

1.3. Szacowanie rzetelności narzędzi pomiarowych przy użyciu pakietu SPSS

W naukach społecznych od lat trwa spór o sposób prowadzenia badań. Antagonistami w tym sporze są zwolennicy badań ilościowych oraz badacze preferujący podejście jakościowe. W ujęciu niektórych metodologów procesy społeczne są „efemerycznymi, płynnymi zjawiskami pozbawionymi egzystencji niezależnej od sposobu konstrukcji i opisu zjawisk, dokonanego przez aktorów działających w tejże rzeczywistości” (Miles, Huberman, 2000, s.2). Co raz częściej jednak odnotować można zmniejszanie się liczby reprezentantów ortodoksyjnego podejścia do uprawiania badań zarówno w paradygmacie jakościowym, jak i ilościowym. Zdecydowana negacja przeciwstawnych paradygmatów metodologicznych zdaje się ustępować miejsca przekonaniu o komplementarności tych podejść. „Na poziomie operacyjnym trudno znaleźć badaczy, którzy rozłożyliby się obozem w jednym ustalonym miejscu stereotypowego kontinuum między relatywizmem i postpozytywizmem” (Miles, Huberman, 2000, s.5).

Wypracowanie metod badawczych, które pozwalają zgromadzić trafne i rzetelne dane empiryczne stało się ważnym zagadnieniem w każdym z wymienionych paradygmatów badawczych. W obu podejściach metodologicznych niekwestionowana jest również rola komputerów w poszerzeniu możliwości analizy danych oraz oceny jakości gromadzonego materiału empirycznego.

W zależności od rodzaju danych inne są sposoby szacowania rzetelności. Jednak zarówno w badaniach ilościowych, jak i jakościowych czyni się starania o zapewnienie wysokiej rzetelności materiału empirycznego (Kirk, Miller, 1986; Denzin, Lincoln, 1994; Konarzewski, 2000).

W naukach społecznych wiele miejsca poświęca się opracowaniu narzędzi badawczych pozwalających na ilościowy pomiar zmiennych nieobserwowalnych (konstruktów teoretycznych). Konstruowane do tego celu testy poddawane są analizie psychometrycznej najczęściej w oparciu o klasyczną (KTT) (Niemierko, 1999; Magnusson, 1981; Brzeziński, 1995) lub probabilistyczną (IRT) teorię testu (Lord, 1980; Hambleton, Swaminathan, Rogers, 1991; Baker, Kim, 2004). Wprowadzenie do szkół obowiązkowych egzaminów zewnętrznych na zakończenie kolejnych etapów kształcenia spowodowało wzrost zainteresowania metodami umożliwiającymi ocenę jakości danych uzyskiwanych w efekcie zastosowania konstruowanych testów kompetencji zarówno wśród nauczycieli, jak i władz oświatowych.

Wysoka rzetelność pomiaru jest jednym z warunków koniecznych jakie musi spełniać prawidłowo skonstruowany test.

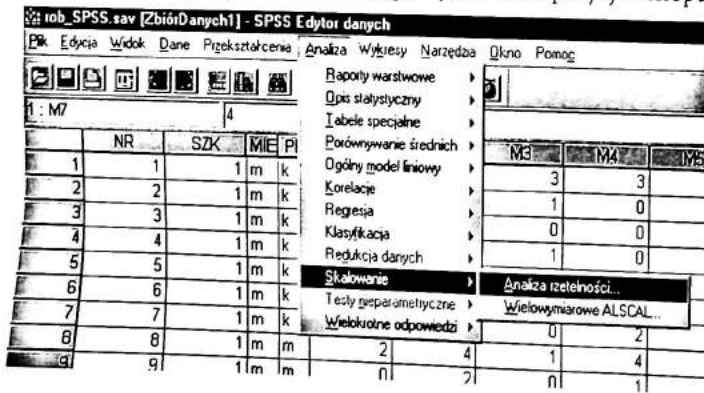
Oszacowanie rzetelności rozumianej, jako dokładności z jaką instrument mierzy wynik prawdziwy, łatwe do zrealizowania w naukach przyrodniczych poprzez wielokrotne powtórzenie pomiaru, w teorii testów wymagało rozbudowanego podejścia. W teorii testów wskaźnik rzetelności określa „wielkość tej części wariancji całkowitej testu, która powstała z wariancji rozkładu wyników prawdziwych” (Magnusson, 1981, s.104). Rzetelność testu „nie jest cechą stałą. Zależy ona zarówno od badanych osób, jak i od warunków, w jakich przeprowa-

