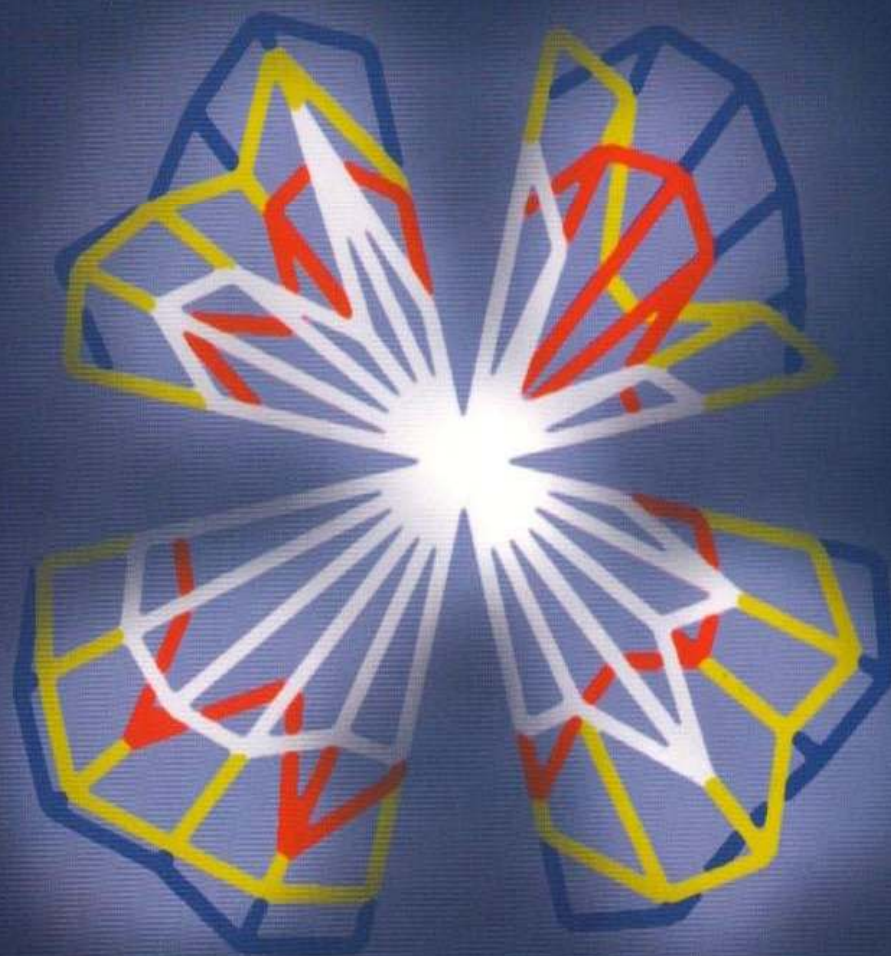


**Stanisław Ubermanowicz**

# **Ewaluacja splotowa InfoKultury**

**Skala dwuważonych ocen**



**WYDAWNICTWO NAUKOWE UAM**

W KSIĄŻCE AUTOR:

- prezentuje teorię i praktykę *Ewaluacji splotowej*, jako model-wzorzec metodologii wartościowania cech lub zjawisk z użyciem narzędzi skalowanych;
- opisuje innowacyjne strategie, metody i techniki badania za pomocą *Skali dwuważonych ocen*, konstruowanej na kryterium konkluzyjności wskaźników;
- podaje nowe, specjalizowane miary statystyczne, służące m.in. do określania rozkładów realnych oraz dynamiki i efektywności procesów;
- objaśnia implementację, weryfikację i normalizację *Kwestionariusza Kultury Informatycznej*, z instrukcją, jak utworzyć własne narzędzie;
- przedstawia obszerny zbiór autorskich definicji związanych z ewaluacją, adresowanych za pomocą skorowidza rzeczowego.

<http://ewaluacja.amu.edu.pl>

UNIwersytet IM. Adama Mickiewicza w Poznaniu

Stanisław Ubermanowicz

# **Ewaluacja splotowa InfoKultury**

## **Skala dwuwazonych ocen**



POZNAŃ 2005

Recenzent  
PROF. DR HAB. KAZIMIERZ WENTA

Wydanie publikacji dofinansowane przez Komitet Badań Naukowych

© Stanisław Ubermanowicz 2005

Projekt okładki  
EWA WĄSOWSKA

Redaktor  
KATARZYNA MUZIA

Książkę wydrukowano z materiałów dostarczonych przez Autora.  
Pakiet dołączony na krążku CD-R stanowi jej integralną obudowę.

ISBN 83-232-1608-8

WYDAWNICTWO NAUKOWE UNIwersytetu IM. ADAMA MICKIEWICZA W POZNANIU  
61-734 Poznań, ul. F. Nowowiejskiego 55, tel. 061 829 3985, fax 061 829 3980  
e-mail: [press@amu.edu.pl](mailto:press@amu.edu.pl)      <http://press.amu.edu.pl>

Wydanie I. Nakład 400 egz. Ark. wyd. 19,5 Ark. druk. 16,75 + CD

Druk i oprawa: WYDAWNICTWO I DRUKARNIA UNI-DRUK s.j.  
62-030 Luboń, ul. Przemysłowa 13

## SPIS TREŚCI

WPROWADZENIE .....	9
Struktura opisu – selektywne czytanie tekstu .....	10
Prawa autorskie – zasady używania pakietu .....	12
Co to jest <i>Ewaluacja</i> – wykładnia metody .....	14
Co to jest <i>InfoKultura</i> – wykładnia indykatum .....	16
TEORIA EWALUACJI SPLOTOWEJ .....	19
Strategie oceniania – symptomy i kwantyfikatory wartości .....	21
Charakterystyka ewaluacji – kluczowe metody i techniki .....	27
Postawa i uświadomienie – ekstrakcja doznania i poznania .....	48
Wzorzec narzędzia pomiarowego – <i>Skala dwuważonych ocen</i> .....	52
Dwa wymiary wypowiedzi – komparacja wartości i ważkości .....	56
Konstruowanie skali – dobór wskaźników i bodźców .....	59
Optymalizacja skali – weryfikacja oznak empirycznych .....	68
Przekroje analityczne – podzbiory cech i przypadków .....	76
Sfery wartościowania – cechy jakościowe, miary ilościowe .....	80
Uprzystępnianie wyników – konwersja poziomów ocen .....	87
Porównywalność prób – indeksy, buforowanie .....	88
Podstawy uogólnień – standardy ewaluatywne .....	91
PRAKTYKA EWALUACJI SPLOTOWEJ .....	95
Jak podjąć ewaluację – projekt badawczy .....	96
Jak dobrać próbę – przekroje statystyczne .....	107
Jak utworzyć narzędzie – budowa kwestionariusza .....	111
Jak zbierać wypowiedzi – technika pomiaru .....	124
Jak kodować wypowiedzi – strukturyzacja danych .....	126
Jak przetwarzać dane – obliczenia statystyczne .....	128
Jak konstruować tabele – strukturyzacja oznak .....	131
Jak interpretować parametry – analiza ilościowa .....	134
Jak odczytywać wykresy – wizualizacja oznak .....	136
Jak interpretować symptomy – analiza jakościowa .....	138
Jak formułować wnioski – synteza rezultatów .....	141

ZMIENNE EWALUACJI INFOKULTURY .....	143
Wskaźniki – argumenty zmiennej losowej .....	144
<i>Wypowiedzi</i> – wartościowanie stwierdzeń .....	145
Składniki – komponenty zmiennych cząstkowych .....	146
<i>Ocena</i>   <i>Aplauz</i> – tak ucza, bo trzeba .....	147
<i>Ambicje</i>   <i>Intencje</i> – chcę się uczyć i używać .....	147
<i>Spokój</i>   <i>Odpór</i> – bez stresu i nałogu .....	147
<i>Osąd</i>   <i>Wgląd</i> – wiem co i dlaczego .....	147
<i>Obycie</i>   <i>Pewność</i> – poćwiczę, to potrafię .....	147
<i>Zdolność</i>   <i>Ogląda</i> – co mogę, a co wolno .....	147
Czynniki – dominanty zmiennych splotowych .....	148
<i>Opinie</i>   <i>Poglądy</i> – intuicja czy wiedza .....	148
<i>Motywacje</i>   <i>Wprawa</i> – aspiracje i zaspokojenie .....	149
<i>Emocje</i>   <i>Rozwaga</i> – impuls czy refleksja .....	149
Wyniki – rezultaty zmiennych globalnych .....	150
<i>Doznanie</i>   <i>Poznanie</i> – afektywność i kognitywność .....	151
<i>Kultura</i> – uogólnienie symptomów .....	151
PARAMETRY SKALI DWUWAŻONYCH OCEN .....	153
Miary chwilowe z pomiaru jednokrotnego .....	158
<i>Jakość</i>   <i>Intensywność</i> – średnie oznak .....	160
<i>Rozziew</i> – dysjunkcja oznak .....	161
<i>Odchylenia realne</i> – dewiacja oznak .....	162
<i>Rozrzut</i> – dyspersja oznak .....	164
<i>Spójność</i> – koniunkcja oznak .....	166
<i>Wartość</i>   <i>Ważkość</i> – średnie przeskalowane .....	167
<i>Baza</i>   <i>Norma</i> – średnie oczekiwane .....	168
<i>Ufność realna</i> – tolerancja średnich .....	168
<i>Poziomy HML</i> – estymacja pozycji .....	170
<i>Skośność</i> – asymetria rozkładu .....	171
<i>Częstość realna</i> – estymacja rozkładu .....	172
Miary dystansowe z pomiaru dwukrotnego .....	178
<i>Zmiana</i> – przesunięcie rozkładu .....	180
<i>Zawilość</i> – nierównomierność zmian .....	181
<i>Bilans</i> – równoważenie rozziewu .....	182
<i>Zbieżność</i> – niwelacja niespójności .....	183
<i>Korelat</i> – współzależność zmian .....	184
<i>Trend</i> – intensywność zmian .....	185

<i>Fluktuacja</i> – wahania wypowiedzi .....	187
<i>Przyrost</i>   <i>Spadek</i> – amplitudy fluktuacji .....	188
<i>Zasięg</i> – obwiednia fluktuacji .....	190
<i>Efekt</i> – dynamika fluktuacji .....	190
Wyznaczniki jakości pomiarowej .....	192
<i>Konkluzyjność</i> – siła wnioskowania .....	193
<i>Selektywność</i> – istotność różnicowania .....	194
<i>Trafność</i> – słuszność wnioskowania .....	195
<i>Rzetelność</i> – zróżnicowanie wewnątrzspójne .....	196
EKSPLORACJA ZASOBÓW EMPIRYCZNYCH .....	199
InfoCult™ Analyser – instalacja i uruchomienie .....	200
Zasoby – menu operowania zbiorami .....	202
<i>Wczytaj</i> – ładowanie zasobów .....	202
<i>Dołącz</i> – kumulowanie danych .....	203
<i>Zapisz</i> – archiwizacja zasobów .....	204
<i>Usuń</i> – kasowanie zasobów .....	204
<i>Twórz</i> – kreowanie sekwencyjne .....	205
<i>Kopiuj</i> – ładowanie do schowka .....	205
<i>Pokaż</i> – odtwarzanie ze schowka .....	206
<i>Drukuj</i> – drukowanie zasobów .....	207
<i>Info...</i> – informacje o programie .....	207
<i>English / Polish</i> – zmiana języka .....	208
<i>Zamknij</i> – wyłączenie programu .....	208
Tabele – menu analizy statystycznej .....	209
<i>Przekroje</i> – wybór zmiennych, miar i prób .....	209
<i>Wskaźniki</i> – przetwarzanie argumentów .....	210
<i>Składniki</i> – obliczanie komponentów .....	211
<i>Czynniki</i> – obliczanie dominant .....	212
<i>Wyniki</i> – obliczanie rezultatów .....	213
Wykresy – menu wizualizacji graficznej .....	214
<i>Opcje</i> – ustawianie atrybutów grafiki .....	215
<i>Średnie</i> – wykres poziomów wartości .....	215
<i>Poziomy</i> – wykres rozkładów HML .....	216
<i>Dyspersje</i> – wykres rozrzutu i rozziwu .....	217
<i>Kierunki</i> – wykres gradientu zmian ML .....	217
<i>Tendencje</i> – wykres bilansu i trendu zmian .....	218
<i>Dynamika</i> – wykres efektu i zasięgu fluktuacji .....	218

---

INTERPRETACJA ZJAWISK EMPIRYCZNYCH .....	219
Względność wypowiedzi – poziomy odniesienia .....	221
Rozproszenie oznak – analiza rozkładowa .....	224
Miarodajność wskaźników – strefa zgodności .....	227
Koherentność oznak – komparacja wskaźników .....	229
Zmienność wypowiedzi – analiza różnicowa .....	231
Współbieżność oznak – komparacja składników .....	232
Harmonijność przemian – analiza gradientowa .....	234
Dopasowanie procesów – bilansowanie przemian .....	235
Wahliwość wypowiedzi – analiza indeksowa .....	237
Rozbieżność symptomów – komparacja czynników .....	238
Efektywność procesów – dynamika przemian .....	240
Powtarzalność wyników – formowanie przemian .....	242
Wiarygodność wniosków – komparacja wyników .....	243
RÓŻNE IMPLEMENTACJE MODELU-WZORCA .....	245
Wykorzystanie wzorca – ewaluacja <i>InfoKultury</i> .....	246
Modyfikacja indykatum – ewaluacja <i>MedioKultury</i> .....	247
Zmiana indykatum – ewaluacja <i>Paleo- i NeoTelewizji</i> .....	249
Przystosowanie modelu – ewaluacja <i>Nowej Matury</i> .....	251
Odmiana modelu – walidacja <i>Kryteriów Oceniania</i> .....	252
SKOROWIDZ z translacją słów kluczowych .....	253
SPIS RYCIN .....	260
SPIS TABEL .....	262
PRZYPISY z literaturą źródłową i rozszerzającą.....	263



## **WPROWADZENIE**

Od wielu lat prowadzę prace studyjne nad problemami wartościującego oceniania zjawisk społecznych towarzyszących ekspansywnej, powszechnej informatyzacji. Wszechobecność komputerów sprawia, że oddziałują one na coraz większe kręgi osób, już nie tylko w obszarze instytucjonalnym (praca, szkoła), lecz także w sferze prywatności (dom). Owo oddziaływanie wywołuje zarówno pozytywne, jaki i negatywne skutki. Jedni w twórczym wykorzystaniu posłusznej maszyny znajdują przyjemność samorealizacji, inni zatracają się w bezmyślnym uleganiu narkotyzującym programom. Dla wtajemniczonych komputer kreuje stanowiska pracy, a dla nienadążających może stać się groźbą utraty zatrudnienia. Jednych cieszy, lecz innych przeraża...

Obok pasjonatów komputeryzacji jest też już spora rzesza infosceptyków znajdujących racjonalne argumenty przeciwko tej nieokiełznanej ekspansji. Wprawdzie procesów tych nie da się powstrzymać, lecz z całą stanowczością należy kształtować ich jakość. Ażeby mieć do tego racjonalne przesłanki, niezbędne jest poznanie całego splotu okoliczności wpływających na pozytywną bądź negatywną ocenę każdego z ogniw procesu. Niestety, zbyt mało jest osób podejmujących się trudu systemowej ewaluacji zachodzących zjawisk. Przyczyna tkwi w tym, że w Polsce nieustannie dominuje tradycja laboratoryjnych, izolowanych od praktyki badań naukowych. Nie ma też literatury dotyczącej empirycznego oceniania wartości procesów informatyzacji. Pustkę tę mógłby wypełnić zaprogramowany metodycznie i obudowany medialnie podręcznik, którego pierwszą implementacją jest niniejsze opracowanie.

Założenia redakcyjne pracy wynikają z nowych nurtów programowania dydaktycznego, które odbiegają od liniowego wpajania tekstu i sprawdzania, czy można już przejść dalej. Obecnie istotą programowania podręcznika jest zorganizowanie jego struktur w taki sposób, aby czytelnik sam podejmował decyzje, jakimi ścieżkami studiować materiał. Jest to szczególnie ważne, ponieważ w pracy umieszczono zagadnienia z różnych dziedzin naukowych (pedagogiki, psychologii, statystyki i metrologii). Zakres treściowy materiału silnie skondensowano, ażeby czytelnik nie został przytłoczony wywodem, lecz

by w jego umyśle powstały *ślady* i *wzorcy*, jako załączki wiedzy strukturalnej. Przypominanie treści przeczytanych rozdziałów ułatwiają „hasłowe abstrakty”. Z budowaniem wiedzy ściśle wiąże się kształtowanie pojęć, dlatego znaczna część pracy to zwięzłe, autorskie definicje. Zrozumienie zagadnień może być łatwiejsze, jeśli połączy się studiowanie materiału z praktycznym działaniem przy komputerze. W załączonym do podręcznika pakiecie czytelnik otrzymuje dane empiryczne, zebrane podczas wieloletnich badań w różnych ośrodkach, oraz specjalistyczny program, za pomocą którego tworzy się tabele z miarami statystycznymi i generuje atrakcyjne wykresy, stanowiące wizualizację zachodzących zjawisk oraz pełną ilustrację zasad interpretacyjnych.

Założenia metodologiczne modelu ewaluacji oparte są na przesłankach psychologicznych i pedagogicznych, dotyczących postaw oraz uświadomienia. Definitywnie są to cechy względnie trwałe, dlatego każda wyrazistsza zmiana wyniku prawdopodobnie z oddziaływania jakichś czynników zewnętrznych. Na tej podstawie – mierząc na przykład dwukrotnie dyspozycje osób, które uczestniczyły w zajęciach – można orzekać, czy dany proces edukacyjny miał wpływ na słuchaczy i czy wpływ ten był pozytywny. Strategia wartościowania, jaką tu propaguję, sytuuje się w tzw. naturalistycznym nurcie *ewaluacji demokratycznej*. Znamienną cechą tego nurtu jest założenie, że to nie opinie ekspertów, lecz bezpośrednie oceny podmiotów danego procesu powinny być materiałem empirycznym do analizy wartościującej.

Założenia pomiarowe i statystyczne bazują na opracowanej przeze mnie *Skali dwuważonych ocen*, która jest pochodną znanej od ponad 70 lat skali Likerta. Jednak, w odróżnieniu od pierwowzoru, kryterium budowania skali dwuważonej nie jest dyskryminacyjność, lecz konkluzyjność bodźców-stwierdzeń, rozumiana jako ich przydatność ze względu na siłę wnioskowania. W moim modelu skali nie chodzi o różnicowanie respondentów, lecz o wartościowanie procesów na podstawie badania skutków. Do tego niezbędne są dwa pomiary z zastosowaniem nowych, specjalnie opracowanych metod statystycznych. Wyróżniającą cechą modelu jest sposób splotowego, hierarchicznego wskaźnikowania indykatum, z dążeniem do znacznej wnioskotwórczości badawczej. *Strategia splotowa* operacjonalizacji zmiennych polega na wielooznakowej, czteropoziomowej ekstrakcji komplementarnych par wyników połówkowych, czynników, składników oraz kontrolnych par wskaźników.

### **Struktura opisu – selektywne czytanie tekstu**

Złożoność procesów i wielorakość metod badawczych sprawiają, że tak, jak nie da się opracować jednego, uniwersalnego modelu ewaluacji, tak samo nie można przystępnie opisać wybrany model-wzorzec w tradycyjnie liniowy sposób. Wprawdzie badania przebiegają w chronologii niedającego się cofnąć

czasu, jednak pewne czynności badawcze mają wyraźnie cykliczny charakter (np. konieczność empirycznej weryfikacji i optymalizacji skal pomiarowych), inne natomiast wymagają futurospekcji bądź predykcji (np. przewidywania możliwych wypowiedzi respondentów i wcześniejszego opracowania schematu interpretacji). Z tego względu, obok wielu liniowych opisów czynności badawczych, spotyka się także struktury w postaci schematów nieliniowych.

Innym problemem w porządkowaniu opisu metodologicznego jest fakt, że ocenianie wartościujące zazwyczaj dotyczy procesów wielowątkowych, biegnących równolegle. Wskażę tu zapewne najważniejszy z podziałów na wątki – chodzi o poznawcze, czynnościowe, aksjologiczne i etyczne aspekty badań. Wątki te wprawdzie ściśle przeplatają się, lecz w pewnych fazach badacz skupia się bardziej na podbudowie teoretycznej, a w innych już tylko na działaniach praktycznych. O ile struktura chronologiczna jest najwłaściwsza do opisu kolejnych czynności badawczych, o tyle opis bazy teoretycznej musi być budowany na implikacyjnej ciągłości wywodu. Stąd wziął się zamysł, ażeby niniejsze opracowanie prezentujące model-wzorzec ewaluacji zbudować na komplementarnych parach wzajemnie uzupełniających się działów.

Pierwszą parę tworzą paralelne ścieżki przejść: *teoretyczna* i *praktyczna*. W części teoretycznej każdy z rozdziałów zawiera najpierw przedstawienie pojedynczego zagadnienia metodologicznego, a następnie objaśnienie sposobu systemowego rozwiązania. Z takiej właśnie przyjętej struktury biorą się dwuczłonowe tytuły rozdziałów, redagowane według klucza: problem → metoda. W części praktycznej rozdziały ułożone są zgodnie z chronologią czynności badawczych, ze szczegółowym omówieniem celu oraz sposobu podejmowania działań. Tytuły zredagowano tam, stosując klucz: zadanie → rozwiązanie.

Drugą parę, najbardziej „pojęciową”, tworzą *zmiennie* i *parametry*. Opis zmiennych przedstawiono w układzie hierarchicznym – od szczegółu do ogółu, tj.: od wskaźników, poprzez składniki i czynniki, aż po wyniki. We wspólnych podrozdziałach połączono ze sobą parami opisy tych zmiennych cząstkowych i splotowych, które pozostają wzajemnie w silnych relacjach. Owo powiązanie wyraża znak graficzny pionowej kreski | spinającej nazwy w podtytułach. Opis parametrów zgrupowano w trzech rozdziałach prezentujących miary chwilowe, dystansowe oraz kwantyfikacyjne, a podrozdziały ułożono według klucza narastającej trudności obliczania (od znanych średnich do nowych wyznaczników skalujących). Również i tam pojawiają się podtytuły z pionową kreską łączącą miary silnie z sobą powiązane, dlatego opisywane łącznie.

Trzecią parę ścieżek przejść, tym razem czynnościowo-umysłową, tworzą części: *eksploracja* i *interpretacja*. W zasadzie eksploracja jest szczegółową instrukcją obsługi oprogramowania służącego do analizy danych zebranych podczas pomiarów ewaluacyjnych. Struktura tej części opisu jest identyczna

z ofertą opcji, jaką daje *menu* programu komputerowego InfoCult™ Analyser. Z kolei opis interpretacji zawiera rozdziały, z których każdy egzemplifikuje po jednym z zachodzących podczas ewaluacji zjawisk empirycznych, wraz z wyjaśnieniem przesłanek rozumowania. Rozdziałom tym nadano dwuczłonowe tytuły według klucza: wykryte zjawisko → proces wnioskowania.

W końcowej części książki ukazano różne *implementacje* modelu-wzorca. Oprócz *InfoKultury* wskazano tam na inne przykłady praktycznego zastosowania metodologii ewaluacyjnej. W pierwszej odmianie, modyfikując jedynie kwestionariusz, rozszerzono obszar badań z kultury informatycznej na medialną, w drugiej – wykorzystano model w eksperymencie różnicowo-dystansowym do wykrycia polaryzacji postaw wobec telewizji tradycyjnej oraz neotelewizji, w trzeciej – całkowicie zmieniono indykator i obszar badań, zajmując się oceną jakości próbnej matury z matematyki w nowo proponowanej formule testowej, w porównaniu z wcześniej przeprowadzaną formułą problemową, w czwartej – przystosowano metodologię do weryfikacji kryteriów oceniania maturalnych wypracowań z języka polskiego. Przykłady tych implementacji ilustrują przydatność metody splotowej w różnych dziedzinach badawczych.

Ze względu na to, że w niniejszej pracy znajduje się wiele specyficznych pojęć związanych z nowatorskim ujęciem ewaluacji, nieodzownym elementem nawigacyjnym oprócz spisu treści okazał się *skorowidz*, ułatwiający czytanie opracowania w sposób nieliniowy. Selektywne studiowanie zawartych tutaj treści jest szczególnie wskazane, gdyż praca podejmuje trudną percepcyjnie materię – zawiera model-wzorzec demokratycznej ewaluacji procesów, pełną metodologię badawczą, znormalizowaną skalę pomiarową, komplet statystyk oraz schemat interpretacji. Niezależnie od wiedzy z zakresu ewaluacji, zachęcam najpierw do zapoznania się z definicjami podanymi we *Wprowadzeniu*. Następnie – zależnie od potrzeb i przygotowania – można czytać część teoretyczną bądź praktyczną, szukając objaśnień trudniejszych pojęć. Przed próbą interpretacji należy jednak zapoznać się z opisami zmiennych i parametrów. Ćwiczenie umiejętności samodzielnej analizy danych empirycznych umożliwia załączony do książki pakiet oprogramowania wraz z zasobami źródłowymi.

### **Prawa autorskie – zasady używania pakietu**

Metodologia *Splotowej ewaluacji InfoKultury*, oparta na *Skali dwuważonych ocen* jest oryginalną koncepcją autora<sup>1</sup>. Odrębną sprawą jest współautorstwo *Kwestionariusza kultury informatycznej*. Pierwotną, anglojęzyczną ankietę<sup>2</sup> zaproponowała Draga Vidakovic (wówczas z Duke University w USA). Dalsze prace nad adaptacją kulturową<sup>3</sup> oraz rozbudową narzędzia<sup>4</sup> prowadził autor niniejszej monografii wraz z Marcinem Paprzyckim (wówczas z University of Texas of the Permian Basin w USA). Kwestionariusz w aktualnej postaci,

udostępniony po wieloletnich zmianach, empirycznej weryfikacji, optymalizacji i normalizacji – znacznie odbiega od pierwowzoru pod względem treści bodźców-stwierdzeń. Różni się też całkowicie strukturą ekstrakcji zmiennych, schematem interpretacji oraz trafnością skalowanych wskaźników.

Przedstawiony w książce dorobek podlega ochronie dóbr intelektualnych w zakresie kompleksowych metod, oryginalnych technik oraz nowatorskich statystyk. Obszar innowacyjności precyzują następujące definicje:

➤ **Metodologia ewaluacji** – to wyprowadzona z przesłanek teoretycznych strategia wartościowania, oparta na kompleksowych modelach i schematach badawczych. Bazą opisanej tu koncepcji są: struktury postaw, standardy ewaluacyjne, skala dwuważona, analiza splotowa, rozkładowa analiza trendów, indeksowa analiza fluktuacji, komparacyjny model interpretacji.

➤ **Techniki ewaluacji** – to praktyczne sposoby pomiaru i oceny, takie jak: dobór wskaźników i przekrojów, pomiar różnicowo-dystansowy, ujęcie dwuwymiarowe, bilansowanie par kontrolnych, konwersja poziomów itp.

➤ **Statystyki ewaluacji** – to parametry ilościowe próby losowej, takie jak: wyznaczniki dyspersji, skośności, spójności, zawilgości, konkluzyjności, trafności, miary bilansu i trendu oraz wektory wzrostu, spadku, zasięgu i efektu.

W sposób szczególny chroniony jest pakiet oprogramowania InfoCult™ służący do zautomatyzowanego przetwarzania danych z ewaluacji *InfoKultury*. Pakiet ten tworzą następujące moduły zawarte w plikach komputerowych: instalator (Install.exe), specjalizowany analizator statystyczny (InfoCult.exe), serwery tabel (Grid.vbx) i grafiki (Graph.vbx) oraz kwestionariusze po polsku (Quest-pl.doc) i angielsku (Quest-en.doc). Czytelnik ma prawo do bezpłatnego użytkowania pakietu w celu ćwiczenia metod analizy i zasad interpretacji, z wykorzystaniem empirycznych danych źródłowych X-x-x.dxt oraz Y-y-y.dxt. Jednakże użycie pakietu do własnych badań wymaga nabycia licencji. Oprogramowanie to efekt wielu lat żmudnej pracy autora, dlatego w celu ochrony wytworu pakiet udostępniany jest na zasadach *shareware*. Oznacza to, że:

- Komplet oryginalnych plików instalacyjnych można bezpłatnie powielać i udostępniać innym osobom, bez prawa dokonywania zmian lub wykorzystywania kodu.
- Można zainstalować pakiet w celu pełnego rozpoznania właściwości użytkowych, a zwłaszcza na potrzeby uczenia się ewaluacji, interpretacji statystyk i wykresów.
- Użytkownicy, którzy chcą wykorzystać moduł *Analysier* do przetwarzania danych pochodzących z własnych pomiarów, powinni uiścić niewielką opłatę licencyjną.
- Wprowadzenie uzyskanego od autora numeru licencyjnego odblokowuje możliwość przetwarzania innych zbiorów danych, a nie tylko dołączonych jako przykładowe.

Wyrażam nadzieję, że trud uzyskania licencji nie zniechęci ewaluatorów pragnących usprawnić przetwarzanie i wizualizację danych, a zwłaszcza zainteresowanych ideą i zaangażowanych w doskonalenie procesów edukacyjnych.

## Co to jest *Ewaluacja* – wykładnia metody

Kluczowe w niniejszym opracowaniu słowo *ewaluacja*<sup>5</sup> nie jest jakimś nowomodnym synonimem badania naukowego, lecz pojęciem oznaczającym nową strategię analizy i oceny rzeczywistości<sup>6</sup>. Z badań naukowych powinny być formułowane raczej ponadczasowe twierdzenia o stanie rzeczy, natomiast w procesie ewaluacji ocenia się najbardziej aktualne, żywotne stany bieżące. Naukowcy budują teorie głównie przez uogólnienia, a ewaluatorzy wspierają praktykę poprzez wyszczególnienia komponentów i wartościowanie procesów. Scharakteryzowana tu definicyjna oś polaryzacji obu strategii badawczych nie jest w rzeczywistości aż tak wyrazista<sup>7</sup>. Wielość nurtów, modeli, metod i technik wyzwała wzajemne przenikanie się najwartościowszych rozwiązań. W konsekwencji oba te komplementarne procesy badawcze stają się równoważne, a wspólną troską w profesjonalnym przygotowaniu zarówno eksperymentu naukowego, jak i ewaluacji powinna być jak najlepsza, starannie i kompleksowo dopracowana metodologia badawcza. O ile wydano już bardzo dużo monografii poświęconych w całości metodom badań naukowych, o tyle niestety ciągle niewiele jest publikacji opisujących teorię i empirię ewaluacji<sup>8</sup>. Niniejsza praca służyć ma upowszechnieniu opracowanego przez mnie oraz weryfikowanego przez szereg lat modelu-wzorca metodologii ewaluacyjnej.

Prezentację modelu rozpocznę od najważniejszych definicji, skupionych wokół następujących zagadnień: *Jaka jest główna strategia metody? W jakim nurcie sytuuje się model? Co stanowi podstawę wartościowania? Jakie zadania poznawcze można stawiać? Jakie zbiorowości tworzą podmiot ewaluacji?*

**Ewaluacja** – to proces badawczy prowadzący do ocenienia wartościującego. W strategii demokratycznej ewaluacji splotowej respondenci (a nie eksperci) wytyczają preferencje normatywne natężenia cech uznawanych za wartości w danym kręgu kulturowo-społecznym. Do owych dynamicznie kreowanych standardów ewaluatywnych, charakterystycznych dla konkretnej populacji, odnoszone są wyniki uzyskane przez poszczególne próby, grupy badawcze.

Ze względu na potrzebę całkowitej jednoznaczności odczytania znaczeń podaje pełniejsze rozwinięcie najważniejszych pojęć tworzących wykładnię ewaluacji:

➤ **Badanie** – to wieloetapowy proces planowych działań, mających na celu rozwiązanie ściśle wytyczonych zadań poznawczych, w rodzaju: odkrywanie i objaśnianie zjawisk, diagnozowanie i prognozowanie, weryfikację hipotez, wartościowanie cech, strukturyzację pojęć, szukanie węzłowych ogniw i sieci powiązań między zdarzeniami, określanie przyczyn i skutków, ustalanie roli czynników, wyznaczanie poziomów, relacji, trendów, dynamiki przemian itp.

➤ **Ocenianie wartościujące** – jest strategią orzekania o jakości indykatum, odmienną od pomiaru sprawdzającego osiągnięty poziom, a także pełniejszą

od pomiaru różnicującego grupy, umożliwia bowiem ocenę tego, czy mierzona cecha jest rzeczywiście wartością w odczuciu społecznym, czy być może tylko w kryteriach założeń i artefaktów ustalonych przez badacza. Oceniane są te jedynie symptomy, które kwalifikowana większość uznaje za wartości.

➤ **Ewaluacja demokratyczna** – powinna być realizacją ściśle wydzielonych, następujących ról i zadań: autor metody określa obszar tego, co powinno być oceniane ( $X_S$ : *szukane*), reprezentacja populacji wytycza standardy tego, czego globalnie należy się spodziewać ( $X_E$ : *oczekiwane*) i na tej bazie ewaluatorzy interpretują to, co odkryli w próbach badawczych ( $X_F$ : *wskaazywane*). Społeczny proces ewaluacji ma większą rangę niż opiniowanie autorytarne, pasujące raczej tylko do decyzyjnej fazy zatwierdzania zwanej walidacją.

➤ **Ewaluacja splotowa** – polega na ocenianiu wartości zjawiska lub obiektu uwikłanego w naturalnym kontekście procesów społecznych. Z jednej strony na wypadkową wartość składają się komponenty addytywne, wzmacniające wynik ogólny, a z drugiej strony – bierze się pod uwagę zarówno czynniki korzystne, jak i niekorzystne, osłabiające końcowy wynik ewaluacji. Istotą ewaluacji splotowej jest interpretacja na kilku poziomach rozwarstwienia zmiennych po to, ażeby trafnie orzekać o tym, co było katalizatorem, a co hamulcem społecznych przemian. Ważne przy tym jest rozsądzanie, czy zaobserwowana fluktuacja cech jest korzystna bądź niekorzystna, czy może raczej jest typowym zjawiskiem w procesie rozwojowym, kiedy to następuje poprawa jednego komponentu wartości kosztem innego.

➤ **Standardy ewaluatywne** – to wyznaczana empirycznie taksonomia cech jakościowych, ocenianych dość zgodnie przez daną zbiorowość jako wartości, koniecznie zespolona z ilościowymi miarami normatywnymi, określającymi poziomy akceptacji owych wartości. W procesie rozwoju cywilizacji standardy te ulegają powolnym, pozytywnym przemianom, jednak w tempie odmiennym w różnych kręgach społeczno-kulturowych.

➤ **Populacja** – to kompletna zbiorowość tych jednostek, które spełniają założone kryterium swoistości, np. uczniowie wszystkich pierwszych klas liceów w Polsce. Mimo że jednostki (osoby) tworzące taką zbiorowość zmieniają się co roku, to jednak uogólnione cechy populacji są w miarę trwałe. Ze względu na realne możliwości pomiary prowadzi się nie na całej populacji, lecz tylko na reprezentatywnych próbach badawczych, dających podstawy do uogólnień.

➤ **Próby, grupy badawcze** – to celowo, losowo bądź okazjonalnie wybrane z populacji zespoły respondentów poddawanych ewaluacji. Zazwyczaj na owe grupy celowo oddziałuje się jakimś kontrolowanym czynnikiem. Jednakże specyfiką ewaluacji splotowej jest to, że żadna z grup nie jest izolowana od środowiska, a wręcz przeciwnie – każda pozostaje pod wpływem oddziaływań wszystkich otaczających ją kontekstów i procesów społecznych.

## Co to jest *InfoKultura* – wykładnia indykatum

Drugie kluczowe w tej pracy pojęcie – *InfoKultura* – powstało w konwencji zapisu węgierskiego, używanego przez informatyków ze względu na zwięzłość. Generalnie opowiadam się za jak najbardziej syntetycznym nazewnictwem stosowanym w ewaluacji, zarówno do określania zmiennych, jak i parametrów, przy czym w opisach sens każdej z krótkich nazw należy precyzyjnie objaśnić. Przyjętą tu nazwę w publikacjach wyników badań można rozwinąć do postaci *Kultura informatyczna*, choć warto zwrócić uwagę, że w ogólnym wzorcu implementacji skali zmienna globalna przyjmuje najkrótszą postać: *Kultura*. Wyjaśniam od razu, że jest to zamierzone pozostawienie miejsca na inny prefiks (np. *MedioKultura*, *EkoKultura*, *PopKultura*) uwarunkowany tym, co jest przedmiotem ewaluacji, nazywanym w literaturze *indykatum* i dosłownie oznaczającym zjawisko „wskazywane” empirycznie. Zaproponowany model skali pomiarowej może bowiem służyć do badań różnych, zupełnie innych zjawisk społecznych, a wszystko zależy wyłącznie od tego, o jakie kwestie pytany jest respondent, a ściślej – z jakich treściowo bodźców-stwierdzeń zbudowany zostanie kwestionariusz. Dla zachowania klarowności teoretycznych wywodów, opis metod odnosić się będzie głównie do ewaluacji kultury informatycznej.

Charakterystykę obranego indykatum najlepiej doprecyzują podstawowe definicje, skupione wokół następujących zagadnień: *Co ma być poddane ewaluacji? Co będzie pośrednim wskaźnikiem cech? Co i jak próbkuje skala pomiarowa? Jakie komponenty wyodrębni się z indykatum? Jakie parametry ilościowe cech da się wyznaczyć? Jaki może być zasięg badawczy ewaluacji?*

**InfoKultura** – to splot mentalnych cech populacji lub prób badawczych, będący odzwierciedleniem stanu uświadomienia informatycznego i postaw wobec komputeryzacji. Ta umowna nazwa dotyczy zagregowanej zmiennej, mierzonej *Kwestionariuszem kultury informatycznej* jako wypadkowy rezultat wartości doznania oraz poznania komputera. Zmiany poziomów *InfoKultury* w skali makro są wyznacznikami trendów rozwoju społecznego, a lokalnie mogą być kwantyfikatorami jakości konkretnych kursów informatycznych.

W powyższej definicji część komponentów uściśla desygnaty obranego indykatum, a inna część precyzuje ogólny cel i obszar badań. Komponenty te są w ewaluacji ważnymi pojęciami, wymagającymi pełniejszego rozwinięcia:

➤ **Cechy mentalne** – to swoiste dla jednostki lub grupy właściwości umysłu i charakterystyczne zróżnicowanie bytów wewnętrznych: struktur i procesów, śladów i wzorców, sposobów myślenia, reprezentacji i wyobrażeń oraz predyspozycji genotypowych bądź dyspozycji zinternalizowanych itp. Ewaluatora interesuje głównie empiryczny sens cech mentalnych na tyle, na ile można im przypisać określoną wartość społeczną i szacować humanistyczno-rozwojową



jakość potencjału ludzkiego. Celem deskryptywnym splotowej ewaluacji cech jest ocena wpływu środowiska i procesów na zmiany właściwości pierwotnie ukształtowanych przez naturę, natomiast celem formatywnym jest poszukiwanie w oddziaływaniach edukacyjnych sposobów harmonizowania emocji i motywacji, będących w ciągłej i silnej interakcji z czynnikami poznawczymi.

➤ **Cecha populacji** – to znamieną dla wybranej zbiorowości osób właściwość, którą można obserwować bezpośrednio lub badać pośrednio. W takim jak tu przypadku ewaluacji cech mentalnych niezbędne jest zastosowanie narzędzia pomiarowego opartego na wskaźnikach pośrednich. W ujęciu jakościowym poddana ocenie cecha mentalna wyrażana jest przez reprezentatywny i odwzorowujący ją zbiór *symptomów*, a w ujęciu ilościowym jej parametry szacowane są za pomocą zbioru *oznak* będących miarami statystycznymi. Na podstawie statystyk z celowo obranej próby przypadków wnioskować można o prawdopodobnych właściwościach i rozkładach cech w całej populacji.

➤ **Uświadomienie informatyczne** – jako cecha sfery poznawczej jest jakością ogólnej wiedzy o komputerach, ujawnianą wprost na podstawie samooceny respondentów co do swych zdolności, umiejętności i doświadczeń, po czym weryfikowaną techniką wnioskowania z kontrolnych wskaźników spójności przekonań, trafności decyzji, oglądy zachowań i roztropności działań. Uświadomienie korzyści i zagrożeń pełni społecznie ważną rolę regulacyjną. Poziom tej świadomości stanowi wyznacznik rozwoju społeczeństwa informacyjnego.

➤ **Postawy** – to predyspozycje człowieka do emocjonalnego reagowania oraz wrodzona skłonność do oceniania wartościującego, co może pobudzać potencjalną chęć do poznawania bądź działania. Są to cechy mentalne względnie trwałe, dlatego wszelkie zmiany przychylności postaw wobec konkretnego obiektu lub zjawiska mogą być pośrednim wskaźnikiem wpływu czynników zewnętrznych. W ujęciu postawy jako cechy internalnej wyrażane oceny oraz obserwowalne reakcje i tendencje są jedynie jej przybliżonymi przejawami.

➤ **Zmienna ogólna, indykatum** – to definicyjnie i operacyjnie najszerszej ujęta cecha jakościowa, którą badamy. Jej parametry ilościowe są bardziej wiarygodne niż wskaźniki tworzących ją komponentów, lecz mniej wnioskotwórcze. Z tego względu w trakcie ewaluacji splotowej należy wyodrębnić i analizować także wszystkie zmienne cząstkowe: składniki i czynniki. Przy tak złożonej poznawczo materii trafniej jest określać nie to, co chcemy badać, lecz to, co udaje się pomierzyć. Dlatego zalecam używanie słowa *indykatum*, oznaczającego konkretnie to coś, co jest wskazywane jako rezultat ewaluacji.

➤ **Wskaźniki, indykatory** – to dające się obserwować elementarne symptomy, tworzące odwzorowanie stanu komponentów zmiennej ogólnej. W pomiarze z użyciem kwestionariusza opartego na *skali dwuważonej* wskaźnikami są uzewnętrznione postawy i stany uświadomienia, wyrażane w wypowiedziach

jako reakcja na pobudzenie specjalnie dobranym zbiorem bodźców-stwierdzeń. Po nadaniu wypowiedziom wag liczbowych i uporządkowaniu w macierzach, tak przygotowane zbiory danych ilościowych stają się argumentami zmiennej losowej, poddawanej następnie wszechstronnej analizie statystycznej.

➤ **Kwantyfikatory** – to wielorakie wyznaczniki wartości, służące głównie do: selekcji pozytywności i negatywności, gradacji poziomów jakości, kwalifikacji optymalnych wskaźników oraz weryfikacji poprawności wnioskowania. Każdy z modeli ewaluacji oparty jest na kwantyfikatorach strategicznych, będących charakterystycznymi dla metody kryteriami podejmowania decyzji dotyczących wyodrębniania symptomów, stopniowania oznak i interpretacji zjawisk.

➤ **Pomiar** – to planowe działanie mające na celu wyodrębnienie zjawisk lub cech jakościowych oraz ustalenie ich parametrów ilościowych, tj.: poziomów, natężeń, różnic, proporcji, dyspersji, zmian, tendencji, dynamiki, efektów itp. W różnicowej ewaluacji dystansowej niezbędny jest dwukrotny pomiar tych samych respondentów w odstępie czasu, za pomocą identycznego narzędzia, natomiast w ewaluacji porównawczej mierzy się te same cechy przynajmniej dwóch obiektów, procesów lub prób badawczych o podobnych właściwościach.

➤ **Kwestionariusz kultury informatycznej** – to zaimplementowane na modelu skali dwuważonej narzędzie pomiarowe w formie testu wyboru z jednorodnych 7-opcyjnych wypowiedzi (od '*absolutnie nie*' ... do '*absolutnie tak*'), zawierające 24 specjalnie dobrane bodźce-stwierdzenia, służące do badania wskaźników afektywnych i kognitywnych, towarzyszących procesom informatyzacji.

➤ **Doznanie** – to emocjonalne oraz motywacyjne stany osobistych odczuć, generowane wewnątrz samoistnie bądź inicjowane bodźcami zewnętrznymi. Okazuje się, iż odczuciowe doznanie obcowania z silnie pobudzającym komputerem jest pierwotne i znacznie intensywniejsze od poziomu *poznania*. W procesach edukacyjnych niezbędne jest badanie oraz harmonizowanie obu tych komponentów i właśnie temu celowi służy *Ewaluacja splotowa* oparta na technice ekstrakcji zmiennych.

