowników wymaga się wykonywania wielu zbędnych czynności.

SKALA H

Moim zdaniem praca, którą wykonuję:

1. W pełni zaspokaja moje ambicje.
2. Pozwala mi czuć się człowiekiem użytecznym i potrzebnym.
3. Nie daje mi dużo satysfakcji.
4. Ma duże znaczenie społeczne.

5. Ogranicza moje zdolności i możliwości.
6. Na ogół jest monotonna i nudna.
7. Pozwala mi wpływać na bieg wydarzeń w zakładzie.
8. Zmusza mnie do przebywania w towarzystwie ludzi, których nie lubię.
9. Jest dla mnie interesująca.
10. Na ogół męczy mnie i ogłupia.
11. Sprawia, że często czuję się mniej człowiekiem, a bardziej maszyną.
12. Daje mi dużo zadowolenia.
13. Byłaby również odpowiednia dla moich dzieci.
14. Często pozbawiona jest sensu.
15. Nie pozwala mi wpływać na cokolwiek.
16. Ma dla mnie więcej zalet niż wad.
17. Zapewnia mi szacunek innych ludzi.
18. Jest bardzo pożyteczna.
19. Często sprawia mi dużo radości.
20. Sprawia, że czuję się jak wół roboczy.
21. Jest godna powszechnego szacunku.

Janusz F. Górski

ON VALIDITY AND RELIABILITY OF MEASUREMENT APPLYING THE LIKERT'S METHOD

The subject of the article is a problem of the Likert's scale validity as a method of measurement. The author on the basis of empirical findings discusses the issue of prognostic credibility of results while measuring the attitudes applying the Likert's method. The basic question concerns the agreement of the prognosis that bases on the attitudes' measurement data, with the actual behavior. Six scales characterizing the different aspects of attitudes toward work conditions have been employed in the investigations. Subsequently, the findings have been compared to the actual workers' efficiency. The result allows for an assumption, that utility of a technique considered as the prognostic tool is very limited. In the majority of cases, the results of measurement differed significantly from the data relating to the actual behavior. Almost satisfactory agreement has been reached in case of one scale only. This testifies inaccuracy and at the same time, unreliability of measurement with that method and requires caution in employing it as the instrument of prediction.