dzono testowanie” (Niemierko, 1975, s.220). Rzetelność testów można estymować stosując różne metody empiryczne informujące o różnych źródłach błędów pomiaru (APA, 1985). Są one mniej lub bardziej przydatne w zależności od rodzaju konstruowanego narzędzia badawczego. Niektóre sposoby szacowania rzetelności zyskują dużą popularność. Zaproponowane przez Cronbacha rozszerzenie współczynników Kuder – Richardsona na testy wielopunktowe jest „uznawane dziś za najlepszy sposób szacowania rzetelności” (Hornowska, 2001, s.54). Współczynnik ten występuje obligatoryjnie we wszystkich aplikacjach komputerowych przeznaczonych do analizy rzetelności i trafności testów (np. SPSS, STATISTICA, STATGRAPHIC, SYSTAT, ITEMAN, GBSTAT).

Profesjonalne pakiety statystyczne zawierają moduły umożliwiające analizę narzędzi pomiarowych i wyników pomiaru pedagogicznego. Niewątpliwą zaletą tych aplikacji jest wyczerpująca lista analiz statystycznych możliwych do przeprowadzenia przy ich użyciu, począwszy od podstawowych statystyk, aż po zaawansowane metody wielozmiennowe. Umożliwiają one również wieloaspektową prezentację wyników w formie tabelarycznej i graficznej.

Poniżej przedstawione zostaną możliwości jakie stwarza pakiet SPSS w zakresie szacowania rzetelności narzędzi badawczych. Analizy zostaną zaprezentowane na przykładzie skali bezradności konstruowanej przez autorkę (Ciżkowicz, 2000). Składa się ona z 20 pozycji zapatrzonych w kafeterię Likerta.

Jak wspomniano powyżej SPSS jest jednym z profesjonalnych pakietów statystycznych przeznaczonych do analizy danych. Składa się on z modułów, wśród których znajduje się *Skalowanie z Analizą rzetelności* (ryc.1). Moduł ten umożliwia analizę pozycji testowych zarówno dichotomicznych, jak i wielokrotnego wyboru oraz pozycji wielopunktowych.



Ryc.1. SPSS - moduł skalowania z analizą rzetelności

Wszelkiego rodzaju właściwości testu, do których zaliczyć można między innymi trafność i rzetelność, związane są bezpośrednio z właściwościami pojedynczych pozycji lub stosunków między nimi (Niemierko, 1975; Brzeziński, 1996; Magnuson, 1981). Dlatego też moduł analityki pojedynczych pozycji.

Statystyki pozycji testu

Zakres dostępnych statystyk przedstawia ryc.2. Dla każdej pozycji możliwe jest wyznaczenie średniego wyniku oraz dyspersji odpowiedzi (*Pozycja testowa*). Pozwala to na ocenę wkładu poszczególnych pozycji do końcowego wyniku badanej zmiennej oraz ustalenie różnicowania odpowiedzi na poszczególne pozycje.