<i>Ślady</i>		<i>Wzorce</i>
<i>Ewaluacja</i> jako proces	.....	ocenianie wartościujące
demokratyzacja ewaluacji	.....	podmiotowość respondentów
metodologia ewaluacji	.....	strategia wartościowania
techniki ewaluacji	.....	sposoby pomiaru i oceny
statystyki ewaluacji	.....	parametry ilościowe próby
ewaluacja splotowa	.....	ocenianie interakcji czynników
<i>InfoKultura</i> jako indykatum	.....	splot uświadomienia i postaw
cechy mentalne	.....	swoiste właściwości umysłu
wskaźniki pośrednie	.....	symptomy odwzorowujące cechy

## TEORIA EWALUACJI SPLOTOWEJ

W niniejszej części przedstawiam **model-wzorzec** metodologii *Ewaluacji splotowej* o uniwersalnych właściwościach. Na założeniach ogólnego modelu możliwa jest budowa różnych narzędzi, służących nie tylko do pomiaru świadomości i postaw wobec dowolnego obiektu, lecz także do ewaluacji zupełnie odmiennych zjawisk bądź też cech osobowości. Wszystko zależy od tego, jakie zagadnienia próbkuje kwestionariusz, jednak z zachowaniem wszelkich zasad, technik i statystyk określonych jako strategię *Skali dwuważonych ocen*. Ważne jest, aby narzędzie testowało rzetelnie i wieloaspektowo, umożliwiając analizę nie tylko zmiennej ogólnej (indykatum), ale także wszystkich warstw zmiennych cząstkowych. Dzięki temu – w zależności od obranego indykatum, od metod badawczych (np. quasi-eksperymentu), od technik pomiarowych (np. różnicowych, dystansowych) i od zastosowanych metod statystycznych (np. testów istotności) – przed ewaluatorami otwiera się wręcz nieograniczona przestrzeń do zajmowania najrozmaitszych obszarów naukowo-badawczych, wytyczania teoretycznych i poznawczych celów ewaluacji oraz formułowania istotnych pytań problemowych, na które warto poszukiwać odpowiedzi<sup>9</sup>.

Trudno jest jednak objaśniać teoretyczny model-wzorzec bez odniesienia do realnego przykładu, dlatego podam w zwartej formie plan zweryfikowanej empirycznie *Ewaluacji InfoKultury*, będącej w niniejszej pracy kanwą opisu:

1. Co zaplanowano zbadać i dlaczego eksplorowano właśnie to zjawisko?
  - Indykatum: *Kultura informatyczna* jako istotna rozwojowo cecha osobnicza.
2. Na co trzeba było rozwarstwić indykatum, aby dało się je rzetelnie zbadać?
  - Warstwy: postawy | uświadomienie, doznanie | poznanie, impuls | namysł.
3. Jaka metoda badano, jakim narzędziem i jaką techniką mierzono?
  - Metoda: *Ewaluacja splotowa* dwuwymiarową *Skalą dwuważonych ocen*.
  - Narzędzie: standaryzowany *Kwestionariusz kultury informatycznej*.
  - Technika: dystansowy pomiar różnic i zmian wypowiedzi respondentów.
4. Jakie wyznaczniki analizowano i jakie sformułowano uogólnienia?
  - Wyznaczniki: wskaźniki, składniki, czynniki, miary chwilowe i dystansowe.
  - Uogólnienia: standardy ewaluatywne *Kultury informatycznej*.

Posługując się wzorcem ewaluacji *InfoKultury* w badaniach naukowych, można zakładać i osiągać różnorakie **cele** teoretyczne i poznawcze, w rodzaju:

- wartościowania globalnych trendów informatyzacji i komputeryzacji;
- odkrywania i objaśniania przyczyn i skutków uzależnień od komputera;
- weryfikacji tezy o pożytkach harmonizowania doznań z poznawaniem;
- określenia cyklu i siły związku między pobudzeniem a zaspokojeniem;
- taksonomii i hierarchizacji komponentów afektywnych i kognitywnych;
- wyznaczania standardu ewaluatywnego *InfoKultury* dla danej populacji;
- ustalenia przyczyn różnic *InfoKultury* w różnych makrośrodkach;
- diagnozowania i prognozowania przemian świadomości informatycznej.

W starannie zaprojektowanej metodologii ewaluacyjnej można formułować i rozwiązywać różnorodne, podane tu jako przykładowe, **problemy** badawcze:

- Czy daje się badać indykator kulturowy za pomocą skali postaw?
- Jakie są standardy ewaluatywne w sferze kultury informatycznej?
- Jak duże są różnice standardów w Polsce względem innych państw?
- Czy istnieje trwała tendencja wzrostu kultury informatycznej?
- Jakie efekty daje harmonia czynników afektywnych i kognitywnych?
- Dlaczego, obok wiedzy, istotne jest formowanie właściwych postaw?
- Jakie są tendencje i relacje pomiędzy motywacją a poznawaniem?
- Czy potwierdza się środowiskowo silne zróżnicowanie społeczne?
- Czy komputer może być perspektywą dla uczniów z niedorozwojem?

Odpowiedzi na tego typu pytania naukowe mogą być pozyskiwane na drodze eksperymentów długofalowych (longitudinalnych), opartych na przemyślanym i ściśle przestrzegającym schemacie badawczym z użyciem testów istotności, na reprezentatywnych próbach uprawniających do uogólnień teoretycznych.

Proponowany model-wzorzec ewaluacji splotowej umożliwia stosowanie dowolnych **planów** badawczych, ujętych w trzy charakterystyczne grupy:

1. Eksperymenty manipulowania, porównywania, kontrolowania i uogólniania, np.: Jaki jest wpływ nadmiernej afektywności na wartościowość postawy?
2. Badania quasi-eksperymentalne nad związkami typu bodziec → reakcja, np.: Jak realizacja zajęć wpłynęła na zmiany osobniczych cech uczniów?
3. Badania przekrojowe z diagnozą właściwości typu obiekt → dyspozycja, np.: Jaką licealiści w danej fazie życia mają świadomość informatyczną?

Niezależnie od wyboru któregoś z powyższych, ogólnych planów badawczych, istotne jest użycie schematu porównawczego z przynajmniej trzema grupami kontrastowymi, a ponadto koniecznie należy zastosować techniki różnicowe pomiaru dystansowego, przeprowadzanego dwukrotnie w odstępie czasu.

Jeśli ewaluacja ma służyć ocenie wartości jakiegoś procesu społecznego, to ważne jest wyznaczenie przemian mentalnych podmiotów tegoż procesu, a więc wykrycie zmian, jakie zaszły pomiędzy pierwszym i drugim pomiarem.

Dzięki rozwarstwieniu zmiennej ogólnej i wykrywaniu zmian zachodzących w każdym z komponentów indykatoru można wnioskować m.in. o tym:

- Jakie symptomy dominowały w procesie, a jakie nikły w splocie oznak?
- Co było katalizatorem dynamizującym proces, a co hamulcem rozwoju?
- Jakie efekty przyniosło formowanie przemian i czy są one powtarzalne?
- Które z czynników oddziaływań edukacyjnych należałoby optymalizować?
- Czy deklarowano dalszą gotowość do poznawania bądź działania?

Jeśli ewaluacja ma służyć formułowaniu praw podstawowych, to można dodatkowo sięgać po wmontowane w ten model-wzorzec techniki wykrywania głębszych symptomów stylu wypowiedzenia się, odzwierciedlających to, czym kierowali się respondenci przy wyborze danej opcji kwestionariusza:

- skryptem gotowych ocen w reakcji na bodziec czy raczej namysłem,
- impulsem emocjonalnym czy też racjonalnie uzasadnioną przesłanką,
- płytką interpretacją bodźców-stwierdzeń czy staranną ich percepcją,
- pochoptością wyborów opcji czy spójną zasadnością powziętych decyzji.

Lista zagadnień badawczych może być tak duża, jak liczne są sploty między zmiennymi cząstkowymi, próbkowanymi przez kwestionariusz. Trzeba jednak wyraźnie zaznaczyć, że celem tej publikacji nie jest szczegółowa prezentacja problemów bądź planów badawczych ani upowszechnianie wyników, lecz jest to kompleksowa prezentacja koncepcji innowacyjnej strategii oceniania.

Podsumowując, zawarta w tej części pracy wykładnia teoretyczna dotyczy głównie metodologii ewaluacji, jest jej uzasadnieniem i podbudową naukową, jest zbiorem przesłanek objaśniających zagadnienia i problemy towarzyszące procesowi oceniania wartościującego, przy czym odniesienia do indykatoru *InfoKultury* stanowią egzemplifikację realnego odwzorowania modelu.

### **Strategie oceniania – symptomy i kwantyfikatory wartości**

Procesy społeczne są tak zawile, że nawet zawężenie pola badawczego do oceny wyłącznie edukacyjnych aspektów informatyzacji wcale nie uprości ewaluację. Najłatwiej byłoby policzyć komputery, ustalić liczbę pracowni, określić wymiar czasu przeznaczanego na kursy i zajęcia informatyczne, lecz przecież ilość nie jest jakimkolwiek wskaźnikiem jakości. Można ponadto rozpoznać jakość zainstalowanego i używanego oprogramowania, lecz byłoby to tylko oszacowanie wartości potencjalnej, a nie faktycznie eksploatowanej. Poza tym komputer może „służyć do wszystkiego” i jako taki w dużej mierze jest używany do zajęć nie mających nic wspólnego z formowaniem wartości. Z tych względów w niniejszej pracy głównym problemem ewaluacyjnym jest ocena procesów informatyzacji, wpływających na *rozwój zasobów ludzkich*, nazwanych tak i określanych jako jeden z priorytetów europejskich funduszy strukturalnych. Nie chodzi więc o *komputeryzację* w rozumieniu nasycenia

technologiami, sprzętem i oprogramowaniem, lecz o *informatyzację* w ujęciu przemian świadomości społecznej. Tę materię efektów trudno jest jakościowo należycie formować, a jeszcze trudniej jest rzetelnie mierzyć i wartościować.

Oceny formułuje się na podstawie zróżnicowanych strategii, m.in. poprzez:

- ocenianie sprawdzające, bazujące na kryteriach standardów wymagań;
- ocenianie różnicujące, wyznaczające empiryczne standardy osiągnięć;
- ocenianie zatwierdzające, wytyczające eksperckie standardy zaleceń;
- ocenianie wartościujące, ujawniające społeczne standardy oczekiwań.

Można wybrać jedną z metod, adekwatną do celu badawczego, lub też można próbować je konsolidować, ponieważ każda ma zarówno zalety, jak i wady<sup>10</sup>.

➤ **Ocenianie sprawdzające** – to najbardziej powszechny sposób wydawania osądów, zwłaszcza w zakresie oceny efektów edukacyjnych uzyskanych przez podmioty procesów kształcenia. W modelowej postaci ustala się arbitralne kryteria progowe wymagań, a następnie sprawdza indywidualnie, jaki poziom osiągnęła każda z osób. Niestety, najczęściej kryteria są wysoce subiektywne, zarówno co do poziomu progów, jak i przede wszystkim co do zakresu treści przedmiotowych (ustne egzaminy wstępne, szkolne odpytywanie, klasówki). Bardzo rzadko są obiektywne, dopracowane przez dużą grupę rzeczoznawców (jak np. wieloletni program *Nowej Matury*) i rzadko są jednolite, porównywalne dla całej populacji (jak np. testy końcowe w szkołach podstawowych i gimnazjach). Zwykle nie dba się o to, aby progi wymagań były możliwe do osiągnięcia przez znaczącą większość podmiotów kształcenia. Ponadto sami eksperci nieustannie spierają się i kwestionują, czy rzeczywiście to, czego nauczano, ma wartość edukacyjno-kulturową i czy właśnie z tego należało odpytywać. W ocenie procesów informatyzacji metoda sprawdzająca wykorzystywana jest do określania poziomu nabytych umiejętności posługiwania się komputerem oraz do weryfikacji przyswojenia treści i rozumienia pojęć informatycznych. Powinna być jednak zawsze uzupełniana ocenianiem wartościującym.

➤ **Ocenianie różnicujące** – to sposób wystawiania klasyfikujących ocen na bazie empiryczno-statystycznej *normalności rozkładu* rzeczywistych osiągnięć podmiotów kształcenia. W założeniach modelowych kryteria progowe ocen powinny być ustalane dopiero po przeprowadzeniu pomiaru (np. egzaminu), przy czym przyjmuje się, że rozkład gęstości uzyskanych ocen powinien być zbliżony do krzywej Gaussa. W tej metodzie niedopuszczalna jest sytuacja, w której wszyscy uczniowie w klasie z jakiegoś sprawdzianu otrzymali ocenę niedostateczną. Wiedza uczniów musi być statystycznie zróżnicowana, a więc, jeśli nauczyciel nie osiągnął zamierzonego efektu dydaktycznego, to powinien anulować oceny. Jeśli zaś sytuacja powtarza się, to... winien zmienić zawód. Oczywiście tak radykalne wnioski można snuć tylko z licznej statystycznie i reprezentatywnej próby badawczej, a wiadomo, że klasa szkolna nie spełnia

tego warunku. Z tego względu ocenianie różnicujące w czystej postaci rzadko stosowane jest w szkołach – częściej próbuje się połączyć obie metody poprzez elastyczną optymalizację progów kryterium oceniania w zależności od tego, jakie przeciętne wyniki uzyskują uczniowie. W badaniach cech osobowości natomiast celowo konstruuje się skale pomiarowe tak, aby posiadały dużą *moc dyskryminacyjną*, czyli zdolność do silnego różnicowania respondentów. Do oceny procesów informatyzacji bardziej przydatne jest różnicowanie grup, a nie pojedynczych respondentów, chodzi bowiem o wykrywanie oddziaływań nauczyciela na całe klasy lub grupy, a nie na wybranych uczniów.

➤ **Ocenianie zatwierdzające (walidacja)** – to formułowanie rozstrzygających osądów i postanowień, głównie na podstawie ekspertyz, zwykle jednak także uzależnione od trendów politycznych, realiów gospodarczych oraz dostępnych poziomów technologicznych. Ten rodzaj oceniania wydaje się być najlepszym kryterium orzekania co do wartości procesów informatyzacji oraz przesłanką podejmowania decyzji dotyczących priorytetów rozwoju społecznego. Niestety, nadmiernie często eksperci nastawieni są na spełnianie celów zdefiniowanych przez decydentów, którzy finansują walidację. Nadto, zarówno rzeczoznawcy, jak i decydenci uwikłani są w zjawiska lobbingsowe, wskutek czego dyrektywy normujące mogą nie być optymalnie dobre dla większości, a jedynie korzystne dla wąskich grup nacisku. Przykładem takiego zjawiska była np. próba wprowadzenia wymogu posiadania *Europejskiego Komputerowego Prawa Jazdy*. Pod pozorem unormowania kompetencji informatycznych zamierzano szkolić i egzaminować z takich treści, które nie miały wartości edukacyjnej, a tylko wspierały monopol twórców oprogramowania oraz tych, którzy z prowadzenia kursów komputerowych czerpią krociowe zyski. Podobne zastrzeżenia można mieć wobec zawirowań wokół nazwy przedmiotu nauczania (*informatyka* czy *technologia informacyjna*), wskutek czego nauczyciele musieli podejmować kosztowne studia podyplomowe. Walidacja edukacji jest zatem ocenianiem autorytarnym nie pozbawionym wad, lecz pełni ogromną rolę w organizowaniu kształcenia, w opracowywaniu programów, w formułowaniu standardów oraz w zatwierdzaniu innowacji. Warto jednak zawsze łączyć ją z oceną aprobaty społecznej dla powziętych decyzji, a więc z ewaluacją demokratyczną.

➤ **Ocenianie wartościujące (ewaluacja)** – to formułowanie osądów na podstawie wielorakich symptomów jakościowych, których wybór zależy od przyjętej strategii wyznaczania tego, co tworzy pożądaną *wartość*. W modelowym ujęciu jest to cykliczny proces, który najpierw obejmuje wstępne pomiary służące do wyznaczenia empirycznych standardów ewaluatywnych, stanowiących później kryteria normujące dla formułowania ocen z kolejnych pomiarów. W ewaluacji demokratycznej eksperci określają jedynie obszar tego, co ma być próbkowane i oceniane, lecz nie mogą narzucać tego, jakie poziomy powinny być osiągane.

W skrajnym przypadku może się zdarzyć, że dany standard ewaluacyjny jest sprzeczny z wartościami uniwersalnymi lub z normami moralno-etycznymi (np. presja społeczna na orzekanie i wykonywanie wyroków śmierci), dlatego wyniki ewaluacji warto konfrontować z konkluzjami walidacyjnymi. Pomimo tej potencjalnej niedoskonałości ewaluacja demokratyczna jest najbardziej pożądaną strategią oceniania wartości procesów edukacyjnych, obejmujących liczne rzesze społeczeństwa. Podmioty kształcenia muszą zatem mieć bardziej znaczący udział w opiniowaniu i formowaniu jakości tak długotrwałych oddziaływań instytucjonalnych, do jakich dochodzi w szkołach i na uczelniach.

Prezentowany w niniejszej pracy model-wzorzec ewaluacji splotowej jest właśnie jedną z bardzo wielu możliwych koncepcji oceniania wartościującego, chociaż sięga w wytyczonym zakresie również do technik znamienych dla innych, wyżej opisanych strategii. Mianowicie – ustala się normujące poziomy i wprowadza skalę ocen, lecz nie są one kryteriami zaliczania; analizuje się rozkłady empiryczne, lecz nie idealizuje i nie nagina do rozkładu normalnego; różnicuje się grupy badawcze, lecz nie różnicuje indywidualnie respondentów; ustala się i weryfikuje komponenty wartości, lecz nie narzuca ich z zewnątrz. Istotą tak rozumianej ewaluacji demokratycznej jest analiza walencyjnych zjawisk i cech, dających się usytuować na osi pozytywności|negatywności, z jednoczesną możliwością oszacowania natężenia danej wartości o większej liczbie poziomów niż tylko dwa skrajnie przeciwstawne bieguny +|-.

Kluczowym dylematem w każdej strategii ewaluacyjnej pozostaje zawsze **trafny wybór symptomów i kwantyfikatorów** wartości. W pierwszym rzędzie chodzi o wyznaczenie zbioru wskaźników tych zmiennych cząstkowych, które są najważniejszymi komponentami zmiennej ogólnej (tj. indykatum), a także o wytyczenie dominujących czynników, wpływających splotowo (w tym również przeciwstawnie) na całkowity wynik ewaluacji. Z tego wstępnie zakreślonego zbioru symptomów do skali pomiarowej wybiera się tylko te wskaźniki, które spełniają przyjęte kryteria selekcyjne – zwane kwantyfikatorami. Niewątpliwie najważniejszym kluczem doboru wskaźników jest to, czy mierzone przejawy są wartością rzeczywistą, czy tylko fasadową. Ze względu na olbrzymią rolę symptomów i kwantyfikatorów zdefiniuję oba te pojęcia bardziej szczegółowo.

**Symptomy wartości** w ewaluacji splotowej to uzewnętrznione, dające się obserwować i mierzyć wszelkie przejawy pozytywnych cech, stanów i zjawisk. Tworzą je komponenty pożądanых wzorców mentalnych: zarówno ich stany chwilowe, jak też dynamiczne procesy przemian postaw i uświadomienia.

W precyzyjnym operowaniu pojęciami zalecam używanie nazwy *symptomy* głównie do aspektów jakościowych, w odróżnieniu od terminu *oznaki*, którego używam dalej do określania przejawów wartości głównie w ujęciu ilościowym.



Symptomy są zbiorem celowo dobranych, reprezentatywnych i koniecznych do przebadania zmiennych jakościowych, konstytuujących indykatum, natomiast oznaki reprezentują empirycznie wyznaczone parametry zmiennych.

Ze względu na stosowaną metodologię splotową, powszechnie używane pojęcia: *zmiennych*, *wskaźników* i *parametrów* musiały zostać uzupełnione. Przyjęto, że symptomami są zarówno wskaźniki elementarne, a więc pewne znamienne wartości uzewnętrzniane wprost w wypowiedziach respondentów, a także wszelkie dostrzegalne prawidłowości w splotach wielu wskaźników, tworzących składniki, czynniki i wyniki połówkowe ewaluacji. To samo dotyczy oznak, które mogą być parametrami wycinkowymi bądź zagregowanymi. Generalnie symptomy wytyczają spektrum właściwości, które współtworzą oczekiwaną wartość, podczas gdy oznaki określają wymiary tejże wartości. Przy interpretacji symptomów w zasadzie wystarczyłaby analiza jakościowa, z ustaleniem znaku walencyjnego +|- każdego z komponentów indykatum, jednakże do pełnej interpretacji oznak niezbędna jest także analiza ilościowa.

**Kwantyfikatory wartości** w ewaluacji splotowej to filtrujące wyznaczniki *dobroci* w ujęciu normatywnym oraz w aspekcie obranej strategii badawczej, określające kryteria optymalnego doboru wskaźników do skali pomiarowej. Stosuje się je do wyodrębniania symptomów obiektywnie kwalifikujących się do kategorii *wartości* oraz do selekcji tylko oznak silnie wnioskotwórczych, które dają najbardziej znamienne, trafne i rzetelne przesłanki interpretacyjne.

Ważkim kwantyfikatorem w demokratycznym modelu ewaluacji procesów edukacyjnych jest **zgodność** wewnętrzna respondentów (podmiotów procesu) potwierdzona akceptacją zewnętrzną walidatorów (sędziów kompetentnych) co do tego, że każdy próbkowany symptom, będący komponentem indykatum, jest faktycznym (empirycznie potwierdzonym) przejawem wartości społecznie pożądanej. Próbkowanie oznacza, że bada się tylko wybrane wycinki większej całości, jako że pomiarem nie da się objąć wszystkich składowych wartości<sup>11</sup>. W profesjonalnej ewaluacji nie ma jednak sensu zajmowanie się oceną takich wartości, co do których istnieje niemalże jednomyślność. Szkoda wysiłku na badanie faktów oczywistych. Z tego powodu konieczne jest użycie kwantyfikatora przeciwstawnego, testującego **zróżnicowanie** wypowiedzi respondentów. Interferencja obu tych kwantyfikatorów powoduje, że w skali pomiarowej (po weryfikacji empirycznej i optymalizacji) pozostają tylko te wskaźniki, które mają swoistą zdolność *polaryzacji* opinii respondentów, polegającej zarówno na wyznaczaniu biegunów nasilenia badanej cechy, jak też na krystalizacji konsensu wokół tego, iż jeden z biegunów jest wartością, a drugi antywartością.

Ewaluacja splotowa nie zajmuje się przypadkami nieoznaczonymi, kiedy to wartością jest pluralizm, ani też wartościami ontologicznie uniwersalnymi.

Zajmuje się jedynie sytuacjami problemowymi i takimi procesami, w których dadzą się zaobserwować i obiektywnie zinterpretować przejawy społecznego formowania oraz kulturowego rozwoju standardów ewaluatywnych. Selekcję wskaźników spełniających opisane powyżej założenia ułatwia zastosowanie empirycznej miary *konkluzyjności*, która wyznacza przydatność pomiarową każdej z osobna pozycji kwestionariusza ze względu na siłę wnioskowania. Ma to szczególne znaczenie, gdy bierze się pod uwagę konieczność racjonalnego ograniczania liczby pozycji zamieszczanych w kwestionariuszu.

W przypadku ewaluacji procesów edukacyjnych niezbędne są ponadto kwantyfikatory empirycznie weryfikujące **selektywność** skali pomiarowej oraz **czułość** na zmiany zachodzące podczas oddziaływań wychowawczych. Wskaźniki zakwalifikowane do narzędzia pomiarowego muszą umożliwiać ocenę nawet niewielkich różnic pomiędzy grupami badawczymi na poziomie statystycznie istotnym, muszą także być wrażliwe na niezbyt wyraziste zmiany wypowiedzi. Dotyczy to zarówno indywidualnej *fluktuacji* wypowiedzi, jak też zmienności uogólnionej. Mierzone cechy (a zwłaszcza postawy) są z natury względnie trwałe, dlatego selektywność skali i czułość na przemiany społeczne weryfikować można tylko poprzez ewaluację długodystansową, z dwukrotnym pomiarem skutków oddziaływań zewnętrznych.

Warto przy okazji zwrócić uwagę na to, ażeby nie ulegać pokusie eliminowania wskaźników, które wykrywają zmiany niekorzystne. Przecież istotą ewaluacji splotowej jest próbkowanie właśnie takich zjawisk, które pozostają w powikłanych, niestałych i często niewspółbieżnych relacjach. Wynika stąd swoista przewaga opisanej tu metodologii w porównaniu z innymi metodami, np. analizą skupień, rzetelności, kanoniczną, czynnikową, dyskryminacyjną, w przypadku których ważniejsza jest silna korelacja poszczególnych pozycji tworzących skalę. W konsekwencji silnego dążenia do statystycznej trafności i rzetelności we wspomnianych powyżej metodach bezwiednie eliminuje się te wskaźniki, które ujawniają czynniki oddziałujące przeciwstawnie lub też wskazują wprost na to, że badane zjawiska zmierzały w różnych kierunkach.

W ewaluacji splotowej oszacowuje się globalną wartościowość procesów, przy czym do ogólnej oceny brane są zarówno wskaźniki przyrostowe, jak też predykcyjnie przewidywane i empirycznie potwierdzone jako spadkowe (np. obniżanie się poziomów motywacji po zaspokojeniu aspiracji). Przy wyborze tych bodźców-stwierdzeń, które są czułe na przemiany zachodzące w czasie, wykorzystuje się empiryczne miary *trendu zmian* i *efektu fluktuacji*. Oprócz kwantyfikatorów selekcyjnych *konkluzyjności* i *selektywności*, a także kumulacyjnej miary *trafności* pomiarowej, do optymalizacji skali przydatny jest też kwantyfikator buforowy *alfa* Cronbacha, uznawany często jako porównawczy wyznacznik jakości badań i traktowany jako miara empirycznej *rzetelności*.

<i>Ślady</i>		<i>Wzorce</i>
ocenianie sprawdzające	.....	standardy wymagań
ocenianie różnicujące	.....	standardy osiągnięć
ocenianie zatwierdzające	.....	standardy zaleceń, walidacja
ocenianie wartościujące	.....	standardy oczekiwań, ewaluacja
symptomy wartości	.....	przejawy cech, stanów i zjawisk
kwantyfikatory wartości	.....	filtrujące wyznaczniki dobroci skal

### **Charakterystyka ewaluacji – kluczowe metody i techniki**

Symptomy oraz kwantyfikatory strategiczne charakteryzują wybraną metodologię badawczą i decydują o tym, jak ma być konstruowane narzędzie pomiarowe. W pierwszej fazie budowy skali są to jednak przede wszystkim struktury pojęciowe, które należy zaimplementować (jak czynią informatycy) lub zoperacjonalizować (jak wolą badacze). W gruncie rzeczy sprowadza się to w obu przypadkach do opracowania zbioru procedur ścisłego postępowania. Generalnie chodzi o to, w jaki sposób zastosować swoiste dla danego modelu metody i techniki badawcze. Podczas implementacji skali pomiarowej, już na etapie konkretyzacji symptomów mierzonego indykatoru, najtrudniejszym zadaniem jest rozczłonkowanie zmiennej ogólnej na takie struktury, których połączenie odwzorowuje w miarę wiernie i kompleksowo cechę pierwotnie zagregowaną. Celowi temu służy *metoda ekstrakcji symptomów splotowych*. W tej samej fazie budowy profesjonalnej skali pomiarowej nieodzowne jest zastosowanie *metody czasoprzestrzennej ekspansji wskaźników*, polegającej na doborze par wskaźników kontrolnych, komplementarnych bądź dychotomicznych, próbujących zmienne cząstkowe w dwóch odmiennych, najlepiej dopełniających się wymiarach czasu lub przestrzeni.

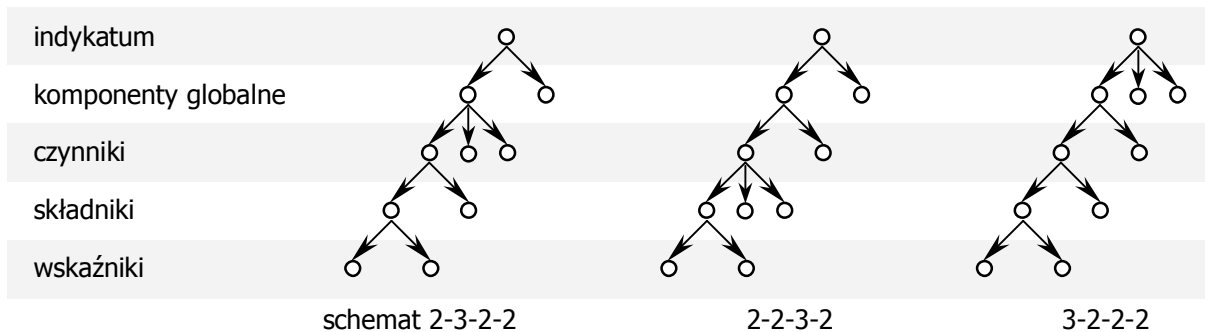
W strategii demokratycznej istotne jest ustalenie, które z pozycji skali są sformułowane w logice pozytywnej (te stwierdzenia respondenci powinni akceptować), a które w logice negatywnej (tym stwierdzeniom respondenci winni zaprzeczać). Wbrew pozorom wcale nie jest to łatwe. Rozpoznanie polaryzacji ułatwia *technika ustalania znaków walencyjnych*. W czynnościach kwantyfikacyjnych niezbędne jest zastosowanie *minimaksowej techniki ważenia zdań*, znamiennej dla propagowanego tu wzorca skali, polegającej na selekcji tylko tych wskaźników, które dają wystarczająco silne przesłanki do wniosku. Ażeby jednak pochopnie nie eliminować zbyt wielu pozycji w projektowanym narzędziu, do optymalizacji przydatna jest *technika justowania estymatorów*, polegająca na modyfikowaniu semantyki bodźców-stwierdzeń i sprawdzaniu, jakie to powoduje zmiany w statystyce oznak empirycznych. Ową czynność weryfikacji wskaźników skali wspomaga *technika dwuwymiarowej analizy*

wypowiedzi, dzięki której ustala się kwantyfikacyjną miarę rozziwu pomiędzy jakością opinii a intensywnością wypowiedzi respondentów. Z kolei *technika komparacji oznak komplementarnych* początkowo także służy do optymalizacji wskaźników, lecz później staje się ważną przesłanką interpretowania zjawisk.

W splotowej ewaluacji procesów edukacyjnych absolutnie niezbędna jest *technika dwukrotnego próbkowania stanów*, polegająca na pomiarach przed rozpoczęciem oraz po zakończeniu celowego i intensywnego oddziaływania na respondentów. Jeśli natomiast celem badań jest ocena właściwości alternatywnych projektów lub wytworów intelektualnych (filmów, multimediów itp.), wówczas korzysta się z *techniki porównywania obiektów podobnych*, której efektem jest wybór wariantu o najlepszych symptomach wartości. Podczas interpretacji oznak warto stosować *metodę koncygowania dwubieżnego*, polegającą na przechodzeniu przez materiał empiryczny w przeciwnych kierunkach, np. raz drogą wnioskowania indukcyjnego, a następnie dedukcyjnego.

Proces tworzenia skali pomiarowej jest trudny i długotrwały. Właściwie dopiero po kilku kompletnych cyklach badań pilotażowych można twierdzić, iż utworzono skalę. Pomiar pilotażowe przydatne są do rozpoznania materii badawczej, lecz przede wszystkim powinny służyć do weryfikacji wskaźników, optymalizacji bodźców-stwierdzeń oraz normalizacji narzędzia. Zaniechanie tych strategicznych powinności dopracowania narzędzia skutkuje aż takim przekłamaniami wyników, że nie pomoże nawet najrzetelniejszy eksperyment. Ze względu na znaczenie metod i technik, charakterystycznych dla obranej metodologii badawczej, ich opis w niniejszej pracy pojawia się wielokrotnie, w różnych ujęciach – od syntetycznych definicji aż po operacyjne desygnaty, wypełniające całe rozdziały zarówno w części teoretycznej, jak i praktycznej. W tym miejscu przedstawię pokrótce swoistość każdej z metod oraz technik obligatoryjnie stosowanych w ewaluacji za pomocą *Skali dwuważonych ocen*.

➤ **Metoda ekstrakcji symptomów splotowych** polega na opracowaniu takiej hierarchiczno-drzewiastej definicji indykatum, w której kolejne, coraz niższe poziomy desygnowania znaczeń są wyszczególnieniem pary, względnie triady symptomów współtworzących poziom definicyjnie ogólniejszy. Znamionym przykładem tego typu triady jest strukturalna definicja *postawy*, w której akcentuje się współlistnienie komponentów: afektywnych, kognitywnych oraz behawioralnych. W modelu-wzorcu ewaluacji splotowej zaleca się zgłębianie aż do czwartego poziomu szczegółowości. Oprócz komponentów globalnych, tworzących bardzo ważne interpretacyjnie tzw. skale połówkowe, wyodrębnia się też takie składowe, które są dominującymi czynnikami modyfikującymi indykatum. Z kolei dla każdego z czynników definiuje się komplementarne pary tworzących je istotnych składników, a dla każdego składnika dobiera kontrolne pary optymalnych, empirycznie weryfikowanych wskaźników (ryc. 1).



Ryc. 1. Trzy wersje ekstrakcji symptomów splotowych dla skali 24-pozycyjnej – dla przejrzystości narysowano tylko pojedyncze wiązki struktury

Możliwe są także inne schematy ekstrakcji, przy czym zaleca się, ażeby na najniższym poziomie umieszczać tylko pary kontrolne wskaźników, a nie triady. Dla schematów 2-4-2-2 lub 2-2-4-2 należy jednak zwiększyć liczbę pozycji w kwestionariuszu do 32, a dla schematu 2-3-3-2 aż do 36.

Strategiczną zasadą w ekstrakcji hierarchicznej jest to, że nie eliminuje się ze skali tych wskaźników, które mogłyby świadczyć o przeciwstawnym wpływie na wynik ogólny (co ma miejsce w skalach bazujących na kryterium trafności). Podczas budowy skal z symptomami splotowymi nie stosuje się znanych miar trafności – rzekomo sprawdzających, czy wskaźniki mierzą to samo indykatum, a de facto wskazujących jedynie na pewne „podobieństwo statystyczne”. Główną przesłanką formułowania treści oraz kwalifikowania bodźców-stwierdzeń do skali jest ich wartość semantyczna oraz implikacje wynikające z logiki formalnej, wiążącej zawieranie się treści zdań testujących symptomy niższego rzędu w taksonomii zdań próbkujących także symptomy nadrzędne, aż po indykatum. Właśnie stąd bierze się potrzeba tak bardzo precyzyjnego definiowania zmiennych, a ponadto – ze względu na strukturę hierarchiczną zmiennych – konieczność opracowania *schematu interpretacji* oznak empirycznych jeszcze przed pomiarami. W fazie analizy i interpretacji wyników stosuje się bowiem proces odwrotny do ekstrakcji, a więc integruje oznaki cząstkowe w wynik ogólny według wcześniej ustalonego schematu.

Ekstrakcja symptomów jest zagadnieniem dość trudnym, gdyż dotyczy rozwarstwienia posplatanych cech osobowych i zawiłych procesów umysłowych. W przypadku ewaluacji *InfoKultury* są to sploty takich dyspozycji, jak: postawy|uświadomienie, doznanie|poznanie, impulsywność|refleksyjność. Problemy biorą się stąd, że są to cechy ukrywane i procesy trudno uchwytnie, dlatego potrzebne są czułe wskaźniki pośrednie. Ponadto zmienne te silnie ze sobą interferują, utrudniając wyodrębnienie symptomów tylko danej cechy. Nieodzowna jest przy tym choćby elementarna wiedza z zakresu psychologii. Jednakże trud budowania skali nagradzany jest dzięki możliwości ewaluacji nie tylko deskryptywnej, lecz także formatywnej, z perspektywą formowania jakości procesów poprzez harmonizowanie czynników dysonansowych.

➤ **Metoda czasoprzestrzennej ekspansji wskaźników** polega na zamierzonym redagowaniu par bodźców-stwierdzeń tworzących skalę w taki sposób, aby próbkowane były cechy, stany i zjawiska z dwóch odmiennych perspektyw. Nie służy ona jednak do nadawania wyższej wagi tym oznakom, które ujawniają stany donioślejsze bądź bardziej niepokojące (jak u Seligmmana<sup>12</sup>), lecz służy do tworzenia bardziej wnioskotwórczego narzędzia pomiarowego, dzięki celowemu dobieraniu parami wskaźników kontrolnych, które później można analizować porównawczo, w tym również pod kątem stylu wypowiedzenia się. Pamiętać przy tym należy, że ewaluacja splotowa nie ma na celu oceniania cech osobniczych jednostek, lecz wyłącznie stanów uogólnionych na grupę. Ekspansja czasoprzestrzenna narzędzia ma szczególne znaczenie w ewaluacji procesów edukacyjnych, z góry bowiem zakłada się zmienność cech i stanów pomiędzy pierwszym i drugim pomiarem. Z tego względu w skali pomiarowej muszą znaleźć się zarówno wskaźniki retrospekcyjne, które próbują stany dotychczasowe, jak też futurospekcyjne, które mogą antycypować zjawiska przyszłe. Nie powinny natomiast znaleźć się bodźce-stwierdzenia mające taką wadę, że już w drugim pomiarze stają się bezprzedmiotowe lub nieaktualne wskutek upływu czasu. Próbkuje się więc splot: tak było | a jak będzie?

Na osi czasu bardzo istotne są też pary kontrolne: chwilowość | stałość. W tej parze jeden ze wskaźników próbkuje, czy badane zjawisko jest w opinii respondentów tylko przejściowe, czy może już uznane za trwałe, przy czym znak pozytywności zależy od treści bodźca-stwierdzenia. Na równi ważne są także sploty wskaźników testujących aspekty zasięgu przestrzennego, takie jak: globalność | lokalność oraz uniwersalność | ograniczoność. Pierwszy splot dotyczy określenia, czy badane zjawisko jest już szeroko rozprzestrzenione, czy może ujawnia się tylko w wydzielonym kręgu społecznym. Drugi splot odnosi się do wyznaczenia, czy dane zjawisko ma charakter wszechstronny, interdyscyplinarny, czy raczej jego zasięg jest wąski a sens interpretowalny jedynie w wybranej dziedzinie wiedzy specjalistycznej.

W ewaluacji splotowej, opierającej się w znacznym stopniu na badaniu cech mentalnych, można jednak zaryzykować twierdzenie, że najistotniejszą parę kontrolną stanowi splot: introspekcja | ekstraspekcja, testujący aspekty personalizacyjne. W większości innych skal pomiarowych dominują pytania próbujące wewnętrzne stany respondenta. Jest to odpowiednie, gdy chodzi o badania mające na celu ustalanie rezultatów indywidualnych. Tymczasem, w demokratycznej ewaluacji procesów edukacyjnych zmiany cech i stanów mentalnych są jedynie pośrednimi wskaźnikami jakości oddziaływań. Stąd na równi ważne z własnymi odczuciami jest to, jak szacujemy poziom zadowolenia innych współuczestników procesu – podmiotów kształcenia. Warto więc, aby każdy ze wskaźników testujących dyspozycje własne uzupełniony

był kontrolnym wskaźnikiem dyspozycji całej grupy, ogółu lub nawet populacji. Takie układy par w szerszym ujęciu reprezentują splot zależności: ja | inni.

Ekspansję czasoprzestrzeni realizuje się w warstwie semantycznej przy opracowywaniu bodźców-stwierdzeń do nowego kwestionariusza. W trakcie optymalizacji skali należy uzupełniać te wskaźniki tworzące pary kontrolne, które w wyniku zastosowania kwantyfikatorów muszą być usunięte. W fazie analizy i interpretacji rezultatów ewaluacji powinno się pamiętać o tym, że odtworzenie zjawisk czasoprzestrzennych jest możliwe jedynie na poziomie wskaźników, a do statystyk z tego poziomu nie można mieć dużego zaufania. O ile więc prosta analiza wskaźników ma znaczenie wtórne, o tyle ekspansja przy tworzeniu narzędzia ma kapitalne znaczenie dla późniejszego uzyskania pełniejszego obrazu powikłanej rzeczywistości. Dzieje się tak dzięki agregacji zróżnicowanych, wielowymiarowych oznak elementarnych w związku splotowe, dostarczające po uogólnieniach wierniejsze odwzorowanie stanu faktycznego, a tym samym dające bardziej rzetelne wyniki ewaluacji.

➤ **Technika ustalania znaków walencyjnych** polega na przypisaniu każdemu ze wskaźników skali znaku + lub –, w zależności od tego, czy dany bodziec-stwierdzenie wyraża i próbkuje pozytywność, czy negatywność. Większość badaczy niesłusznie uważa tę czynność za najłatwiejszą i niemal automatyczną, co jednak świadczy o nieuprawnionym i autorytarnym podejściu do ewaluacji. Przecież badacz może się pomylić, co w wypadku zamiany znaku walencyjnego spowoduje w konsekwencji sformułowanie wniosków dokładnie przeciwstawnych niż wynikałoby to z materiału empirycznego. W ewaluacji demokratycznej za kluczową przesłankę decyzji zwykle przyjmuje się opinię większości, jednakże w wielu sytuacjach wcale to nie wystarcza, aby jednoznacznie stwierdzić: co jest wartością, a co antywartością. Wyjaśniałem to zagadnienie już wcześniej, przy opisywaniu symptomów i kwantyfikatorów wartości. W tym miejscu skoncentruję się głównie na taktyce postępowania.

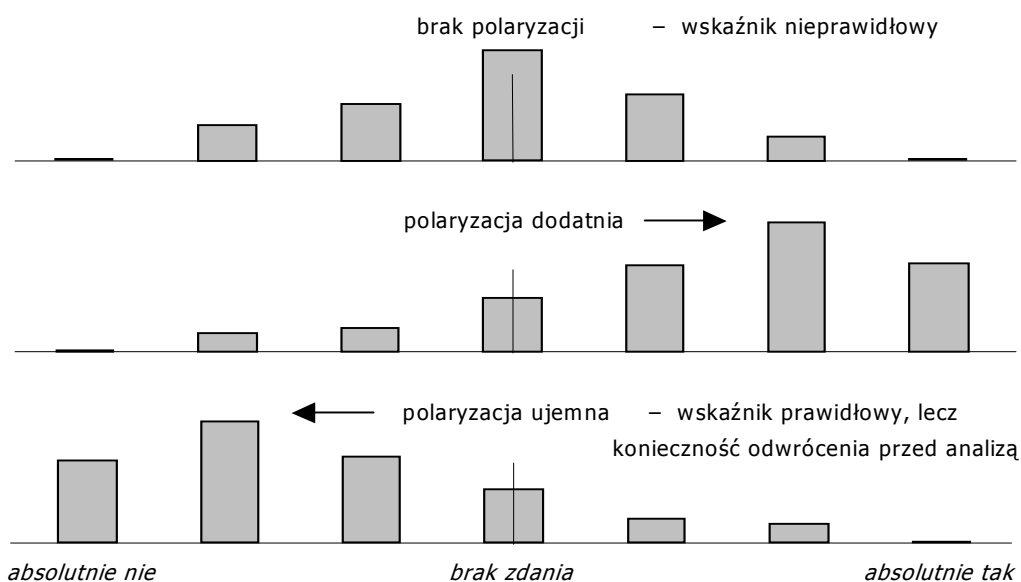
Podczas konstruowania kwestionariusza układa się specyficzne zdania, mające takie właściwości, że stwierdzenia w nich zawarte nie tylko pobudzają do wypowiedzi akceptującej bądź negującej, lecz także niosą w semantycznej warstwie pewne przesłanie wartościujące – czasem jawne, a częściej ukryte. Pozycje kwestionariusza o takich cechach zwiemy *bodźcami-stwierdzeniami*. Nie są one pytaniami, lecz swoistymi tezami, do których respondenci powinni ustosunkować się podczas pomiaru. Ze względu na potrzebę wyeliminowania automatyzmu wypełniania kwestionariusza stosuje się taką zasadę, że część zdań formułowanych jest na założeniu tez fałszywych, którym badani winni zaprzeczać\*. Sprawa byłaby prosta, gdyby stwierdzenia dotyczyły zagadnień

---

\* Przykładem bodźca spolaryzowanego ujemnie, wyzwającego zaprzeczenie respondentów, jest sformułowanie – *Komputer (hardware) może radzić, jak rozwiązywać życiowe problemy*.

oczywistych, jednakże badania prowadzi się właśnie wokół kwestii problemowych, wobec których trudno jest wskazać kierunek wartości | antywartości\*.