Ryc.2. Statystyki pozycji skali

Z punktu widzenia rzetelności najbardziej pożądane są zadania o przeciętnej łatwości. Należy przypomnieć, że wariancja wyników testu jest sumą wariancji pozycji i podwojonej kowariancji między pozycjami (Ferguson, Takane, 1997). Większej wariancji testu zaś odpowiada jego większa rzetelność. Stąd, z uwagi na rzetelność pomiaru pożądane są wysokie wartości zarówno wariancji pozycji, jak i ich kowariancji. W przypadku zadań wielokrotnego wyboru punktowanych 0 - 1 średni wynik jest równy łatwości zadania.

Program umożliwia też obliczenie statystyk skali po wyłączeniu określonej pozycji (*Skala po wykluczeniu pozycji*). W efekcie wybrania tej opcji otrzymujemy dane zawarte w tabeli 1, w której średnia, wariancja oraz współczynnik rzetelności α Cronbacha są statystykami jakie posiadałyby skala po wyeliminowaniu rozpatrywanej pozycji. Dla każdej pozycji skali wyznaczana jest również moc dyskryminacyjna (moc różnicująca), której wskaźnikiem jest współczynnik korelacji liniowej Pearsona wyników pozycji i wyników testowania (z wyłączeniem analizowanej pozycji) przyjmujący wartości z przedziału od -1 do +1. Wymagana jest wysoka wartość współczynnika mocy różnicującej (najlepiej powyżej 0,4). Jednocześnie niedopuszczalna jest jego ujemna wartość.

Tabela 1. Statystyki pozycji Ogółem^{*}

Nr pozycji	Średnia skali po usunięciu pozycji	Wariancja skali po usunięciu pozycji	Korelacja pozycji Ogółem	Alfa Cronbacha po usunięciu pozycji
M1	21,24	97,159	,293	,847
M2	20,49	92,554	,444	,841
M3	21,11	96,141	,321	,846
M4	21,36	93,022	,464	,839
.....
M18	21,37	94,850	,380	,843
M19	20,93	92,869	,614	,834
M20	20,88	91,145	,652	,832

Zestawione w tabeli 1 statystyki są szczególnie ważne na etapie konstrukcji narzędzia badawczego. Poprzez analizę mocy różnicującej oraz porównanie współczynników α Cronbacha zawartych w tabeli z jego wartością dla całej skali można stwierdzić, które pozycje źle współpracują z pozostałą częścią skali. Powinno to skutkować powtórnią analizą logiczną pozycji i ich korektą lub usunięciem ze skali.

^{*} Wszystkie tabele zamieszczone w tekście są oryginalnymi tabelami dostępnymi w raportach pakietu SPSS

Opcje *Między pozycjami/ Korelacje i Kowariancje* (ryc.2) pozwalają wyznaczyć macierz korelacji i/lub kowariancji między pozycjami. Wysokie wartości tych współczynników wskazują na homogeniczność pozycji testu. Jak wspomniano powyżej kowariancja między pozycjami stanowi jeden ze składników całkowitej wariancji testu. Im wyższa interkorelacja pozycji tym wyższa wariancja testu, a w konsekwencji wyższa rzetelność. Wysoka interkorelacja jest szczególnie ważna dla wewnętrznej zgodności testu (Guilford, 1989a). Na podstawie wyznaczonych tu korelacji i kowariancji wyliczana jest średnia korelacja i kowariancja między pozycjami (ryc.2 *Podsumowania*). W efekcie dostępne są informacje zamieszczone w tabeli 2. Identyczna struktura informacji edytowana jest przez program w przypadku każdej z czterech statystyk dostępnych w *Podsumowaniu*. Wariancja zamieszczona w szóstej kolumnie informuje o zróżnicowaniu każdej z tych miar wśród pozycji skali.

Tabela 2. Podsumowujące statystyki pozycji

Statystyki	(1) Średnia	(2) Min	Maksimum	(4) Rozstęp	Max/ Min	(6) Wariancja	Liczba pozycji
Średnie z pozycji	1,114	0,497	2,259	1,762	4,541	0,165	20
Wariancje z pozycji	1,010	0,535	1,606	1,071	3,000	0,077	20
Kowariancja międzypozycyjna	0,220	-0,055	0,658	0,713	-11,853	0,017	20
Korelacja międzypozycyjna	0,223	-0,056	0,625	0,681	-11,269	0,016	20

Wszystkie przedstawione powyżej statystyki pozwalają wyłonić pozycje źle współpracujące z pozostałymi pozycjami w celu przeprowadzenia ich ewentualnej korekty szczególnie jeśli narzędzie badawcze jest dopiero konstruowane.

Rzetelność testu

Rzetelność testu może być szacowana na kilka sposobów: α Cronbacha pozwala ocenić rzetelność metodą zgodności wewnętrznej, wzór Spearmana – Browna - metodą połówkową. Współczynnik rzetelności Rulona i jego wygodniejsza postać podana przez Guttmana zakładają również podział testu na dwie części. Wydzielone połówki traktowane są jako wersje równoległe testu. Każde z tych podejść stawia określone wymagania danym, na podstawie których ma być szacowana rzetelność. I tak, metoda zgodności wewnętrznej (α Cronbacha) traktująca każdą pozycję jako odrębny test równoległy, zakłada jednakowe średnie i jednakowe interkorelacje z pozostałymi pozycjami. Do szacowania rzetelności metodą połówkową, w której test dzielony jest na dwie części, stosowanie wzoru Spearmana – Browna wymaga by dwa wydzielone testy były tak samo trudne (równe średnie) i tak samo różnicujące (równość wariancji). W podejściu Guttmana nie zakłada się równości wariancji połówek testu. Jeżeli jednak wariancje te są równe, oszacowanie rzetelności współczynnikiem Spearmana – Browna i Guttmana da takie same wartości. Kiedy wariancje połówek nie są wyrównane oszacowanie rzetelności współczynnikiem Spearmana – Browna daje systematycznie wyższe wartości niż współczynnik Guttmana (Magnuson, 1981, s. 165).

Dostępny w tym module test T^2 Hotellinga umożliwia weryfikację hipotezy o równości średnich dla wszystkich pozycji skali (Lomax, 1992). Wartość testu Hotellinga wyliczona dla 20 pozycji przykładowej skali pozwala odrzucić hipotezę o równości średnich. Średnie aprotaty pozycji różnią się istotnie na poziomie $\alpha < 0,001$ (T^2 Hotellinga = 1369,295; $F = 68,733$).

Uwzględniony został też w programie szczególny przypadek szacowania rzetelności metodą zgodności wewnętrznej z wykorzystaniem analizy wariancji zaproponowany przez Hoyta. Dostępny w module test addytywności Tukey'a (ryc. 2) umożliwia przetestowanie założenia o braku multiplikatywnej interakcji między pozycjami.

Rzetelność półkownika

Ten rodzaj estymacji rzetelności informuje o niezależności wyników pomiaru od specyficznej treści zawartej w pozycjach skali (Hornowska, 2001, s. 55). Używając tej metody oceny rzetelności należy podzielić test na dwie części, które są traktowane jako wersje równoległe. Przydział pozycji do odpowiedniej części testu może odbywać się na trzy sposoby: poprzez losowy podział pozycji, poprzez dzielenie pozycji z uwzględnieniem ich treści i trudności oraz najbardziej popularny podział na części zawierające pozycje o numerach parzystych i nieparzystych (Hornowska, 2001, s. 52; Konarzewski, 2000).

W poniższym przypadku skalę rozdzielono na połowy przypisując do jednej z nich pozycje o numerach nieparzystych, do drugiej – o numerach parzystych. SPSS dokonuje podziału pozycji na dwa testy w kolejności ich wprowadzania do okna dialogowego.

W efekcie przeprowadzonych obliczeń program zwraca wyniki zamieszczone w tabeli 3. Szacowana jest rzetelność każdej części testu metodą zgodności wewnętrznej. Rezultatem jest wartość współczynnika α Cronbacha dla części 1 i 2. Podana jest też liczba pozycji testu przydzielonych do każdej części oraz u dołu tabeli – ich nazwy.