W procesie ewaluacji wartościowanie jest celem nadrzędnym, dlatego po empirycznej weryfikacji narzędzia eliminuje się ze skali w pierwszym rzędzie te bodźce-stwierdzenia, które wykazują brak wyrazistej *walencji*, rozumianej jako zdolność do polaryzowania wartości. Można te przypadki przyrównać do sytuacji pata w głosowaniu, kiedy to każdą z opcji popiera bardzo zbliżona liczba zwolenników. Po usunięciu tych niemal dosłownie „bezwartościowych” wskaźników, pozostałym pozycjom skali nadaje się znak polaryzacji zależny od tego, czy większość respondentów akceptowała tezę [+], czy negowała [-]. Przy podejmowaniu tego typu decyzji warto skorzystać z histogramów (ryc. 2).



Ryc. 2. Ustalanie polaryzacji bodźców-stwierdzeń na podstawie histogramów

Weryfikacja empirycznych rozkładów wypowiedzi jest niezwykle istotną fazą ustalania polaryzacji bodźców-stwierdzeń. Wyznaczone w ten sposób znaki walencyjne porównuje się z przewidywaniami wynikającymi z analizy warstwy semantycznej zdań. Jeśli oczekiwania, związane z tym, jakie naszym zdaniem opinie powinni wyrażać respondenci, różnią się od ich preferencji, to konieczna jest wnikliwa analiza treści danego bodźca-stwierdzenia. Przede wszystkim należy rozstrzygnąć, o czym w głównej mierze świadczy wskaźnik:

- czy o jakości pożądanej cechy mentalnej, czy raczej o stanie świadomości odnoszącej się do obiektywnie zróżnicowanej cechy zjawiska zewnętrznego♦;

\* Np. stwierdzenie – *Pewnego dnia komputery zniewolą ludzi* – implikuje różne wypowiedzi, gdyż wiemy, że wciąż rosną rzesze uzależnionych, lecz chcielibyśmy, aby tak się nie działo.

♦ Np. stwierdzenie – *Telewizja mnie nudzi* – implikuje wypowiedź TAK, choć te same osoby chętnie oglądają TV. Trudno więc przyjąć wskaźnik ten za symptom negatywnych postaw.



- czy o negatywnej postawie wobec obiektu autentycznie wartościowego, czy raczej o tym, że nie jest to w zasadzie obiekt wart oczekiwanej akceptacji;
- czy o faktycznie negatywnej dyspozycji gotowości do działania, czy jedynie o upozorowanej niedyspozycji ze względu na i tak już silne zaangażowanie\*;
- czy o niskim poziomie poznawczym w czasie pierwszego pomiaru (jeszcze przed procesem edukacyjnym), co przecież ulega poprawie wskutek zajęć;
- czy o rzeczywiście niekorzystnej niszy emocjonalnej, czy raczej o stanie, który świadczy o pozytywnej intencji nadrobienia dystansu wobec innych♦;
- czy o stresie spowodowanym zbyt małą sprawnością obsługi komputera, czy może o pełnej świadomości faktu, że jest to medium silnie uzależniające.

To zaledwie nieliczne przykłady dylematów, z jakimi spotyka się twórca skali podczas ustalania związków pomiędzy wskaźnikami a symptomami wartości.

W skali splotowej wskaźniki postaw mają zwykle nieco inne właściwości niż wskaźniki uświadomienia, podobnie jak symptomy cech emocjonalnych różnią się od wyznaczników komponentów poznawczych. Jednak w pewnych sytuacjach, kiedy to bodziec-stwierdzenie pobudza jednoczesną reakcję silnie zespolonych składników afektywnych i kognitywnych, mogą wręcz wystąpić nie tylko trudności z ustaleniem, czego wyznacznikiem jest dany wskaźnik, lecz mogą też pojawić się kłopoty z ustaleniem znaku walencyjnego. W tych pełnych wątpliwości przypadkach warto wyeliminować wskaźnik ze skali lub zmodyfikować warstwę językową bodźca-stwierdzenia. Kończącą bowiem fazą ustalania polaryzacji każdego wskaźnika musi być jednoznaczna zgodność znaku walencyjnego ustalonego empirycznie ze znakiem wyprowadzonym na drodze definicyjno-semantycznej, a co więcej – całe narzędzie zaakceptować powinni kompetentni sędziowie w procesie zwanym *walidacją skali*.

➤ **Minimaksowa technika ważenia zdań** polega na wyznaczaniu optymalnych wskaźników tworzących skalę, ze względu na ich przydatność wnioskotwórczą, głębię problemową, a jednocześnie rzetelność pomiarową. Procedurę *ważenia* wypowiedzi respondentów przeprowadza się w każdym cyklu weryfikacyjno-optymalizacyjnym tworzenia nowej skali jako strategiczną metodologicznie czynność, dającą przesłanki do wyodrębnienia empirycznie najdoskonalszych bodźców-stwierdzeń. Zawsze bowiem podczas tworzenia narzędzia układa się nadmiar logicznych, choć intuicyjnych tez, tworzących pozycje kwestionariusza, po czym – na postawie analizy rezultatów uzyskanych z kilku cykli pomiarów przedwstępnych – eliminuje się tezy nieznajdujące potwierdzenia w opiniach

---

\* Np. stwierdzenie – *Chcę więcej czasu przeznaczyć na naukę* – spowoduje wypowiedź NIE właśnie u najbardziej pracowitych uczniów, pomimo ich wysokiej gotowości do uczenia się.

♦ Np. stwierdzenie – *Czuję się źle, gdy inni wiedzą więcej ode mnie* – wcale nie musi być wskaźnikiem niepokojącym, może bowiem oznaczać odczuwanie stymulacji do działania.

respondentów. To jednak w profesjonalnej ewaluacji nie wystarcza. Konieczny jest kwantyfikatory jakościowy bodźców-stwierdzeń tworzących skalę. Funkcję tę wobec każdego z osobna wskaźnika pełni wyznacznik tzw. *konkluzyjności*.

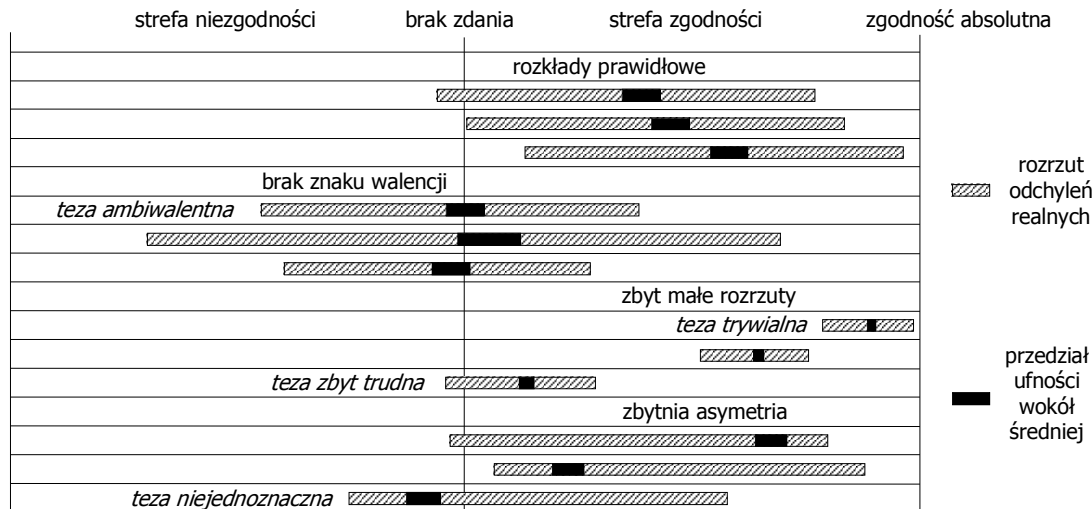
Pomińmy na razie statystyczne znaczenie tego parametru i spróbujmy zrozumieć, na czym polega technika ważenia minimaksowego. Otóż z jednej strony chodzi o sprawdzenie właściwości każdej z osobna pozycji kwestionariusza tak, ażeby trafnie dawało się ustalać preferencje respondentów co do znaku walencyjnego. W tej sprawie najlepsza byłaby jednomyślność wypowiedzi. Jednak z drugiej strony, ze względu na wartość poznawczą wyników badań niezbędne jest jak największe zróżnicowanie uzyskanych wypowiedzi. Powstaje konflikt oczekiwań wobec tak dysonansowych właściwości splotu: zgodność | rozziew zdań. W sytuacji, gdy przyrost jednego parametru powoduje obniżenie innego, lecz na równi ważnego parametru – sięga się właśnie po *technikę minimaksową*, która umożliwia znalezienie optymalnego kompromisu.

Poszukiwanie tzw. *minimaksa* jest żmudne, gdyż odbywa się w długofalowym cyklu selekcji i preredagowywania bodźców-stwierdzeń, z empiryczną weryfikacją efektów statystycznych, jakie wywołuje owo dostrajanie narzędzia. Przede wszystkim w pierwszym podejściu buduje się większą liczbę bodźców-stwierdzeń niż wynika to z wymaganej pozycji skali. W środkowej fazie warto podjąć próbę modyfikacji treści stwierdzeń oraz przeprowadzić optymalizację zwaną *techniką justowania estymatorów*. Sama bowiem eliminacja pozycji ze skali szybko prowadzi do tego, że nie tylko zabraknie stwierdzeń, lecz także zabraknie pomysłów na inne sformułowania tych samych treści. W końcowej fazie ważenia zdań odrzuca się wszystkie te pozycje skali, które nie generują wyraźnego znaku polaryzacji walencyjnej, a także wszystkie te tezy, wobec których wypowiedzi są zbyt jednomyślne lub bez wyjątku zgodne. Ponadto niekorzystne są również takie bodźce-stwierdzenia, które dają w efekcie silnie asymetryczne rozkłady wypowiedzi. W podejmowaniu prawidłowych decyzji co do ustalenia znaku walencyjnego lub odrzucenia wskaźnika niezbędny jest wykres rozrzutów empirycznych, bazujący na odchyleniach realnych (ryc. 3)\*.

Dobrze sformułowane bodźce-stwierdzenia kwestionariusza dają w efekcie rozrzuty wypowiedzi rozpostarte w miarę szeroko w strefie zgodności, przy czym dopuszczalne jest częściowe zachodzenie także na strefę niezgodności. Niedopuszczalna jest natomiast sytuacja, kiedy rozrzut obejmuje w podobnym zakresie obie strefy: zgodności i niezgodności. Nie można wówczas wyznaczyć standardu ewaluatywnego, a bez znaku walencyjnego nie wolno przyjmować, jakoby mierzona oznaka była przejawem wartości ukształtowanej społecznie.

---

\* Wykresy pasmowe rozrzutu ilustrują, w jakim zakresie mieści się około 68% respondentów. Sposób interpretacji tego typu rycin objaśniam m.in. w rozdziale: *Jak odczytywać wykresy...*



Ryc. 3. Ważenie zdań na podstawie empirycznego rozrzutu wskaźników

Przypadki zbyt małych rozrzutów świadczą o tym, że próbkowana właściwość jest oczywista, więc nie ma sensu jej badać. Wyjątkowo, w sytuacji skupienia blisko osi 'brak zdania', może okazać się, iż próbkowano problem tak trudny, że badana zbiorowość respondentów nie jest w stanie niczego rozstrzygnąć. Brak symetrii rozkładów w zasadzie nie obliguje do odrzucania wskaźników, chyba że zachodzi szczególnie niekorzystny przypadek niejednoznaczności.

➤ **Technika justowania estymatorów** polega na starannym modyfikowaniu treści bodźców-stwierdzeń tworzących skalę w celu uzyskania optymalnych rozkładów i parametrów pozycyjnych, oczekiwanych od każdego wskaźnika. Jest to niezwykle istotna czynność dostrajania skali, silnie związana z *mini-maksową techniką ważenia zdań*, połączona z empiryczną weryfikacją statystycznych skutków, jakie wywołuje doskonalenie warstwy językowej twierdzeń. Istotne przy tym jest zrozumienie, że nie chodzi tutaj o wpływanie na wyniki ewaluacji, lecz o kalibrację polegającą na dopasowaniu zakresu pomiarowego skali do rzeczywistego rozrzutu wartości wypowiedzi. Rozwinę ten problem szerzej, gdyż jest on przyczyną często popełnianych błędów metodologicznych w sondażach analizujących zjawiska społeczne. Otóż nagminnie prowadzone są quasi-badania, polegające na ułożeniu pewnej liczby pytań i po zebraniu odpowiedzi opublikowanie, że daną opcję w pewnej kwestii wybrało np. 76% a w innej kwestii aż 84% respondentów, z płynącym stąd wnioskiem, jakoby „to coś drugie było lepsze”. Osoby układające ankiety nie zadają sobie trudu analizy językowej zdań, aby wykluczyć sytuacje, w których wybór opcji zależy od właściwości pytania, a nie od rzeczywistego wskaźnikowania indykatum. Badacze bezwiednie ulegają fasadowej magii procentów oraz manipulacyjnej socjotechnice, jaka stosowana jest z premedytacją np. podczas sondowania opinii publicznej przez mass media. Po pierwsze – pytania dotyczące identycznej kwestii, lecz różniące się nawet nieznacznie (choćby jednym wyrazem)

powodują odmienną reakcję respondentów, a to wpływa na ich wypowiedzi. Po drugie – brak odniesienia do jakichkolwiek norm nie uprawnia badacza do wnioskowania, że większość 76% bądź 84% głosów wystarcza. Przecież może właśnie w konkretnej kwestii (zwłaszcza wobec wartości uniwersalnej) powinno być ponad 90%? Po trzecie – celowe uwikłanie treści pytania w kontekst problemowy obniża poziom uważany powszechnie za normę społeczną\*. Wszystko to powoduje, że nie można beztrudno przenosić mocy indagującej pytania z warstwy stylistycznej na poziom parametrów statystycznych. Jeśli wypowiedziom w formie twierdzeń słownych przypisujemy rangi, to mają one tylko sens porządkujący; jeśli zaś nadajemy im umowne wagi, to mogą one służyć wyłącznie do porównywania zjawisk w różnych, lecz równoważnych próbach badawczych (jak w eksperymencie), bądź w dwóch pomiarach tej samej grupy osób (jak w panelu). Niestety, w wielu raportach z badań spotyka się ilościowe porównywanie efektów oddziaływania różnych pytań i liczbowe wyrażanie wyników, mimo braku przesłanek uprawniających do takich działań.

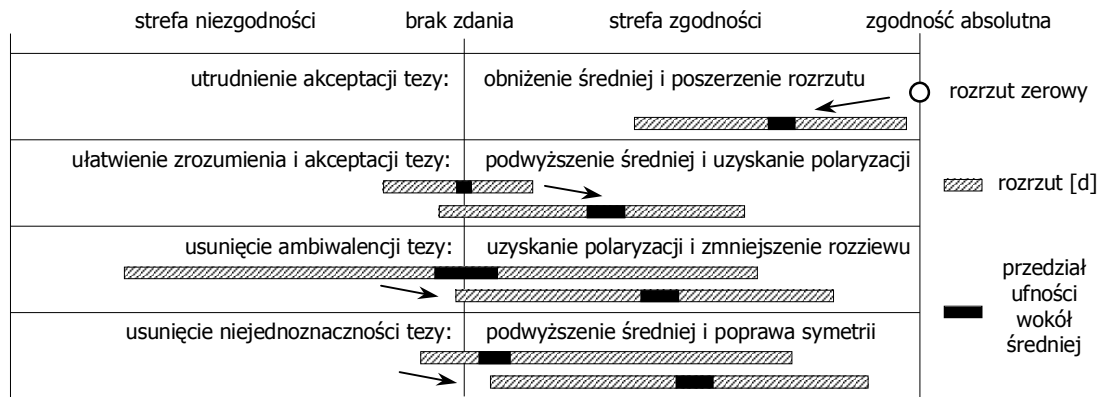
Skoro zatem poziomy bezwzględne cech mentalnych są niemierzalne, a ponadto wagi przypisane tym cechom mają całkowicie umowny charakter, to dopuszczalne jest (i wręcz niezbędne na etapach optymalizacji skali) takie doskonalenie językowe bodźców-stwierdzeń kwestionariusza, aby skutkowało ono optymalnym usytuowaniem miar statystycznych, będących rezultatami pomiarów empirycznych. W konsekwencji takiego typu zabiegów justowania każdy wskaźnik skali pomiarowej dopasowywany jest możliwie jak najlepiej do rzeczywistej właściwości konkretnej cechy, którą ów wskaźnik próbkuje. Możliwe jest to dzięki silnemu sprzężeniu zwrotnemu pary: bodziec|reakcja. W zasadzie cała istota justowania *estymatorów* (czyli wartości statystycznie oczekiwanych) opiera się na bogactwie językowych środków wyrazu, wykorzystywanych podczas redagowania treści bodźców-stwierdzeń. Wyjaśniam to dokładnie w rozdziałach dotyczących budowy i optymalizacji kwestionariusza. Tutaj skoncentruję objaśnianie techniki justowania na jej zasadniczym celu, a mianowicie na osiągnięciu zbliżonych do ideału charakterystyk statystycznych.

W prezentowanym wzorcu skali zakłada się pomiar takich cech, które mają znaki walencyjne wartości|antywartości, w związku z czym wypowiedzi mogą układać się w obu strefach zgodności|niezgodności ze standardem. Standardy ewaluatywne powstają w ten sposób, że kwalifikowana większość respondentów opowiada się właśnie za nimi. Jeżeli jednak reakcja respondentów na bodziec-stwierdzenie w zdecydowanej większości jest nadmiernie akceptująca, to średni poziom akceptacji zbliża się zbyt do górnego kresu

---

\* Zważmy na przykład, jak bardzo zaniżana jest ocena wartości życia w przypadku osobistej tragedii, gdy bliska osoba cierpi fizycznie, walcząc przez długi czas z nieuleczalną chorobą.

skali. W skrajnym przypadku może dojść do zgodności absolutnej, kiedy to wszyscy w próbie badawczej wybiorą skrajną opcję 'absolutnie...' (zob. ryc. 4). Takie sytuacje zerowego rozrzutu wypowiedzi są niedopuszczalne ze względu na analizę rozkładową stosowaną podczas statystycznej interpretacji oznak\*. Należy wówczas zmienić treść tezy w taki sposób, ażeby trudniej było z nią się zgadzać. Średni poziom akceptowania zmodyfikowanej tezy w kolejnym pomiarze powinien obniżyć się, dając wypowiedzi prawidłowo zróżnicowane.



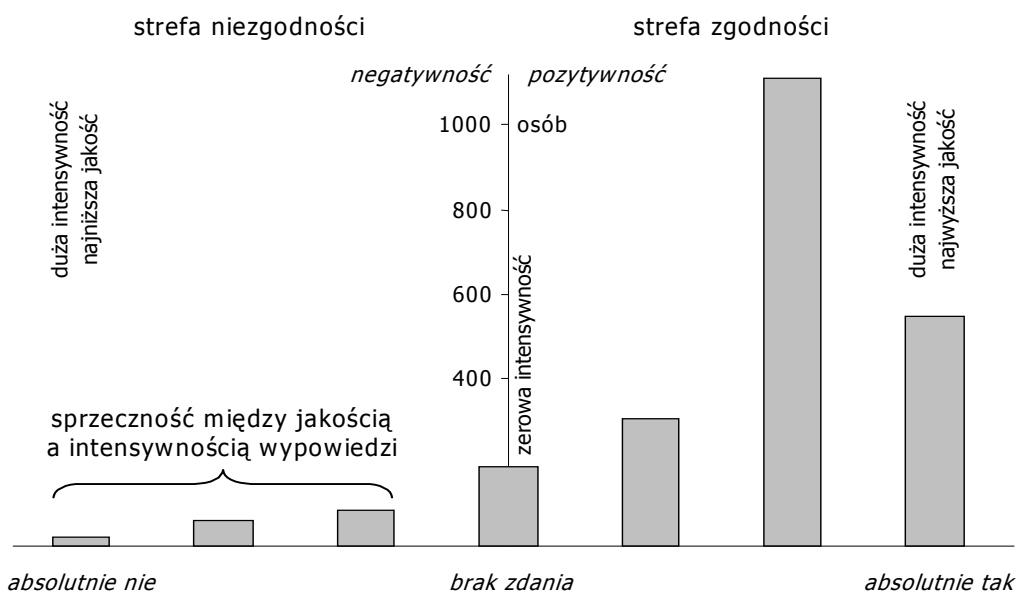
Ryc. 4. Justowanie estymatorów w celu unormowania wskaźników

Inne zjawisko niekorzystnie usytuowanego rozkładu empirycznego zdarza się wtedy, gdy teza jest zbyt trudna do ustosunkowania się przez respondentów, bądź też gdy stwierdzenie znajduje tyle samo zwolenników, co przeciwników. W takiej sytuacji należy językowo doprecyzować sformułowanie tezy, a przede wszystkim zadbać o to, ażeby stwierdzenie wyraźnie polaryzowało przewagę osób opowiadających się za opcjami zgodnymi z symptomami pozytywności. Optymalnie dobrana treść bodźców-stwierdzeń indukują takie wypowiedzi, że ich rozkłady wypełniają równo strefę zgodności, a średnie arytmetyczne z wag przypisanych opcjom wypowiedzi usytuowane są w pobliżu środka tej strefy.

Badania właściwe można podejmować dopiero po zakończeniu procesu optymalizacji i normalizacji skali pomiarowej. Od tej pory nie wolno już modyfikować treści bodźców-stwierdzeń, gdyż prowadziłoby to do nieuprawnionej manipulacji wynikami. Gdyby jednak w przyszłości zachodziła konieczność modyfikacji, wynikająca na przykład ze zmiany standardów ewaluacyjnych lub ze zmiany znaków walencyjnych, wówczas należy odpowiednio przystosować narzędzie, jednakże bez możliwości sięgania do danych uzyskanych w pomiarach za pomocą wcześniejszej wersji skali. Nawet bowiem niewielka zmiana treści bodźca-stwierdzenia lub choćby inna kolejność pozycji w kwestionariuszu może spowodować istotne zmiany parametrów statystycznych.

\* Empirycznie wyznaczony rozrzut wypowiedzi [d] nie powinien być zerowy, gdyż – jako miara standaryzująca rozkład – jest w wielu wzorach statystycznych umieszczany w mianowniku.

➤ **Technika dwuwymiarowej analizy wypowiedzi** polega na porównywaniu rzeczywistej jakości merytorycznej wyborów dokonanych przez respondentów z poziomem intensywności ich samooceny tego, na ile pewni są swych racji. Neutralną opcją wypowiedzi w skali dwuważonej jest '*brak zdania*' – taki wybór oznacza zerową intensywność. Najbardziej natomiast intensywne są wybory skrajne, świadczące o całkowitej akceptacji lub negacji danego stwierdzenia – zarówno opcje '*absolutnie tak*', jak i '*absolutnie nie*'. Intensywność wypowiedzi respondenta może jednak pozostawać w sprzeczności z niesłusznością zajmowanego stanowiska. Taką sytuację, kiedy to wybór opcji jest niezgodny ze znakiem walencyjnym pozytywności, nazywać będziemy *rozziewem* (ryc. 5).



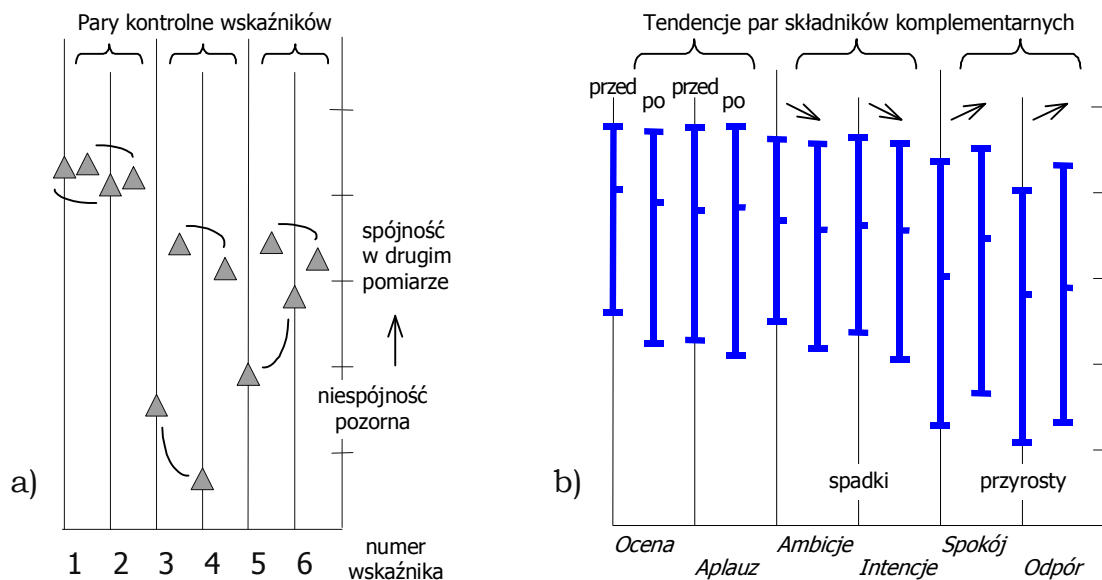
Ryc. 5. Histogram ilustrujący zjawisko rozziewu wypowiedzi

Z rozziewem wobec konkretnego wskaźnika mamy do czynienia wówczas, gdy część respondentów nie zgadza się z tezą akceptowaną przez wystarczającą większość badanych. W zależności od tego, jak liczna jest owa opozycja oraz jak intensywne są sprzeciwy wobec treści danego bodźca-stwierdzenia, mamy do czynienia z mniejszą bądź większą rozbieżnością opinii. Przyjmuje się, że opinia kwalifikowanej większości respondentów tworzy normę wartościującą, zwaną *standardem ewaluatywnym*. Respondenci mają prawo do odmiennych zdań, lecz w ujęciu wartościującym ich wypowiedziom przypisuje się niższą jakość, co może oznaczać niski poziom świadomości lub negatywną postawę.

Nasilenie zjawiska dysjunkcji wyraża statystyczny parametr *Rozziewu*, który z jednej strony charakteryzuje osobniczą niezgodność próby badawczej, a z drugiej służy jako ważny komponent *Konkluzyjności*, będącej miarą siły wnioskotwórczej wskaźników. W tym strategicznym wyznaczniku zawarta jest cała istota dwuwymiarowej analizy wypowiedzi. Po sprawdzeniu rozkładów

każdego wskaźnika, po weryfikacji graficznej – ostatecznym kwantyfikatorem rozstrzygającym o poprawnym sformułowaniu tez w bodźcach-stwierdzeniach jest empirycznie wyznaczona siła wnioskowania każdej z osobna pozycji skali. Im wyższa konkluzyjność wskaźnika, tym lepiej sformułowano stwierdzenie w warstwie językowej. W praktyce parametr ten przy prawidłowych wskaźnikach osiąga wartość do 0,7 i najczęściej mieści się w przedziale od 0,5 do 0,7. Dla narzędzi wzorowanych na *Skali dwuważonych ocen* w zasadzie przyjmuje się wartość progową nie niższą niż 0,5. Jednak w pojedynczych przypadkach dopuszczalne jest pozostawienie wskaźnika o nieco niższej konkluzyjności, jeśli uzasadnia to jakaś ważna przesłanka wynikająca z założeń badawczych lub gdy rozkład zdań jest niekorzystny tylko w pierwszym pomiarze. Powstaje stąd konieczność przeprowadzania prac weryfikacyjno-optimalizujących skalę w kompletnych cyklach dwupomiarowych, w pełnej procedurze pilotażowej. Warto ponadto zwrócić uwagę na silne powiązania wyżej opisanych technik, przy czym – o ile justowanie estymatorów polega na operowaniu bodźcami, o tyle analiza dwuwymiarowa ujawnia reakcje właśnie na owe pobudzenia.

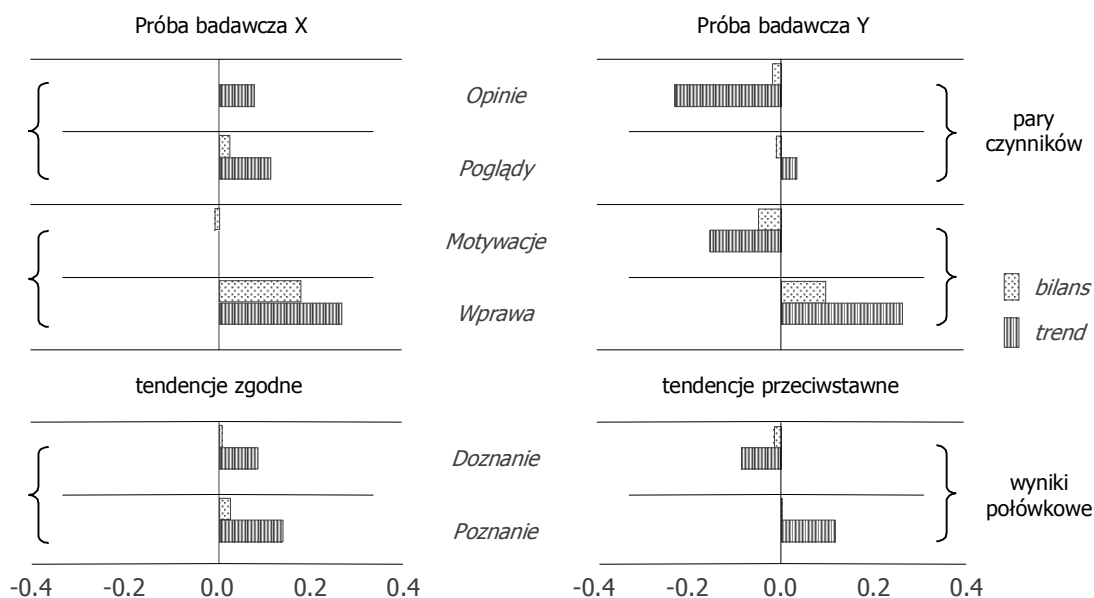
➤ **Technika komparacji oznak komplementarnych** polega na porównywaniu spójności empirycznie wyznaczonych statystyk par wskaźników kontrolnych. Na etapie weryfikowania skali pomiarowej celem komparacji jest ustalenie, czy ułożone w pary bodźce-stwierdzenia implikują wypowiedzi wewnętrznie zgodne, a więc takie, w których – gdyby analizować pojedynczo respondentów – okazałoby się, że wybierane przez nich opcje są logicznie konsekwentne. W pierwszej kolejności jest to więc czynność ustalania, czy kontrolna para wskaźników trafnie wyznacza jakiś mający wspólne cechy przejaw indykatum, np. zgodność poziomów bądź zbieżność zmian (porównaj parę 1-2 na ryc. 6a).



Ryc. 6. Komparacja: a) poziomu wskaźników, b) współbieżności składników

Oczywiście pamiętać należy o specyfice ewaluacji splotowej, w której tworzy się pary wskaźników celowo próbujących dwie odmienne perspektywy zjawisk, stąd bardzo istotne znaczenie ma analiza, czy ewentualna niespójność jest skutkiem niekonsekwencji wypowiedzi, czy raczej wynika zupełnie naturalnie z przeciwstawnych aspektów czasoprzestrzennych. W tym drugim przypadku należy uznać niespójność za pozorną, zazwyczaj zniwelowaną już w pomiarze końcowym (porównaj pary wskaźników 3-4 i 5-6 na ryc. 6a).

W kolejnej fazie komparacji oznak sprawdza się wewnętrzną spójność par składników, gdyż właśnie one tworzą komponenty następnego poziomu zmiennych. Weryfikuje się więc przez porównywanie, czy komplementarne składniki mają podobne właściwości i podlegają współbieżnym przemianom, a przez to – czy trafnie dobrane są jako symptomy czynników wpływających na indykatum (ryc. 6b). Jednakże już na tym poziomie analizy należy być szczególnie ostrożnym w oponowaniu przeciwko ewentualnej niespójności, ze względu na celową zasadę wieloaspektowego próbkowania powikłanych cech splotowych. Właśnie z tego powodu, na najwyższych poziomach interpretacji czynników i wyników połówkowych, każde dostrzegalne zróżnicowanie parametrów par komplementarnych bądź odmienne kierunki przemian stanowią przesłanki do pełniejszego i rzetelniejszego wnioskowania (ryc. 7).



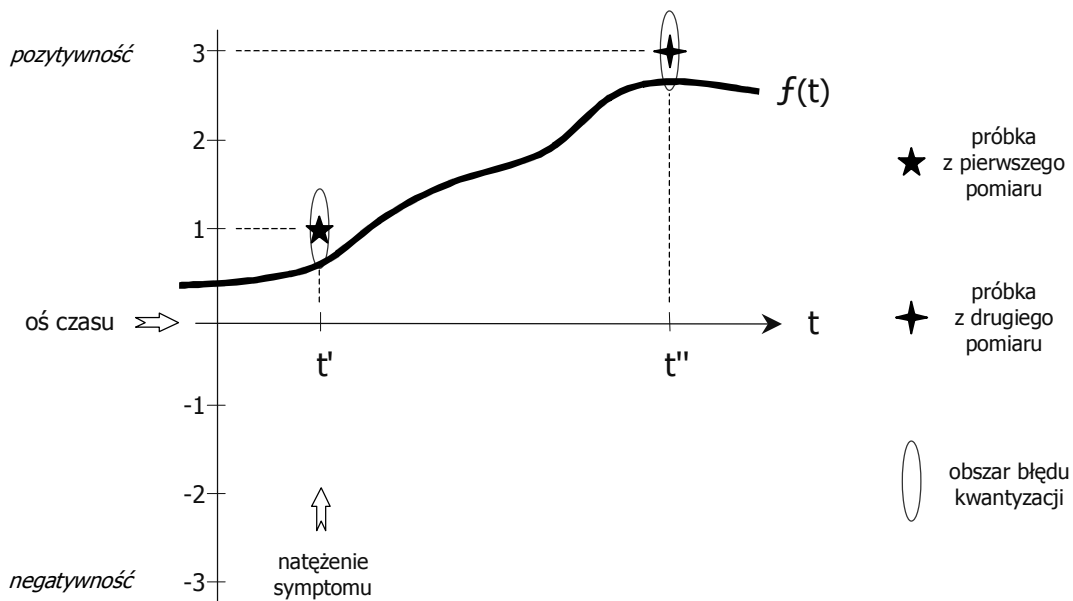
Ryc. 7. Komparacja tendencji par czynników i wyników połówkowych

Strategię komparacji można uogólnić następująco. W trakcie agregacji oznak splotowych (a więc procesu odwrotnego do ekstrakcji) pozytywnym zjawiskiem jest spójność na poziomie wskaźników i składników, natomiast na poziomie czynników i wyników najbardziej korzystne w sensie jakości pomiarowej jest wykrywanie każdej wielkości różnic poziomów chwilowych



oraz wszelkich zmian – zarówno współbieżnych, jak i rozbieżnych. Technika komparacji oznak jest więc immanentną procedurą z jednej strony na etapie weryfikacji narzędzia pomiarowego, a z drugiej strony, najważniejszą taktyką na etapie interpretowania wyników badań empirycznych, przy czym aspekty dynamiczne komparacji uchwytne są tylko dzięki pomiarom dwukrotnym.

➤ **Technika dwukrotnego próbkowania stanów** polega na powtórzeniu pomiaru za pomocą identycznego kwestionariusza po upływie pewnego czasu, przy czym musi być przestrzegany warunek ułożenia parami obu formularzy wypełnionych przez te same osoby. Próbkowanie w ewaluacji oznacza wycinkowe pobranie chwilowych wartości tych wskaźników, które według badacza są najbardziej znaczącymi symptomami indykatum. Kwantuje się wówczas natężenie badanych cech, osiągane właśnie w momencie pobierania próbek. W rezultacie przyjęcia ściśle wyznaczonej liczby progów próbkowania (tutaj 7) uzyskuje się zmienną losową dyskretną (skokową), która jest przybliżonym oraz statycznym jedno- lub dwukrotnym odwzorowaniem splotu symptomów o właściwościach zmiennej losowej ciągłej. Kwantyzacja wprowadza skokowe stopniowanie funkcji natężenia badanych cech, zaś próbkowanie chwilowe staje się przyczyną nieciągłości na osi czasu (ryc. 8).



Ryc. 8. Próbkowanie dwukrotne z kwantyzacją progów natężenia cechy

Oczywiste jest, że na podstawie próbek jednorazowych ustalić można wyłącznie to, jaki był stan dyspozycji deklarowanych przez respondentów w momencie pomiaru. Ze względu na fakt, że każda z badanych osób ma swój własny, różniący się intensywnością skrypt werbalnego wyrażania nastrojów wewnętrznych, pojedyncze pomiary postaw trudno uznawać za miarodajne. Mierzmy bowiem tylko wskaźniki pośrednie, zakładając relacje semantyczne

między wyborami wypowiedzi respondentów a domniemanym natężeniem ich stanów emocjonalnych i poznawczych. Relacje te są cechami osobniczymi, stąd najbardziej rzetelną normą standaryzującą jest porównanie wypowiedzi uzyskanych wcześniej od tej samej osoby wobec identycznej kwestii. Uważam, że dwukrotne próbkowanie cech mentalnych powinno stać się nieodłączną techniką w wielu metodologiach pomiarów skalowalnych. Wydaje się słuszne przeniesienie z teorii próbkowania oraz z praktyki pomiarowej takich zasad, które obligują do próbkowania dwa razy częstszego niż częstotliwość zdarzeń oraz do powtórzenia pomiaru w celu upewnienia się, iż nie popełniono błędu.

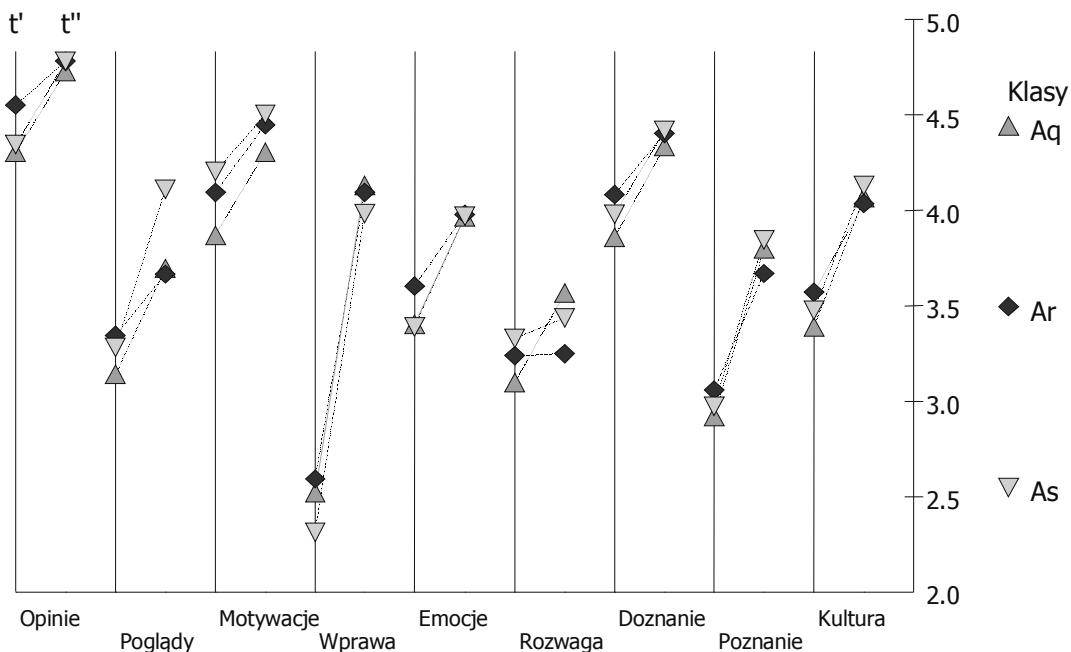
Kardynalnym błędem wielu przeprowadzanych sondaży diagnostycznych jest wykonywanie wyłącznie jednego badania właściwego za pomocą doraźnie utworzonych ankiet. Tymczasem nowo powstałe narzędzie koniecznie musi być poddane weryfikacji empirycznej, a do tego potrzebny jest przynajmniej jeden pomiar wstępny. Dane pochodzące z fazy weryfikacji narzędzia jedynie wyjątkowo mogą być wykorzystane do ewaluacji, jeśli okaże się, że nie trzeba było niczego zmieniać w kwestionariuszu i grupa badawcza była odpowiednia. W przeciwnym wypadku próbek uzyskanych z pomiaru wstępnego nie wolno dołączać do danych z pomiaru właściwego. Jakakolwiek modyfikacja narzędzia badawczego powoduje, że próbki pochodzące z wcześniejszych pomiarów stają się zupełnie nieprzydatne do porównywania z nowymi próbkami.

Z inną sytuacją mamy do czynienia po całkowitym zakończeniu czynności optymalizujących skalę pomiarową. Gromadzone za pomocą dopracowanego, identycznego narzędzia dane, pochodzące z różnych środowisk i gromadzone przez dłuższy okres badań, stają się wartościową podstawą normalizacji skali. Ze względu na to, że przyjmujemy – jako obowiązujący standard – konieczność dwukrotnego próbkowania, uzyskujemy zawsze zasób danych empirycznych, które pochodzą z różnych lat i okresów. Naturalnym okresem pomiarów w wypadku ewaluacji procesów edukacyjnych jest czas trwania kształcenia, czyli dystans pomiędzy pierwszym i ostatnim uczestnictwem w zajęciach. Z moich badań empirycznych wynika, że bardzo wyraziste zmiany w drugim pomiarze uzyskuje się nawet wtedy, gdy intensywny kurs trwa zaledwie przez kilka dni. I przeciwnie – długotrwałe kursy komputerowe wcale nie gwarantują osiągnięcia korzystniejszych efektów przyrostu kultury informatycznej. Zależy to przede wszystkim od jakości kursu i dynamiki prowadzenia zajęć. Wynika stąd, że można badać oraz porównywać procesy zarówno krótko, jak i długotrwałe, przy czym oddziaływanie edukacyjne nie powinno trwać krócej niż 20 godzin.

Warto w tym miejscu raz jeszcze podkreślić rolę powtarzania pomiarów. W fazie przygotowywania nowego narzędzia lub podczas kulturowej adaptacji skali utworzonej w innym obszarze społecznym – ewaluację właściwą należy bezwzględnie poprzedzać próbkowaniem wstępnym. Podczas badania jakości

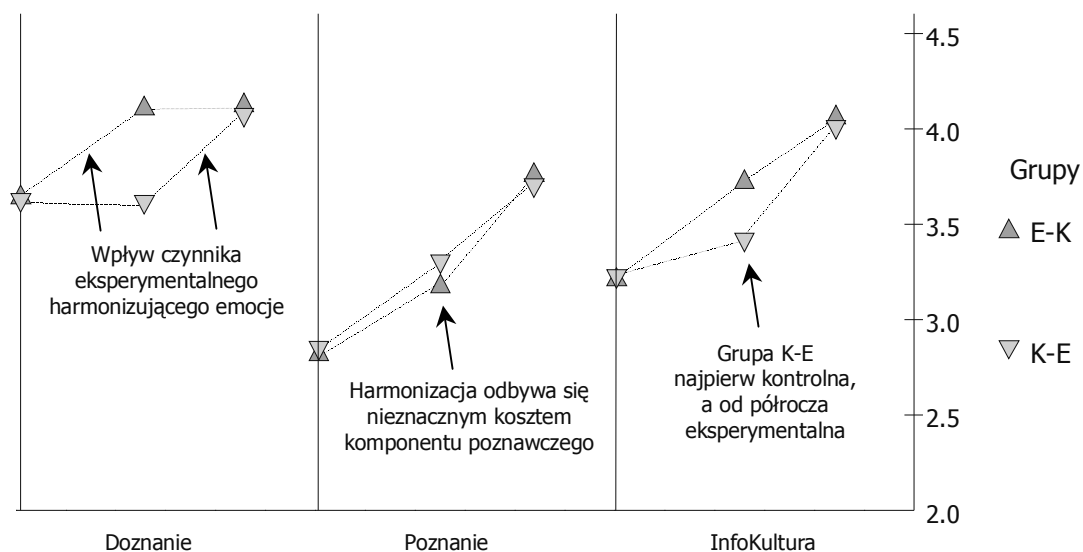
procesów społecznych nieodzowny jest pomiar dystansowy z próbkowaniem stanów wejściowych i wyjściowych. Liczne próbki empiryczne, uzyskane poprzez powtarzanie pomiarów w reprezentatywnych dla populacji środowiskach, są źródłem miar bazowych i normujących. Pierwsze z nich wyznaczają, jakiego stanu można spodziewać się na początku procesu, a drugie wytyczają standardowy cel, jaki wypadałoby osiągnąć w rezultacie całości oddziaływań intencjonalnych (wpływ kursu) i okazjonalnych (wpływ środowiska).