Tabela 3. Statystyki rzetelności półkownikowej

Alfa Cronbacha	Część 1	Wartość	,699
		Liczba pozycji	10(a)
	Część 2	Wartość	,763
		Liczba pozycji	10(b)
Całkowita liczba pozycji			20
Korelacje międzypołówkowe			,742
Współczynnik Spearmana-Browna	Równej długości		,852
	Nierównej długości		,852
Współczynnik podziału półkownikowego Guttmana			,841

a Pozycjami są: M1, M3, M5, M7, M9, M11, M13, M15, M17, M19.

b Pozycjami są: M2, M4, M6, M8, M10, M12, M14, M16, M18, M20.

Korelacja międzypołówkowa (tab. 3) ma wartość 0,742. Jest to bardzo wysoka zależność między wynikami uzyskiwanymi przez badanych w półkawkach testu. Współczynnik ten został obliczony dla dwóch skal o połowę krótszych od oryginalnej dlatego może być traktowany, jako oszacowanie rzetelności połowy skali. Do obliczenia rzetelności całej skali należy więc zastosować wzór proroczy Spearmana – Browna (Guilford, 1988b, Niemierko, 1975, s. 248), w którym n , oznaczające stosunek liczby pozycji zawartych w dłuższej skali do liczby pozycji w krótszej skali, przyjmuje wartość równą 2. Ponieważ wzór Spearmana – Browna można stosować tylko dla połówek skali o wyrównanych wariancjach sprawdzono spełnienie tego warunku. W pierwszej połowie wariancja jest równa 5,11, a w drugiej - 5,71. Jak podaje B. Niemierko „w praktyce dopuszcza się różnice wariancji połówek do 10 – 20% mniejszej wariancji” (1975, s. 248). W analizowanym przypadku wyższa wariancja mieści się w dopuszczalnym przedziale i ma wartość większą o 12% niższej wariancji. Zatem warunek dotyczący wyrównanych wariancji jest spełniony. Tak więc współczynnik rzetelności całej skali szacowany na podstawie korelacji połówek przyjął wartość 0,852, co można uznać za zadowalającą rzetelność.

Wyświetlane dwukrotnie wartości współczynnika Spearmana – Browna dla równej i nierównej długości są identyczne (por. tab. 3) dla testów o parzystej liczbie zadań. Wartości te mogą różnić się jedynie dla testów o nieparzystej liczbie zadań. W tym przypadku program oblicza wartość współczynnika Spearmana – Browna dla testów o nierównej długości oraz dla testów równej długości, którą osiąga przez włączenie środkowej pozycji do każdej części testu.

Zgodność wewnętrzna

Szacowanie rzetelności metodą zgodności wewnętrznej informuje w jakiej mierze poszczególne pozycje skali mierzą to, co cała skala (Niemierko, 1999, Brzeziński, 1996).

Ocenę zgodności wewnętrznej skali można przeprowadzić wykorzystując do tego celu analizę wariancji (APA, 1888). Hoyt jako pierwszy zastosował analizę wariancji ANOVA do szacowania rzetelności. Wykorzystał do tego celu dwuczynnikową analizę w układzie krzyżowym, w której jednym czynnikiem jest liczba pozycji testu, drugim – liczba badanych (liczebność kratki $n=1$). Metoda Hoyta jest jednym z układów ANOVA stosowanych do szacowania rzetelności testów. Chociaż zarówno M. Choynowski, jak i J. Brzeziński są zdania, że „analiza wariancji daje najlepszą estymację zgodności wewnętrznej testu psychometrycznego”, to zaproponowana przez Hoyta metoda nie znajduje akceptacji J. Brzezińskiego (1984, s. 13, 1996). Do szacowania rzetelności proponowane jest stosowanie innych planów eksperymentalnych. W podejściu Winer a i Brzezińskiego wykorzystano do tego celu jednoczynnikową analizę wariancji z powtórzonymi pomiarami, gdzie czynnikiem są pozycje testu. Zdaniem A. Machowskiego „należy wyraźnie podkreślić, że różnica między układem krzyżowym „osoby badane x pozycje testowe” a układem o jednym czynniku: pozycje testowe z powtarzającymi pomiarami w obrębie tego czynnika są pozorne i można je sprowadzić do różnic terminologicznych” (1993, s. 38).