➤ **Technika porównywania obiektów podobnych** odnosi się do dwóch zupełnie różnych sytuacji badawczych. Pierwsza ściśle wiąże się z pomiarem cech mentalnych, kiedy to porównuje się cechy klas lub grup pochodzących z tego samego środowiska społecznego. Zakłada się wówczas wspólne właściwości tych prób badawczych i próbuje wyjaśniać wszelkie ewentualne różnice. Najczęściej poddawany ewaluacji obiektami podobnymi są klasy równoległe bądź kolejne roczniki uczniów w tej samej szkole, przy czym klasy te zwykle obowiązują identyczny program kształcenia, a ponadto przedmiot specjalistyczny (np. informatykę) zazwyczaj realizuje ten sam nauczyciel. Jeśli celem ewaluacji jest pośrednia ocena jakości procesu edukacyjnego, to porównanie skutków identycznego oddziaływania na podobne grupy wzmacnia pewność co do trafności wniosków końcowych. Oczywiście także i w tej technice konieczne dla każdej z grup/klas przeprowadza się pomiary dwukrotne [ $t'$  i  $t''$ ], po czym należy porównywać parami kwestionariusze wypełnione dokładnie przez tych samych respondentów – podmiotów danego kształcenia (ryc. 9).



Ryc. 9. Porównanie skuteczności oddziaływania tego samego nauczyciela na trzy klasy w kolejnych trzech latach szkolnych

W regułach przedstawionej tutaj techniki ściśle mieści się klasyczny eksperyment. Jednej z dwóch homogenicznych zbiorowości przypisuje się wówczas rolę tzw. grupy eksperymentalnej, a drugiej rolę grupy kontrolnej. W zasadzie na obie grupy należy oddziaływać podobnie, wprowadzając jednak tzw. kanon jedynej różnicy. Powinien to być tylko jeden czynnik modyfikujący oddziaływanie, po to, ażeby można było stwierdzić, czy ów czynnik wpłynął w statystycznie istotny sposób na wybraną próbę badawczą. W eksperymentach dotyczących procesów edukacyjnych najczęściej weryfikuje się innowacyjne formy metodyki nauczania. Ażeby grupa kontrolna nie była w tym względzie pokrzywdzona, a zwłaszcza dla zachowania podobieństwa oddziaływań na obie grupy, zalecana jest technika z zamianą grup (skrzyżowaniem). Otóż w eksperymentach trwających przez dłuższy czas (np. cały rok szkolny) w połowie tego okresu odwraca się role grup, usuwając czynnik eksperymentalny z pierwszej i wprowadzając go do drugiej grupy. W takim planie badawczym konieczne jest przeprowadzenie trzykrotnego pomiaru oraz porównanie trzech stanów: na początku, w połowie i na końcu eksperymentu (ryc. 10).



Ryc. 10. Eksperyment z grupami skrzyżowanymi i trzema pomiarami

Warto jednak zwrócić uwagę na to, że ewaluacja splotowa nie zajmuje się ocenianiem sprawdzającym osiągnięcie zadanego poziomu wiedzy bądź umiejętności, ponieważ komponent poznawczy indykatum nie powinien być w żaden sposób związany z obowiązującymi treściami kształcenia. Z tego powodu procesy edukacyjne, modyfikowane eksperymentalnie i poddane ewaluacji, powinny być równolegle kontrolowane tradycyjnymi metodami oceny efektów nauczania. W zadania szkoły wpisane jest jednak kształtowanie właściwych postaw, stąd ewaluacja cech mentalnych jest też jak najbardziej wskazana. Ściślej rzecz ujmując – musi to zawsze polegać na porównywaniu szerokiego spektrum zmian, jakie zachodzą w umysłowych sferach podmiotów edukacji.

Z zupełnie odmienną sytuacją poznawczą mamy do czynienia wówczas, gdy indykatum nie dotyczy cech mentalnych. Ewaluacja splotowa może być bowiem stosowana również do oceny wartości tzw. wytworów intelektualnych. Chodzi tu o rozmaite formy dóbr kultury: dzieła pisane, audycje, widowiska, spektakle, filmy, koncerty, wystawy, a także o te produkty zaawansowanych technologii informacyjnych, które na razie pozostają poza domeną kultury, jak: multimedia, prezentacje, programy komputerowe i serwisy internetowe. W ramach konkretnej formy dostępne są wytwory alternatywne o zbliżonych właściwościach, z których ze względu na sztukę i artyzm trudno jest wybrać najlepsze. Pomóc w tym może ewaluacja splotowa z techniką porównywania obiektów podobnych. Respondenci pełnią wówczas rolę *jury*, odnosząc się do wyekstrahowanych komponentów wartości każdego z ocenianych wytworów.

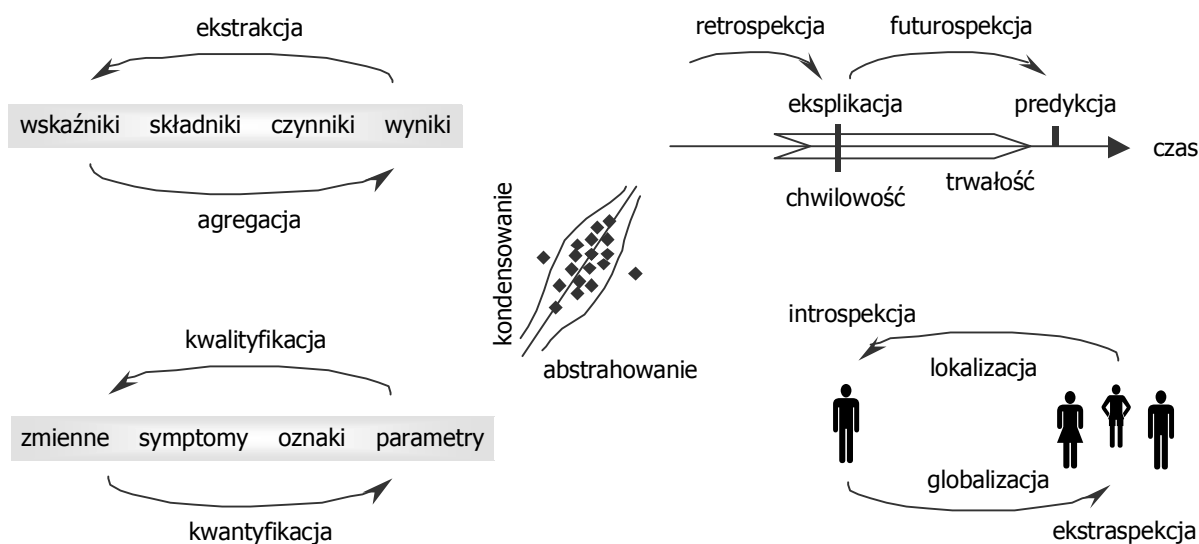
W technice tej olbrzymie znaczenie ma kolejność poznawania obiektów. Zalecane jest pełne permutowanie porządku oceniania obiektów podobnych\*, jednak liczba permutacji rośnie z szybkością  $n!$  ( $n$ -silnia, gdzie  $n$  jest liczbą obiektów ocenianych), dlatego do końcowej ewaluacji powinno się wstępnie zakwalifikować tylko kilka wytworów. Zależy to głównie od liczebności grupy oceniających, a ponieważ jeden *juror* (lub *sędzia kompetentny*) ocenia dany wytwór tylko jednokrotnie, metodologicznie optymalnymi rozwiązaniami są strategie następujące: 3 obiekty i  $6k$  jurorów, bądź 4 obiekty i  $24k$  jurorów, gdzie  $k = 1, 2, \dots$  jest niską liczbą naturalną oznaczającą ewentualną krotność. Dla przykładu podaję tabelę ze schematem pełnej permutacji  $4! = 24$ .

Numer porządkowy nadany każdemu z 24 jurorów																							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
↓	Numery czterech obiektów, ułożone w kolumnach w kolejności oceniania przez danego jurora																						
1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4
2	2	3	3	4	4	3	3	4	4	1	1	4	4	1	1	2	2	1	1	2	2	3	3
3	4	2	4	2	3	1	4	1	3	3	4	1	2	2	4	1	4	2	3	1	3	1	2
4	3	4	2	3	2	4	1	3	1	4	3	2	1	4	2	4	1	3	2	3	1	2	1

Warto podkreślić, że – zgodnie ze strategią ewaluacji demokratycznej – role jurorów oceniających wytwory powinni pełnić przede wszystkim odbiorcy dzieł. Dotyczy to zwłaszcza mediów funkcjonujących jako środki dydaktyczne w samokształceniu. Każdy respondent może być sędzią kompetentnym, jeśli rozumie stwierdzenia kwestionariusza, a jego wypowiedzi nie są niedorzeczne bądź rażąco sprzeczne z opinią ogółu. Natomiast narzędzie ewaluacji wytworów powinno być opracowane przez kompetentnego znawcę danej dziedziny.

\* Ocena dzieła o średniej wartości bezpośrednio po wcześniej ocenianym dziele o niższej wartości jest zawyżana wobec tej, jaką wystawiono by, gdyby najpierw oceniano dzieło lepsze.

➤ **Metoda koncyptowania dwubieżnego** odnosi się do każdego etapu procesu ewaluacji i polega na uporządkowaniu procesów myślowych w dychotomiczne, dwukierunkowe, dwutorowe lub dwufazowe pary inferencyjne<sup>13</sup>. Typową tego rodzaju parą jest na przykład dedukcja i indukcja. Koncyptowanie dwubieżne powinno towarzyszyć: opracowywaniu koncepcji, kreowaniu narzędzia, eksploracji zasobów oraz interpretacji zjawisk. Fazę wnioskowania poprzedza zawsze faza badania – interpretujemy to, co wcześniej odkryliśmy, a jeśli do wyjaśnienia symptomów jeszcze czegoś brakuje, to poszukujemy dalszych oznak. Tak więc para eksploracja|interpretacja tworzy nadrzędny cykl inferencyjny. Cyklicznie przeplatają się przy tym fazy przeciwbieżne, np.: analiza różnicowa zmian z syntezą trendów, analiza indeksowa fluktuacji z syntezą efektów. Przyjrzyjmy się owym ścieżkom koncyptowania w ewaluacji (ryc. 11).



Ryc. 11. Schematy koncyptowania dwubieżnego

1) Ekstrakcja|agregacja dotyczy operacji przechodzenia na różne poziomy wyszczególnienia i uogólnienia zmiennych. W fazie ekstrakcji indykatum poszukuje się komponentów najbardziej znamienych, tj.: głównych czynników, istotnych składników i mierzalnych wskaźników. W fazie agregacji scala się wszystkie składowe stany zmiennych cząstkowych w jeden stan wynikowy zmiennej ogólnej, a także w dwa potrzebne do porównań rezultaty połówkowe.

2) Kwantyfikacja|kwalifikacja dotyczy strategii przechodzenia pomiędzy jakościowym a ilościowym ujmowaniem indykatum. Podczas kwantyfikacji symptomom cech i zjawisk niefizycznych przypisuje się umowne wartości liczbowe, które poddawane są ilościowej analizie statystycznej. W zwrotnym procesie kwalifikacji<sup>14</sup> na podstawie analizy uzyskanych parametrów i syntezy oznak definiuje się umowne desygnaty jakościowe, wyrażane semantycznie.

3) Abstrahowanie|kondensowanie to czynności przeciwstawne, z których pierwsza polega na pomijaniu przypadków mało istotnych, a druga oznacza

uwzględnianie wszystkich elementów. Obliczanie większości statystyk jest abstrahowaniem od indywidualnych danych na rzecz parametrów zbiorczych. W ewaluacji jedna niekorzystna oznaka nie może być podstawą do uogólnień: warto np. odrzucić kwestionariusz wypełniony w pośpiechu, bez przeczytania treści bodźców-stwierdzeń, kiedy to przeważają wypowiedzi 'brak zdania'.

4) Retrospekcja|futurespekcja to wątki analizy aspektów czasowych, które odnoszą się do kwestii: *tak było, a jak będzie?* Podbudowę analizy dają pary wskaźników kontrolnych próbujących relacje czasoprzestrzenne. Wnioski formułuje się na podstawie wypowiedzi respondentów – to oni wyrażają opinie o stanie dotychczasowym i o potencjalnych tendencjach w przyszłości. Istotne jest to, czy zjawiska niekorzystne uznają oni jako przejściowe, czy jako trwałe.

5) Eksplikacja|predykcja to wątki syntezy w fazie formułowania wniosków, polegające najpierw na opisaniu i wyjaśnieniu odkrytych zjawisk, a następnie na przewidywaniu stanów przyszłych. Synteza jest już rolą ewaluatora – to on wartościuje stan, jaki zastał, a także prognozuje, w jakim kierunku będą zmierzać procesy. Ocenia zwłaszcza trwałość zjawisk i powtarzalność efektów.

6) Introspekcja|estraspekcja to wątki analizy aspektów osobowościowych, odnoszących się do relacji: ja ↔ inni. Także i tu podstawą analizy są kontrolne pary wskaźników próbujących samoocenę respondenta na tle oceny innych. Istotne jest to, czy respondenci doceniają swe własne zdolności, czy może uważają, że i tak nie są w stanie przyswoić tego wszystkiego, co inni potrafią. Introspekcja jest szczególnie ważna dla nauczycieli prowadzących zajęcia.

7) Lokalizacja|globalizacja to wątki syntezy aspektów przestrzennych w fazie wnioskowania formatywnego. Zawężonym celem ewaluacji może być ocena wartości jedynie procesów lokalnych, kiedy to nauczyciel bada swój wpływ na własnych uczniów. Ważna wówczas jest analiza pojedynczych przypadków. Dla badaczy istotniejsze są uogólnienia i rozstrzygnięcia – czy zjawisko jest charakterystyczne tylko w danej próbie u jednego wykładowcy, czy może też powtarzalne dla wielu grup, u innych wykładowców i w innych środowiskach.

<i>Ślady</i>		<i>Wzorce</i>
ekstrakcja symptomów	.....	rozwarstwienie indykatum
ekspansja czasoprzestrzeni	.....	pary kontrolne wskaźników
ustalanie znaków walencyjnych	.....	polaryzacja bodźców-stwierdzeń
minimaksowe ważenie zdań	.....	weryfikacja wskaźników
justowanie estymatorów	.....	normalizacja wskaźników
analiza dwuwymiarowa	.....	rozziew jakości i intensywności
komparacja oznak	.....	spójne składniki, niespójne czynniki
próbkowanie stanów	.....	kwantyzacja natężenia cech
porównywanie obiektów	.....	cechy grup, wartości wytworów
koncypowanie dwubieżne	.....	ścieżki i pary inferencyjne

### **Postawa i uświadomienie – ekstrakcja doznania i poznania**

Sama postawa, choćby najbardziej przychylna, wcale nie musi być przejawem pożądaných, autentycznie wartościowych relacji użytkownika wobec informatyzacji i komputeryzacji. Przecież nadmierna ekscytacja bogactwem możliwości programów komputerowych oraz serwisów internetowych może wprost prowadzić do autodestrukcyjnego, nałogowego nadużywania. I wcale nie pomaga – w zwalczaniu tych nasilających się ostatnio uzależnień – nawet najgłębsza wiedza o komputerze. Niezbędny jest pewien stan uświadomienia o bilansie korzyści względem zagrożeń. Z tego powodu nie wystarcza jedynie mierzenie samych postaw, tak chętnie przeprowadzane przez ewaluatorów. Wprawdzie w strukturalnej definicji postaw<sup>15</sup> zawarte są trzy fundamentalne dla edukacji komponenty: emocjonalne, poznawcze i działaniowe, jednakże – w zależności od priorytetów konkretnej dyscypliny naukowej – w rozmaity sposób interpretuje się owe składniki<sup>16</sup>. Przez psychologów akcentowany był przede wszystkim *afekt*, jako atrybut niezbędny, ażeby w ogóle móc mówić o cesze, iż jest to postawa. Pedagodzy nadal na pierwszym miejscu sytuują komponent poznawczy, upatrując w nim wsparcia dla interesujących ich procesów wychowawczych – czasem wręcz zacierając różnice między wiedzą a składnikiem kognitywnym i jakby zapominając, że ów komponent postaw jest raczej tylko dyspozycją gotowości do poznawania. Z kolei dla behawiorystów najważniejsze byłyby zachowania – choć one mogą, lecz wcale nie muszą być przejawem postaw. Ze względu na trudności z ustalaniem wzajemnych zależności pomiędzy trzema komponentami oraz na to, że niektóre z nich mogą nie występować, obecnie w psychologii społecznej dominuje pogląd, iż nie są to definicyjne składniki, lecz jedynie czynniki źródłowe genezy postaw<sup>17</sup>. Większość zgadza się jedynie z najogólniejszą definicją postawy jako skłonności człowieka do walencyjnego ustosunkowywania się do obiektu lub idei.

Model-wzorzec ewaluacji splotowej osadzam na tzw. predylekcyjnej, czyli skłonnościowo-upodobaniowej wykładni pojęcia *postawy*, wzorowanej nadal na tradycyjnej definicji strukturalnej, chociaż bardziej od niej doprecyzowanej:

**Postawa** jest predylekcyjną strukturą mentalną, złożoną z komponentów:

- *afektywnych*, tj. predyspozycji do emocjonalnego reagowania i oceniania;
- *kognitywnych*, tj. potencjalnej subdyspozycji do podjęcia poznawania;
- *behawioralnych*, tj. potencjalnej subdyspozycji do podjęcia działania.

Rozmyślnie wprowadzam przedrostki *pre-* i *sub-*, aby zaakcentować, że emocje pobudzające ocenianie są pierwotnymi elementami gotowości immanentnej, instynktownej, natomiast przy dwóch następnych komponentach struktury celowo akcentuję to, iż są one jedynie wtórnymi deklaracjami gotowości potencjalnie możliwej, a rzadziej koherentnej. Człowiek reaguje emocjonalnie



i ocenia *obiekt postaw* niezależnie od tego, czy dany obiekt wcześniej poznał, przy czym może w ogóle nie mieć zamiaru poznawać ani działać, a jeśli nawet wyrazi takie intencje, to będzie to tylko deklaracja, nie zawsze wiarygodna.

Właśnie ze względu na ową możliwość „oceniania nawet bez poznania” nie wystarcza badanie procesów wyłącznie za pomocą samych skal postaw. Konieczne jest z jednej strony wartościowanie postaw, a z drugiej testowanie wskaźników poziomu świadomości<sup>18</sup> podmiotów danego procesu w zakresie meritum wartości kierunkowych, np. edukacji medialnej<sup>19</sup> lub ekologicznej<sup>20</sup>. Chodzi m.in. o świadomość, że mass media manipulują umysłami odbiorców, świadomość bytów wirtualnych w mediach, świadomość znaczenia równowagi biosfery, świadomość globalnych zagrożeń dla przyrody itd. Nie polega to jednak na przyswojeniu głębokiej wiedzy z tych dziedzin, lecz przede wszystkim na uwewnętrznieniu rangi problemów i zdawaniu sobie sprawy ze znaczenia regulacyjnej roli pożądanych zachowań w danym zakresie. W odróżnieniu od pary pojęć świadomość | podświadomość, jako ponad- i podprogowych stanów psychiki, przyjmijmy tę zinternalizowaną warstwę nazywać *uświadomieniem*:

**Uświadomienie** jest cechą sfery poznawczej, nabytą przez doświadczenie lub internalizację elementarnych prawidłowości z wybranego obszaru wiedzy, a przejawiająca się w wyrażanych poglądach, przekonaniach, intencjach i inklinacjach oraz w ujawnianych zasadach regulujących własne postępowanie.

W ewaluacji bezpośrednimi wskaźnikami uświadomienia respondentów są:

- *poprawność* merytoryczna wypowiedzi wobec tych bodźców-stwierdzeń, które próbują obiektywne obszary znajomości ogólnoprzedmiotowej;
- *trafność* osobistych osądów, przeświadczeń, zamysłów oraz upodobań, w odniesieniu do kryterium regulacyjnego standardów kulturowych.

Ponadto oznakami pośrednimi podczas analizy szczegółowej mogą stać się:

- *refleksyjność*, mająca swe odbicie w zróżnicowanym sposobie wypowiedzi, bez pomijania pozycji problemowych poprzez wybór opcji '*brak zdania*';
- *roztropność* wyboru skrajnych opcji skali bez pochopnego automatyzmu, polegającego na zbytnej pewności siebie i nadużywania opcji '*absolutnie*';
- *spójność* wewnętrzna wypowiedzi respondenta wobec kontrolnych par bodźców-stwierdzeń, oznaczająca wypowiedzianie się w zgodzie z sobą.

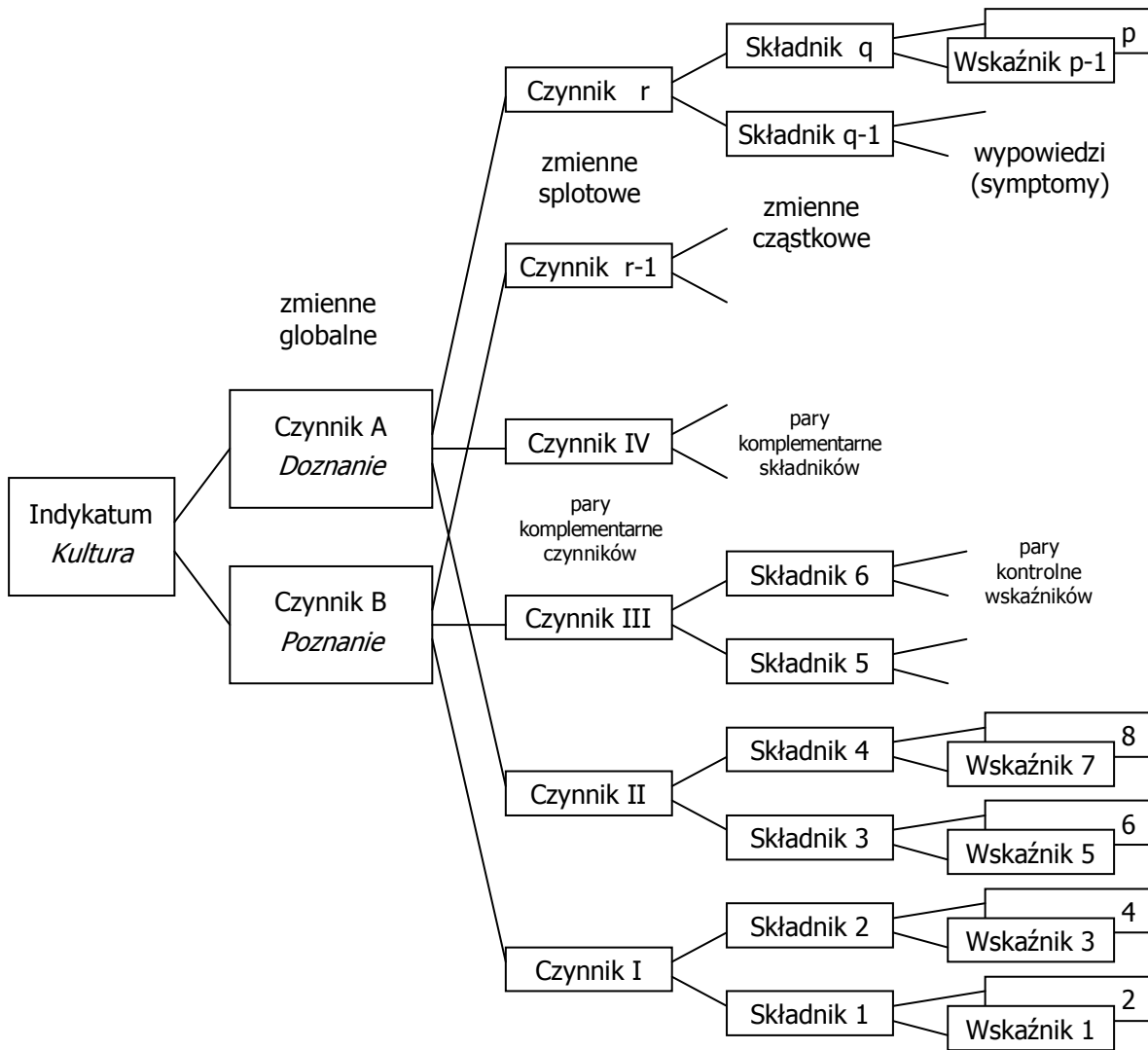
O silnym związku obu omawianych tu cech mentalnych świadczy fakt pojawiania się poplątanych definicji w różnych publikacjach, a mianowicie – przy objaśnianiu pojęcia *uświadomienie* niektórzy autorzy włączają do niego komponenty postaw jako konstytutywne składowe. Podobnie jest z innymi definicjami pojęć psychologicznych (np. funkcjonującymi w teorii osobowości) – wszędzie przewijają się odwołania z najczęstszym podziałem na czynniki lub składowe emocjonalne, poznawcze i działaniowe. Skoro więc struktury

umysłu są tak bardzo powikłane, a cechy mentalne trudne do rozdzielenia, to warto oprzeć całą koncepcję ewaluacji na wyodrębnianiu i badaniu dwóch najczęściej przywoływanych komponentów, tj.: afektywnych i kognitywnych. W propagowanym tutaj modelu komponent kognitywny postaw nie jest jednak tożsamy z wiedzą, a uświadomienie też jest bardziej regulatorem niż wiedzą. Wspólną właściwością postaw i uświadomienia jest to, że w obu cechach mentalnych szacować można poziomy natężenia oraz ustalać wartościujący znak walencyjny, oznaczający bieguny pozytywności | negatywności.

Uświadomienie korzyści bądź strat, jakie mogłyby być następstwem wyrażania określonych postaw, modyfikuje i filtruje to, co jest uzewnętrzniane. Wpływa to na zachowania, które regulowane są przez konteksty społeczne, normy postępowania, kulturę osobistą i różne konwenanse<sup>21</sup>. Ewaluatorom to zbytnio nie przeszkadza, a co ciekawsze – właśnie sam sposób udzielania wypowiedzi (np. przejaw *certacji*) może być dla nich cenną przesłanką wnioskowania, o ile stosuje się adekwatną technikę wskaźnikowania i interpretacji.

W przeciwieństwie do scjentyistycznych eksperymentów badawczych, proponowana tu metoda ewaluacji procesów zakłada ocenianie wartościujące zachodzących zjawisk w ich naturalnym bogactwie kontekstów społecznych. W takim pomiarze ani nie izoluje się grup eksperymentalnych od żadnych czynników, ani nie dzieli się cech badanych na zmienne niezależne | zależne. Istotą strategii jest podział cech na równoważne przekroje komplementarne, a następnie bilansowanie zarówno statycznych, jak i dynamicznych relacji obustronnych między zmiennymi cząstkowymi, ułożonymi w pary kontrolne. Znamienne przy tym jest to, że wzajemne związki oznak cech paralelnych wykazują silną zależność wytwarzającą taki układ, w którym jedna zmienna cząstkowa wobec drugiej raz jest czynnikiem modyfikującym, a następnie role mogą się odwrócić, tj. staje się ona wynikiem-rezultatem modyfikacji. Rozwarstwianie cechy ogólnej na kolejne pary symptomów zwiemy ekstrakcją.

W pierwszej fazie **ekstrakcji** wyodrębnia się dwie główne grupy symptomów, będących wyznacznikami najbardziej ogólnej pary komplementarnej. W przyjętym modelu badawczym indykatum kulturowe tworzą dyspozycje afektywne (odczuciowo-oceniające i emocjonalno-motywacyjne komponenty postaw) oraz dyspozycje kognitywne (osądowo racjonalne, merytorycznie dojrzałe stany uświadomienia). Wytyczają one dwie zmienne globalne: *Doznanie* i *Poznanie*. Podczas analizy traktuje się je jako czynniki modyfikujące zmienną ogólną, a podczas syntezy interpretuje jako odrębne rezultaty, będące połówkowymi wynikami ewaluacji<sup>22</sup>. Z tego właśnie względu, w programie komputerowym przetwarzającym dane surowe, parametry statystyczne doznania i poznania obliczane są jako *Czynniki*, natomiast relatywistyczny opis tej splecionej pary zamieszczono w rozdziale pt. *Wyniki – rezultaty zmiennych globalnych*.



Ryc. 12. Schemat ekstrakcji czynników, składników i wskaźników indykatum  
– minimalna liczba pozycji w kwestionariuszu  $p = 2q = 4r$

W kolejnych fazach ekstrakcji wyodrębnia się komplementarne pary czynników dominujących w zmiennej ogólnej oraz pary składników będących komponentami czynników. Na każdym etapie rozwarstwiania zmiennej ogólnej, podczas obróbki statystycznej sięga się do symptomów źródłowych. Wskaźnikami źródłowymi w proponowanej metodzie są wypowiedzi respondentów, a ilościowo – umowne wagi liczbowe przypisane opcjom wypowiedzi. Podczas interpretacji wyników empirycznych postępuje się odwrotnie, a mianowicie – jakościowe wnioskowanie syntetyzujące buduje się na oznakach hierarchicznie coraz bardziej uogólnianych i agregowanych.

Z powyższych wywodów wynika, że w metodologii badawczej ważny jest nie tylko dobór odpowiednich wskaźników zmiennej ogólnej, będących mierzalnymi oznakami indykatum, ale nade wszystko istotne jest zbudowanie narzędzia pomiarowego o odpowiedniej strukturze, umożliwiającej ekstrakcję i analizę wszystkich przekrojów zmiennych cząstkowych, a następnie syntezę

oznak na podstawie ściśle zdefiniowanego schematu interpretacji. Narzędzia badawcze tworzone *ad hoc* mają zazwyczaj wątpliwą jakość, z tego względu podczas konstruowania nowego kwestionariusza najlepiej opierać się na kompleksowo dopracowanych oraz zweryfikowanych wzorcach<sup>23</sup>. Oczywiście jest, że za pomocą jednego narzędzia, choćby najbardziej rozbudowanego, nie da się mierzyć wszystkiego. Można jednak zastosować metodę próbkowania najbardziej charakterystycznych symptomów wartości procesu społecznego, poprzez pomiar i analizę reprezentatywnie wyselekcjonowanych wskaźników zjawisk<sup>24</sup>, po to, ażeby następnie metodą syntezy i uogólniania wnioskować o szerszych aspektach rzeczywistości. Stanowi to istotę badań wartościujących za pomocą specyficznie skonstruowanych, standaryzowanych i znormalizowanych narzędzi pomiarowych, zwanych *skalami*.

### **Wzorzec narzędzia pomiarowego – Skala dwuważonych ocen**

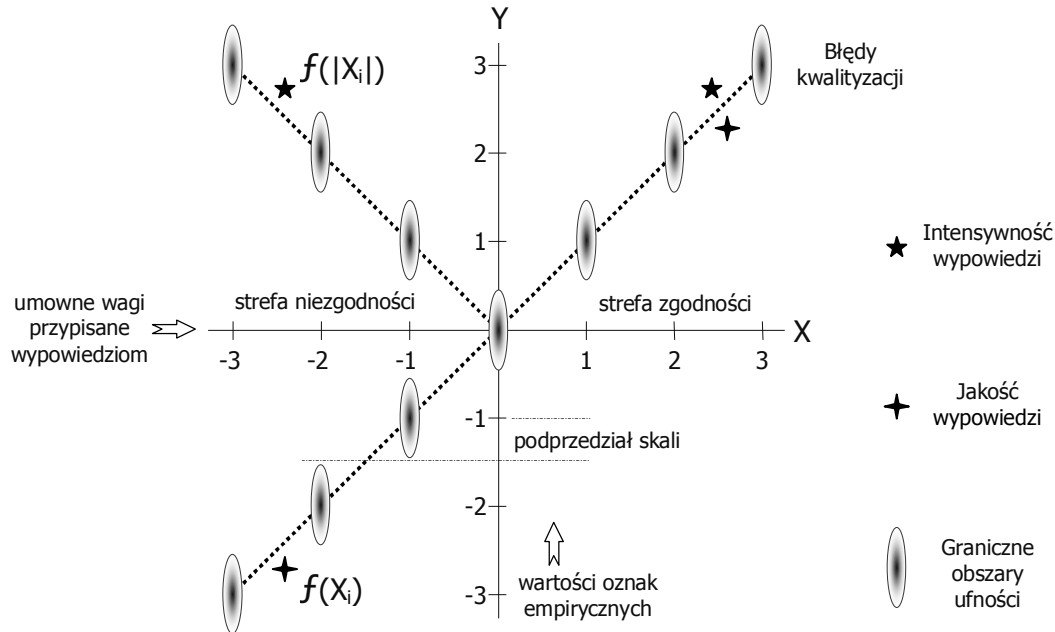
Dużą popularnością cieszą się wciąż narzędzia badawcze wzorowane na skali Likerta<sup>25</sup>. Dobór bodźców, kategorie wypowiedzi i zasady interpretacji wyników stały się przedmiotem wielu analiz metodologicznych i publikacji. Istotą tego typu skal jest swoista zgodność semantyczna stwierdzeń, które – na podstawie ścisłych kryteriów empirycznych – wylania się z większej liczby sformułowań, pierwotnie redagowanych intuicyjnie. W konsekwencji wzrasta pewność, że każdym bodźcem-stwierdzeniem mierzymy jedynie różne aspekty tego samego indykatum, co uprawnia do scalania wskaźników cząstkowych w rzetelniejszy wynik ogólny. Cecha addytywności uzasadnia używaną nazwę *Skale sumowanych ocen*. Jednakże zsumowany wynik ogólny to zbyt mało, stąd mój zamysł przedstawienia w niniejszej pracy innowacyjnej koncepcji kompleksowej strategii pomiarowej, nazwanej *Skalą dwuważonych ocen*.

Przy każdym zamiarze budowy narzędzia pomiarowego należy określić, z jakim typem skal mamy do czynienia. Uważam, że w taksonomiach klasyfikujących skale brakuje specyficznej grupy używanej do badania dyspozycji, a zwłaszcza postaw. Proponuję więc wprowadzenie odrębnej kategorii ogólnej o nazwie *skala dwuważona*, którą charakteryzuje następująca definicja:

**Skala dwuważona** wyróżnia się tym, że jedna z kategorii próbkujących wyznacza neutralną lub ambiwalentnie zneutralizowaną jakość badanej cechy, a pozostałe próbki wskazują na polaryzację i gradację wartości, choć nie można określić podziałki jednostkowej o równych przyrostach. Istotne jest to, że skala ma bipolarnie symetryczne kategorie próbkowania, a każda z próbek odzwierciedla dwa wymiary: jakość oraz intensywność cechy.

Obydwa te wymiary nie są współbieżne, co oznacza, że w układzie współrzędnych z osią X skalowaną argumentami jakości  $X_i$  funkcja intensywności

$f(|X_i|)$  nie jest monotoniczna w całym zakresie próbkowania (zob. ryc. 13). Ponadto – w każdym z głównych przedziałów skali, wytyczonych gęstością próbkowania, daje się empirycznie wyodrębnić dwa sąsiednie podprzedziały, wobec których test potwierdza istotność różnic i jednokierunkowość zmian, lecz także bez możliwości wykazania, czy narastanie cechy jest równomierne.



Ryc. 13. Graficzna reprezentacja wzorca skali dwuważonej

Skala dwuważona pod względem mocy pomiarowej sytuuje się między skalami *interwałową* a *porządkową*. Wprawdzie pozornie wyznaczany jest punkt zerowy, jednak jest to wyłącznie zerowa intensywność wypowiedzi, a nie zerowa ani najniższa jakość badanej cechy. W pomiarze postaw wybór przez respondenta opcji '*brak zdania*' oznaczać może zarówno brak pobudzenia emocjonalnego wobec danego bodźca-stwierdzenia (a bez komponentu afektywnego w ogóle nie ma podstaw sądzić, iż jest to postawa), albo też może oznaczać powstrzymanie się z jakiegoś powodu od ujawnienia doznań (np. gdy są one intensywne, lecz ambiwalentne i dysonansowe). Ze względu na brak zera absolutnego, narzędzi zbudowanych według przedstawionego modelu nie można zaliczać do najmocniejszej grupy *skal ilorazowych*. Nie należy ich także zaliczać do grupy najbardziej zbliżonych *skal interwałowych*, mimo że występują przedziały, które jednak nie mają równomiernej podziałki. Można natomiast empirycznie wykazać, że przy dwuwymiarowo-bilansowej technice analizy wypowiedzi są one silniejsze od *skal porządkowych*, dzięki wyznaczaniu rozdzielczych progów narastającej jakości danej cechy, kalibrujących skalę na rozłączne podprzedziały. Dzięki tej właściwości *skala dwuważona* nawet przy nielicznych próbach jest wystarczająco czuła. Rozdzielczość tę ilustruje tabelka teoretycznych przedziałów ufności, zależnych od liczebności

próby (liczby respondentów) oraz od odchylenia standardowego wypowiedzi [ $\sigma$ ], przy założeniu prawdopodobieństwa popełnienia błędu nie większego niż 5%:

Liczebność próby	15	35	60	375	600	1070
Przedziały ufności	$\pm 0,51 \cdot \sigma$	$\pm 0,33 \cdot \sigma$	$\pm 0,25 \cdot \sigma$	$\pm 0,10 \cdot \sigma$	$\pm 0,08 \cdot \sigma$	$\pm 0,06 \cdot \sigma$

Dalsze, choćby nawet bardzo duże zwiększanie liczebności prób badawczych powoduje już tylko niewielkie zawężanie przedziałów ufności. Natomiast, ze względu na to, że podczas syntezy zmiennych cząstkowych odchylenia maleją w miarę przechodzenia od wskaźników do wyników uogólnionych, najbardziej wiarygodne i rzetelne stają się zagregowane rezultaty mierzonego indykatum:

Poziom uogólnienia	wskaźniki	składniki	czynniki	wyniki połówkowe	wyniki ogólne
Przeciętne odchylenia	$\sigma \approx 1,32$	$\approx 1,04$	$\approx 0,83$	$\approx 0,63$	$\approx 0,55$

Właśnie z tego powodu do ewaluacji indykatum o złożonej strukturze zaleca się tworzenie skalowanych narzędzi agregujących znaczną liczbę wskaźników.

Każda skala pomiarowa musi być nieodłącznie obudowana metodologią konstruowania, stosowania i wnioskowania. Wszystko to tworzy specyficzną strategię postępowania, bez stosowania której narzędzia w ogóle nie wolno używać. Im dokładniej dopracowana i opisana jest metodologia skalowania, tym bardziej standardowe, rzetelne i porównywalne są wyniki uzyskiwane za pomocą danej skali. W tym celu opis ewaluacji splotowej osadziłem wyłącznie na jednym odwzorowaniu modelu o nazwie własnej *Skala dwuważonych ocen*:

**Skala dwuważonych ocen** jest kompleksową strategią pomiarową, opartą na analizie dwuwymiarowych wskaźników dobranych ze względu na dużą moc wnioskotwórczą. Kwantyfikatorem selekcyjnym optymalnych bodźców-stwierdzeń jest konkluzyjność, ustalana poprzez ważenie opinii kwalifikowanej większości respondentów. Kwantyfikator próbkujący tworzą kategorie (7) standardowych wypowiedzi. Kontrolne pary (24) wskaźników wyznaczają znak walencyjny, jakość oraz intensywność dwukrotnie mierzonych cech.

Liczby kategorii wypowiedzi i wskaźników mogą być inne, dlatego umieściłem je w nawiasach jako propozycję najbardziej zalecanych rozmiarów skali ocen.

Charakterystykę narzędzia skalowanego określają trzy główne założenia kwantyfikacyjne: 1) indeksacja – sposób doboru wskaźników, 2) kwantyzacja – sposób próbkowania cech, oraz 3) kalibracja – sposób wymiarowania skali.

➤ **Indeksacja** ma na celu dobór optymalnych wskaźników konstytuujących indykatum. Wskaźnikami są wypowiedzi respondentów wobec tez zawartych w bodźcach-stwierdzeniach, dlatego indeksacja polega na selekcji stwierdzeń, bądź na korygowaniu ich treści tak, aby uzyskać pożądane oznaki empiryczne. Indeksacja jest zatem procesem cyklicznym na etapie tworzenia nowej skali. Obejmuje czynność empirycznej weryfikacji wskaźników przyjętych intuicyjnie podczas ekstrakcji symptomów jako najlepiej wyznaczające stan indykatum.

Wstępny kryterium weryfikacji wskaźników są: prawidłowe rozkłady na histogramach, walencyjność oraz miary spójności i zbieżności par kontrolnych. Końcowym kryterium selekcyjnym jest parametr *Konkluzyjności* [č], określający moc wnioskotwórczą każdej pozycji skali. Jeśli wartość liczbowa [č] jest zbyt niska (<0,5), to trzeba przereagować treść tezy lub usunąć ją ze skali. Indeksacja jest ostatecznym sprawdzeniem, czy tezy były słuszne, a co więcej – czy stawianie i badanie takich tez jest zasadne i poznawczo istotne. Jednak nie umożliwia ona sprawdzenia, czy wskaźniki trafnie obrazują indykatum.

➤ **Kwantyzacja** ma na celu pozyskanie danych empirycznych o jednorodnej postaci, niezbędnych do agregowania wskaźników i do obliczania zbiorczych parametrów indeksowych. W fazie budowy skali bardzo istotnym zagadnieniem jest ustalenie odpowiednich kwantyfikatorów próbkujących wypowiedzi. Bodźce-stwierdzenia tworzące skalę wywołują u respondentów urozmaicone reakcje-wypowiedzi, które trzeba zmierzyć. Z owej „rozmaitości” wypowiedzi potrzebne jest tylko odtworzenie *jakości* i *intensywności* próbkowanej cechy, dlatego ogranicza się możliwość wypowiedzi do wyboru jednej z siedmiu opcji. Każdej opcji podczas wprowadzania danych nadaje się rangę porządkującą [R]:

	absolutnie nie	nie	raczej nie	brak zdania	raczej tak	tak	absolutnie tak				
R	1	2	3	4	5	6	7				
W	-3	-2	-1	0	1	2	3				
O			1,5	2	2,5	3	3,5	4	4,5	5	5,5

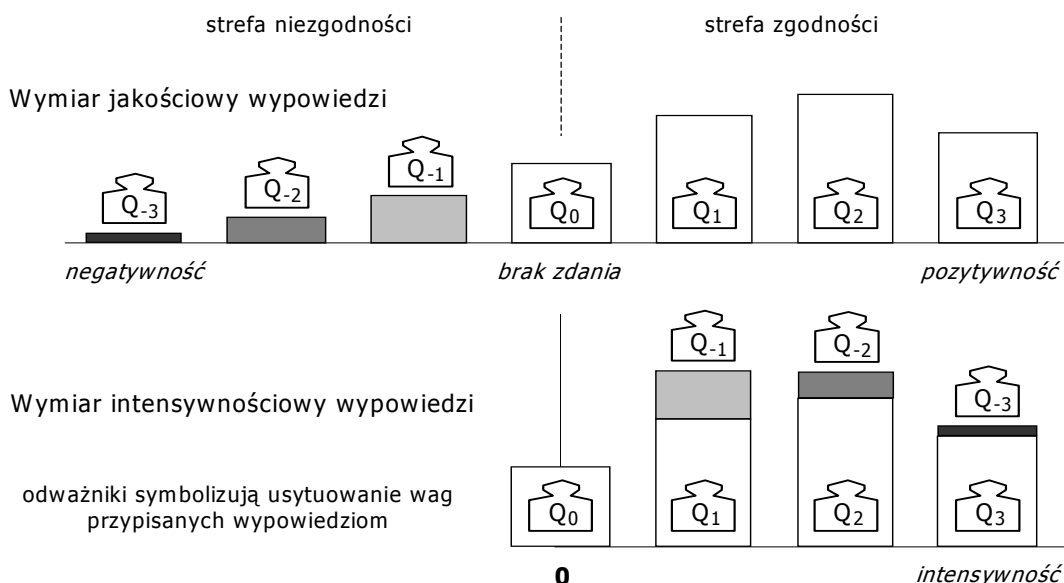
➤ **Kalibracja** w fazie opracowywania danych jest kontynuacją kwantyzacji. Każdej wypowiedzi przypisuje się umowne wagi liczbowe [W] z zakresu (-3; 3). Liczby ujemne oznaczają niezgodność, a liczby dodatnie oznaczają zgodność z opinią większości. Należy przy tym odwrócić te pozycje skali, które spolaryzowane są ujemnie (gdy teza jest fałszywa i respondenci winni jej zaprzeczać). Waga zerowa rozdziela skalę na dwie strefy, a każda z wag oznacza stopniowo narastające natężenie cech. Taki sposób wymiarowania nazwijmy *skalą wag*. Dalszym etapem kalibracji jest konwersja na *skalę ocen* [O], co ma na celu ułatwienie interpretacji wyników oraz porównywanie efektów między grupami. W fazie dostrajania skali kalibracja jest także proces justowania estymatorów. Dzięki niemu, po indeksacji i kalibracji w skali pozostają wyłącznie znormalizowane bodźce-stwierdzenia, które implikują rozkłady wypowiedzi rozmieszczone w miarę optymalnie w strefie zgodności ze standardem ewaluatywnym.