Poniżej przedstawiono zastosowany w pakiecie SPSS, zaproponowany przez Hoyta, układ dwuczynnikowej analizy wariancji. Wariancje potrzebne do szacowania rzetelności skali metodą Hoyta (MS_A i MS_R) przedstawia tabela 4. Wartość testu Fishera wyznaczono, jako stosunek wariancji między pozycjami skali do wariancji reszt ($F=MS_B/MS_R$) (Brzeziński, 1984, s. 17).

Tabela 4. ANOVA z testem Tukey'a na nieaddytywność

		Suma kwadratów	df	Sredni kwadrat (MS)	F	Istotność
Zmienność międzyobiektowa (A)		2019,127	389	5,191		
Zmienność wewnątrzobiektowa	Między pozycjami (B)	1221,136	19	64,270	81,38	0,000
	Nieaddytywność	43,919(a)	1	43,919	56,02	0,000
	Równowaga	5793,496	7390	,784		
	Ogółem (R)	5837,414	7391	,790		
	Ogółem	7058,550	7410	,953		
Ogółem		9077,677	7799	1,164		

Srednia ogólna = 1,11

a Oszacowanie Tukey'a dla wykładnika potęgi, do której obserwacje muszą być podniesione, aby osiągnąć addytywność = ,585.

df – liczba stopni swobody

F – wartość testu Fishera

MS – wariancje

p – poziom istotności wnioskowania

Zastosowanie testu addytywności Tukey'a pozwala uwzględnić konsekwencje odstępstwa od przyjętego modelu addytywnego analizy wariancji (Brzeziński, 1996). Stosując podany przez Hoyta związek między rzetelnością a wariancją osób i wariancją reszt $r_{ii} = (MS_A - MS_R) / MS_A$ otrzymano współczynnik rzetelności $r_{ii} = 0,848$. Przyjęte przez Hoyta założenie o addytywności modelu ANOVA prowadzi do niedoszacowania współczynnika rzetelności. Zdaniem J. Brzezińskiego podejście Hoyta daje dolne oszacowanie rzetelności danego narzędzia. Test addytywności Tukey'a pozwala skorygować wartość współczynnika rzetelności. Poprzez wprowadzenie odpowiedniej transformacji danych zapewniona zostaje addytywność modelu (Brzeziński, Stachowski, 1984). Uwzględnienie w powyższym wzorze na rzetelność (r_{ii}) zamiast całej wariancji reszt (MS_R) jej części ($MS_{Równowaga}$) daje wartość współczynnika 0,849 wyższą zaledwie o 0,01 do wcześniej obliczonego.

Program SPSS w części dotyczącej szacowanie rzetelności narzędzia badawczego z wykorzystaniem analizy wariancji oferuje trzy rodzaje testów statystycznych, które należy stosować w zależności od rodzaju danych uzyskiwanych w efekcie zastosowania analizowanego narzędzia pomiarowego (por. ryc.2).

I tak, moduł umożliwia oszacowanie rzetelności dla danych ilościowych stosując test F (Fishera – Snedecora). Dla danych porządkowych w postaci rang program wylicza χ^2 Friedmana i współczynnik zgodności W Kendalla (Ferguson, Takane, 1997). Test χ^2 zastępuje zwykły test F w tabeli ANOVA (tab.5). Program udostępnia również dla danych dychotomicznych χ^2 Cochрана. W tym przypadku zwykły test F w tabeli ANOVA zastępowany jest wyświetlaną w tablicy statystyką Q Cochрана.