Przedstawiona w niniejszej pracy *Skala dwuważonych ocen* – konstruowana przede wszystkim na podstawie kryterium doboru wskaźników pod względem maksymalnej konkluzyjności (czyli mocy wnioskotwórczej) – jest tylko jedną z wielu potencjalnych implementacji modelu ogólniejszego, który umożliwia stosowanie także innych kryteriów selekcji bodźców-stwierdzeń (np. trafności).

Proponowana nazwa dla tego typu skal wywodzi się od techniki podwójnego „ważenia racji”, którego istotą jest porównywanie każdej wypowiedzi respondentów w dwóch wymiarach: szacowania wartości i wyznaczania ważkości.

### Dwa wymiary wypowiedzi – komparacja wartości i ważkości

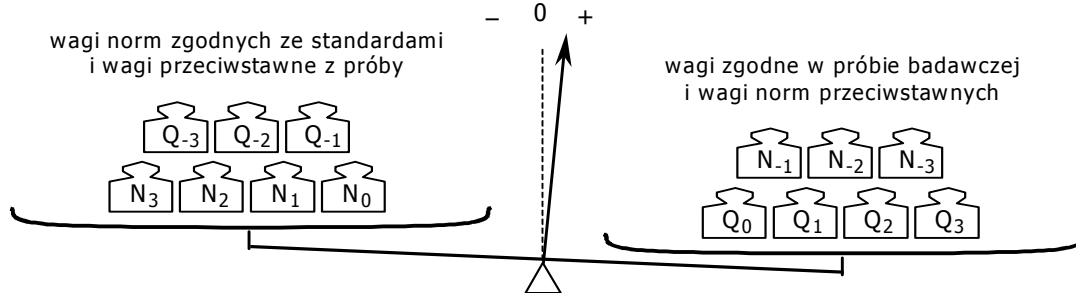
Z symetrii opcji wyboru wypowiedzi na *Skali dwuważonych ocen* bierze się intuicyjnie wyczuwana dwuwymiarowość. Najłatwiej jest nie mieć zdania, co w dużym uproszczeniu oznacza zerową intensywność procesów decyzyjnych. W miarę podejmowania coraz bardziej pewnych decyzji wzrasta stopień wagi wypowiedzi. Zwykle dominują decyzje oparte na racjonalnych przesłankach. Przesłanki wyboru decyzji mogą być różne, a każdy ma prawo do własnego zdania, toteż w konsekwencji oba skrajnie zdecydowane wybory w wymiarze *ważkości* są równo cenne. Jednakże demokratycznym odniesieniem wartościującym wypowiedzi są ukształtowane w danym środowisku tzw. standardy ewaluatywne<sup>26</sup>. Można stwierdzić, czy dany respondent akceptuje wartości uznawane przez większość, a co istotniejsze – czy przypadkiem nie myli się w kwestiach zasadniczych. W tych unormowanych środowiskowo i w miarę wyrazistych zagadnieniach daje się więc oceniać także *wartość* wypowiedzi. Każda wypowiedź respondenta ma równocześnie wymiar intensywnościowy oraz jakościowy (zob. ryc. 14). Tę symptomatyczną dwuwymiarowość najlepiej oddaje izofora: – *Głośno krzyczy, lecz czy ma rację?* Zdarza się bowiem często, że mówca w swoim mniemaniu wypowiada ważne słowa, lecz w ocenie opinii społecznej jego twierdzenia nie mają wartości. Porównanie obu tych wymiarów odgrywa istotną rolę w ewaluacji za pomocą skali dwuważonej. Bierze się to stąd, że strategie skalowania oparte są na opiniach kwalifikowanej większości.



Ryc. 14. Odmienność histogramów jakości i intensywności wypowiedzi



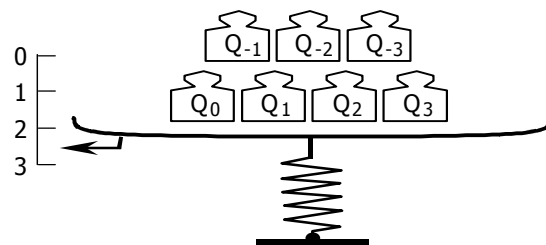
➤ **Wartość wypowiedzi** mierzona jest na takiej zasadzie, jak porównywanie masy na uchylniej wadze dwuszałkowej. Na jednej „szali” umieszcza się wórzec, jakim są standardy ewaluatywne, a na drugiej – opinię badanej grupy. Jednakże zarówno w próbie badawczej, jak i w zbiorowości tworzącej normę część respondentów wypowiada się przeciwstawnie do większości, dlatego też przypadki te należy usytuować na przeciwległych szalkach „wirtualnej wagi”.



Ryc. 15. Wartość wypowiedzi a standardy ewaluatywne

Uchyły „wagi” rzetelnie wyznaczają poziom indykatoru, jeśli odnoszone są do miar normalizujących skalę (patrz *Baza* i *Norma*), o ile oczywiście ewaluator dysponuje w miarę aktualnymi standardami. Jeśli natomiast ich nie posiada, wówczas może wyznaczyć je empirycznie, biorąc do badań dużą statystycznie próbę, reprezentatywną dla danej populacji. Warto tu jednak podkreślić, że standardy są niezbędne tylko do uogólnień i analizy badania jednokrotnego, natomiast różnicowy, dwukrotny pomiar tych samych grup osób eliminuje konieczność posiadania norm i licznych prób (patrz *Miary dystansowe*).

➤ **Ważkość wypowiedzi** wyznaczana jest względem poziomu zerowego, co upodabnia sytuację do zasady określania ciężaru na jednoszałkowej wadze sprężynowej. Ten wymiar intensywności wypowiedzi nie wymaga odnoszenia do żadnych norm, gdyż jest wprost bezpośrednią miarą stopnia zdecydowania respondentów co do swych racji, a ponadto współwyznacznikiem miary dysjunkcji wypowiedzi (*Rozziewu*) oraz podstawą kryteriów służących do weryfikowania bodźców-stwierdzeń (patrz *Konkluzyjność*). Trzeba jednak zwrócić

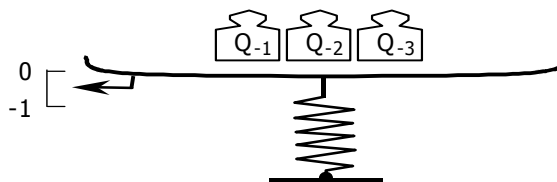


Ryc. 16. Ważkość wypowiedzi

uwagę na to, że miara *Ważkości* jest silnie uzależniona od cech osobniczych respondentów tworzących daną próbę badawczą. Są przecież osoby, które nadużywają wypowiedzi zdecydowanych (*absolutnie!*) oraz takie, które nigdy owych skrajnych sformułowań w kwestionariuszu nie zakreślą, bo nie leżałoby to w ich stylu wysławiania się. Tę niedokładność pomiarową całkowicie eliminuje dwukrotne badanie panelowe różnic dokładnie u tych samych osób.

➤ **Rozziew wypowiedzi** interpretacyjnie oznacza pojawianie się wypowiedzi przeciwnych opinii większości. Jeśli w danej próbie badawczej nie ma żadnego respondenta, który miałby odmienne zdanie, wówczas stwierdzamy, iż rozziew nie występuje. Zatem o statystycznym parametrze *Rozziewu*  $[-b]$  decyduje wyłącznie to, czy w ewaluacji uczestniczyły osoby, które negowały tezę zawartą w bodźcu-stwierdzeniu. Wartość liczbowa parametru określa, na ile stanowczo osoby te przeciwstawiały się standardom ustanowionym przez kwalifikowaną większość. Brak zdania bądź celowe wstrzymanie się od wyrażenia opinii  $[Q_0]$  nie wpływa na wielkość parametru  $[-b]$ .

Można więc rozziew interpretować jako wagę głosów przeciwnych standardom. Jest to zjawisko niekorzystne, dlatego do oznaczenia tegoż parametru trwale przypisany jest znak minusa, a jego wartość jest ujemna bądź zerowa.



Ryc. 17. Rozziew wypowiedzi

➤ **Bilans zmian wypowiedzi** jest porównaniem, na ile zmienił się rozziew wyznaczony w drugim pomiarze względem rozziewu z pomiaru początkowego. Mamy przy tym do czynienia z podwójną komparacją parametrów. Rozziew bowiem oznacza porównanie ważkości wypowiedzi z jej wartością (w ujęciu parametrycznym jest ich różnicą), natomiast bilans oznacza zestawienie zbiorcze przemian zachodzących w obszarach niezgodności ze standardami. Przy interpretacji pamiętajmy o zasadach ewaluacji demokratycznej – każdy ma prawo do osobistych opinii, dlatego nie wolno oceniać indywidualnych wypowiedzi respondentów, a zwłaszcza ich postaw. Podstawą wnioskowania ewaluacyjnego może być wyłącznie reprezentatywny zbiór wielu wypowiedzi. Tę wymaganą wielość wypowiedzi pozyskuje się w dwojaki sposób: z jednej strony poprzez licznosc grup badawczych i reprezentatywnosc respondentów, a z drugiej – poprzez budowę skali pomiarowej zawierającej optymalnie dużo (ale racjonalnie niezbyt dużo) specjalnie zróżnicowanych pozycji testujących, próbujących komponenty indykatorum w różnych aspektach i perspektywach.

<i>Ślady</i>		<i>Wzorce</i>
postawy, doznanie i ocenianie	.....	predyspozycje afektywne
uświadomienie, poznanie	.....	dyspozycje kognitywne
ekstrakcja symptomów	.....	czynniki, składniki, wskaźniki
skala dwuważona	.....	strefy, przedziały, kwantyfikatory
strategie kwantyfikacyjne	.....	indeksacja, kwantyzacja, kalibracja
wartość wypowiedzi	.....	wymiar jakościowy
ważkość wypowiedzi	.....	wymiar intensywnościowy
rozziew i bilans zmian	.....	komparacja dwu wymiarów skali

### **Konstruowanie skali – dobór wskaźników i bodźców**

Najtrudniejszym zadaniem w przygotowaniu nowego, własnego projektu ewaluacji jest utworzenie narzędzia badawczego o odpowiedniej jakości. Owa jakość zależy wyłącznie od tego, z jakich wskaźników zbudowana jest skala. Po pierwsze – wskaźniki muszą być rzeczywistymi wyznacznikami tego, co zdefiniowaliśmy jako indykatum i co zamierzamy zbadać, po drugie – muszą w miarę możliwości jak najpełniej próbkować składową strukturę ocenianej cechy mentalnej bądź różne potencjalne symptomy ewaluowanego procesu. Po trzecie – wskaźniki muszą być addytywne, aby można było je agregować i tworzyć bardziej wiarygodne oznaki. Po czwarte – muszą być wewnętrznie spójne, a jednocześnie komplementarne tak, by testowały aspekty osobowe i czasoprzestrzenne z obydwu dychotomicznych perspektyw (np. ja | inni). Po piąte – wskaźniki powinny nieść możliwie znaczny ładunek informacyjny, co łącznie ze zróżnicowaniem opinii współtworzy moc wnioskotwórczą skali. Po szóste – skala powinna być semantycznie i stylistycznie zharmonizowana.

W precyzyjnym posługiwaniu się pojęciami metodologicznymi potrzebne jest jeszcze pewne uściślenie definicji. Otóż *skalę dwuważoną* traktujemy tu jako abstrakcyjną konstrukcję odwzorowującą model teoretyczny, podczas gdy *Skala dwuważonych ocen* jest ukonkretnionym wzorcem implementacji. Fizycznym natomiast narzędziem do przeprowadzania pomiarów konkretnego indykatum jest *kwestionariusz*. Tak więc, gdy przedstawiałem powyżej sześć właściwości wskaźników warunkujących jakość narzędzia, to odnosiłem je do metodologicznie ogólniejszego poziomu skalowania. Na tym poziomie skale tworzą właśnie wskaźniki, jako kategoria szersza. Natomiast na konkretnym poziomie pomiarowym, gdy ustalili się już sposób wskaźnikowania indykatum, wówczas właściwą parę pojęć tworzy nazwa narzędzia oraz nazwa rodzaju próbkowania – w naszym przypadku *kwestionariusz* i *bodźce-stwierdzenia*. Precyzuje to dość jednoznacznie, iż narzędziem nie jest tu test sprawdzający, a nadto, że pomiar polega na zupełnie czymś innym niż na stawianiu pytań. Podczas pomiaru respondenci otrzymują zwarte komunikaty w formie zdań oznajmiających, wobec których ustosunkowują się. Zdania te odnoszą się do kwestii ważnych nie tylko dla ewaluatora, lecz także do tez zagadnieniowo nieobcych i emocjonalnie nieobojętnych dla osób badanych. Pobudza to ich do *reakcji-wypowiedzi*, które w skalach celowo są ograniczane i ujednolicane. Zbiór wypowiedzi wobec tez zawartych w kwestionariuszu tworzy wskaźniki.

W niniejszej publikacji przyjmijmy szeroki obszar interpretacji tego, co nazywamy *wskaźnikami*. Wyjaśnię to na przykładzie pomiaru temperatury. Wskaźnikami pośrednimi mogą być m.in. rozszerzalność cieplna lub zmiana przewodnictwa. Do określenia stanu tych cech fizycznych potrzebne są różne próbki (słupek rtęci, bimetal lub termistor) oraz odpowiednie wskaźniki

(skala wzdłużna, skala dookólna i wskazówka, bądź wyświetlacz cyfrowy). Zatem jako wskaźniki można uznać: 1) pewne symptomy fizyczne świadczące o wyższej lub niższej temperaturze, takie jak rozszerzalność i przewodnictwo; 2) określone oznaki stanu próbników, takie jak długość słupka rtęci, stopień skrętu bimetalu lub oporność termistora; 3) konkretne parametry liczbowe wskazywane na skali lub wyświetlane cyfrowo.

Wskaźnikowanie cech mentalnych lub stanu procesów społecznych jest daleko bardziej skomplikowane<sup>27</sup>. Wskażę tu jednak na pewne podobieństwo podejścia pomiarowo-wartościującego. Otóż w profesjonalnych prognozach pogody pojawia się ostatnio także pojęcie *temperatury odczuwanej*. Oprócz precyzyjnego określenia w jednostkach fizycznych podawana jest też ważna informacja o doznaniach, jakie będą skutkiem oddziaływania danej temperatury w połączeniu z innymi czynnikami pogodowymi. W odbiorze społecznym ta informacja wydaje się być cenniejsza i bardziej komunikatywna niż dane czysto fizyczne. Potwierdza to wzrastającą rangę ewaluacji naturalistycznej, budowanej na tzw. *miękkich* wskaźnikach, uważanych przez zachowawczych scjentyistów za zbyt subiektywne. Moim zdaniem wskaźniki powinny dotyczyć zawsze zarówno komponentów *kwalitatywnych* (odnoszących się do jakości, usytuowanych np. na osi ciepło|zimno), jak też wyznaczników *kwantytatywnych* (odnoszących się do ilości, wyrażonych np. w stopniach Celsjusza). Strategia ewaluacji demokratycznej zakłada przy tym, że to właśnie odczucia kwalifikowanej większości respondentów są podstawą kalibrowania skali.

Przyjmijmy najbardziej uogólnioną i syntetyczną definicję wskaźników:

**Wskaźniki**, to zmienne najniższego poziomu ekstrakcji, wyrażone poprzez:

- elementarne symptomy reprezentujące wybór komponentów indykatum,
- elementarne oznaki stanu cech i zjawisk w postaci parametrów bazowych,
- argumenty zmiennej losowej ułożone w macierzach danych źródłowych.

Z ujęcia tego wynika, że wskaźniki w skali dwuważonej wytyczają zarówno jakościowy obszar badawczy (próbkując elementarne zmienne, które należy zmierzyć), jak i wyznaczają ilościowe stany zmiennych (ukazując parametry, jakie udało się zmierzyć w ściśle ustalony sposób). Podczas tworzenia nowej skali pomiarowej cyklicznie przeplatają się dwie fazy: 1) doboru wskaźników ze względu na najlepsze próbkowanie koniecznych do zbadania symptomów indykatum oraz 2) weryfikacji i optymalizacji wskaźników ze względu na statystycznie najlepsze wskaźnikowanie oznak empirycznych. Opis doboru optymalnych wskaźników i ich weryfikacji (w tym i w następnym rozdziale) ograniczę do przesłanek teoretycznych, a jako że podejmowanie konkretnych działań polega na formowaniu treści bodźców-stwierdzeń, dlatego sposoby redagowania kwestionariusza objaśniam w części praktycznej (str. 114).

Grupy podstawowych właściwości, które decydują o pomiarowej dobroci wskaźników i tym samym o jakości skali, można sklasyfikować następująco:

Trafność	Komplementarność	Konkluzyjność
- ekstrakcyjna	- treściowa	- merytoryczna
- operacjonalizacyjna	- osobowa	- statystyczna
- predykcyjna	- czasowa	
Kompletność	Addytywność	Rzetelność
- strukturalna	- kwalitatywna	- metodologiczna
- eksplikacyjna	- kwantytatywna	- statystyczna

Te z powyższych właściwości, które związane są przede wszystkim z budową i eksploracją nowego narzędzia, scharakteryzuję w bieżącym rozdziale, a te, które dotyczą głównie fazy weryfikacji statystycznej, objaśnię w następnym.

➤ **Trafność wskaźników** oznacza, czy są one faktycznym odzwierciedleniem zmiennych elementarnych, tworzących badaną zmienną ogólną. Trafność tę powinno się określać we wszystkich aspektach: teoretycznych, empirycznych i treściowych, jednak owe aspekty nie powinny być próbą klasyfikacji metod sprawdzania trafności (jak czyni to wielu autorów), lecz integralną składową każdej z metod. Nie jest słuszne nazywanie trafności „teoretyczną”, jeśli i tak należy ją potwierdzić empirycznie. Analogicznie metody oceniania trafności zaklasyfikowane do grupy „empirycznych” nie mogą być przeprowadzane bez osadzenia na założeniach teoretyczno-metodologicznych. Wreszcie, spotykane w publikacjach odżegnywanie się metrologów-scjentystów od rzekomo „fasadowych” metod formowania i weryfikowania trafności treściowej powoduje, że twórcy narzędzi badawczych przykładają zbyt małą wagę do użytych słów, a przecież są one źródłem wszelkiej jakości lub bylejakości skali pomiarowej. Nie ma jakiegś jednej uniwersalnej metody gwarantującej uzyskanie trafności wskaźników, ani też nie ma pewnego sposobu sprawdzania owej właściwości, dlatego konieczne jest zastosowanie określonych zasad konstruowania skali, aby uzyskać trafność potencjalnie i prawdopodobnie wysoką. Przedstawię te najbardziej istotne i brzemienne w skutkach uwarunkowania jakościowe:

*Trafność ekstrakcyjna.* Głównym źródłem trafności całej skali jest prawidłowa ekstrakcja indykatum, zarówno pod względem optymalnego rozwarstwienia zmiennej ogólnej na zmienne cząstkowe, jak też ze względu na właściwe rozdysponowanie poziomów ekstrakcji, czyli ustalenie: co ma być czynnikiem, co składnikiem, a co elementarnym wskaźnikiem. Odbywa się to w fazie budowy modelu teoretycznego oraz w fazie definiowania zmiennych, a polega na skonstruowaniu trafnej struktury schematu ekstrakcji i interpretacji. Do tego celu wykorzystuje się dotychczasowy dorobek dziedziny naukowej, która zajmuje się tym, co stanowi obiekt ewaluacji. Jeśli indykatum dotyczy cech mentalnych, to sięga się po teorie, twierdzenia i aksjomaty psychologii. Jeśli

obiektem ewaluacji jest proces edukacyjny, czerpie się z dorobku różnych nauk społecznych. Poszukuje się ukształtowanych już i sprawdzonych, a przede wszystkim uzgodnionych między uczonymi pierwiastków konstytuujących to pojęcie teoretyczne, którego sens empiryczny zamierzamy poddać ewaluacji<sup>28</sup>. Najbardziej przydatne do wykorzystania podczas tworzenia skal są te teorie przedstawione w formie struktur pojęciowych, które zawierają niejako gotowe schematy ekstrakcji i gotowe definicje komponentów. Badacz może korzystać z nich odtwórczo, może je doskonalić lub tylko wzorować się na nich, tworząc własny model strukturyzacji i deskrypcji zmiennych. Trafność ekstrakcyjną skali kształtuje się więc i weryfikuje poprzez odniesienie do uznanych teorii.

*Trafność operacjonalizacyjna.* Ta właściwość oznacza stopień odpowiedniości wskaźnikowania pośredniego i uwarunkowana jest prawidłową konwersją z cech bezpośrednio niemierzalnych na płaszczyznę wskaźników empirycznie wyznaczalnych. Niezależnie od tego, czy model indykatum wyrażony w formie struktury pojęciowej jest odtworzony, czy zupełnie nowo utworzony, konieczne jest przejście od ujęcia teoretycznego do praktycznego. Zmienne w warstwie jakościowej są pojęciami, a pojęcie to jedynie umysłowa abstrakcja, umowny symbol, werbalne bądź wyobrazeniowe odwzorowanie obiektu, cechy, stanu, zjawiska, procesu, relacji itp. Wszystko to zależy od przyjętego indykatum. Zmienne elementarne tworzące indykatum są pierwotnie zawsze pojęciami\*, które wymagają ścisłego zdefiniowania. W definiowaniu tego, co zamierzamy zbadać, najczęściej wykorzystujemy inne pojęcia, których sens interpretacyjny jest już społecznie bardziej uzgodniony. Rzadziej użyć można wprost definicji operacyjnej, objaśniającej sens pojęcia w formie opisu czynności dochodzenia do właściwej interpretacji poprzez osiągnięcie stanu doświadczalnie reprezentującego dane pojęcie<sup>29</sup>. Tymczasem w budowaniu skali pomiarowej konieczna jest właśnie *operacjonalizacja*, rozumiana jako przejście z poziomu pojęciowo-teoretycznego na poziom empiryczno-obszewacyjny. Jest to niezbędne tym bardziej, że ewaluacje dotyczące cech mentalnych oraz procesów społecznych z natury rzeczy opierać się muszą na pośrednich wskaźnikach indykatum. Od poprawności użycia wskaźników pośrednich, od wierności odwzorowania definicyjnych komponentów indykatum za pomocą obserwowalnych quasi-desygnatów zależy w ogóle trafność pomiaru. Trafność tę jednak nie tylko najtrudniej jest uzyskać, lecz także najtrudniej jest zweryfikować. Wszystko to bowiem rozgrywa się w warstwie semantycznej komunikowania społecznego. W pierwszym rzędzie twórca skali musi skonstruować w jego przeświadczeniu najlepsze „próbki pośrednie” – w naszym modelu są to bodźce-stwierdzenia.

---

\* W publikacjach spotkać można nieścisłość polegającą na podawaniu kategorii *Wzrost* jako przykładu zmiennej ilościowej. A przecież pierwotnie jest to zmienna jakościowa wymagająca zdefiniowania, czy chodzi o cechę, stan chwilowy, czy o zjawisko, proces, kierunek trendu itp.

Może poddać ich treść pod ocenę sędziów kompetentnych, opiniujących to, czy skala rzeczywiście mierzy indykatum. Musi je zweryfikować empirycznie, czy wzajemne relacje są zgodne z założeniami teoretycznymi. I może próbować rzeczy mało realnej – porównać wyniki z uzyskanymi za pomocą innej skali.

*Trafność predykcyjna.* W ewaluacji splotowej (za pomocą narzędzia skalowanego) metodologia badawcza oparta jest na wcześniej opracowanych teoriach i modelach z góry zakładających występowanie określonych symptomów. Z tego względu wskaźniki powinny mieć taką właściwość, że przeprowadzone na ich podstawie badania empiryczne potwierdzą przewidywania lub pozwolą obalić daną teorię. Oczywiście nie chodzi tu o budowanie skali „spełniającej życzenia” badacza, lecz o uzyskanie takiej trafności prognostycznej, która przy zachowaniu wszelkich metodologicznych prawideł rzetelnego pomiaru pozwoli na podjęcie istotnych statystycznie decyzji o przyjęciu bądź odrzuceniu tez. Każdy bowiem wskaźnik w *Skali dwuważonych ocen* jest wyznacznikiem stopnia akceptacji elementarnych tez zawartych w bodźcach-stwierdzeniach. Pierwszym warunkiem trafności predykcyjnej jest więc to, czy sformułowane wcześniej tezy elementarne znalazły empirycznie jednoznaczne potwierdzenie w argumentach rozstrzygających kwestie badawcze. Jeśli zachodzi przypadek taki, że dane stwierdzenie jest zgodne z modelem, z intuicją twórcy skali oraz z opinią sędziów kompetentnych, a jednak respondenci przeczą stwierdzeniu, to strategia metodologii skalowanej nakazuje odrzucenie tegoż wskaźnika. Jeden bowiem wskaźnik nie wystarcza do obalenia standardu ewaluatywnego. Z tego powodu sprawdza się również drugi warunek trafności predykcyjnej. Mianowicie analizuje się, czy wzajemne relacje wskaźników zagregowanych (przede wszystkim na poziomie czynników) są zgodne z oczekiwanym modelem powiązań i zmienności wynikających z teorii. Jeśli także zależności pomiędzy czynnikami empirycznymi okażą się niezgodne z modelem, to wówczas mamy do czynienia albo z nietrafnymi wskaźnikami, albo z nieprawidłową podstawą teoretyczną (lub z niereprezentatywną próbą badawczą i z wieloma innymi przyczynami, które mogą pojawić się w zawiłym procesie ewaluacji splotowej). Takie przypadki zupełnie eliminują możliwość uznawania narzędzia jako skali. Skalowanie bowiem opiera się na spełnianiu ściśle wyznaczonych kryteriów, a jeśli doświadczenie nie potwierdza przewidywań, to gdzieś musi tkwić błąd. Do podjęcia decyzji o trafności bądź nietrafności predykcyjnej konieczna jest empiryczna weryfikacja skali połączona z analizą statystyczną. Na tym polu sprawdza się, czy zgodne z oczekiwaniami były następujące oznaki pomiarowe: 1) identyczne ze standardami znaki walencyjne, 2) spójne i zbieżne składniki, 3) stosowne do struktury modelu relacje czynników, 4) przyrosty świadomości, 5) korekta nadmiernej afektywności, 6) równomierne tendencje wewnątrz prób, 7) jednomodalne rozkłady, 8) zgodne wyniki porównania trendów i efektów.

➤ **Kompletność wskaźników** oznacza, na ile reprezentatywnie wypełniają one taksonomię komponentów tworzących indykatum i czy zbiór oznak empirycznych będzie wystarczający do wyjaśnienia zachodzących zjawisk. Właściwość ta z jednej strony odnosi się do całościowego odwzorowania struktury modelu, a z drugiej – do wielości zagadnień badawczych, jakie planuje się rozstrzygnąć w rezultacie pomiaru za pomocą skali. Nie chodzi tu jednak o dużą liczbę wskaźników, lecz o ich zupełność kategorialną, relacyjną oraz problemową. Muszą one kompleksowo próbować strukturę ewaluowanej cechy mentalnej bądź szerokie spektrum właściwości ocenianych obiektów i zjawisk, a także wielorakie symptomy jakościowe oddziaływań, stanów chwilowych i procesów. W praktyce budowania skal kompletność wskaźnikowania jest daleko idącym kompromisem pomiędzy wygórowanymi potrzebami eksploracyjnymi a silnie limitowaną złożonością kwestionariusza. W przypadku skali o 24 pozycjach na zmienne połówkowe składa się po 12 wskaźników. Pary wskaźników komplementarnych próbują w zasadzie tę samą cechę, tylko z innej perspektywy, toteż rezultat połówkowy tworzony jest z 6 komponentów albo z 3 czynników. Optymalne są również skale złożone z 32 lub 36 pozycji, przy czym *pozycjami* ogólnie nazywa się składowe próbujące kwestionariusza, niezależnie od tego, czy są to pytania, czy bodźce-stwierdzenia, czy opcje do wyboru, czy też jakieś inne formy wypowiedzi, np. graficzne. W przypadku modelu skal dwuważonych liczba pozycji kwestionariusza odpowiada dokładnie liczbie wskaźników skali.

*Kompletność strukturalna.* Właściwość ta dotyczy zupełności odwzorowania modelu teoretycznego za pomocą zbioru wskaźników w taki, w miarę pełny sposób, ażeby próbkowane były zarówno wszystkie najważniejsze komponenty indykatum, jak też reprezentowane były wszystkie poziomy struktur definicyjnych, określających wskaźniki, składniki, czynniki oraz wyniki połówkowe, dające się zagregować w jeden wynik ogólny. Kompletność strukturalna silnie związana jest z trafnością ekstrakcyjną – obie właściwości dotyczą rozwarstwienia zmiennej ogólnej na cząstkowe, przy czym kompletność uzyskuje się przez abstrahowanie | kondensowanie pojęć oraz przez atomizację zmiennych.

*Kompletność eksplikacyjna.* Ta właściwość oznacza ujęcie w skali wszystkich tych wskaźników, których pomiar jest niezbędny do osiągnięcia wytyczonych celów badawczych i wyjaśnienia zakreślonego obszaru problemów. Wskaźniki podczas obu pomiarów przyjmują różne wartości, dzięki czemu uzyskuje się dynamiczny obraz relacji zachodzących pomiędzy komponentami indykatum. W ewaluacji splotowej celowo próbkuje się czynniki zarówno współbieżne, jak i przeciwstawne. Nie eliminuje się ich ze skali (jak w innych strategiach), lecz świadomie wprowadza do narzędzia badawczego, w celu uzyskania pełnej mocy wyjaśniającej. Kompletność eksplikacyjna wiąże się z trafnością predykcyjną – należy przewidzieć, co da się wyjaśnić, a czego w ogóle nie można zmierzyć.



➤ **Komplementarność wskaźników** oznacza prawidłowe skompletowanie par kontrolnych bodźców-stwierdzeń o takiej treści, aby testowały stany i zjawiska bipolarnie, dychotomicznie, przestrzennie, wieloaspektowo, perspektywicznie, a jednocześnie były spójne, zwarte i stanowiły harmonijnie dobraną całość. Oczywiście komplementarność dotyczy też składników, czynników i wyników połówkowych, lecz wiąże się to z procesem ekstrakcji, natomiast na poziomie wskaźników właściwość tę kształtuje się poprzez ekspansję aspektów czasoprzestrzennych, w warstwie semantycznej i stylistycznej bodźców-stwierdzeń.

*Komplementarność treściowa.* Kwestionariusz składa się ze zbioru stwierdzeń, które podczas czytania przez respondentów powinny układać się w pewną logiczną ciągłość. Dwa wskaźniki próbujące tę samą cechę umieszcza się obok siebie, natomiast obie pary tworzące czynnik nie muszą już być razem. Kolejność wynika z ogólnego toku wyводу, choć tezy są na tyle różne, że trudno o ciągłość narracyjną. Możliwość udzielenia wypowiedzi nie może być uzależniona od innych, wcześniej przeczytanych tez. O komplementarności decyduje także wprowadzenie tzw. *przeplatania polaryzacyjnego*, które polega na nieregularnym umieszczeniu w skali zarówno tez prawdziwych, jak i fałszywych, co umożliwia wykrycie przypadków kwestionariuszy wypełnionych nierzetelnie, na przykład bez czytania stwierdzeń. Tezy fałszywe redaguje się tak, aby nie były zdaniem przeczącymi i nie zawierały partykuły 'nie'. Każda z tez powinna być wyrażona zwężym i zrozumiałym językiem.

*Komplementarność osobowa.* Jeśli skala dotyczy pomiaru cech mentalnych, to większość stwierdzeń powinna być redagowana jako wypowiedź w pierwszej osobie ('ja'). Jednak dla pełnego poznania warto, aby kilka pozycji skali próbowało to, jak dana cecha wygląda u innych. Obok wskaźników introspekcyjnych wprowadza się więc również ekstraspekcyjne, choć te ostatnie są mniej wiarygodnym wypowiedzeniem się „za innych”. Jeśli skala ma mierzyć głównie jakość procesów, to wskaźników introspekcyjnych może być mniej. Krańcowo może ich nie być w ogóle, gdy skala ma służyć do oceny wartości wytworów. Należy jednakże zawsze próbować lokalne oraz globalne aspekty wartości.

*Komplementarność czasowa.* Właściwość tę uzyskuje się poprzez wprowadzenie zarówno wskaźników retrospekcyjnych, diagnozujących dotychczasowy stan zjawisk, jak też futurospekcyjnych, prognozujących stany przyszłe. Ponadto kontrolne bodźce-stwierdzenia powinny umożliwiać wnioskowanie, czy stany te są tylko chwilowe i przejściowe, czy może długotrwałe bądź utrwalone już na stałe. Możliwość oceny aspektów czasowych ma olbrzymie znaczenie interpretacyjne. Komplementarność tego rodzaju obliuguje jednakże do takiego zredagowania stwierdzeń, ażeby formy czasownikowe były logiczne zarówno w pomiarze początkowym, jak i końcowym. Żadna z tez nie może być bezprzedmiotowa ze względu na upływ czasu lub zakończenie procesu oddziaływań.

➤ **Addytywność wskaźników** oznacza możliwość agregowania zmiennych elementarnych w jedną zmienną ogólną. Skalę złożoną z wielu wskaźników buduje się po to, ażeby mieć większą pewność co do rzetelności pomiarowej wyniku najbardziej uogólnionego. W klasycznej metodologii badawczej formułuje się cele i zadania badawcze, problemy ogólne i problemy szczegółowe, definiuje się zmienne i operacjonalizuje do postaci wskaźników. Wszystko to prowadzi do powstania zbioru pytań, na które badacz poszukuje odpowiedzi drogą pomiarów empirycznych. Większość badań społecznych prowadzona jest za pomocą narzędzi składających się z kilkudziesięciu pytań, lecz niestety – względy semantyczne i kwantyzacyjne powodują, że wypowiedzi zazwyczaj nie dają się połączyć w całość. W konsekwencji każdą ze zmiennych cząstkowych analizuje się i relacjonuje w sprawozdaniu odrębnie, pomimo że wnioskowanie na podstawie wypowiedzi wobec pojedynczego pytania nie ma dostatecznej mocy wyjaśniającej, aby można było uznać to za rezultat badania. Prace doktorskie i magisterskie ujawniają brak umiejętności kumulowania cech jakościowych, a nawet jeśli zmienne cząstkowe z natury są addytywne, to niejednorodność opcji wypowiedzi powoduje trudności z agregowaniem. W narzędziach skalowanych wszystkie wskaźniki obligatoryjnie muszą być addytywne i jednolite, aby można było je łączyć w bardziej wiarygodne oznaki. Odnosi się to zarówno do kumulowania jakościowego, jak i do ilościowego.

*Addytywność jakościowa.* Właściwość ta umożliwia kumulowanie zmiennych w sferze jakościowej i wiąże się bezpośrednio z trafnością ekstrakcyjną. Jeśli model teoretyczny rozwarstwienia indykatum jest prawidłowy, to agregacja jako proces odwrotny ma z góry ustalony schemat i nie wymaga empirycznej weryfikacji. Co więcej – w procesach społecznych pewne czynniki wzmacniają jakość indykatum, a inne obniżają, gdyby więc eliminować te, które działają przeciwnie (jak w innych skalach), to wynik ogólny byłby nieprawdziwy. O addytywności jakościowej decyduje to, czy elementarne próbkowane symptomy mieszczą się w zbiorze pojęć definiujących indykatum, co w pełni już uprawnia do scalania składników i czynników w wynik ogólny.

*Addytywność kwantytatywna.* Scalanie ilościowe staje się możliwe dzięki temu, że zakładamy monotoniczne charakterystyki wskaźników jakości i na nich przeprowadzamy kwantyzację. Cechy niefizyczne oraz niemiarowe symptomy zjawisk społecznych próbkujemy w zunifikowany sposób, a próbkom przypisujemy umowne wagi liczbowe, które wyrażają narastające natężenie jakości. Unifikacja pomiaru za pomocą skal i konsolidacja próbek polega na tym, że respondentom podaje się zbiór zwartych tez (lub pytań), wobec których mogą wyrazić swoje zdanie, lecz wyłącznie w formie skategoryzowanych wypowiedzi. Bodźce-stwierdzenia wywołują u respondentów odmienne reakcje-wypowiedzi, jednak opcje wyboru ograniczone są tylko do wymiarów, jakie mamy zmierzyć.

Przedstawione tu właściwości wskaźników pozostają w ścisłej zależności od dobroci pomiarowej próbników. Najkorzystniejszymi próbnikami zalecanymi dla skal dwuważonych są tzw. *bodźce-stwierdzenia*, definiowane następująco:

**Bodźce-stwierdzenia** to zbiór zdań oznajmujących istotne operacjonalnie tezy i jednocześnie próbujących opinie respondentów poprzez pobudzenie ich do wypowiedzi w danej kwestii. Jako bodźce muszą być emocjonalnie nieobojętne, a jako stwierdzenia – treściowo zrozumiałe i implikacyjnie trafne, gdyż są źródłem pośredniego wskaźnikowania komponentów indykatum.

Definicja ta wymaga rozwinięcia. Otóż, gdy nazywam stwierdzenia teza-  
mi istotnymi operacjonalnie, to mam na myśli ważność ze względu na proces operacjonalizacji zmiennych. Każdą z elementarnych zmiennych wyraża się w postaci tezy czynnościowej, którą respondenci w mniejszym lub większym stopniu akceptują bądź odrzucają. Oczywiście nie są to tezy ani hipotezy rozumiane tradycyjnie, a więc nie są to twierdzenia rozstrzygające podstawowe problemy badawcze. Co więcej – niezwykle rzadko treść tezy bezpośrednio dotyczy tej kwestii, jaką mamy zbadać. Paradoksalnie lepsze w skalach są takie bodźce-stwierdzenia, które semantycznie nie ujawniają wprost tego, co jest próbkowane. Jeśli konkretna pozycja skali ma mierzyć na przykład poziom motywacji, to lepiej, ażeby słowa *motywacja* w ogóle w niej nie było. Wszystkie stwierdzenia w skali powinny składać się z prostych sformułowań i potocznie zrozumiałych słów. W rezultacie tego powstaje wrażenie, jakoby kwestionariusz składał się wręcz ze stwierdzeń trywialnych, tymczasem ich głębszy i niejako ukryty sens musi być trafnie uformowany przez twórcę skali. Każde ze stwierdzeń jest elementarną tezą, stanowiącą część indykatum. Owe części niosą ze sobą *ładunek walencyjny* pozytywności/negatywności. Elementarne tezy są potwierdzane bądź odrzucane przez respondentów, a to stanowi strategię wartościowania i formowania standardów ewaluatywnych.

Drugim istotnym wyznacznikiem dobroci pomiarowej bodźców-stwierdzeń jest niesiony przez ich treść *ładunek emocjonalny*. Zdania oznajmujące muszą zawierać treści, wobec których respondenci nie pozostaną obojętni. Ważne jest to przede wszystkim ze względu na potrzebę uzyskania wyraziście spolaryzowanych opinii, a wiąże się z tym, że badani mają do dyspozycji całkowicie neutralną opcję '*brak zdania*'. Gdyby treść stwierdzeń była dla nich obojętna, to wybieraliby w większości właśnie tę opcję, mimo że oznacza ona brak opinii, a nie obojętność. Takie pomiary nie dostarczałyby konkluzyjnych rezultatów. Dwuczłonowa nazwa *bodźce-stwierdzenia* jest zasadna także z innego względu. Otóż skala próbkuje cechy mentalne, a warunkiem ich ujawnienia jest właśnie pobudzenie do reakcji. Bodźce muszą być jednak subtelne, ażeby reakcja odzwierciedlała badane cechy, a nie była złośliwym odreagowaniem na badanie.

## **Optymalizacja skali – weryfikacja oznak empirycznych**

Istotą dostrajania skali pomiarowej jest formowanie jakości kwestionariusza, pierwotnie – poprzez perfekcyjną redakcję treści bodźców-stwierdzeń, a następnie – poprzez selekcję oraz modyfikację sformułowań na podstawie wyników badań weryfikujących, prowadzonych w różnych środowiskach na licznych i reprezentatywnych próbach badawczych. Praktycznie zawsze przed ewaluacją właściwą nieodzowne jest przeprowadzenie badania wstępnego, a przy użyciu skali zwykle potrzeba aż kilku cykli optymalizacji i weryfikacji. Empiryczna weryfikacja narzędzia jest kluczowym procesem w strategiach przygotowujących trafne i rzetelne badania oparte na pomiarach skalowanych. Z jednej strony chodzi o sprawdzenie, jak w praktyce funkcjonują próbniaki, a z drugiej – jakie otrzymuje się wskaźniki. Jeśli próbnikami w danej skali są bodźce-stwierdzenia, to sprawdza się, czy rzeczywiście wszystkie zawierały prawidłowe ładunki operacyjne, walencyjne oraz emocjonalne. Wynika to z charakterystyki uzyskanych wskaźników, a w naszym modelu konkretnie ze statystycznych rozkładów reakcji-wypowiedzi. Zdefiniujmy ostatnie pojęcie:

**Reakcje-wypowiedzi** to zbiór elementarnych, ściśle skwantyfikowanych wyborów opcji, jakich dokonali respondenci w wyniku pobudzenia celowo dobranym zbiorem bodźców-stwierdzeń. Owe wybory są ustosunkowaniem się do tez zawartych w skali pomiarowej, a pośrednio są także wskaźnikami uzewnętrzniającymi próbkowane stany mentalne uświadomienia i postaw.

Uważny czytelnik może mieć słuszne wątpliwości co do tego, czy ostatnie zdanie definicji nie jest niepotrzebnym zawężeniem ogólnego modelu skali. Przecież *Skalę dwuważonych ocen* można stosować nie tylko do pomiaru cech mentalnych. Warto więc w tym miejscu zwrócić uwagę na znamienne piętno, którego źródłem w badaniach są stany mentalne respondentów. W sytuacji, gdy obiektem ewaluacji są wytwory intelektualne, a wypowiedzi pochodzą od osób pełniących rolę jurorów, ich osobiste cechy niestety wpływają na oceny. Przykładowo, w ocenianych multimediami lub stronach internetowych istnieje granica porcjowania treści i optymalnie funkcjonalna forma graficzna, tymczasem brak świadomości owych kardynalnych zasad budowy komunikatów prowadzi do błędnej oceny: im więcej „esencji” i lepsza „graficzka”, tym lepiej. Podobnie aprobująca postawa dla wirtualnej przemocy na ekranie prowadzi do powszechnego, wyższego wartościowania filmów lub gier komputerowych o jak najbardziej realistycznym odwzorowaniu „totalnej masakry”. Są to fakty empiryczne, które nie mogą być wyznacznikami standardów ewaluacyjnych oraz niepokojące stany mentalne, które mogą zakłócać pełnienie ról jurorów. Z całą pewnością są to jednak bezpośrednie wskaźniki niskiego poziomu kultury medialnej. Dlatego też twórca skali pomiarowej musi zweryfikować to, co

miały według jego intencji próbkować bodźce-stwierdzenia, a co rzeczywiście wyrażają reakcje-wypowiedzi: czy cechy mentalne, czy wartość wytworów, czy może jeszcze coś innego<sup>30</sup>. Demokratyczną weryfikację skali (na próbie będącej podmiotem ewaluacji) powinna uzupełniać ekspercka walidacja narzędzia.