Tabela 5. ANOVA z testem Friedmana i testem Tukey'a na nieaddywność

		Suma kwadratów	df	Średni kwadrat (MS)	χ^2 Friedmana	Istotność	
Zmienność międzyobiekto		2019,127	389	5,191			
Zmienność wewnątrzobiekto	Między pozycjami	1221,136(a)	19	64,270	1281,937	,000	
	Reszta	Nieaddywność	43,919(b)	1	43,919	56,021	,000
		Równowaga	5793,496	7390	,784		
		Ogółem	5837,414	7391	,790		
Ogółem		7058,550	7410	,953			
Ogółem		9077,677	7799	1,164			

Srednia ogólna = 1,11

a Współczynnik zgodności Kendalla W = ,135.

b Oszacowanie Tukey'a dla wykładnika potęgi, do której obserwacje muszą być podniesione, aby osiągnąć addywność = ,585.

W pakiecie SPSS zastosowano też do badania zgodności wewnętrznej współczynnik α Cronbacha. Współczynnik rzetelności dla przykładowej skali ma wartość 0,848. Należy podkreślić, że zbieżność oszacowania rzetelności metodą Hoyta i przy użyciu α Cronbacha nie jest przypadkowa. Choć są to inne podejścia do problemu szacowania rzetelności to proste przekształcenia pozwalają wykazać równość efektów końcowych (Brzeziński, 1984, s. 14).

Wartości wyznaczonych współczynników rzetelności reprezentujących trzy zgoła inne podejścia do zagadnienia rzetelności dały w efekcie dość zbliżone, zadowalające rezultaty. Taka rzetelność narzędzia pozwala na stosowanie go do oceny różnic interpersonalnych.

Podsumowanie

Wysoka rzetelność wyników uzyskiwanych w efekcie zastosowania narzędzia badawczego jest warunkiem koniecznym zgromadzenia wiarygodnych danych empirycznych. W badaniach ilościowych dzięki automatyzacji procedury badawczej możliwe jest wykorzystanie statystycznych metod analizy danych, a szacowanie rzetelności zostało w pełni zalgorytmizowane. Stąd przydatność komputerów w badaniach ilościowych jest niekwestionowana. Do czasu ich upowszechnienia pracochłonność obliczeń przeprowadzanych zarówno na etapie opracowywania narzędzi badawczych, jak i na etapie analizy danych stanowiła poważną przeszkodę w pełnym wykorzystaniu zgromadzonego materiału empirycznego. Była też bezpośrednią przyczyną pozostawania w sferze zastosowań wyłącznie teoretycznych wielu metod analizy danych. Dotyczy to szczególnie analiz wielozmiennowych oraz iteracyjnych metod dopasowywania modeli adekwatnych do danych empirycznych.

Coraz częściej dostrzegana jest rola komputera również w badaniach jakościowych (Denzin, Lincoln, 1994; Miles, Huberman, 2000). Ułatwia on kodowanie tekstów przechowywanych na nośnikach cyfrowych. Pozwala na szybkie wyszukiwanie oraz zliczanie wystąpień słów i fraz w tekście. „Komputer pomaga empirycznie kategoryzować dane (wiązać ze sobą jednostki kodu) i znajdować związki między kategoriami” (Konarzewski, 2000, s. 177).

Tak więc umiejętne wykorzystanie aplikacji komputerowych do analizy danych stanowi znaczne ułatwienie w skutecznym prowadzeniu przedsięwzięć badawczych.