Sposoby sprawdzania trafności wskaźników przedstawiłem w poprzednim rozdziale. W tym miejscu objaśnię dwie dalsze właściwości skali, istotne ze względu na proces skalowania, weryfikacji i optymalizacji narzędzia. Pierwsza właściwość jest znamiona i niejako konstytutywna dla *Skali dwuważonych ocen*. Chodzi o *konkluzyjność* wskaźników. Druga ma znamiona właściwości uogólniającej, uznawanej jako wyznacznik *rzetelności* narzędzia i pomiaru.

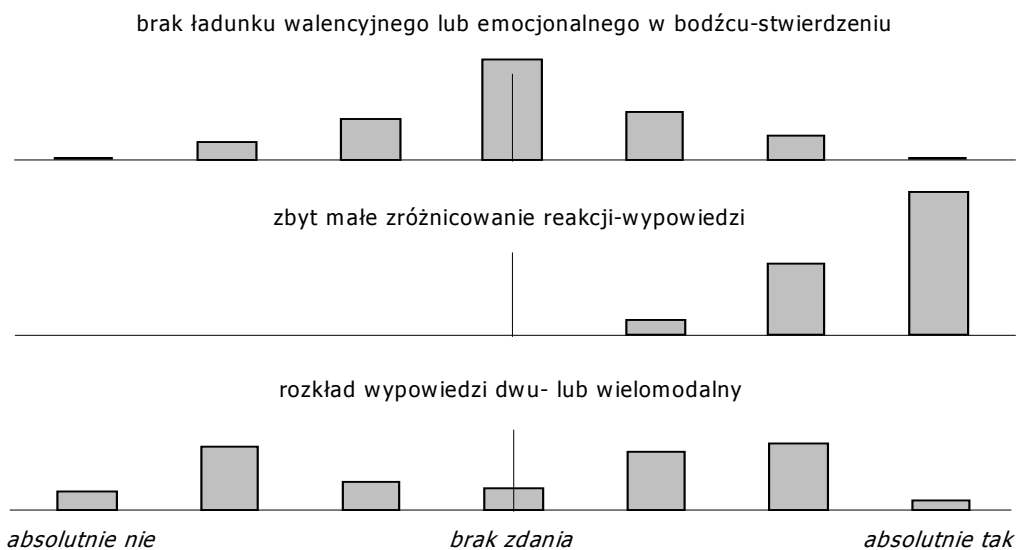
➤ **Konkluzyjność wskaźników** oznacza ich przydatność ze względu na siłę wnioskowania. W oczywisty sposób łączy się ona z trafnością i kompletnością wskaźników. Okazuje się jednak, że spektrum tego, co wyrażają sobą reakcje-wypowiedzi, jest szersze niż zwykle rozstrzygnięcie elementarnych tez. Oznaki empiryczne są znacznie pełniejsze od jednobitowej informacji prawda/fałsz. Rozstrzygnięcie prawdziwości stwierdzenia jest potrzebne do ustalenia znaku walencyjnego tezy, lecz dodatkowo, na podstawie wzajemnych relacji pomiędzy wskaźnikami tworzącymi skalę, możliwe jest formułowanie istotnych ewaluacyjnie wniosków o stanie komponentów indykatum w ujęciu merytorycznym i statystycznym, w aspektach zarówno analizy jakościowej, jak też ilościowej.

*Konkluzyjność merytoryczna.* Właściwość ta odnosi się do wnioskotwórczości w zakresie: odkrywania symptomów wartości, stwierdzania jakości cech lub procesów, oceny jakości oddziaływań itp., a więc do kwalitatywnej warstwy ewaluacji indykatum. O konkluzyjności świadczy zmienność cech, które ze swej natury powinny wzbogacać się jakościowo; świadczą też tendencje do społecznego formowania takich standardów ewaluatywnych, co do których nie było pewności, czy się uformują i w jakim kierunku będą zmierzać. O sile wnioskowania decydują m.in.: spójność i współbieżność oznak, harmonijność przemian, efektywność procesów, powtarzalność oraz wiarygodność wyników, przy czym wnioski muszą być istotne problemowo i ważne poznawczo.

*Konkluzyjność statystyczna.* Ten rodzaj wnioskotwórczości kształtowany jest poprzez minimaksowe ważenie zdań i justowanie estymatorów. Polega to na cyklicznej modyfikacji treści bodźców-stwierdzeń i pomiarach próbnych tak długo, aż analiza statystyczna wykaże optymalne dopasowanie parametrów i rozkładów empirycznych do zakresu pomiarowego skali. Narzędzie można uznać za prawidłowo wyskalowane dopiero wtedy, gdy rozkład cech mierzonych za jego pomocą mieści się w zakresie, a o konkluzyjności świadczy to, czy doświadczalnie wyznaczone miary pozycyjne wskaźników mają tendencje do pełnego rozpostarcia – od zera lub neutralnego punktu skali aż po jej kres. Siłę statystycznego wnioskowania z pojedynczego wskaźnika wyraża wprost

parametr *Konkluzyjności*, natomiast wnioskotwórczą zdatność zagregowanych oznak empirycznych wyznaczają pośrednio różne miary: dyspersji, dysjunkcji, ufności, selektywności, zawilości, trafności oraz tendencji i dynamiki.

Konkluzyjność jest podstawowym kryterium doboru wskaźników do skali, a tym samym – głównym kwantyfikatorem bodźców-stwierdzeń. W pierwszej fazie weryfikacji kwestionariusza usuwa się z niego wszystkie te stwierdzenia, które nie spełniają warunków kreowania konkluzyjności. Reakcje-wypowiedzi muszą być: 1) spolaryzowane większościowo ze względu na znak walencyjny, 2) wyraźnie zróżnicowane, przede wszystkim w całej strefie zgodności opinii, 3) jednomodalne, walencyjnie niesprzeczne i wzajemnie nie wykluczające się.

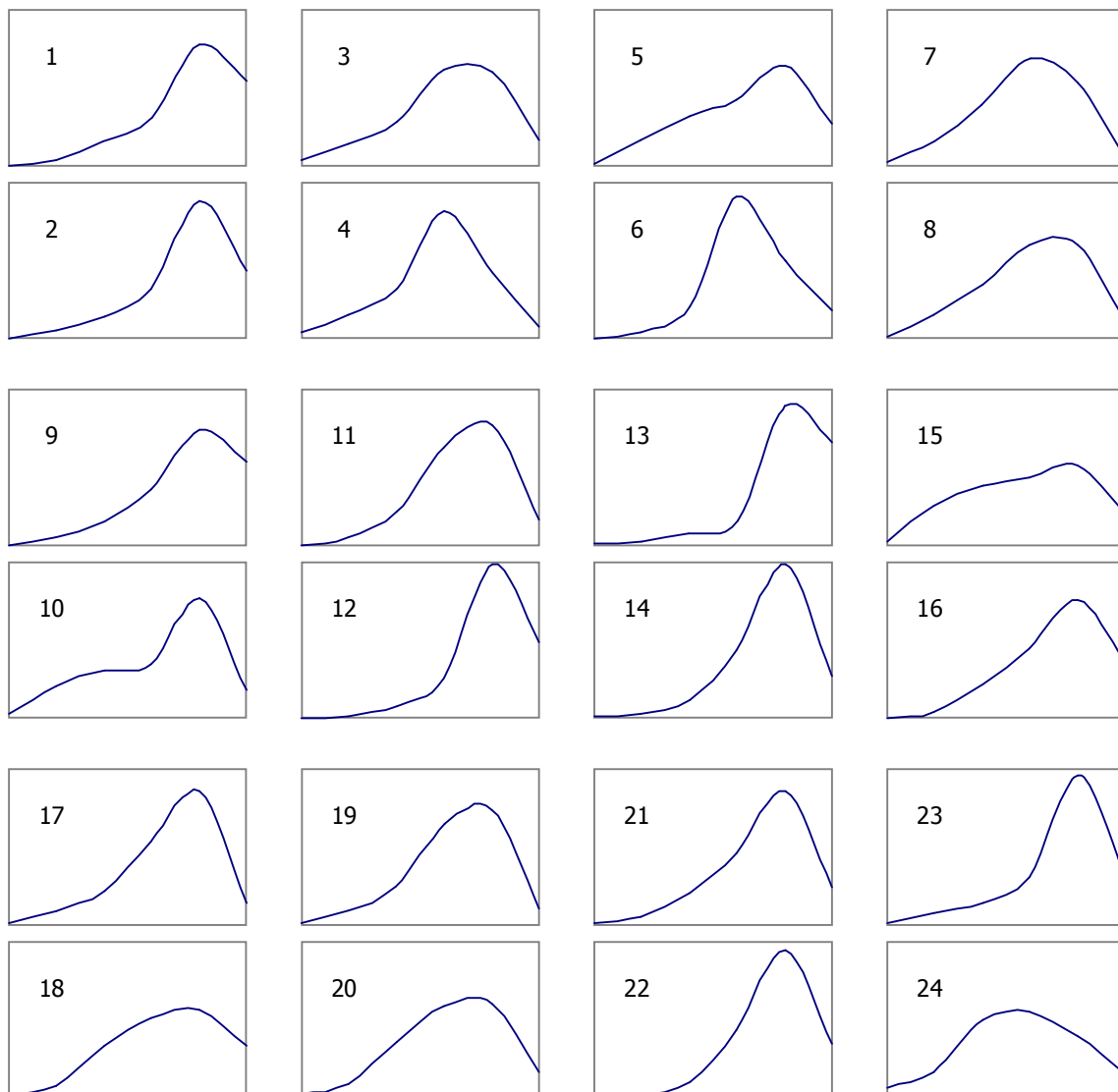


Ryc. 18. Rozkłady obligujące do odrzucenia niekonkluzyjnych wskaźników

Ze względu na konieczność rozstrzygnięcia kwestii tez elementarnych należy usunąć te stwierdzenia, wobec których nie było kwalifikowanej większości ('tak' lub 'nie') bądź większość wyraziła brak zdania. Jeśli jednak zbyt wielu respondentów wybrało skrajną opcję 'absolutnie', to oznacza, że teza jest trywialna i nie warto jej badać. Usunąć należy także tezy niejednoznaczne, wobec których częściową rację mają zarówno wybierający 'za', jak i 'przeciw'.

Usuwanie bodźców-stwierdzeń, które empirycznie nie sprawdziły się, jest posunięciem zasadnym, jednak zbyt radykalnym. W wielu przypadkach lepiej jest podjąć próbę modyfikacji treści stwierdzenia, tym bardziej, że wymyślenie zupełnie innego próbnika tej samej tezy to trudne zadanie. W praktyce okazuje się, że modyfikacja treści zdań poprzez zmianę czasu lub trybu orzeczeń, podmianę podmiotów lub dopełnień, stopniowanie przydawek, uzupełnianie okoliczników bądź syntetyzowanie i uogólnianie treści – wszystko to ma duży wpływ na zmianę reakcji zwrotnych, wyrażanych w wypowiedziach. W konsekwencji możliwa staje się daleko idąca optymalizacja bodźców-stwierdzeń, prowadząca do wyższej konkluzyjności, a także do swoistej normalizacji skali.

Cykl optymalizacji stwierdzeń i weryfikacji wypowiedzi można przeprowadzać, posługując się pasmowymi wykresami rozrzutów realnych (ryc. 3 i 4), histogramami (ryc. 18) lub wykresami rozkładów aproksymowanych (ryc. 19). Analiza rozkładowa ułatwia podejmowanie decyzji co do sposobu modyfikacji warstwy treściowej, lecz nie zastąpi pomysłowości, intuicji oraz daru trafnego operowania logiką zdań oraz semantyką języka w próbkowaniu zmiennych. Na to nie ma jakichś gotowych algorytmów, a można jedynie określić ogólny cel do osiągnięcia, wskazać zarys ścieżek i warunków dochodzenia do celu. Oczywiście nadrzędnym celem jest trafna operacjonalizacja zmiennych, której jednak nie da się zweryfikować bezpośrednio. Z tego powodu głównym celem optymalizacji i weryfikacji skal dwuważonych jest objaśnione już wcześniej *justowanie estymatorów* (str. 35), a warunkiem – *konkluzyjność wskaźników*. Przyjrzyjmy się efektom justowania i charakterystycznym cechom rozkładów.



Ryc. 19. Aproksymowane rozkłady wskaźników *InfoKultury*

Rzeczywiste rozkłady empiryczne wskaźników *InfoKultury* (ryc. 19) mają zróżnicowane kształty, pomimo justowania. Dominują asymetrie lewoskośne, przy czym właściwości rozkładów są znamienne dla pewnych typów wskaźników. I tak, krzywe wskaźników afektywnych są bardziej smukłe, a kognitywnych bardziej spłaszczone. Wartości modalne i zasadnicze skupiska gęstości usytuowane są z prawej strony, w strefach zgodności z tezami. Z wyjątkiem wskaźnika nr 24 graficznie ustalić można polaryzację znaku walencyjnego. Większość rozkładów ma korzystnie duży rozrzut, najmniej wskaźniki 12 i 13. Wobec kilku stwierdzeń wystąpił zbyt duży rozrzew wypowiedzi (m.in. 15 i 24). Stwierdzenia futurospekcyjne są akceptowane mniej intensywnie (m.in. 6 i 24).

Wskaźnik nr 24 nie spełnia warunków zakwalifikowania do skali, a mimo to nie został odrzucony, co wymaga uzasadnienia. Otóż zdarzają się sytuacje zmuszające do kompromisu pomiędzy koniecznością próbkowania konkretnej cechy szczegółowej a potencjalną możliwością wnioskowania z uzyskanego rozkładu empirycznego. Dotyczy to zwłaszcza futurospekcyjnych komponentów poznawczych. Dzieje się tak, gdy bodziec-stwierdzenie próbkuje ważką problemowo tezę predykcyjną, natomiast pomiar ujawnia brak polaryzacji opinii i brak zaprzeczenia tezy prognozującej niekorzystną przyszłość. Owe oznaki zupełnego braku formowania standardu ewaluatywnego niepokoją mnie, gdyż nie można pozostawać obojętnym wobec niemal katastroficznej tezy nr 24, która brzmi tak: *Pewnego dnia komputery zniewolą ludzi*. Wydaje się, że brak zaprzeczenia nie jest oznaką obojętności, lecz raczej wskaźnikiem obawy o to, iż pozostające jak dotąd w sferze *science fiction* zagrożenie będzie realne. Na szczęście kontrolny, komplementarny wskaźnik nr 23 ma rozkład korzystny, gdyż respondenci przeczą tezie: *Obawiam się uzależnienia od komputera*.

Dużą niespodzianką okazał się spłaszczony rozkład nr 15, który świadczy o tym, jak bardzo wielu respondentów wciąż uważa, że: *W szkole komputer winien być używany tylko na „Informatyce”*, w ogóle nie zdając sobie sprawy z istotnej roli mediów i technologii informacyjnych w edukacji. Z drugiej strony przeceniają rolę bezdusznej maszyny, akceptując tezę: *Komputer może radzić, jak rozwiązywać życiowe problemy*, co ujawnia rozkład nr 18. Oznacza to, że pojęcie *komputer* jest utożsamiane bardziej z systemami niż ze sprzętem.

Analiza rozkładowa wskaźników, jak wynika to z powyższych przykładów, daje zarówno podstawy do wniosku wartościującego, jak też umożliwia podjęcie decyzji o tym, czy usunąć dane stwierdzenie, czy pozostawić, czy może jeszcze zmodyfikować treść. Rozkłady wskaźników kognitywnych i futurospekcyjnych zawsze wyróżniają się większym rozrzutem i rozziwem, toteż tej właściwości nie należy niwelować, o ile polaryzacja wypowiedzi jest wyraźna. Justowanie estymatorów ma na celu przede wszystkim skupienie gęstości rozkładu w strefie zgodności, a dominanty – możliwie blisko środka tej strefy.



W każdym indywidualnym przypadku decyzje w sprawie uznania konkretnego bodźca-stwierdzenia jako odpowiedniego do skali podejmuje się ostatecznie uwzględniając łączne kryteria zarówno wartości semantycznej zdań, ważkości próbkowanego problemu, jak też wartości parametrów rozrzutu i rozziewu.

Analiza parametryczna konkluzyjności polega na technice balansowania pomiędzy optymalnym *Rozrzutem* a minimalnym *Rozziewem*. Rozziew jest tu miarą niezgodności opinii jednostek z ogółem w kwestii elementarnego standardu ewaluatywnego, dlatego najlepiej, aby był zerowy. Rozrzut natomiast oznacza standardową rozbieżność zdań w tej samej sprawie i najlepiej, jeśli jest tak duży, że wypowiedzi wypełniają całą strefę zgodności  $\langle 0; +3 \rangle$ . Wynika to z konieczności uproblemowienia stwierdzeń, ponieważ szkoda wysiłku na badania trywialne. Jednakże zbyt ni wzrost rozrzutu zwiększa rozziew ponad dopuszczalne granice. Metody osiągania najlepszego kompromisu zwanego *minimaks*em są żmudne. W praktyce trzeba kilkakrotnie przeprowadzić proces weryfikacyjno-optymalizacyjny bodźców-stwierdzeń, zmieniając początkowo zdania, a później nawet pojedyncze słowa. Każda modyfikacja tezy wpływa na wypowiedzi, a te są wskaźnikami elementarnej zmiennej. W całym procesie konstruowania nowego narzędzia, jego weryfikacji i dostrajania najważniejsze jest to, ażeby równoważyć oraz uwzględniać zarówno kryteria warunkujące konkluzyjność merytoryczną, jak też konkluzyjność statystyczną. Nie może jednak być tak, jak w innych skalach, gdy kryteria statystyczne automatycznie eliminują tezy istotne poznawczo. Analiza rozkładowa oraz parametryczna to jedynie techniki wspierające problemową, merytoryczną analizę ewaluacyjną. Zsynchronizowanie wszystkich aspektów podwyższa rzetelność badawczą.

➤ **Rzetelność wskaźnikowania** oznacza najbardziej uogólnioną dobroć pomiarową, wynikającą zarówno z solidnej podbudowy teoretycznego modelu skali, z prawidłowej implementacji realnego odwzorowania, z trafnej operacjonalizacji zmiennych, z opracowania konkluzyjnych wskaźników, jak też z prawdziwości wyników, co znajduje się już poza wpływem twórcy narzędzia, gdyż zależy od należytej staranności i uczciwości przy wypełnianiu kwestionariusza przez respondentów. Oczywiście zakłada się, że sam ewaluator jest także rzetelny, a więc nie dokonuje żadnych manipulacji ze zbiorem danych empirycznych, nie pomija w sprawozdaniu części wyników świadczących o procesach niekorzystnych, nie dobiera celowo lepszych prób badawczych, które znajdują się tylko na wyższych poziomach indykatorum itp. Od razu jednak zastrzec trzeba, iż rzetelność ewaluatora nie oznacza bezkrytyczności. Zdarzają się bowiem takie przypadki, gdy brak interwencji oznaczać będzie właśnie nierzetelność wskaźnikowania. Chodzi o to, że niektóre parametry statystyczne są bardziej wrażliwe na pojedyncze wypowiedzi. Jeśli badamy z całą pewnością dobrze przeprowadzone zajęcia komputerowe, a jeden z respondentów zakpił sobie

z ewaluacji, wypełniając kwestionariusz celowo przewrotnie, to pozostawienie jego zbioru wypowiedzi doprowadziłoby do przekłamania wyników całej grupy. Pojedynczy retorsyjny przypadek nie powinien zaniżać poziomu standardów, stąd zasadne jest wykluczenie. Uogólniając – rzetelność wskaźnikowania jest splotem prawidłowości metodologicznej<sup>31</sup> oraz prawdziwości empirycznej.

*Rzetelność metodologiczna.* Właściwość ta kształtowana jest poprzez spełnienie wszystkich procedur charakterystycznych dla obranej strategii ewaluacyjnej:

1) W fazie koncepcyjnej określa się przedmiot ewaluacji, tj. obiera indykatum oraz wytycza podmiotowo zbiory przypadków będących próbami badawczymi. Następnie formułuje się cele teoretyczne i praktyczne oraz definiuje problemy wymagające rozstrzygnięcia. Ustala się przy tym realny zasięg tego, co da się ocenić: czy tylko zjawiska lokalne i szczegółowe, czy globalne i uogólnione na populację. Ponadto ustala się formę: czy ma to być ewaluacja konkluzyjna (diagnostyczna), czy dyrektywna (z zaleceniami), czy formatywna (kształtująca).

2) W fazie implementacji buduje się model ekstrakcji symptomów z rozwarstwieniem indykatum na rezultaty połówkowe, czynniki, składniki i wskaźniki. Definicje pojęciowe zmiennych elementarnych należy zoperacjonalizować. Do próbkowania przygotowuje się walencyjne bodźce-stwierdzenia, ułożone w pary kontrolne z ekspansją czasoprzestrzeni. Próbniki te weryfikuje się empirycznie poprzez analizę reakcji-wypowiedzi, będących pośrednimi wskaźnikami cech. Wskaźniki muszą być przede wszystkim trafne, kompletne i konkluzyjne, do czego prowadzą techniki kwantyfikacji, optymalizacji tez i justowania skali.

3) W fazie eksploracji przeprowadza się dwa pomiary zasadnicze: na początku i na końcu procesu oddziaływań. Próbę badawczą stanowią podmioty procesu, przy czym istotna jest reprezentatywność i liczebność grup. Wypowiedzi analizuje się w dwóch wymiarach: jakości i intensywności. Opinie kwalifikowanej większości stanowią podstawę wartościowania. Dane empiryczne poddaje się obróbce analitycznej w przekrojach przez zmienne i przez przypadki, z odtworzeniem struktury komponentów indykatum oraz struktury grup badawczych. Metodami statystycznymi wyznacza się parametry chwilowe i dystansowe oznak oraz ustala miary bazowe i normujące, a także wyznaczniki jakości pomiarowej.

4) W fazie interpretacji prowadzi się ilościową analizę oznak i na niej opiera jakościową analizę symptomów. Ocenia się interakcję czynników splotowych, spójność składników i trafność wskaźników. W ewaluacji cech mentalnych określa się relacje komponentów afektywnych i kognitywnych, wartościując postawy i stan uświadomienia podmiotów. W ewaluacji wytworów ocenia się wpływ każdej z elementarnych właściwości na wartość ogólną. W ewaluacji procesów porównuje się rezultaty uzyskane przez różne grupy, poddawane celowo odmiennym lub podobnym oddziaływaniom. W każdym rodzaju ewaluacji wnikliwie ocenia się wyznaczniki norm wartości i dobroci pomiarowej.

*Rzetelność statystyczna.* Właściwość tę można zweryfikować na różne sposoby: 1) poprzez ustalenie korelacji między wynikami 'przed' i 'po', 2) poprzez porównanie rezultatów połówkowych, 3) poprzez analizę konsystencji macierzy danych. Każda z tych metod ma jednakże pewne słabe strony i ograniczenia. Obliczanie współczynnika korelacji pomiędzy pierwszym i drugim pomiarem, z zachowaniem zasady, że zawsze składa się parami obie wypowiedzi tych samych osób, jest jak najbardziej zalecane. Jednak nie można stale oczekiwać identycznych wyników, ponieważ chodzi tu właśnie o ewaluację zmienności. Z tego powodu współczynnik korelacji osiąga najczęściej wartości rzędu 0,6 i pośrednio oznacza to próg zupełnie wystarczającej rzetelności pomiarowej. Porównanie rezultatów połówkowych nie może być prowadzone w klasyczny sposób: z podziałem na pozycje parzyste i nieparzyste, ponieważ model skali zakłada wprowadzenie par nierównoważnych wskaźników kontrolnych. Jeśli pierwszy wskaźnik jest retrospekcyjny, a drugi futurospekcyjny, to poziomy wag wypowiedzi z natury rzeczy będą niższe dla prognozowania na przyszłość. Trudno jest więc skalę splotową podzielić na dwie równoważne części, dlatego analizę połówkową prowadzi się głównie dla pary rezultatów globalnych, jednak nie w celu weryfikacji rzetelności, lecz uzyskania wyższej konkluzyjności. Analizę konsystencji macierzy danych przeprowadza się przede wszystkim po obliczeniu współczynnika *alfa* Cronbacha, przyjętego dość powszechnie jako miarę uniwersalny wyznacznik rzetelności, porównawczy dla różnych skal. W publikacjach spotyka się rozmaite nietrafne próby nazwania tego, czym tak naprawdę jest współczynnik *alfa*, a ponieważ wyznacza on proporcję między dwiema właściwościami skali – rozdzielczością i koherentnością – dlatego proponuję definiowanie tego parametru jako *zróźnicowanie wewnętrzne*. Miara ta ma jednak trafne zastosowanie tylko w takich badaniach, w których zmienne elementarne pojedynczych przypadków z definicji mają być spójne, natomiast respondenci muszą być zróźnicowani. Warunek ten spełniony jest w ewaluacji opartej na próbkowaniu cech mentalnych. Bardziej szczegółowe objaśnienie podaję przy opisie statystycznego parametru *Rzetelności* (str. 196).

<i>Ślady</i>		<i>Wzorce</i>
próbkowanie indykatum	.....	wskaźniki, bodźce-stwierdzenia
odzwierciedlenie indykatum	.....	trafność wskaźników
zupełność próbkowania	.....	kompletność wskaźników
pary kontrolne próbek	.....	komplementarność wskaźników
agregowanie próbek	.....	addytywność wskaźników
odtworzenie indykatum	.....	reakcje-wypowiedzi, symptomy
kwantyfikacja wypowiedzi	.....	konkluzyjność wskaźników
optymalizacja wskaźników	.....	justowanie estymatorów
konsystencja wypowiedzi	.....	rzetelność wskaźnikowania

### Przekroje analityczne – podzbiory cech i przypadków

Efektom pomiarów za pomocą skal dwuważonych jest zbiór wskaźników. W praktyce – najpierw otrzymuje się jedynie zakreślone formularze. Po drugim pomiarze kompletuje się w pary obydwie kwestionariusze wypełnione przez te same osoby. Następnie każdej z kolejnej wypowiedzi przypisuje się rangę porządkową 1÷7, w zależności od wybranej przez respondenta opcji. Powstałe w ten sposób zbiory liczb będą zawierać w sobie cechy wskaźników tylko pod warunkiem żelaznej konsekwencji w budowaniu struktury danych surowych. Proponuję przyjąć taki format danych, w którym każdy wiersz zawiera komplet wypowiedzi 'przed' i 'po' jednego respondenta. Podzbiór ten w statystyce nosi nazwę *przypadku*. Kolejność przypadków ma znaczenie tylko takie, że należy uporządkować je w próbach badawczych, nazywanych tutaj *grupami* i *klasami*.

O ile kolejność osób nie ma znaczenia, o tyle grupowanie przypadków jest istotną procedurą strukturyzacji danych. W naturalny sposób uczniowie z tej samej klasy szkolnej mogą tworzyć *klasę* jako próbę badawczą, natomiast uczniowie kilku klas, z którymi zajęcia prowadzi ten sam nauczyciel, mogą stanowić *grupę*. Wszystko to jednak zależy od planu badawczego: na przykład jedna z klas szkolnych może być grupą eksperymentalną, a druga kontrolną. Ważne jest oznaczenie identyfikacyjne prób badawczych. I tak przynależność do większej próby, jaką jest grupa, oznacza pierwszy znak w szeregu danych (duża litera 'A' na ryc. 20). Drugi znak identyfikuje klasy, będące próbami elementarnymi (tutaj litery 'b' i 'c'). Poniższy zbiór danych jest fragmentem z grupy 'A' – zawiera cztery przypadki z klasy 'b' i trzy przypadki z klasy 'c'.

```
Ab=325435363446272644375727-215545542346174633666623
Ab=236536563266172677745623-216626671155171777636623
Ab=225565322244254622362522-225536513255266632654522
Ab=224255335266175665656624-126525651167162665636622
Ac=563264325655456422653433-226626652556354642636534
Ac=324154225555264566634644-225526662466262666266624
Ac=464554353654263662653546-325535642246266661373546
```

Ryc. 20. Zwarta struktura danych surowych w formacie .dxt

Jeszcze bardziej istotną procedurą strukturyzacji danych jest uporządkowanie elementarnych zmiennych w układzie kolumn. Jakakolwiek pomyłka z usytuowaniem wskaźników prowadzi do całkowitego przekłamania wyników. Abstrahując od oznaczenia grup i klas oraz separatorów, musimy zadbać o to, ażeby kolejność rang (liczb jednocyfrowych) w każdym wierszu odpowiadała dokładnie kolejności pozycji z kwestionariusza, a ponadto, aby w lewej części znalazły się wyłącznie wskaźniki z pomiaru pierwszego, a w prawej – z drugiego.

Z liczby rang wpisanych do jednego wiersza (ryc. 20) wynika, że dane dotyczą skali złożonej z 24 pozycji. Pierwsza w pionie kolumna rang za separatorem (=) to zbiór wypowiedzi 'przed' wobec pierwszej tezy skali, czyli wskaźnik  $X'_{11}$ . Analogicznie pierwsza za separatorem (-) kolumna ze zbioru wypowiedzi 'po' to wskaźnik  $X''_{11}$ . Ostatnia kolumna z prawej strony danych to wskaźnik  $X''_{24}$ . Po uogólnieniu zbioru danych na dowolną liczbę zmiennych elementarnych można ich strukturę zdefiniować jako macierz dwuwymiarową o wierszach zawierających przypadki oraz kolumnach reprezentujących wskaźniki cech. Ściślej – są to zawsze dwie macierze z pomiaru pierwszego  $X'$  i drugiego  $X''$ .

$$\begin{array}{l} \text{przed} \\ X' = \end{array} \begin{bmatrix} X'_{11} & X'_{12} & \dots & X'_{1p} \\ X'_{21} & X'_{22} & \dots & X'_{2p} \\ \dots & \dots & \dots & \dots \\ X'_{n1} & X'_{n2} & \dots & X'_{np} \end{bmatrix} \quad \Delta \quad \begin{array}{l} \text{po} \\ X'' = \end{array} \begin{bmatrix} X''_{11} & X''_{12} & \dots & X''_{1p} \\ X''_{21} & X''_{22} & \dots & X''_{2p} \\ \dots & \dots & \dots & \dots \\ X''_{n1} & X''_{n2} & \dots & X''_{np} \end{bmatrix}$$

Ryc. 21. Organizacja zmiennych  $X'$  i  $X''$  w macierzach dwuwymiarowych  
 $n$  – numer przypadku (respondenta),  $p$  – numer pozycji skali  
 $\Delta$  – operator różnicy między zbiorami, z def.:  $\Delta X = X'' - X'$

Formalnie operator różnicy (delta) pomiędzy zbiorami  $X'$  i  $X''$  nie jest potrzebny ani jednoznaczny, jednak zamieściłem go, aby zasygnalizować, że najczęściej bada się właśnie zróżnicowanie elementów obu macierzy.

Operowanie na wszystkich elementach macierzy prowadzi do obliczenia wyniku ogólnego dla całej zbiorowości respondentów, którą to zbiorowość nazywać będziemy *ogółem* badanych. Jednakże w ewaluacji splotowej wynik ogólny to zbyt mało w sensie odkrywczym – strategia ta bowiem umożliwi ustalenie także tego, jakie czynniki wpłynęły na dany wynik, jakie składniki cech dominowały, jakie efekty osiągnęły poszczególne grupy, klasy itp. Z tego względu w fazach eksploracji zasobów oraz interpretacji oznak z całej macierzy wyodrębnia się podzbiory danych zwane *przekrojami analitycznymi*.

**Przekroje analityczne** to kwerendy eksploracyjno-interpretacyjne, służące do wyodrębniania wycinkowych oznak z całościowego splotu zjawisk empirycznych i do rozwiązywania szczegółowych problemów badawczych. Przekrój przez zmienne oznacza ekstrakcję czynników, składników lub wskaźników, a przekrój przez przypadki to wyodrębnienie klas lub grup respondentów.

Częstą wadą narzędzi badawczych jest zbyt rozbudowana metryczka zbierająca dane etnograficzne (kto, skąd, płeć, wiek, wykształcenie, praca itp.), choć z danych tych nie robi się użytku. Powodem jest to, że w podręcznikach nieadekwatnie nadużywa się pojęcia *zmiennej niezależnej* zamiast *grupującej*. Tymczasem właśnie cechy osób mogą prowadzić do zasadnego grupowania i do oceny zależności między rezultatami osiągniętymi przez poszczególne grupy.

Strategię grupowania powinno się określać już w fazie projektowania badań, tym niemniej także podczas eksploracji można dzielić badanych na podgrupy ze względu na dodatkową cechę, na przykład odkrytą w rezultacie pomiaru. Do podziału potrzebna jest jednak odpowiednio liczna próba badawcza.

Przypomnę, że w przyjętym modelu ewaluacji nie zajmujemy się analizą pojedynczych przypadków, z wyjątkiem eliminowania przypadków całkowicie nierzetelnego wypełnienia kwestionariusza. Najmniejszą próbą badawczą jest więc *klasa*, licząca minimum 15 a optimum ponad 25 osób. Jeśli pomiarem objęliśmy bardziej liczną *grupę* (minimum 30 a optimum ponad 50 osób), to – przy braku innej koncepcji podziału – możemy przepołowić tę próbę choćby ze względu na uzyskany wynik ogólny. Porównanie zmienności i fluktuacji wypowiedzi respondentów sytuujących się w górnej i dolnej strefie rezultatów jest jednym z wielu możliwych środków wzbogacania konkluzyjności badań.

Tworzenie przekroju przez przypadki polega na wybraniu z macierzy źródłowej ściśle określonych wierszy, które są zbiorem wypowiedzi pochodzących od jednej i tej samej klasy lub grupy respondentów. Zapis formalny wyboru dwóch pierwszych klas, oznaczonych jako [a i b], usytuowanych od początku macierzy i tworzących jedną wspólną grupę [A], przedstawia się następująco:

$$\begin{array}{c}
 X'_{Aa} \\
 \\
 \\
 \\
 X'_{Ab} \\
 \\
 \\
 \\
 \\
 \\
 \end{array}
 \left[ \begin{array}{cccc}
 X'_{11} & X'_{12} & \dots & X'_{1p} \\
 \dots & \dots & \dots & \dots \\
 X'_{a1} & X'_{a2} & \dots & X'_{ap} \\
 \hline
 X'_{a+1;1} & X'_{a+1;2} & \dots & X'_{a+1;p} \\
 \dots & \dots & \dots & \dots \\
 X'_{a+b;1} & X'_{a+b;2} & \dots & X'_{a+b;p} \\
 \hline
 \dots & \dots & \dots & \dots
 \end{array} \right]
 \quad
 \begin{array}{c}
 X''_{Aa} \\
 \\
 \\
 \\
 X''_{Ab} \\
 \\
 \\
 \\
 \\
 \\
 \end{array}
 \left[ \begin{array}{cccc}
 X''_{11} & X''_{12} & \dots & X''_{1p} \\
 \dots & \dots & \dots & \dots \\
 X''_{a1} & X''_{a2} & \dots & X''_{ap} \\
 \hline
 X''_{a+1;1} & X''_{a+1;2} & \dots & X''_{a+1;p} \\
 \dots & \dots & \dots & \dots \\
 X''_{a+b;1} & X''_{a+b;2} & \dots & X''_{a+b;p} \\
 \hline
 \dots & \dots & \dots & \dots
 \end{array} \right]$$

Ryc. 22. Przekrój przez przypadki z wyodrębnieniem grupy [A] i dwóch klas [a, b]  
 $a, b$  – liczebności klas [a, b]       $p$  – numer pozycji skali

Każdą z elementarnych prób tworzy tyle wierszy, ile wynosi liczebność klas. Jeśli na przykład  $a=25$  a  $b=30$ , to przypadki reprezentujące klasę [b] rozpoczynają się w wierszu o numerze  $a+1=26$ , a kończą na wierszu  $a+b=55$ . Przekrój celowo pochodzi z początku macierzy źródłowej, ażeby uprościć zapis i objaśnienie. Tylko w takim wypadku, gdy pierwsza wyodrębniona klasa jest również pierwszą w macierzy źródłowej, jej indeksy zaczynają się od jedności. Wyodrębniane klasy bądź grupy nie muszą być równoliczne, chyba że wynika to z innych przesłanek metodologicznych, dotyczących porównywania prób.

Wybór potrzebnego przekroju wiąże się niekiedy z koniecznością przesunięcia przypadków. Dzieje się tak, gdy chcemy na przykład zgrupować osoby ze względu na odkrytą u nich cechę. Do automatyzacji przegrupowania warto

wykorzystać arkusz kalkulacyjny, choć należy przekonwertować format danych. Osoby biegle informatycznie mogą napisać odpowiednie makro lub procedurę, mniej wprawnym pozostaje ręczne przenoszenie wierszy w edytorze tekstowym.

Tworzenie przekroju przez zmienne to operacja na kolumnach macierzy. Zgodnie z przyjętą tu konwencją ciąg liczb ułożonych pionowo w pojedynczej kolumnie reprezentuje wskaźnik. Ciąg taki ma różną długość w zależności od wyboru przypadków. Można wyodrębnić wskaźniki klas, grup lub ogółu badanych. Sąsiadujące ze sobą w kwestionariuszu pozycje {1∩2; 3∩4; ...} stanowią kontrolne pary wskaźników poddawanych analizie porównawczej. Ponadto, po nadaniu odpowiednich wag i obliczeniu średnich arytmetycznych tworzy się z tych par zmienne cząstkowe, zwane *składnikami* {1⊕2; 3⊕4; ...}. Uogólniony, formalny zapis przekrojów przez zmienne wygląda następująco:

$$\begin{array}{ccc}
 \text{Wskaźniki} & & \text{Składniki } k = 2 \\
 \begin{array}{cc}
 \textit{przed} & \textit{po} \\
 \left[ \begin{array}{c} X'_{1p} \\ X'_{2p} \\ \dots \\ X'_{np} \end{array} \right] & \Delta & \left[ \begin{array}{c} X''_{1p} \\ X''_{2p} \\ \dots \\ X''_{np} \end{array} \right]
 \end{array} & & 
 \begin{array}{cc}
 \textit{przed} & \textit{po} \\
 \left[ \begin{array}{cc} X'_{1p-1} \oplus X'_{1p} \\ X'_{2p-1} \oplus X'_{2p} \\ \dots & \dots \\ X'_{np-1} \oplus X'_{np} \end{array} \right] & \Delta & \left[ \begin{array}{cc} X''_{1p-1} \oplus X''_{1p} \\ X''_{2p-1} \oplus X''_{2p} \\ \dots & \dots \\ X''_{np-1} \oplus X''_{np} \end{array} \right]
 \end{array}
 \end{array}$$

Ryc. 23. Przekroje potrzebne do wyodrębnienia wskaźników oraz składników  
 $p$  – dowolnie wybrany numer pozycji skali,  $p$  – parzysty numer pozycji  
 $\oplus$  – operator średniej arytmetycznej  $k$ -elementów zbioru, z def.:  $1/k \sum X_j$

Warto zwrócić uwagę na to, że w analizie statystycznej argumentami są najczęściej pojedyncze ciągi liczb. Jeśli zamierzamy analizować składniki, to najpierw z dwóch ciągów (z kolumny nieparzystej i parzystej) należy utworzyć jeden ciąg. W odróżnieniu od skali Likerta nie sumuje się wag cząstkowych, lecz oblicza z nich średnią arytmetyczną. Nie dotyczy to jedynie statystycznej analizy spójności wskaźników oraz rzetelności skali. Czynność agregowania wskaźników obowiązuje także w analizie czynników, rezultatów połówkowych i wyników ogólnych. Z formalnym, uogólnionym zapisem macierzy czynników jest pewna trudność, ponieważ składowe czynniki nie muszą być kolejnymi pozycjami skali. Z tego względu zamiast macierzami lepiej jest posługiwać się *schematem ekstrakcji* w dowolnej formie – graficznej, tabelarycznej lub czysto tekstowej. Niezależnie od przyjętej formy schemat ten szczegółowo informuje o strukturze ekstrakcji bądź agregacji komponentów indykatorum. Każda ze zmiennych musi być w schemacie powiązana z odpowiednimi pozycjami skali, a więc z właściwymi numerami wskaźników tworzących tę zmienną. Oprócz jednoznacznego adresowania do wskaźników, w schemacie warto umieścić zwarte nazwy wszystkich składników, czynników i rezultatów połówkowych.

Najbardziej przejrzystym schematem ekstrakcji jest tabela hierarchiczna, w której każdy wiersz odpowiada coraz głębszemu poziomowi rozwarstwienia:

<i>InfoKultura</i>											
<i>Doznanie</i>						<i>Poznanie</i>					
<i>Opinie</i>		<i>Motywacje</i>		<i>Emocje</i>		<i>Poglądy</i>		<i>Wprawa</i>		<i>Rozwaga</i>	
<i>Ocena</i>	<i>Aplauz</i>	<i>Ambicje</i>	<i>Intencje</i>	<i>Spokój</i>	<i>Odpór</i>	<i>Osąd</i>	<i>Wgląd</i>	<i>Obycie</i>	<i>Pewność</i>	<i>Zdolność</i>	<i>Ogląda</i>
13×14	1×2	11×12	21×22	9×10	23×24	15×16	17×18	3×4	5×6	7×8	19×20

Schemat ten jest ściśle związany z układem bodźców-stwierdzeń w *Kwestionariuszu kultury informatycznej*, a więc nie jest jakąś uniwersalną strukturą zalecaną dla innych skal, lecz szablonem interpretacji konkretnego narzędzia. Zmienne cząstkowe są w tabeli uporządkowane, natomiast zmienne elementarne (wskaźniki) powiązane są w pary, lecz w innym układzie niż kolejność stwierdzeń w kwestionariuszu. Z pozycji respondentów ważniejsza jest pewna ciągłość treści zdań, a także to, aby obok siebie nie były skupione aż cztery tezy trudne recepcyjnie i problemowo. Z kolei dla ewaluatora podczas analizy najważniejsze jest w pełni prawidłowe agregowanie wskaźników i odtworzenie całej struktury zmiennych cząstkowych, zgodnie ze schematem interpretacji. Właśnie do tego celu niezbędne są przekroje analityczne przez zmienne. Program InfoCult™ Analyser wyodrębnia potrzebne przekroje automatycznie.

### **Sfery wartościowania – cechy jakościowe, miary ilościowe**

Kiedy wartościujemy jakąś ogólniejszą cechę, to chcielibyśmy wiedzieć, które z cząstkowych komponentów tej cechy są lepsze bądź gorsze. Jednakże nie wystarcza tylko stwierdzenie faktu, że „coś jest dobre”. Zazwyczaj potrzebujemy także oszacowania, „na ile to coś jest dobre”. Wyodrębniamy więc cząstkowe cechy jakościowe i dla każdej z nich wyznaczamy specjalizowane miary statystyczne, reprezentujące parametry ilościowe. Cechą jakościową w ewaluacji demokratycznej jest każda z badanych zmiennych cząstkowych, którą respondenci w procesie normalizacji skali uznali za wartość, natomiast miarą ilościową jest stopień nasilenia akceptowania tej cechy jako wartości.

Najtrudniejszą sprawą podczas wnioskowania ewaluacyjnego jest to, że w zasadzie przy ocenie każdej z cech i miar stosować trzeba zupełnie inne przesłanki wartościowania<sup>32</sup>. Najogólniej rzecz ujmując – w sytuacji jakościowej analizy zmiennych o dobrych wynikach można mówić wówczas, gdy ich wzajemne relacje i tendencje układają się w zgodnej z modelem harmonii. Sam fakt stwierdzenia w efekcie pomiaru, że występuje wartościowy symptom, będący komponentem ogólniejszej wartości indykatum, może być podstawą do formułowania jakościowych wniosków o pozytywnym wyniku ewaluacji<sup>33</sup>. Przy ilościowej natomiast analizie parametrów każdą z miar należy najpierw odnieść do jej sensu statystycznego i dopiero wtedy można oceniać zjawisko.



Platforma wnioskowania statystycznego często sprawia kłopoty, o ile bowiem każdy rozumie istotę średniej arytmetycznej, o tyle mało kto przyswaja sobie i prawidłowo interpretuje choćby to, co oznacza współczynnik korelacji. W tej dziedzinie jedno jest ważne – nie trzeba w ogóle znać wzorów statystycznych, gdy obliczenia wykonuje komputer, lecz przed interpretacją miar ilościowych należy koniecznie zrozumieć istotę tego, o czym świadczy dany parametr.