Streszczenie

Wysoka rzetelność pomiaru jest jednym z warunków koniecznych dla zapewnienia wiarygodności gromadzonego materiału empirycznego. Poniżej przedstawione zostały różne metody szacowania rzetelności stosowane w badaniach ilościowych (metoda połówkowa, zgodności wewnętrznej, zgodności sędziów kompetentnych) oraz możliwości ich realizacji przy użyciu pakietu SPSS. Metoda oparta na badaniu zgodności sędziów kompetentnych, powszechnie używana w badaniach ilościowych, znajduje coraz szersze zastosowanie również w badaniach jakościowych.

Abstract

High reliability of measurement is one the necessary prerequisites to ensure credibility of the collected empirical data. The paper discusses different methods of reliability estimation used in qualitative research (split-half reliability, odd-even reliability; internal reliability; coefficient of concordance) and possibilities of implementation using SPSS package. The coefficient of concordance method, universally used in research quantitative, is also more commonly used in qualitative research.

Bibliografia:

1. American Psychological Association, APA, Standardy dla testów stosowanych w psychologii i pedagogice, [w:] Biblioteka Psychologa Praktyka, T. I., Warszawa 1985.
2. Baker F.B., Kim S., Item Response Theory. Parameter Estimation Techniques, Marcel Dekker, Inc., New York 2004.
3. Brzeziński J., Badanie testu psychometrycznego metodą analizy wariancji, [w:] Wybrane zagadnienia z psychometrii i diagnostyki psychologiczne, pod redakcją J. Brzezińskiego, Poznań 1984.
4. Brzeziński J., Elementy metodologii badań psychologicznych, Warszawa 1980
5. Brzeziński J., Metodologia badań psychologicznych, Warszawa 1996
6. Brzeziński J., Rzetelność i trafność testów psychologicznych – jak jest i jak być powinno, [w:] Z zagadnień diagnostyki psychologicznej, pod redakcją J. Brzezińskiego, Poznań 1995.
7. Brzeziński J., Teoria testów psychologicznych: ABC psychometrii, [w:] Psychologia. Podręcznik akademicki (t.1), pod redakcją J. Strelau, Gdańsk 2000.
8. Ciżkowicz, B., Diagnoza bezradności intelektualnej uczniów klas VI, [w:] Edukacja w perspektywie integracji, pod redakcją M. Ochmańskiego Lublin 2000.
9. Denzin N, Lincoln Y., Handbook of Qualitative Research. Sage Publications, London 1994.
0. Ferguson G., Takane Y., Analiza statystyczna w psychologii i pedagogice, Warszawa 1997
- . Frankfort- Nachmias Ch., Nachmias D., Metody badawcze w naukach społecznych, Poznań 2001.
- Guilford J., Rzetelność i trafność pomiaru. [w:] Problemy teorii, rzetelności, konstrukcji i analizy wyników testów psychologicznych. Pod redakcją J. Brzezińskiego, Warszawa 1988.

13. Guilford J., Teoria i konstruowanie testów psychologicznych, [w:] Problemy teorii, rzetelności, konstrukcji i analizy wyników testów psychologicznych. Pod redakcją J. Brzezińskiego, Warszawa 1988.
14. Hambleton R., Swaminathan H., Rogers H., Fundamentals of Item Response Theory, Newbury Park 1991.
15. Kirk J., Miller M., Reliability and validity in qualitative research, Beverly Hills 1986
16. Konarzewski K., Jak uprawiać badania oświatowe, Warszawa 2000.
17. Lomax R., Statistical Concepts. A second course for education and the behavioral sciences, New York 1992.
18. Lord F., Introduction to item response theory. Hillsdale, New York 1980.
19. Machowski A., Rzetelność testów psychologicznych. Dwa ujęcia modelowe, Warszawa 1993.
20. Magnusson, D., Wprowadzenie do teorii testów, Warszawa 1981.
21. Miles, M., Huberman, A., Analiza danych jakościowych, Białystok 2000.
22. Morrison, D., Wielowymiarowa analiza statystyczna, Warszawa 1990.
23. Niemięko B., Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe, Warszawa 1975.
24. Niemięko B., Pomiar wyników kształcenia, Warszawa 1999.