W procesie ewaluacji splotowej bada się zawsze ściśle określone indykatorum złożone z elementów cząstkowych. Poprawność wnioskowania jakościowego zależy od tego, na ile prawidłowo ewaluator interpretuje definicyjny sens każdej ze zmiennych składowych oraz na ile trafnie rozumie operacyjny sens wskaźników. Nie zawsze bowiem ewaluator jest twórcą narzędzia, a wręcz przeciwnie – dobre skale pomiarowe wykorzystywane są przez wielu badaczy. Osoba używająca narzędzia powinna znać nie tylko schemat interpretacji, lecz również powinna posiadać podstawową wiedzę z dziedziny, którą eksploruje. Jeśli skala próbkuje na przykład cechy mentalne, to nauczyciel z racji wymaganego przygotowania pedagogicznego, obejmującego również zarys psychologii, jest potencjalnie kompetentny do analizy rezultatów pomiaru tych cech. Nie powinny więc być mu obce tak ważne w edukacji pojęcia, jak postawy, emocje, motywacje itp. Co więcej – zna teorie dotyczące relacji zachowań, prawidłowości występowania cech oraz metody ich subtelnej kształtowania. Zna też swoich uczniów, może więc być najlepszym ewaluatorem-analitikiem zjawisk lokalnych i jednocześnie osobą zwrotnie formującą pożądane efekty. Wszystko jednak zależy od tego, czy potrafi być obiektywnym i wszechstronnym analitykiem zarówno cech jakościowych, jak też miar ilościowych.

W ewaluacji z użyciem skali analiza jakościowa i analiza ilościowa silnie się przenikają. W zasadzie trudno mówić o fazie czystej analizy statystycznej, gdyż i tak każdy z parametrów dotyczy konkretnej zmiennej. Przetwarzanie danych odbywa się w przekrojach przez zmienne, a więc wnioskowanie jest w naturalny sposób integralnym ujęciem cech i miar. Jakkolwiek inferencja jakościowa i ilościowa przebiegają niemal równocześnie, tym nie mniej każda z nich ma odmienne strategie i inne obiekty analizy. Z tego powodu lepiej jest posługiwać się pojęciami oznaczającymi swoiste kontinuum zarówno w sferze przestrzeni eksploracyjnej substancji indykatorum, jak i w zakresie polaryzacji faz czynności prowadzonych podczas całego złożonego procesu ewaluacji. To kontinuum na osi jakość | ilość można wyrazić następująco: w koncypaniu dwubieżnym ewaluator cyklicznie przemierza obszar w jedną i drugą stronę pomiędzy wartościowaniem aksjologicznym i wartościowaniem statystycznym, przy czym – poprzez procesy kwantyfikacji i kwantyzacji zbliża się bardziej do bieguna *sferę kwantytatywnej*, odnoszącej się do ilości, a poprzez procesy kwalifikacji i kwalizacji<sup>14</sup> wraca do *sferę kwalitatywnej*, dotyczącej jakości.

Charakterystykę procesów i obiektów inferencyjnych, które są podłożem wartościowania, przedstawię z podziałem na sfery kwalitatywne i kwantytatywne, rozpoczynając od procesów związanych z operacjonalizacją, pomiarem i analizą statystyczną, a więc czynności zmierzających do ujęcia wymiernego.

**Kwantyfikacja** to proces przechodzenia od definicyjnego sensu zmiennych do empirycznego zasobu obserwowalnych wskaźników, parametrów i charakterystyk statystycznych, odzwierciedlających ilościowy wymiar indykatum.

Proces ten obejmuje: operacjonalizację, indeksację, kwantyzację i kalibrację oraz obróbkę danych i generowanie parametrów w przekroju przez przypadki. Z pojęciowego poziomu zmiennych przechodzi się najpierw do operacyjnego zbioru symptomów, a następnie dobiera optymalne i skalowane wskaźniki. W trakcie pomiaru próbkuje się natężenie symptomów, próbkom przypisuje odpowiednio skalibrowane wagi i z nich oblicza miary niezbędne do analizy. Dla odróżnienia od tak szeroko zdefiniowanego pojęcia *kwantyfikacji* podam uzupełniającą definicję znacznie bardziej skumulowanej czynności *kwantyzacji*:

**Kwantyzacja** to procedura obejmująca: 1) ustalenie gradacji próbkowania, progów natężeń i rozpiętości przedziałów oraz 2) empiryczne wyznaczenie, do którego z przedziałów wpada natężenie cechy w momencie pomiaru. Produktem jest zmienna losowa dyskretna, poddawana obróbce statystycznej.

Procedura ta jest swoistym porcjowaniem poziomu cech oraz unifikowaniem wartości wskaźników tworzących liczbowe zasoby danych empirycznych.

Przetwarzanie statystyczne danych ma na celu uzyskanie obrazu różnych właściwości zmiennej losowej (liczbowej), odzwierciedlających odpowiednie symptomy jakościowe zmiennych cząstkowych indykatum. W rezultacie obliczeń statystycznych uzyskuje się zarówno parametry wyrażone przez pojedynczą liczbę, jak też miary nieparametryczne, do wyrażenia których potrzebne są zbiory liczb, wektory, funkcje, a nawet wykresy. Analiza ilościowa może więc oznaczać koncyptowanie z samych liczb, lecz także konieczność interpretacji graficznej, jak to się dzieje np. w analizie kształtu krzywej rozkładu gęstości. Szczegółowe objaśnienia każdej z miar skal dwuwazonych umieściłem w części statystycznej, zatem w tym miejscu podam tylko krótkie charakterystyki:

➤ **Poziom** – chwilowa wartość cechy, wyrażona w skali wag lub skali ocen, zależna zarówno od aktualnego stanu cechy, lecz niestety także od operacjonalnej trafności i semantycznej optymalności treści bodźca-stwierdzenia. Z tego powodu poziomy muszą być odnoszone do norm standaryzujących skalę.

➤ **Ufność** – przedział tolerancji, w jakim prawdopodobnie mieści się średnia oczekiwana dla populacji, oszacowana na podstawie wybranej próby losowej, przy założeniu progu błędu nie większego niż określony poziom istotności.

➤ **Intensywność** – natężenie wynikające z emocjonalnego ładunku bodźca-stwierdzenia i z nasilenia reakcji-wypowiedzi, a także z osobistych inklinacji do wypowiedzi zdecydowanych ('*absolutnie tak*') lub stonowanych ('*raczej tak*').

➤ **Dysjunkcja** – rozdziew pomiędzy opiniami respondentów zgadzających się z elementarną tezą a mniejszością wyrażającą zdanie odrębne. Właściwość ta jest podstawą polaryzacji walencyjnej i kwantyfikatorem wartościowania.

➤ **Koniunkcja** – spójność par wskaźników kontrolnych, zależna nie tylko od wewnętrznej zgodności wypowiedzi wobec dwu sąsiednich pozycji kwestionariusza, lecz także od zróżnicowania aspektów czasoprzestrzennych skali.

➤ **Dyspersja** – rozproszenie wypowiedzi, miara ta z jednej strony wyraża zróżnicowanie próbkowanej cechy u respondentów, lecz z drugiej strony jest znamieną dla danej grupy, w zależności od niejednorodności osobniczej.

➤ **Dewiacja** – para realnych odchyłeń od średniej, właściwość wyznaczająca smukłość i asymetrię rozkładu empirycznego, używana do rozwarstwienia grup na 3 podgrupy o wysokim, średnim i niskim poziomie badanych cech.

➤ **Rozkład** – wyrażona w definicyjnej postaci funkcji albo w formie wykresu charakterystyka zróżnicowania próby, odzwierciedlająca: gęstość prawdopodobieństwa, liczebność lub częstość przypadków o takim samym natężeniu cech.

➤ **Skośność** – asymetria rozkładu empirycznego, znamieną dla pomiarów skalą dwuważoną ze względu na większe skupisko przypadków usytuowanych w strefie zgodności, obligująca do niestandardowych modeli testów istotności.

➤ **Zmiana** – przesunięcie poziomów natężenia cech w czasie, wyznaczone z różnic pomiędzy uśrednionymi rezultatami pomiaru końcowego i początkowego, obejmującymi całą próbę badawczą, a nie pojedyncze przypadki.

➤ **Korelacja** – zależność pomiędzy obiema wypowiedziami tych samych osób w pierwszym i drugim pomiarze, oznaczająca to, na ile niezmiennie pozostało uporządkowanie respondentów według rankingu wyników indywidualnych.

➤ **Tendencje** – pozytywne bądź negatywne kierunki przemian stwierdzonych w pomiarze końcowym względem stanu początkowego oraz wymiar trendów uwzględniający zmiany na trzech poziomach HML: górnym, średnim i dolnym.

➤ **Zawiłość** – nierównomierność i niewspółbieżność przemian na trzech poziomach HML, oznaczająca takie zjawisko, że każda z podgrup respondentów podzielonych według poziomów cech uzyskuje różne przyrosty lub spadki, co może pośrednio świadczyć o pozytywnym lub negatywnym oddziaływaniu.

➤ **Fluktuacja** – wahania pomiędzy pierwszą a drugą wypowiedzią tych samych osób, określające zbiór indywidualnych, rzeczywistych przewartościowań odniesionych do przemian potencjalnie możliwych w zakresie skali.

➤ **Amplituda** – uogólniona na próbę badawczą miara intensywności zmian wypowiedzi, będąca sumą wahań indywidualnych, liczonych względem kresu górnego skali, oznaczająca intensywność oddziaływań zewnętrznych.

➤ **Dynamika** – uogólniona na daną próbę miara efektywności oddziaływań, obliczana z różnicy pomiędzy pozytywnymi oraz negatywnymi zmianami cech kształtowanych i próbkowanych na zbiorze indywidualnych przypadków.

➤ **Konkluzyjność** – wyznaczona empirycznie wnioskotwórczość każdego z elementarnych wskaźników, świadcząca o tym, że bodźce-stwierdzenia zawierają nietrywialne tezy i ładunek emocjonalny, pobudzający do reakcji-wypowiedzi.

➤ **Selektywność** – rozdzielczość między dwoma poziomami cech testowanych ze względu na zróżnicowanie; oznacza szacunkową odległość średnich, którą z określonym prawdopodobieństwem uznać można jako różnicę istotną.

➤ **Trafność** – prawidłowość potwierdzająca dobroć pomiarową w taki sposób, że rezultaty liczone za pomocą odmiennej techniki są zgodne. Porównuje się różnice miar zagregowanych z indeksami względnymi różnic indywidualnych.

➤ **Rzetelność** – prawidłowość konsystencji danych, wskazująca na dobroć pomiarową poprzez potwierdzenie zróżnicowania wewnątrzspójnego, oczekiwanego w pomiarach cech mentalnych. Skala powinna różnicować przypadki, lecz elementarne składowe cech u tej samej osoby powinny być spójne.

Warto zwrócić uwagę na to, że cechy mentalne są trudne do ilościowego wyrażania wartości bezwzględnej, dlatego niezbędne jest zastosowanie technik pomiaru różnicowego, porównywania i względnego obliczania indeksowego. Różnicuje się rezultaty pierwszego i drugiego pomiaru, porównuje się grupy badawcze wzajemnie i w odniesieniu do standardów, a ponadto każdą ze zmian wypowiedzi odnosi się do zmian granicznych ze względu na kres skali. Właśnie w podejściu względnym do miar statystycznych, w ich różnicowaniu, porównywaniu i odnoszeniu do wyznaczników normujących tkwią podstawy analizy jakościowej oznak i wartościowania symptomów empirycznych.

Proces aksjologicznego wnioskowania z doświadczenia jest zwieńczeniem ewaluacji, tymczasem w wielu monografiach dotyczących metodologii badań ta faza jest niestety przedstawiana w zdawkowy sposób lub wręcz pomijana. Co więcej – nie ukształtowało się odpowiednio trafne słownictwo określające czynności przechodzenia z obszarów empiryczno-ilościowych na teoretyczno-jakościowe. W słownikach nie ma wyrazów komplementarnych do funkcjonujących pojęć *kwantyfikacja* i *kwantyzacja*, dlatego wydaje się uzasadnione podjęcie próby wprowadzenia neologizmów *kwalifikacja* i *kwalizacja*, które oznaczałyby procesy przeciwbieżne, zmierzające do uogólnienia jakościowego.

**Kwalifikacja** to proces przechodzenia ze statystycznego obszaru oznak empirycznych na poziom pojęciowo-definicyjny symptomów weryfikujących twierdzenia, modele i systemy określające jakościowy wymiar indykatum.

W fazie tej przenosi się wymiar liczbowy na sferę wartościowania w kategoriach pozytywności/negatywności, dynamiki/statyki, świadomości/nieświadomości,

postaw pożądaných/niepożądaných, przyrostu/spadku, spójności/niespójności, współbieżności/rozbieżności, efektywności/nieefektywności itp. Jeśli ewaluacja ma cel naukowy, to zadaniem jest ocena zjawisk globalnych, uogólnianie na populację, formułowanie twierdzeń i budowanie teorii. W celu praktycznym chodzi głównie o ocenę zjawisk lokalnych i o formowanie dalszych przemian. Proces kwalifikacji obejmuje: analizę oznak w przekrojach przez zmienne, deskrypcję, eksplikację, predykcję, kwalifikację, generalizację oraz idealizację.

➤ **Analiza oznak** to czynność polegająca na wyodrębnianiu odpowiednich przekrojów przez wszystkie poziomy ekstrakcji zmiennych, połączona z oceną różnorodnych (zwłaszcza nietypowych) parametrów i relacji statystycznych. Na poziomie *wskaźników* weryfikuje się przede wszystkim jakość ich wyjustowania, czyli dopasowanie rozpiętości badanych cech do zakresu pomiarowego skali i związaną z tym perspektywę konkluzyjnego wnioskowania. Dodatkowo można odtworzyć aspekty czasoprzestrzenne kontrolnych par wskaźników, lecz interpretacja tych wymiarów jest konieczna tylko wtedy, gdy wyjątkowo silnie ujawni się niespójność wypowiedzi. Na poziomie *składników* analizuje się głównie wpływ poziomu poszczególnych komponentów na rezultat ogólny, a ponadto określa zbieżność i relacje współbieżności komplementarnej pary tworzącej czynniki. Z kolei na poziomie *czynników* najważniejsze jest ustalenie tego, co dominowało: intuicja czy wiedza, pobudzenie czy zaspokojenie, impuls czy namysł. Na tym etapie z jednej strony sprawdza się poprawność modelu, a z drugiej strony weryfikuje trafność narzędzia pomiarowego. Jeśli zachodzi zgodność między relacjami teoretycznymi modelu i empirycznymi czynników, to sprawa jest prosta. Jeśli jednak nie ma zgodności, to ewaluator musi ocenić, czy model był niewłaściwy, czy raczej nietrafna była ekstrakcja lub operacjonalizacja zmiennych. Pamiętać jednak trzeba o tym, że w ewaluacji splotowej niektóre czynniki właśnie z założeń teoretycznych mogą być niewspółbieżne (np. nieliniowe zależności emocji i motywacji bądź motywacji i zaspokojenia). Na poziomie *rezultatów połówkowych* ocenia się przede wszystkim wzajemne tendencje wzrostowe lub spadkowe, z rozwarstwieniem na podgrupy respondentów o wysokich, średnich i niskich natężeniach cech. W ten sposób można ustalić zawilść przemian i pośrednio ocenić jakość procesów oddziaływań. W proponowanym modelu ekstrakcji na komponenty afektywne i kognitywne siłą rzeczy najważniejsza jest ocena harmonii procesów doznania i poznania. Dodatkowo można podjąć próbę analizy tego, czy respondenci podczas wypełniania kwestionariusza kierowali się bardziej skrytem intuicyjnych reakcji, czy raczej odtwarzali zinternalizowane wzorce, czy też sami kreowali standardy. Wszystkie powyższe elementy analizy oznak empirycznych tworzą jakościowy obraz *wyniku ogólnego*. Zagregowany poziom indykatorum jest najbardziej wiarygodny, lecz pełnię ewaluacji wyznacza opis oznak na wszystkich poziomach.

W analizach jakościowych na równi ważne są opisy diagnostyczne oznak empirycznych, jak też objaśnienia symptomów oraz prognozowanie trendów. Te procedury analityczne nazywamy: deskrypcją, eksplikacją i predykcją:

➤ **Deskrypcja** – faza rozpoznawania prawidłowości lub nietypowości oznak, połączona z opisywaniem tych zjawisk w kategoriach aspektów jakościowych. Rola ewaluatora sprowadza się do stwierdzenia stanu: *Jest tak a tak...*

➤ **Eksplikacja** – faza wyjaśniania przyczynowego i zależnościowego oznak, połączona z formułowaniem konkluzji i uogólnień wartościujących. Ewaluator określa relacje symptomów i wnika w ich genezę: *Jest tak, ponieważ...*

➤ **Predykcja** – faza estymowania i przewidywania, polegająca na szacowaniu poziomu standardów oczekiwanych, wytyczaniu trendów i stanów przyszłych. Z norm empirycznych ewaluator formułuje prognozę: *Powinno być tak...*

O ile powyższe trzy procedury są typowe i stosowane w większości badań, o tyle w przyjętym tutaj modelu strategią znamioną jest także kwalityzacja:

**Kwalityzacja** to procedura odtworzenia jakości indykatorum z empirycznych próbek dyskretnych, chwilowych i progowych. Produktami są konwencyjne, skryptowe poziomy i dedukcyjne relacje wytyczające standardy ewaluatywne.

Przykładem kwalityzacji jest konwersja poziomów na skalę ocen – odniesienie do utrwalonego społecznie skryptu skalowania poprzez przejście na kategorie poziomów jakościowych (bardzo dobry, dobry, dostateczny, niedostateczny).

Dwie końcowe procedury oparte są bardziej na syntezie niż na analizie, przy czym dotyczą one głównie ewaluacji na potrzeby badań podstawowych:

➤ **Generalizacja** – faza agregowania symptomów i uogólniania wniosków na populację. Podstawą generalizacji z jednej strony jest reprezentatywność prób badawczych, a z drugiej – powtarzalność tych samych zjawisk empirycznych w różnych pomiarach oraz społeczna zgodność co do przesłanek formowania wartości. W fazie tej produktem finalnym są zweryfikowane twierdzenia.

➤ **Idealizacja** – faza abstrahowania od wymiaru empirycznego z przejściem na poziom teoretyczny. Obszarem inferencji na powrót staje się definiowanie i modelowanie, lecz już na wyższym pułapie rozwojowym. Produktami finalnymi są nowe struktury pojęciowe, doskonalsze modele i systemy teoretyczne.

<i>Ślady</i>		<i>Wzorce</i>
przekroje analityczne	.....	kwerendy zmiennych i przypadków
struktury danych	.....	macierze i schematy ekstrakcji
sfery wartościowania	.....	kwantyfikacja a kwalifikacja
analiza ilościowa	.....	znaczenie i właściwości statystyk
analiza jakościowa	.....	deskrypcja, eksplikacja, predykcja
synteza jakościowa	.....	generalizacja, idealizacja

## Uprzystępnianie wyników – konwersja poziomów ocen

Nagminnie ferowane w sprawozdaniach z badań rozkłady procentowe traktować należy raczej jako publikację danych surowych, a nie jako wyniki. Podanie informacji tylko o tym, ile procent badanych wybrało jakąś opcję, jest rzuceniem na czytelnika obowiązku przetwarzania. Aby nie poprzestać tylko na krytyce, przedstawiam propozycję zabiegu ułatwiającego interpretację wyników pomiaru *Skalą dwuważonych ocen*. Chodzi o odniesienie parametrów do tak powszechnego schematu poznawczego, jakim jest zakorzeniony w nas skrypt ewaluacyjny tradycyjnych ocen szkolnych {2, 3, 4, 5}. Kategorie nazw tej starszej skali {ndst, dst, db, bdb} są bardziej jednorodne w porównaniu z nowszą skalą {1÷6}, lepiej pasują do procentowego ustalenia progu wymagań na zaliczenie w szkolnictwie wyższym i dlatego są tam nadal stosowane.

Przyjrzyjmy się uważnie przedstawionej tu propozycji konwersji z opcji wypowiedzi w *Skali dwuważonych ocen* na kategorie i wartości stopni szkolnych:

<u>raczej nie</u>	<u>brak zdania</u>	<u>raczej tak</u>	<u>tak</u>	<u>absolutnie tak</u>
niedostateczny	dostateczny	dobry	bardzo dobry	
... 1,5 {2! 2 2+}	2,5 {3- 3 3+}	3,5 {4- 4 4+}	4,5 {5- 5 5!}	5,5

Zauważmy, że opcje wypowiedzi (wskaźniki) są zmienną dyskretną, kategorie stopni wytyczają rozłączne przedziały osiągniętych poziomów, a reprezentacje liczbowo-symboliczne tworzą dodatkowe podprzedziały, które ostatecznie – wskutek liczenia średnich – przechodzą w zmienną ciągłą. Zjawisko to jest typowe w procesie odtwarzania cechy pierwotnie ciągłej z jej dyskretnych próbek.

Pojawienie się ocen 2! i 5! nie zdziwi tych, którzy już dawno spotkali się z intuicją nauczycieli podpowiadającą, że są przypadki nie mieszczące się w standardach edukacyjnych (stąd dzisiejsze stopnie 1 i 6). W uzyskanych rezultatach badań może się zdarzyć, że pojedynczy wskaźnik wypadnie poza kresem skali, przy znormalizowanym narzędziu jest to jednak rzadkie, toteż dla tych wyjątków nie warto rozciągać skali kosztem wyrazistości wyników. Warto natomiast przyjąć, że każdy z przedziałów ocen ma jednostkową rozpiętość  $\pm$  pół stopnia. W ten sposób powstaje quasi-realny, a wykorzystywany do obliczeń efektów względnych, absolutny kres górny osiągnięć (5,5). W pewnym sensie oznacza to przypadek, gdy student ma do indeksu wpisana słowną ocenę 'celujący', lecz ze względów formalnych ocenę liczbową tylko 5.

Węzłem gordyjskim wypowiedzi w kwestionariuszu jest 'brak zdania'. Odpowiada to sytuacji, gdy trzeba podjąć trudną decyzję: zaliczyć czy nie (2,5). Oczywiście dotyczy to założenia, że skalę wyrażamy w ocenach akademickich. Dowolne odchylenie w lewo oznacza poziom niedostateczny, co w *Skali dwuważonych ocen* jest wyrazem rozbieżności między wartościami uznawanymi przez jednostkę a standardami ewaluacyjnymi. Kolejny punkt krytyczny to

strefa niepewności '*raczej tak*' (3,5), która jest progiem przejścia na poziom dobry, wytyczając środek zarówno cztero- jak i sześciostopniowej skali ocen. Graniczna jest też wypowiedź '*tak*' (4,5), rozdzielająca poziom dobry od bardzo dobrego. Zdobyć najwyższej oceny (5,5) warunkowane jest pewnością siebie w zwrocie '*absolutnie tak*', oczywiście w dodatniej polaryzacji stwierdzeń.

Konwersja na oceny szkolne przydaje się w praktyce upowszechniania wyników, lecz czy jest uprawniona teoretycznie? Oznacza przecież dodanie do każdej z wag pewnej wartości stałej (*const.* = 2,5). Umożliwia to kardynalna właściwość skali dwuważonej, jaką jest skalowalność. Dzięki temu można szacować estymatory, wyznaczać empiryczne poziomy oznak i przesuwając je o wartość stałą, nawet jeśli badana cecha ma charakterystykę krzywoliniową. Przesunięcie oznacza jedynie zmianę kryteriów oceniania. Liczy się również dowolne średnie, lecz nie wolno wnioskować, że 4 jest dwa razy „lepsze” od 2, analogicznie jak ocena dobra nie jest „dwa razy” lepsza od niedostatecznej. Podobnie jak ze średnimi ocen szkolnych, tak i tu przedstawiany choćby z dokładnością do dwóch miejsc po przecinku poziom jakości cech mentalnych jest jedynie wartością czysto umowną, wymagającą odniesienia i porównania.

Podsumowując, konwersja obejmuje przejście z siedmiopunktowej skali wypowiedzi na sześciopredziałową skalę wartości o środkach 0, 1, 2, 3, 4, 5 i rozpiętości działek  $\pm 0,5$ , z czego do końcowej prezentacji wyników bierze się cztery górne przedziały odniesione do tradycyjnych osiągnięć akademickich, tj. poziomy: niedostateczny, dostateczny, dobry i bardzo dobry. Zastosowanie tej ściśle skodyfikowanej procedury transpozycji znakomicie ułatwi percepcję wyników ewaluacji i wymianę wniosków, zwłaszcza pomiędzy nauczycielami, którzy uczą w środowiskach homogenicznych, badawczo podobnych. Samo jednakże posługiwanie się poziomami skalowalnymi nie wystarcza do analizy porównawczej osiągnięć grup różniących się społecznie, etnicznie i kulturowo.

### **Porównywalność prób – indeksy, buforowanie**

W skali dwuważonej nie można w zbyt uproszczony sposób interpretować przedziałów ocen i poziomu wyników. Aby zrozumieć to zagadnienie, zastanówmy się, czy poprawa szkolnej oceny z trójki na czwórkę jest tyle samo warta, co z czwórki na piątkę? Otóż trudno w tej sprawie cokolwiek orzekać, pomimo że w obu sytuacjach nastąpiła zmiana dokładnie o jeden stopień. Sprawa komplikuje się jeszcze bardziej przy próbie porównywania efektów jakiegoś procesu, zachodzącego w odmiennych środowiskach. Spróbujmy na przykład porównać efektywność zajęć komputerowych w klasach licealnych o profilu informatycznym z klasami o profilu humanistycznym. Kontekst jest nieporównywalny, zarówno pod względem poziomu *InfoKultury* na początku kursu, jak też pod względem możliwości percepcyjnych w toku kształcenia.



Innego rodzaju trudność porównawczego wartościowania efektywności procesów edukacyjnych wynika z faktu, że skala pomiarowa ma ograniczoną liczbę kategorii wypowiedzi, a więc krańcowe wybory ograniczone są kresem dolnym i górnym. W konsekwencji – dla startujących z niższego poziomu uprzednich doświadczeń dostępna przestrzeń wyrażania pozytywnych przemian jest szersza. Natomiast uczniowie zaawansowani mogą już w pierwszym pomiarze wybrać wypowiedź skrajną ('*absolutnie...*'), bez możliwości uzyskania przyrostu w pomiarze końcowym. Warto przy okazji uzmysłowić sobie, jak ważną funkcję pełnią owe, pozornie redundantne, skrajne opcje wyboru. Są one niezwykle korzystnym rozszerzeniem zakresu pomiarowej skali. Innym korzystnym zabiegiem jest zwiększenie rozdzielczości skali poprzez dodanie opcji wypowiedzi niezbyt pewnych, dopuszczających wątpliwości ('*raczej...*'). Jednakże nawet opcja umożliwiająca powstrzymanie się od wyrażenia opinii ('*brak zdania*') nie załatwia problemu z brakiem poziomu zerowego, do którego można byłoby odnosić natężenie cechy, ażeby porównywać rezultaty.

Proponowanym przeze mnie rozwiązaniem problemów z analizą porównawczą jest wykorzystanie względnych parametrów statystycznych, zwanych *indeksami*<sup>34</sup>. Liczenie względne polega na odnoszeniu osiągnięć konkretnych grup do tego, co dla wszystkich jest wspólne, a mianowicie do kresu górnego skali. Rezultat względny jest stosunkiem sumy zmian rzeczywistych wobec sumy przyrostów potencjalnych, teoretycznie możliwych do osiągnięcia jako maksymalne. Tego rodzaju ilorazowo-wielomianowy sposób liczenia sprawia, że indeksy są bardziej czułe na minimalne fluktuacje wypowiedzi w górnym obszarze skali, a mniej czułe na poziomie dolnym. Niweluje to nierówność swobody wyboru i umożliwia porównywanie efektów uzyskanych w dowolnie zaawansowanych grupach, poddawanych w zróżnicowanych środowiskach odmiennym procesom edukacyjnym. Możliwe jest zestawianie rezultatów nie tylko grup rówieśniczych, lecz także wyników kolejnych roczników w ramach badań wieloletnich. Indeksy dotyczą jednakże tylko zmian, a nie poziomu cech.

Jeśli chodzi o porównywanie poziomów cech, to oczywiście najbardziej wnioskotwórcze jest badanie grup równoległych, poddawanych podobnemu procesowi. W ocenie zróżnicowania poziomów uzyskanych przez dwie grupy pomaga wyznacznik ufności statystycznej. Miara ta opiera się na założeniu, że przy wzroście liczebności prób badawczych coraz bardziej zawęża się przedział, w którym z określonym prawdopodobieństwem mieści się rzeczywisty poziom średni cech danej grupy. Teoria estymacji przedziałowej ma jednak mniejsze znaczenie w praktyce pomiarów grup niezbyt licznych, jak to się dzieje w ewaluacji procesów zachodzących np. w zbiorowości klas szkolnych. W takich lokalnych pomiarach bezpośrednie porównywanie poziomów cech jest jak najbardziej uzasadnione ze względu na homogeniczność środowiska.

W ocenie tego, czy dane dwie grupy są choćby w pewnym wycinkowym zakresie porównywalne, pomaga wmontowany do skali *wskaźnik buforowy*.

**Wskaźnik buforowy** to celowo wprowadzony do narzędzia pomiarowego próbnik takiej cechy, która nie zależy od indykatum, lecz może wpływać na indykatum i jest bardzo istotna w szerszym kontekście badanej dziedziny. Wskaźnik ten służy do weryfikacji podobieństwa grup i rzetelności badań.

Definicja ta wymaga znacznego rozwinięcia, ponieważ ze znajomością techniki buforowania nie jest najlepiej. Zauważmy, że wskaźnik buforowy z założenia nie mierzy wprost żadnego z komponentów indykatum, a jednak jego poziom jest kumulowany do wyniku ogólnego. W *Skali dwuważonych ocen* przyjąłem taką właśnie zasadę, gdyż próbkowanie cechy neutralnej i porównywanie jej w kontrolnej parze z cechą zależną od indykatum jest silnie wnioskotwórcze. Konkretnie – w *Kwestionariuszu InfoKultury* wskaźnik buforowy (pozycja 11) testuje ogólną gotowość do uczenia się, podczas gdy wskaźnik paralelny (12) wyznacza gotowość uczenia się o komputerach. Bodziec-stwierdzenie nr 11 jako jedyna pozycja skali nie dotyczy komputeryzacji, lecz edukacji. Narzędzie służy do pomiaru *InfoKultury*, zaś dyspozycja do uczenia się z pewnością jest częścią kultury osobistej, choć nie wprost *Kultury informatycznej*. Wskaźnik ten nie wypacza wyniku ogólnego, natomiast jest niezwykle istotny poznawczo:

- 1) w ujęciu globalnym ukazuje, że w badanych populacjach chęć uczenia się o komputerach jest wyższa niż ogólna gotowość do uczenia się (ryc. 19);
- 2) w pomiarach lokalnych weryfikuje, czy poddane ewaluacji bądź eksperymentowi grupy są porównywalne pod względem dyspozycji kognitywnej;
- 3) w porównywaniu wyników ewaluacji wyraźnie zróżnicowanych środowisk umożliwia odniesienie cech informatycznych do różnicy cechy neutralnej;
- 4) w pomiarach zupełnie innego indykatum użycie identycznego wskaźnika buforowego pozwala na weryfikację reprezentatywności próby badawczej.

Wymieniłem tu tylko część korzyści wynikających z techniki buforowania. Szczególnie przydatne jest umieszczanie w różnych narzędziach tego samego wskaźnika buforowego, zweryfikowanego wcześniej empirycznie na reprezentatywnej próbie i upowszechnionego z unormowanym rozkładem statystycznym. W badaniach pedagogicznych, zwłaszcza w ewaluacji procesów edukacyjnych, dyspozycja do uczenia się jest istotnym czynnikiem, którego pomijać nie można, stąd nieporozumieniem jest zalecany w książkach czysto losowy dobór osób do grup w eksperymencie, bez zapewnienia jakiegokolwiek porównywalności. Losowanie jest przyczyną zróżnicowania grup kontrolnej i eksperymentalnej właśnie ze względu na tego typu istotne czynniki. Jeśli respondenci nie są celowo dobrani do grup równoległych za pomocą innego narzędzia, to należy zweryfikować ich homogeniczność choćby za pomocą wskaźnika buforowego.

## Podstawy uogólnień – standardy ewaluatywne

Każdy proces ewaluacji powinien mieć z góry wytyczony cel. Może to być albo dążenie do rozwikłania problemów naukowych, albo czysto praktyczny zamiar oceny zjawisk lokalnych. Jakkolwiek w tym drugim wypadku w ogóle nie ma zamiaru ani potrzeby uogólniania wniosków na populację, to jednak niektóre uwarunkowania tzw. *dobroci pomiarowej* powinny być przestrzegane także w ewaluacji praktycznej, której celem jest wyłącznie wyszczególnianie i porównywanie rezultatów wybranych grup oraz klas. O ile więc uogólnianie zachodzi jedynie w sytuacji badań naukowych, a te z uwagi na swą specyfikę unikalności nie można ująć w jakieś uniwersalne ramy gotowych zaleceń co do postępowania, o tyle wspólną przesłanką wszelkich koncepcji badawczych jest właśnie stosowanie pewnych strategii determinujących uzyskanie jakości badawczej. Strategie, metody oraz techniki charakterystyczne dla ewaluacji za pomocą skal dwuważonych zostały już wcześniej objaśnione, toteż tutaj przedstawię w formie podsumowania zbiór podstawowych zasad jakościowych, warunkujących możliwość uogólniania wniosków poza zakres badanych prób. Uwarunkowania te ujmę w grupach jakości eksploracyjnej i interpretacyjnej.

➤ **Jakość eksploracyjna** obejmuje wszystkie te zasady postępowania, które mieszczą się w fazach: od opracowania koncepcji, sformułowania problemów, ekstrakcji indykatorum, skonstruowania narzędzia, doboru prób badawczych, aż do przeprowadzenia pomiarów. Są to te fazy, w których ewaluatorzy mają jeszcze istotny wpływ na to, jak cenne zasoby empiryczne zostaną pozyskane i czy na podstawie tych zasobów wolno będzie formułować twierdzenia ogólne. Uogólnienie jest dopuszczalne i sensowne tylko wtedy, gdy problem badawczy jest poznawczo znaczący, grupy respondentów są liczne i reprezentatywne, narzędzie trafne i unormowane, a pomiary rzetelne, z powtarzalnymi wynikami.

*Jakość problemów.* Uwarunkowanie to dotyczy doniosłości podjętej problematyki badawczej z perspektywy rozwoju dziedziny wiedzy i postępu naukowego. Chodzi o to, czy wytyczone cele ewaluacji są w aspekcie poznawczym wystarczająco istotne, ażeby ich realizacja i upowszechnienie rezultatów stanowiło przyczynek rozwojowy. Zatem, zadania badawcze nie mogą być trywialne, lecz chociażby w skromnym zakresie powinny być namiastką swoistego odkrycia. Formułuje się więc i rozstrzyga takie szczegółowe problemy badawcze, które dotąd nie były wystarczająco rozwikłane. Wymaga to rozpoznania aktualnego stanu wiedzy w eksplorowanej dziedzinie – dodatkowo także i z tego powodu, ażeby zapewnić taksonomiczną prawidłowość rozwarstwienia indykatorum na właściwe komponenty i relacje. Problemy bowiem dość często formułowane są jako szukanie zależności, stąd tak ważna jest trafność ekstrakcyjna zmiennej ogólnej, kompletność strukturalna definicyjno-relacyjnego modelu składników oraz kompletność eksplikacyjna w poszukiwaniu czynników i rezultatów.

*Jakość narzędzia.* Kolejne uwarunkowanie jest ściśle związane z poprzednim, jako że poziom pojęciowo-teoretyczny problemów musi być wiernie przełożony na konkretną implementację empiryczno-obserwacyjną. Przy założeniu trafnej ekstrakcji komponentów indykatorum, dobroć pomiarowa skali zależy przede wszystkim od jakości wskaźnikowania. Największą niepewność budzi zawsze to, czy udało się osiągnąć wystarczającą trafność operacjonalizacyjną elementarnych oznak indykatorum. Podstawą uogólnień mogą być tylko takie oznaki empiryczne, które są autentycznymi przejawami tego, co pierwotnie zdefiniowano jako substancję ewaluacji (problemy, obiekty, zjawiska itp.), a wtórnie wyrażane jest jako rezultaty badawcze w formie wniosków i twierdzeń. Formułowanie praw podstawowych wymaga jednak wcześniejszego przewidywania rezultatów, stąd tak ważna jest trafność predykcyjna. Ponadto uogólnienia są wiarygodne jedynie wtedy, gdy zjawiska próbkowano w komplementarnym ujęciu czasoprzestrzeni, a skala była standaryzowana i normalizowana.

*Jakość prób.* Uwarunkowanie to odnosi się do źródeł próbek empirycznych, a więc do przekroju respondentów, od których pochodzą reakcje-wypowiedzi. Jeśli celem ewaluacji jest uogólnienie wniosków na populację, to w pierwszym rzędzie należy doprecyzować, o jaką zbiorowość chodziło. W wielu raportach z badań opis prób jest zbyt powierzchowny, ażeby wyniki mogły być uznane za podstawę do uogólnień. Ponadto często w sposób nieuprawniony utożsamia się charakterystykę próby badawczej z zakresem definiowania populacji. Samo zawężenie określenia populacji na przykład do '*młodzieży gimnazjalnej*', jakkolwiek precyzuje dość wąsko przedział wiekowy respondentów, to jednak nie upoważnia do uogólnień, jeżeli próbki pochodzą wyłącznie ze środowiska wielkomiejskiego. Trzeba wówczas zawęzić zakres tego, co definiuje populację, albo należy pobrać próbki także z mniejszych środowisk. Warunkiem ufności do miar statystycznych jest wystarczająca liczebność respondentów, natomiast podstawą wiarygodności twierdzeń musi być przekrojowość prób i reprezentatywność społeczna: demograficzna, etniczna, środowiskowa, kulturowa itp.

*Jakość pomiarów.* W badaniach za pomocą skal wpływ ankietera jest zminimalizowany, toteż sam pomiar nie musi być przeprowadzany przez ewaluatora. Wręcz korzystniej jest, gdy w pomiarach jakości zajęć kwestionariusze rozdaje i zbiera wykładowca, a nie osoba z zewnątrz. Ważny jest jednak wybór najwłaściwszego momentu próbkowania. W ewaluacji procesów edukacyjnych pierwszy pomiar powinien być wykonany przed rozpoczęciem intencjonalnego oddziaływania, natomiast drugi – blisko końca (o ile jest to możliwe, przed wystawieniem ocen końcowych, gdyż mogą one wpłynąć na zmianę postaw). Istotny wpływ na rzetelność pomiaru ma chwilowy kontekst nastroju respondentów (np. stres przed czekającym ich kolokwium z innego przedmiotu). Tego typu uwarunkowania zewnętrzne nie powinny być lekceważone, gdyż

mogą być przyczyną silnego wypaczenia rezultatów przez czynniki nie mające nic wspólnego z indykatum. Zdarza się, że próbkowanie lepiej jest przełożyć, dlatego warto zaplanować drugi pomiar na przedostatnie zajęcia. Poddanie się badaniu i konieczność wypowiedzania się same w sobie są dyskomfortem, toteż duże znaczenie ma objaśnienie celu ewaluacji oraz taktowne zachęcenie do rzetelnego, przede wszystkim samodzielnego wypełniania kwestionariusza. Respondenci bowiem wykazują trudne do pohamowania (wyniesione ze szkoły) inklinacje do uzgadniania „prawidłowych odpowiedzi” z osobą siedzącą obok.

➤ **Jakość interpretacyjna** obejmuje te uwarunkowania wnioskotwórcze, które wynikają z wartości posiadanych zasobów empirycznych, z wszechstronności analiz statystycznych, z poprawności wyjaśniania oznak i prawidłowości formułowania twierdzeń. Uogólnianie w fazie interpretacji jest zasadne wtedy, gdy konsystencja danych daje podstawy do estymacji rozszerzającej wnioski na populację. Podstawę stanowią rozkłady, parametry i wyznaczniki jakości, o ile spełnione były wcześniejsze uwarunkowania jakości eksploracyjnej.

*Jakość zasobów.* W początkowym stadium interpretacji ewaluator podejmuje decyzję o tym, czy pozyskane z pomiarów dane są wystarczające do uogólnień ze względu na licznosc, reprezentatywnosc i przekrojowosc prób. Jeśli tak, to przystępuje do oceny wewnętrznej konsystencji danych pod kątem spójności i rzetelności wypowiedzi, przy założeniu, że ewentualne pojedyncze przypadki skrajnie niefrasobliwego wypełnienia kwestionariuszy zostały już wcześniej wyeliminowane. Ponadto w tej fazie przyjmuje się, że skala pomiarowa jest już w pełni zweryfikowana, stąd wszelkie oznaki są odzwierciedleniem zjawisk empirycznych, a nie wadą narzędzia. Jakość zasobów z jednej strony wynika z wielości i różnorodności ujawnionych oznak, a z drugiej strony kontrolowana jest pod względem konkluzyjności, zbieżności oraz powtarzalności wypowiedzi. Najbardziej wartościowe zasoby otrzymuje się wówczas, gdy respondenci są podmiotami ocenianego procesu, a dane pochodzą z pomiarów wieloletnich.

*Jakość wniosków.* Uwarunkowanie to zależy od kompetencji i dociekliwości ewaluatora. Przede wszystkim konieczna jest umiejętność posługiwania się różnymi metodami analiz statystycznych. Z analizy rozkładowej wyprowadzić można uogólnienia dotyczące istotności rezultatów i ufności do nich. Z analizy gradientowej odczytuje się dominujące kierunki zmian. W analizie różnicowej ustala się tendencje, a w analizie indeksowej dynamikę i efekty przemian. Za pomocą statystycznych wyznaczników jakości pomiarowej weryfikuje się te właściwości zasobów, od których zależy możliwość uogólniania. I tak, poprawność wskaźnikowania i siłę wnioskowania stwierdza się na podstawie miary konkluzyjności, istotność różnicowania grup testuje się miarą selektywności, słuszność wnioskowania sprawdza się miarą trafności, a konsystencję danych pod względem spójności wypowiedzi kontroluje wyznacznikiem rzetelności.

Oprócz uogólnień budowanych na teoriach estymacji i testach istotności, ważnym rezultatem procesu oceniania wartościującego powinno być tworzenie **standardów ewaluatywnych**. W pewnym sensie wartościowanie przebiega samoistnie, kiedy to w toku rozwoju cywilizacji, na podstawie generalizowania niemal jednomyślnych postaw, kreowane są wartości uniwersalne. Wobec kwestii złożonych wartościowanie formowane jest przez kontekst społeczny, a zwłaszcza poprzez procesy edukacji. Ewaluacja demokratyczna umożliwia doświadczalne wyznaczenie szczegółowych norm wartości w wąskim zakresie próbkowania indykatum. Każda z elementarnych tez tworzących skalę wyraża mikroprzesłanie, będące częścią wartościowania o spolaryzowanym ładunku walencyjnym. Na licznych próbach badawczych dostrzec można zjawisko auto-normalizacji wskaźników, oznaczające dochodzenie do konsensusu społecznego wokół tego, co jest wartością. Oprócz znaku walencyjnego z pomiarów uzyskuje się także konkretny wymiar stopnia akceptacji owych mikroprzesłań. Poziom akceptacji zależy od treści stwierdzeń, a więc od samego narzędzia, jednak w badaniach długofalowych ujawniają się wyraźne trendy wzrostowe, co świadczy o rozwoju respondentów i również o rewaluowaniu standardów.

Ewaluacja procesów oparta jest na dwóch pomiarach, dlatego potrzebne są podwójne standardy. Jedne standardy wytyczają nominalny poziom bazowy na początku procesu, a drugie określają potencjalną normę docelową, jaką powinno się osiągnąć na końcu. Odniesienie do bazy umożliwia oszacowanie wstępnego poziomu cech konkretnej grupy, co daje podstawę do optymalizacji oddziaływań edukacyjnych. Porównanie rezultatów końcowych danej grupy z normami dla populacji w pewnym stopniu weryfikuje jakość oddziaływań. Dla większej pewności co do rzeczywistych oddziaływań potrzebne są pomiary wielokrotne, w których uzyskuje się powtarzalność osiągniętych przyrostów. Wyznaczanie standardów ewaluatywnych jest procesem dynamicznym, wymagającym badań cyklicznych z ocenianiem deskryptywnym i formatywnym. Ustala się bieżące standardy normalizujące skalę, które ulegają powolnym przemianom. Z tego powodu co pewien czas konieczna jest aktualizacja norm, a przy chęci użycia skali w innych nacjach, także adaptacja kulturowa.

<i>Ślady</i>		<i>Wzorcy</i>
uprzystępnianie wyników	.....	poziomy ocen akademickich
dyskretyzacja poziomów	.....	progi, przedziały i kresy skali
porównywalność efektów	.....	indeksy i przedziały ufności
porównywalność grup	.....	technika buforowania
źródła jakości eksploracyjnej	.....	problemy, narzędzia, próby, pomiary
źródła jakości interpretacyjnej	.....	zasoby, analizy, normy, standardy
stany bazowe i normujące	.....	oceny deskryptywne i formatywne

## **PRAKTYKA EWALUACJI SPLOTOWEJ**

Przedstawienie teoretycznego modelu-wzorca metodologii ewaluacyjnej jest sensowne tylko wtedy, gdy możliwe jest przejście od koncepcji do praktycznej implementacji. Wprawdzie strategie metodologii badawczej są ze swej natury w większości wyrażane operacyjnie – jako procedury postępowania – tym niemniej zawsze niezbędne jest zejście z poziomu modelowej ogólności na poziom opisu konkretnych, szczegółowych czynności<sup>35</sup>. Ze względu na złożoność procesów, przed przystąpieniem do czynności empirycznych, eksploracyjnych i interpretacyjnych warto poznać pojęciowo-teoretyczną podbudowę ewaluacji. Jest to istotne choćby z tego powodu, ażeby mieć pewność, czy zaplanowana metodologia jest właściwa dla rozwikłania obranych problemów badawczych.

Teoria i praktyka ewaluacji pozostają w ścisłym, obustronnym związku. Z jednej strony ocenianie wartościujące musi być osadzone na przesłankach naukowych, co oznacza, że podstawę koncepcji ewaluacyjnej stanowią teorie i modele, które sprawdza się doświadczalnie. Owe przesłanki dotyczą zarówno problemów i twierdzeń związanych z samym indykatum, jak też odnoszą się do teoretycznych podstaw obranej strategii pomiarowej. Optymalny scenariusz praktycznego przebiegu ewaluacji zależy dokładnie od wybranej metodologii. Z drugiej jednak strony oznaki empiryczne mogą być podstawą formułowania dodatkowych twierdzeń, które nie były przewidziane ani jako cel badawczy, ani jako strategia pomiarowa. W rezultacie tego typu wniosków praktycznych rozwijana jest teoria, wskutek nabywania wiedzy o indykatum, i doskonalona metodologia samej ewaluacji, dzięki weryfikacji metod i technik badawczych.

Aspekty teoretyczne i praktyczne przenikają się w poszczególnych fazach ewaluacji, chociaż w obu tych warstwach na co innego rozłożone są akcenty. Domeną teorii jest przede wszystkim wyjaśnianie i formowanie, podczas gdy na użytek praktyki najważniejsze jest instruowanie. W kolejnej części książki postaram się odpowiedzieć przede wszystkim na pytanie o to, jak postępować, ażeby zaprojektować i przeprowadzić ewaluację. Także i tu odniesieniem będzie informatyzacja, a więc jedynie wąski fragment edukacji, chociaż ewaluacja bywa czasem ujmowana jako ogólna ocena jakości organizacji kształcenia<sup>36</sup>.

## **Jak podjąć ewaluację – projekt badawczy**

Ewaluacja spłotowa z użyciem narzędzia skalowanego jest złożonym procesem, który wymaga szczegółowego zaplanowania i starannego przygotowania. Niezależnie od tego, czy zamiarem jest realizacja badań podstawowych, czy jedynie pomiar lokalny służący praktyce – zawsze podstawą powinno być opracowanie projektu badawczego. Właśnie w projekcie tym precyzuje się:

- zamierzenia finalne, wyrażone operacyjnie jako cele ewaluacji;
- zadania szczegółowe, prowadzące do rozwikłania problemów badawczych;
- założenia co do dziedziny i meritum wytyczającego przedmiot ewaluacji;
- założenia co do populacji i doboru prób będących podmiotami ewaluacji;
- kryteria uogólniania określające zasięg przestrzenny ocenianych zjawisk;
- kryteria formułowania wniosków i wytycznych w aspekcie czasowym;
- strategię wyboru sposobów wskaźnikowania i skalowania narzędzia;
- strategię planu badawczego, metod próbkowania i techniki pomiarowej;
- założenia eksploracyjno-deskryptywne określające metody analizy;
- założenia interpretacyjno-formatywne określające metody syntezy.

Projekt badawczy tworzy się poprzez udzielenie odpowiedzi na pytania:

- Jakie cele teoretyczne i praktyczne zamierzamy osiągnąć dzięki ewaluacji?
- Jakie problemy ogólne i szczegółowe planujemy w efekcie rozstrzygnąć?
- Co stanowi przedmiot badań, z jakich składa się komponentów i struktur?
- Kto będzie podmiotem oceniania, jak liczne to próby i z jakiej populacji?
- Czy rezultaty mają być uogólniane, czy jedynie wyszczególniane lokalnie?
- Czy wnioski mają opisywać stan aktualny, czy również wytyczać rozwój?
- Jakie narzędzie posłuży do pomiaru i co będzie wskaźnikiem zmiennych?
- Jaki plan badawczy i schemat grup porównawczych będzie zastosowany?
- Czego poszukiwać się będzie w zasobach empirycznych i jak analizować?
- Co stanowić będzie kryteria interpretacji i przesłanki syntezy?

Sformułowanie odpowiedzi w tego typu kwestiach jest fundamentem ewaluacji. Najpełniejszą wykładnię podać musi sam twórca koncepcji badawczej, co jednak nie zwalnia badaczy wykorzystujących gotową skalę od konieczności nabycia pełnej wiedzy w tej materii i od potrzeby określenia kierunków planowanego przez siebie doświadczenia. Oznacza to, że nie wolno używać narzędzi umieszczanych jako załączniki w raportach z badań, o ile nie ma tam wystarczającego opisu podstaw metodologicznych ani schematu interpretacji. Każde doświadczenie pomiarowe jest w pewnych szczegółach niepowtarzalne, dlatego zbyt ogólnikowe projekty badawcze mogą prowadzić do uzyskiwania sprzecznych rezultatów, mimo zastosowania rzekomo identycznych strategii. Inni są respondenci pod względem cech osobniczych, odmienny społecznie, kulturowo i technologicznie może być teren badań, upływa też czas, który ma znaczenie w formowaniu dynamicznego poziomu standardów ewaluacyjnych.



Wszystko to ma wpływ na rezultaty pomiarów, a więc powinno być ujęte nie tylko w projekcie ewaluacji, lecz również opisane w sprawozdaniach z badań. Z jednej strony ułatwia to zadanie tym osobom, które chciałyby wykorzystać opracowaną koncepcję do porównawczych pomiarów w lokalnym środowisku, a z drugiej strony umożliwia badaczom rozwijanie i uzupełnianie metodologii oraz poszerzanie zakresu problematyki objętej eksploracją empiryczną. Okazuje się bowiem, że za pomocą tego samego narzędzia pomiarowego, lecz przy zastosowaniu różnych planów badawczych, metod i technik, można osiągać dodatkowe cele i rozwiązywać kolejne problemy szczegółowe. Przyjrzyjmy się sposobom budowy projektu badawczego i przykładom formułowania założeń.

➤ **Cele ewaluacji** wyrażają w zoperacjonalizowany sposób dążenie do osiągnięcia pewnego stanu dokonanych czynności badawczych. Wprawdzie pojęcie celu kojarzy się z rezultatem końcowym, jednakże sama ewaluacja jest procesem dynamicznym, a często także cyklicznym, toteż cel nie powinien być formułowany w kategoriach domkniętych i zadaniach ostatecznie spełnionych, tym bardziej, że w przypadku eksperymentu z posłużeniem się testami istotności można nie mieć statystycznych podstaw do powzięcia rozstrzygających decyzji. Przede wszystkim jednak przyjąć należy, że nawet ewaluacja *formatywna*, której rolą jest towarzyszenie procesom poddawanych ocenianiu, nie powinna mieć na celu bezpośredniego formowania rezultatu końcowego, lecz jedynie powinna służyć wykrywaniu czynników, które kształtują proces w pożądanym, wartościowym kierunku. Generalnym celem ewaluacji jest więc poznanie wartości ściśle określonego wycinka rzeczywistości, przy czym owo poznanie służyć może zarówno celom naukowym, jak też może być podstawą dyrektyw praktycznego postępowania ulepszającego rzeczywistość.

*Cele teoretyczne.* Do kategorii tej należą założenia ukierunkowujące badania podstawowe lub interdyscyplinarne, służące rozwojowi nauki. Mieszczą się tu również wszelkie te zamierzenia, które służą weryfikacji naukowych podstaw metodologii pomiarowej. Cel teoretyczny może więc być wyznaczony jako badanie indykatum lub jako sprawdzanie strategii, zwłaszcza wtedy, gdy opracowuje się innowacyjną koncepcję wskaźnikowania pośredniego. Przykładowo można obrać cel *wyodrębnienia* struktury komponentów indykatum, bądź *weryfikacji możliwości wyodrębniania* komponentów, które są ze sobą silnie powiązane. Wszystko zależy od tego, na jakim etapie rozwojowym znajduje się badana problematyka. Sformułowanie celów ewaluacji musi więc być poprzedzone wnikliwym rozpoznaniem aktualnego stanu wiedzy o indykatum. Jeśli dziś podważa się np. sens strukturalnej definicji *postawy* tylko dlatego, że jej komponenty afektywne i kognitywne pozostają w silnie powikłanych i niestałych związkach, to nadal celowe są badania podstawowe weryfikujące owe zależności, niezależnie od mody na badanie tego, jak kształtować postawy.

Wybrane przykłady sformułowań teoretycznych celów ewaluacji *InfoKultury* podałem na stronie 20, zatem tutaj scharakteryzuję jedynie metodologiczne przesłanki ich redagowania. Podstawowym założeniem badań naukowych jest możliwość *uogólniania*, dlatego główne cele teoretyczne powinny odnosić się do wartościowania zjawisk globalnych, obejmujących populację. Kolejnym założeniem jest konieczność *deskrypcji* i *eksplikacji* – stąd biorą się cele wyrażone jako opisy i objaśnienia zjawiska, zwłaszcza w ujęciu przyczyn i skutków. Równie ważnymi przesłankami formułowania celów są *inspekcja* i *predykcja*, zobowiązujące do zdiagnozowania stanów bieżących i prognozowania zmian. Prowadzi to do uzyskania dynamicznych standardów właściwości oraz dyspozycji podmiotów ewaluacji. Istotnymi elementami ewaluacji splotowej są też procesy *strukturyzacji* i *hierarchizacji*, co wytycza cele przynajmniej w zakresie budowania taksonomii, z definiowaniem kategorii oraz relacji komponentów indykatorum. Domeną badań eksperymentalnych jest *weryfikacja* hipotez, gdzie cel stanowi testowanie siły związku pomiędzy oddziaływaniem a skutkami. Realizacja tak różnych celów wymaga stosowania różnych metod i technik.

*Cele praktyczne.* Znamienną odmianą w tej kategorii założeń jest *wyszczególnianie*, polegające na wartościowaniu zjawisk lokalnych. Rezultaty pomiarów pozostają interpretowane jedynie na poziomie tych klas i grup, które poddane były ocenie. Wyniki nie są uogólniane na populację, a jedynie porównywane ze standardami. W takiej sytuacji nie ma potrzeby testowania istotności różnic ani estymowania poziomów i przedziałów ufności. Celem natomiast może być *aktualizacja* standardów, co wymaga jednak dysponowania wieloma wynikami reprezentatywnych prób badawczych. W formułowaniu celów praktycznych najważniejszą przesłanką jest *użyteczność* rezultatów. Zmierza się do tego, ażeby określić efekty konkretnych oddziaływań na konkretne grupy, ale bez eksperymentów izolujących od czynników zewnętrznych. Nie testuje się więc hipotez i nie buduje twierdzeń, lecz jedynie wyznacza stany chwilowe oraz dynamikę zmian, bez prognozowania. Oto przykładowe cele praktyczne, jakie można wytyczać i realizować, posługując się kwestionariuszem, programem do analizowania danych oraz metodami i technikami ewaluacji *InfoKultury*:

- ustalanie różnic oraz trendów indykatorum w lokalnych środowiskach;
- pomiar skuteczności zabiegów równoważących doznanie i poznanie;
- ocenianie jakości zajęć z informatyki lub technologii informacyjnej;
- optymalizacja autorskich programów nauczania tychże przedmiotów;
- weryfikacja standardów wymagań wobec realnych osiągnięć uczniów;
- ustalenie siły osobistego oddziaływania nauczyciela na swych uczniów;
- wykrywanie i niwelowanie niekorzystnych uzależnień od komputera;
- samokształcenie, polegające na poznaniu zjawisk mikrospołecznych;
- przyswajanie pojęć badawczych i ćwiczenie technik statystycznych.

Każdy z wymienionych celów realizuje się w nieco odmienny sposób, z akcentami położonymi na innych technikach. I tak, do ustalenia różnic porównuje się grupy w miarę możliwości homogeniczne, a do analizy trendów potrzebne są cykliczne pomiary dystansowe. Ocenę jakości zajęć odczytuje się wprost ze wskaźników, natomiast optymalizacji programu nauczania można dokonać pośrednio poprzez ocenę skutków wdrożenia innowacji. Weryfikację realności wymagań na tle możliwości realizacyjnych ocenia się poprzez rozwarstwienie klas na podgrupy uczniów o wysokim, średnim i niskim poziomie recepcji, po czym analizuje parametr zawilości jako wskaźnik dopasowania oddziaływań. Siłę oddziaływań nauczyciela pośrednio wyznacza miara amplitudy fluktuacji, a syndrom uzależnienia można bezpośrednio odczytać ze wskaźników. Cele samodoskonalenia ewaluatorów to zadania wtórne, lecz nie mniej ważne.

➤ **Problemy ewaluacji** w formie pytań badawczych wyrażają fundamentalne kwestie wymagające rozstrzygnięcia. Zagadnienia te charakteryzuje dylemat wyboru: *Jest tak, czy inaczej?* Nie wystarcza to jednak jako problematyka badawcza. W ślad za prostymi rozstrzygnięciami podążać muszą szczegółowe opisy jakościowe i ilościowe, odpowiadające na pytania: *Co na to się składa? W jaki sposób to przebiega? Co jest przyczyną, a co skutkiem? Dlaczego tak się dzieje? Jakie będą tendencje? Na ile jest to trwałe? Jak szeroki ma zasięg? Na jakim jest poziomie? Jak silne są relacje? Czy jest to istotne statystycznie?* Wielość i wnikliwość pytań w zasadzie nie jest ograniczana metodologicznie. Zaleca się redagowanie pytań w zwartej formie, przy czym każda z kwestii powinna dotyczyć wyłącznie pojedynczego problemu. Zamiast pytań treściowo złożonych lepiej jest utworzyć większą liczbę sformułowań prostszych. Z tego powodu zwykle redaguje się tylko kilka problemów ogólnych, odpowiadających celom ewaluacji, po czym rozdziela się je na problemy szczegółowe, odpowiadające odrębnym zadaniom badawczym. Już sama nazwa '*problemy*' oznacza, że ich rozwiązanie stanowi pewną trudność, dlatego rozbitcie zadań złożonych na elementarne ułatwia ewaluację, zwłaszcza na etapie interpretacji zasobów i opracowywania wyników. Wnioski interpretacyjne tworzone są bowiem jako odpowiedzi na wcześniej postawione pytania. Przesłanki formułowania problemów są dokładnie takie same, jak przy wytyczaniu celów. Podstawowym założeniem jest podział na problemy metodologiczne, teoretyczne oraz praktyczne. Pierwsze z nich dotyczą przede wszystkim rozstrzygnięć, czy przyjęto poprawny model badawczy struktury indykatorum i czy model ten znajduje potwierdzenie empiryczne. Problemy naukowe mogą być redagowane zarówno tak, ażeby weryfikować i uzupełniać funkcjonujące już teorie i twierdzenia, bądź też tak, aby doświadczalnie sprawdzać nowe koncepcje i hipotezy. Przykłady formułowania konkretnych problemów teoretycznych, możliwych do rozwiązania poprzez splotową ewaluację modelu *InfoKultury*, przedstawiłem na stronie 20.

W tym miejscu podam przykładowe problemy praktyczne, jakie stawiać można wobec lokalnego procesu edukacyjnego i orzekać na podstawie dwukrotnego, dystansowo-różnicowego pomiaru stanów na początku oraz na końcu zajęć:

- Czy przeprowadzony kurs komputerowy był jakościowo dobry?
- Jakie czynniki wpłynęły na osiągnięcie słabszych/lepszych wyników?
- Czy dopasowano skalę trudności zajęć do możliwości uczniów?
- Czy kurs był trafny, zbyt trudny, czy może raczej zbyt mało ambitny?
- Jak duże są podobieństwa oceniania wykładowcy przez różne grupy?
- Z jakiego poziomu startowała konkretna grupa ćwiczeniowa?
- Jakie były początkowe, a jakie końcowe stany uświadomienia?
- Jak silna okazała się dynamika wzrostu kultury informatycznej?
- Na ile dodatnie trendy dotyczą najslabszych/najlepszych uczniów?

Z podanych wyżej przykładów wynika, że jeśli tylko problematyka mieści się w obszarze objętym indykatem, to kwestia dalszych potencjalnych możliwości stawiania zadań badawczych zależy przede wszystkim od intuicji ewaluatora oraz od inwencji twórczej w praktycznym przeprowadzeniu wartościowania.

➤ **Przedmiot ewaluacji** precyzuje materię tego, co stanowi obiekt oceniania wartościującego. W badaniach społecznych najczęściej jest to rzecz niematerialna, np.: cecha osobnicza lub populacyjna, stan rozwoju intelektualnego, stan emocjonalny, zjawisko socjologiczne, proces przemian cywilizacyjnych, interakcje społeczne, uwarunkowania i efekty oddziaływań, jakość organizacyjna procesów edukacyjnych, przemiany mentalne i kulturowe, dyspozycje psychiczne, postawy, motywacje, aktywności i wiele innych tego typu kwestii dotyczących działania człowieka w środowisku konkretnej zbiorowości. Odrebnym nurtem ewaluacji może być ocena wartości wytworów intelektualnych, rozumianych jako kulturotwórcze przesłania niesione przez tradycyjne lub nowe media i mass media. Pozornie wartość tych wytworów można byłoby opisać za pomocą kryteriów jakości fizycznej, jednakże nie odzwierciedlałoby to rzeczywistej wartości w aspektach aksjologicznych, kulturowych i edukacyjnych. *Przedmiotu ewaluacji* nie należy więc traktować jako desygnatu pojęcia, lecz jest to opis: obiektu zainteresowania, materii poddawanej wskaźnikowaniu, substancji o ukrytych właściwościach oraz zmiennych o szacowanych parametrach. W takim ujęciu przedmiot ewaluacji jest pojęciem szerszym niż indykatem, które dotyczy jedynie wprost zmiennej ogólnej, mierzonej skalą. Za pomocą skali mierzącej to samo indykatem można pośrednio przeprowadzać ocenianie wartościujące wielu dodatkowych zjawisk, które w projekcie badawczym ujmuje się właśnie jako szczegółowy opis przedmiotu ewaluacji. Przykładowo, gdy indykatem stanowi *InfoKultura*, to jest ona podstawowym obiektem pomiaru, choć obiektem wartościowania może być także określony proces intencjonalnego bądź spontanicznego formowania owej właściwości.

Tak więc, podczas badania *InfoKultury* przedmiot ewaluacji definiuje się jako:

- cechy mentalne wyrażane w postawach i w świadomości informatycznej;
  - struktury umysłu wyrażane w skryptach postępowania wobec komputera;
  - stany dyspozycji ujawniane poprzez gotowość do poznawania i działania;
  - stany motywacji ustalane w fazach pobudzenia i zaspokojenia poznawczego;
  - stany emocjonalne ujawniane jako stres lub uzależnienie od komputera
- oraz inne tego rodzaju właściwości osób poddanych badaniom, lecz także jako:
- oceny społeczne globalnych procesów komputeryzacji i informatyzacji;
  - oceny poziomu rozwoju oraz przydatności technologii informacyjnych;
  - poziomy i składowe standardów ewaluatywnych w sferze *InfoKultury*;
  - tendencje w przemianach świadomości społeczeństwa informacyjnego;
  - oceny lokalnych kursów komputerowych wyrażane przez uczestników;
  - modele nowatorskich metodyk zajęć laboratoryjnych z komputerami;
  - kryteria wymagań i treści kształcenia przedmiotów informatycznych.

W powyższych przykładach przedmiotów ewaluacji warto zwrócić uwagę na pozorny pleonazm „oceniają ocen”. Wyjaśniam więc, że często obiektem ocen dokonywanych przez ewaluatora jest właśnie to, jak respondenci oceniają dane zjawisko. Wynika to ze specyfiki ewaluacji demokratycznej, w której o wartości świadczą wypowiedzi badanych, a sam badacz wartościuje wystawione oceny, odnosząc je do kryteriów uznawanych jako standardy.

Opis przedmiotu ewaluacji daje odpowiedź na pytanie: *Co jest oceniane?* Wynika to wprost z celów oraz problemów badawczych, jednakże w zależności od tego, co jest obiektem badań, w opisie odmiennie rozkłada się akcenty:

*Główny przedmiot ewaluacji* wymaga wyrażenia w postaci kompletnego modelu struktury pojęciowej. Oznacza to potrzebę nie tylko nazwania indykatoru, lecz także konieczność zdefiniowania komponentów tworzących strukturę, łącznie z opisem relacji zachodzących pomiędzy składowymi. Zalecana jest też próba prognozowania tego, co prawdopodobnie uzyska się w wyniku badań i w jaki sposób można to będzie wyjaśniać. Nie wystarcza więc samo wyszczególnienie próbkowanych kategorii, lecz potrzebny jest też szkic sieci powiązań i schemat interpretacji z wmontowanymi aksjomatami i twierdzeniami eksplikacyjnymi. Podłożem strukturalnego definiowania indykatoru są więc teorie i taksonomie.

*Oboczny przedmiot ewaluacji* wymaga opisu w kategoriach praktycznych efektów stosowania zabiegów metodologicznych, planów dystansowo-panelowych i technik statystycznych. Badając zmiany postaw, które z definicji powinny być w miarę trwałe, można oceniać procesy wychowawcze; mierząc uzależnienia od komputera można weryfikować skuteczność terapii; porównując rezultaty grupy kontrolnej i eksperymentalnej można testować koncepcje dydaktyczne; obserwując tendencje zmiany opinii można wartościować procesy społeczne. Przedmiotami ewaluacji są wówczas czynniki oddziaływania na indykator.

➤ **Podmioty ewaluacji** to określenie odnoszące się do grupy tych osób, które w istotny sposób związane są z przedmiotem badań. Jeśli ewaluacja dotyczy cech mentalnych, to wówczas podmiotami w pierwszym rzędzie są próby badawcze poddawane eksploracji pod kątem poziomu owych cech. W podejściu uprawniającym do uogólnień podmiotem staje się również populacja, z której wyłoniono próby. W ewaluacji lokalnych procesów edukacyjnych podmiotami są wyłącznie respondenci będący jednocześnie uczestnikami ocenianych zajęć, natomiast w badaniach podstawowych o zasięgu globalnym upodmiotowienie może być rozciągnięte na szerszą zbiorowość niż ta, którą poddano pomiarom. W szczególnym przypadku ewaluacji wytworów intelektualnych podmiotami są jurorzy oceniający, przy czym zaleca się, ażeby pochodzili oni ze środowisk będących w intencji twórców głównymi adresatami i odbiorcami, a nie medio- i rzeczoznawcami z całkowicie odmiennej kategorii społecznej. Dla ekspertów stosowniejszą rolą jest walidacja (ocenie zatwierdzające), a nie ewaluacja. W ewaluacji demokratycznej upodmiotowienie respondentów stanowi sposób sięgania do źródeł opinii o odczuciu osobistego 'ja' uwikłanego w środowisko. Oczywiście w dobrej skali pomiarowej wmontowane są dodatkowo wskaźniki kontrolne relacji typu 'ja' ↔ 'inni' oraz 'inni' ↔ 'proces', jednakże pierwiastek własnych odczuć i świadomości jest zasadniczym kryterium wartościowania.

Opis podmiotów ewaluacji jest odpowiedzią na pytanie: *Kto wyraża opinie?* Już na etapie projektowania należy sprecyzować specyfikę prób badawczych pod kątem rozwarstwienia na takie właściwości, które są istotne w założonej koncepcji eksploracyjnej. Przede wszystkim wyodrębnia się respondentów ze względu na ich wiek, określając w miarę wąski przedział, zwłaszcza w okresie szybkiego rozwoju młodzieży szkolnej. Dla tych grup optymalnym podziałem jest kategoryzacja na okresy nie dłuższe niż 3-letnie, odpowiadające etapom na II, III i IV poziomie kształcenia w systemie oświaty, czyli klasom 4÷6 szkół podstawowych, klasom 1÷3 w gimnazjach oraz klasom 1÷3 w liceach. Niekiedy jednak potrzebne jest bardziej szczegółowe rozróżnienie, ponieważ uczniowie pierwszej klasy liceum różnią się istotnie poziomem świadomości od kolegów z klas maturalnych. Nie mniej ważnym aspektem opisu podmiotów ewaluacji jest charakterystyka środowiska, skąd pochodzą respondenci. Zwykle stosuje się podział jedynie ze względu na wielkość aglomeracji, chociaż nie wystarcza to do opisu próby. Można bowiem wykazać empirycznie, że cechy intelektualne mieszkańców z różnych dzielnic w tej samej miejscowości różnią się znacznie, podobnie jak silnie zróżnicowane są cechy osób mieszkających w zbliżonych liczebnie skupiskach, lecz w innej części kraju. W opisywaniu respondentów dorosłych potrzebne jest podanie specjalności wyuczonej lub wykonywanej, gdyż od osobistych predyspozycji zależy oczekiwany poziom cech mentalnych. Jeśli na przykład planujemy pomiar *InfoKultury* w środowisku studenckim, to

istotne jest to, jaki kierunek studiują respondenci. U osób, których specjalnością jest informatyka lub inne nauki ścisłe można spodziewać się wyższego poziomu *InfoKultury* niż u studiujących nauki humanistyczne. Zróżnicowanie wynikające z osobistych predyspozycji, a przekładające się na właściwości grupy badawczej, ujawnia się już wcześniej, choćby w profilowanych liceach i technikach. Z tego powodu dobór właściwych grup pomiarowych jest trudnym wyzwaniem przy planowaniu badań uogólniających, gdyż respondenci muszą stanowić w miarę liczną, w pełni reprezentatywną próbę z populacji. W ewaluacji lokalnej nie ma problemu doboru podmiotów badawczych i prób, które w naturalny sposób powstają z grup ćwiczeniowych lub klas szkolnych.

➤ **Zasięg przestrzenny** ewaluacji oznacza wytyczenie obszaru eksploracyjno-interpretacyjnego. Z jednej strony stanowi opis terenu badań, zarówno w aspekcie wielkości i charakterystyki obszaru objętego próbkowaniem, jak też ze względu na wewnętrzną wnikliwość lub zewnętrzną ogólność wskaźnikowania. Z drugiej strony wyznacza ramy potencjalnego wnioskowania, prowadzącego do formułowania praw podstawowych lub do konkluzji jedynie okazjonalnych. Zasadniczym punktem planowania ewaluacji jest ustalenie, czy otrzymane rezultaty pomiarów mają służyć wyłącznie do oceny cech i zjawisk lokalnych, czy przede wszystkim do budowania teorii z wyprowadzeniem i dowodzeniem twierdzeń ogólnych. Założenia dotyczące zasięgu przestrzennego wpływają na inne elementy projektu badawczego, takie jak: dobór prób, techniki pomiaru i metody analizy. Zasięg przestrzenny to odpowiedź na pytanie: *Gdzie mierzymy i jak dalece uogólniamy?* W każdym rodzaju badań ustala się lokalizację, natomiast tylko w badaniach naukowych zmierza się do ujęcia globalnego.

➤ **Zakres czasowy** ewaluacji precyzuje zarówno kluczowe momenty, w których przeprowadza się dwukrotne próbkowanie, jak też definiuje formę wnioskowania co do aspektów czasowych. Jest odpowiedzią na pytanie: *Kiedy mierzymy i jak dalece prognozujemy?* W pomiarach jakości procesów edukacyjnych lokalizacja czasowa pomiarów ściśle wiąże się z tokiem oddziaływań. Edukacja jest procesem ciągłym, w którym równolegle z samokształceniem ustawicznym występują fazy oddziaływań instytucjonalnych o zmiennej intensywności. Z tego powodu wybór właściwych momentów próbkowania stanów chwilowych ma istotne znaczenie w prawidłowym odtworzeniu dynamiki procesów. Naturalnymi okresami pomiarów są cykle kształcenia, stąd najczęściej ewaluację przeprowadza się na początku oraz przy końcu roku szkolnego. W badaniach eksperymentalnych próbkuje się stany przed i po wprowadzeniu czynników oddziałujących na zmienną zależną. Odstęp czasu jest wówczas krótszy, gdyż często uwzględniać trzeba także zjawisko zapominania i osłabiania efektów oddziaływań z upływem czasu. Wprawdzie komponenty postaw są bardziej trwałe, lecz chęć ich rzetelnego ujawnienia zależy też od chwilowego nastroju.

Z wielu powodów wybór optymalnego dnia, a nawet pory dnia pomiaru jest istotny ze względu na solidność i szczerść wypowiedzi. Dotyczy to zwłaszcza drugiego pomiaru, szczególnie pod koniec roku szkolnego, gdy inne czynniki emocjonalne, nie związane z indykatum, mogą dominować. Dlatego w planach ustalających momenty próbkowania warto przewidzieć alternatywne terminy. Ponadto warto z góry powziąć decyzję, czy zamierza się przeprowadzić tylko jeden kompletny, dwupomiarowy proces ewaluacji, czy może będzie to szereg czasowy z wieloma pomiarami tych samych grup badawczych, względnie cykl kilkuletni z wymianą roczników respondentów należących do tej samej populacji. Ze względu na dynamikę procesów, badania cykliczne są najkorzystniejsze.

W opisie czasowego zakresu ewaluacji istotne jest ustalenie, czy ocenianie i wnioskowanie ma być tylko konkluzywne, czy także formatywne. Chodzi o to, czy rezultatem badań ma być diagnoza stanu dotychczasowego, czy również prognoza stanów przyszłych. Od tego zależy dobór odpowiednich wskaźników retro- i futurospekcyjnych oraz zastosowanie ewentualnych technik predykcyjnych. We wnioskach z każdego rodzaju ewaluacji konieczne jest opisanie i objaśnienie zaobserwowanych zjawisk empirycznych, natomiast dodatkowo w ewaluacji formatywnej konieczne jest przewidywanie i oszacowywanie, jak zmieniać się będą podmioty badań pod wpływem intencjonalnych oddziaływań. W założeniach precyzuje się zakres poszukiwania tego: *Jak jest?* względnie *Jak będzie?* lub *Jak powinno być?* W tym ostatnim przypadku niezbędnym elementem poszukiwań badawczych jest ustalenie, w jaki sposób to osiągnąć.

➤ **Narzędzia pomiarowe** to określenie odnoszone do zespołu instrumentów służących pobieraniu próbek empirycznych. W projektach ewaluacji powinny znaleźć się nie tylko puste nazwy narzędzi (skala, ankieta, kwestionariusz, arkusz obserwacji itp.), które są rozumiane zupełnie niejednoznacznie, lecz przede wszystkim podać należy: 1) założenia metodologiczne leżące u podstaw budowy modelu narzędzia, 2) pełne odwzorowanie formularza, jaki otrzymują respondenci, 3) kompletny schemat interpretacji wskaźników empirycznych. Nie wystarcza więc jedynie odpowiedź na pytanie: *Czym mierzymy?* Konieczne jest wyjaśnienie, na podstawie czego mamy prawo tak właśnie mierzyć oraz tak, a nie inaczej odczytywać symptomy. W pierwszym rzędzie ustala się, czy narzędzie ma być skalowane, czy jednak nie. W obu przypadkach uzasadnić trzeba dobór wskaźników, a przy skalach podać należy podstawowe kryterium kwantyfikacyjne. Następnie podejmuje się decyzję, czy i które wskaźniki będą addytywne, co oznacza możliwość ich agregowania i łącznego interpretowania. Jest to szczególnie istotne ze względu na fakt, że pojedynczy wskaźnik jest zbyt niepewnym probierzem zmiennej. Na kolejnym etapie ustala się formułę indagowania (np. pytania czy stwierdzenia) oraz formułę wypowiedzi (wybór opcji, grupowanie, rangowanie, szacowanie ocen itp. czy wypowiedzi otwarte).



W zależności od przyjętych formuł próbkowania i założeń co do addytywności, planuje się jednorodną bądź niejednorodną kategorię wypowiedzi, przy czym jednorodność wcale nie musi oznaczać identyczności. Zasadniczym kryterium jest to, czy dane empiryczne po wstępnym przetworzeniu (np. po nadaniu wag) można będzie uporządkować w macierzach i przetwarzać statystycznie w przekrojach przez kolumny i wiersze. Chociaż cechom, które posiadają bipolarną wartościowość zazwyczaj daje się przyporządkować *dyferencjał semantyczny* (opisowe, przysłówkowo-przymiotnikowe stopniowanie natężenia), to jednak dwie cechy wcale nie muszą być taksonomicznie addytywne. Ostatecznie więc o możliwości sumowania wskaźników decyduje model ekstrakcji indykatorów, a ściślej – zbudowany na jego podbudowie *schemat interpretacji* symptomów. Schemat ten powinien być opracowany jeszcze przed podjęciem pomiarów. Definiuje się w nim zmienne cząstkowe, określa strukturę powiązań składników addytywnych, ustala poziomy czynników niewspółbieżnych, wskazuje komponenty rezultatów połówkowych, par kontrolnych i komplementarnych. Bez schematu interpretacji narzędzie pomiarowe jest zupełnie bezużyteczne.

➤ **Strategia pomiaru** to określenie zawężone do sposobów pobierania próbek. Projektując ewaluację, z góry zakłada się realizację precyzyjnego scenariusza postępowania w każdej fazie procesu badawczego. Tym niemniej faza pomiaru wymaga najbardziej skodyfikowanego, schematycznego podejścia, zwłaszcza do eksperymentu. Z tego powodu w ogólnym projekcie ewaluacji definiuje się *plan badawczy* ze szczegółowym opisem procedur i schematem doboru prób. Najpierw należy podjąć decyzję co do rodzaju planu badawczego. Następnie z planu wyłania się wątki wymagające zastosowania ściśle określonych metod i technik. W przypadku planowania klasycznego eksperymentu strategicznymi wątkami są: manipulowanie, porównywanie, kontrolowanie i uogólnianie. Do realizacji strategii eksperymentu niezbędne jest pozyskanie minimum dwóch grup respondentów o podobnych cechach, co osiąga się poprzez różne techniki doboru prób. W oddziaływaniu czynnikiem eksperymentalnym zachowuje się sztywny podział na grupę eksperymentalną i kontrolną, względnie stosuje techniki krzyżowania grup. Do porównywania służą techniki statystyczne: estymacja, indeksacja, testy istotności różnic i relacji, natomiast do kontrolowania: techniki buforowe, testy pedagogiczne i psychologiczne. Uogólnianie uzależnione jest od doboru reprezentatywnych, wystarczających prób i od powtarzalności wyników, potwierdzonych wyznacznikami jakości pomiarowej. W przypadku badań quasi-eksperymentalnych stosuje się podobne, lecz mniej rygorystyczne procedury, przy czym w poszukiwaniu związków przyczynowo-skutkowych istotne są techniki pobudzania i ujawniania reakcji na bodźce. Najmniej sformalizowane są badania przekrojowe, podczas których główną rolę pełnią techniki diagnozowania osobistych dyspozycji podmiotów ewaluacji.

W badaniach panelowych najważniejszy jest pomiar różnicowo-dystansowy, połączony z dwukrotnym próbkowaniem stanów procesu i właściwości tych samych osób, przy czym ocenia się nie jednostki, lecz przemiany w czasie i różnice pomiędzy grupami. Podczas wartościowania wytworów medialnych stosuje się technikę porównywania obiektów podobnych z rotacją jurorów.

Opis strategii pomiaru przede wszystkim stanowi odpowiedź na pytanie: *Jak mierzymy?* Uzupełnia go jednak szczegółowe rozwinięcie dotyczące tego: *Jak obieramy próby i czy na nie oddziałujemy?* Od tego zależy nie tylko dobór technik samego pomiaru, lecz także role przypisane grupom badawczym oraz późniejsze metody analizy zasobów empirycznych i podstawy syntezy wyników.

➤ **Metody analizy** to zespół środków eksploracyjno-deskryptywnych, które wytyczają sposoby poszukiwania i opisywania oznak empirycznych. Już podczas projektowania ewaluacji przewidzieć należy, jak zamierza się przetwarzać dane. Analiza odnosi się przede wszystkim do szczegółowego poziomu oznak, a więc do zmiennych cząstkowych, powstałych z rozwarstwienia indykatum. Zasadniczym obszarem poszukiwań są przekroje statystyczne przez zmienne i przez przypadki, przy czym eksploracja dotyczy odkrywania i odtwarzania aspektów na równi ilościowych i jakościowych. Metody statystyczne dostarczają znaczący zasób parametrów, które wymagają opisanie wartościującego. Dzięki nim uzyskuje się liczbowe oszacowania poziomów, trendów i efektów, choć wartości ilościowe są w dużej mierze umowne. Metody taksonomiczne posiłkują się głównie ujęciem kategoryjnym-relacyjnym, dostarczającym wiedzy jakościowej o hierarchii i zależnościach pomiędzy komponentami indykatum. Opis metod analizy ewaluacyjnej sprowadza się do odpowiedzi na pytanie: *Czego i jak szukamy w zasobach danych, w jaki sposób to opisujemy?*

➤ **Metody syntezy** to zespół środków interpretacyjno-formatywnych, które wytyczają sposoby wnioskowania z uogólnianiem symptomów empirycznych. Synteza ewaluacyjna jest procesem generalizacji, a niekiedy także idealizacji. W każdym przypadku polega na agregowaniu wielości rozmaitych zjawisk wykrytych podczas badań, przy czym zakłada się kondensowanie wszystkiego lub abstrahowanie od tych oznak, których nie udało się wyjaśnić. Zależy to od celu badań i przeznaczenia wyników. Istotną rolę odgrywają teoretyczne podstawy uogólnień i metody weryfikacji jakości pomiarowej. Najczęściej wykorzystuje się techniki komparacyjne (porównawcze) oraz metody indeksacji. Rzadziej sięga się po teorie estymacji bądź predykcji. Wybór metod wnioskowania wpływa na sposób redakcji sprawozdań, toteż projekt ewaluacji winien zawierać opis tego, co będzie formą finalną badań. Chodzi o to, czy wnioski będą tylko diagnostyczne, czy może również formatywne, ze zbiorem zaleceń i standardów ewaluacyjnych. Opis metod syntezy wyników to odpowiedź na pytania: *Jak wyjaśniamy? Na podstawie czego uogólniamy? Co formułujemy?*

## **Jak dobrać próbę – przekroje statystyczne**

W strategii ewaluacji demokratycznej respondenci są upodmiotowieni tak dalece, że to właśnie ich wypowiedzi tworzą podstawy formowania norm wartościujących. Z tego powodu niezwykle istotny jest wybór osób reprezentujących badaną populację. Zbiorowość jednostek tworzących populację jest jednak całkowicie dynamiczna, zarówno pod względem ujęcia definicyjnego, jak i czysto technicznego. Po pierwsze – w zależności od przyjętych kryteriów podmiotowych i przestrzennych, ustalających zasięg eksploracyjny, populacja może oznaczać liczną zbiorowość globalną lub nieliczną lokalną. Po drugie – osoby desygnowane z definicji do danej populacji z upływem czasu przestają odpowiadać założonym kryteriom. Punktem wyjścia musi więc być powzięcie decyzji co do charakterystyki podmiotów tworzących populację oraz zasięgu przestrzennego i zakresu czasowego ewaluacji. Trzeba najpierw odpowiedzieć na pytanie odnoszące się do populacji: *Kogo, gdzie i kiedy zamierzamy badać?* Dopiero następnym krokiem jest ustalenie strategii: *W jaki sposób wyłonić reprezentatywne próby poddawane pomiarowi?* oraz *Jak liczne będą próby?*

Charakterystykę podmiotów ewaluacji, opis miejsca i czasu pomiarów umieszcza się w projekcie badawczym, którego sposób konstruowania przedstawiłem w poprzednim rozdziale. Przypomnę, że populację tworzyć powinny zbiorowości w podmiotowy sposób związane z ewaluowanym zjawiskiem lub procesem. Zaleca się, ażeby badania przeprowadzać z rozwarstwieniem na w miarę jednorodne grupy społeczne, zwłaszcza wtedy, gdy mierzone są cechy osobowe. Istotne jest wówczas określenie takich swoistych cech badanej populacji, jak: przedział wieku, poziom wykształcenia, pochodzenie społeczne, profil zawodowy i inne tego rodzaju właściwości możliwe do wyrażenia jako uogólnione znamiona, wspólne dla zbiorowości będącej obiektem ewaluacji. Zbiorowości te funkcjonują w określonych środowiskach społecznych, różnie oddziałujących na jednostki i odmiennie kształtujących profile populacyjne, dlatego potrzebne jest określenie zasięgu przestrzennego z umiejscowieniem badań oraz wyznaczeniem obszaru dopuszczalnych uogólnień. Ze względu na dynamikę procesów konieczne jest ustalenie zakresu czasowego pobierania próbek, z ewentualnym wytyczeniem planu prognozowania tendencji.

W zależności od powyższych założeń oraz od celów i problemów ewaluacji, wybór odpowiednich prób badawczych jest zadaniem łatwym bądź trudnym. W najprostszym przypadku, gdy chodzi wyłącznie o pomiary lokalne bez uogólnień wychodzących poza te ramy, zespół respondentów jest tożsamy z ogółem osób, które zamierzaliśmy badać. Jeśli jednak planujemy porównywanie grup, to grupy te winny być w miarę podobne w aspekcie szerszym niż indykatum. Jeśli wyraźnie różnią się innymi cechami, niebędącymi obiektem pomiaru, to należy ten fakt wziąć pod uwagę podczas wnioskowania i w sprawozdaniu.

Sprawa doboru prób komplikuje się, gdy planujemy badania z uogólnianiem wniosków na populację. Respondenci są wówczas jedynie przedstawicielami większej zbiorowości, nie mogą więc być desygnowani zupełnie przypadkowo. Aby z pomiarów można było wysnuć trafne wnioski, niezbędne jest właściwe wyłonienie prób odzwierciedlających właściwości populacji. Z jednej strony chodzi o reprezentatywność<sup>37</sup>, lecz z drugiej o optymalną liczbę przypadków. Techniki wyłaniania prób są szczegółowo opisywane w literaturze dotyczącej metodologii badań, dlatego podam jedynie krótkie ich charakterystyki.

Najczęściej zalecanym sposobem wyboru respondentów jest losowanie. Niestety, milczeniem pomija się fakt, że ma ono uzasadnienie tylko w sytuacji, gdy operuje się bardzo dużymi liczebnie populacjami i próbami. Tymczasem niezbędne do losowania pełne listy dużych populacji są trudne do zdobycia. Z tego powodu częściej stosuje się strategie mieszane, łączące dobór losowy z doborem celowym. Aczkolwiek wybór celowy obarczony jest subiektywnym piętnem decyzji badacza, to jednak kierują nim bardziej racjonalne przesłanki niż czysto losowa taktyka *na chybił trafił*. W strategii celowego doboru prób ustala się jawne kryterium selekcyjne, które w jakikolwiek ukryty sposób nie powinno wpływać na ulepszanie wyników, natomiast potencjalnie powinno prowadzić do uzyskania wyników bardziej rzetelnych. Kryteria te najczęściej dotyczą rozwarstwienia populacji na mniejsze zbiorowości, które bez potrzeby dokonania pomiaru można pogrupować ze względu na znane wcześniej cechy powiązane istotnie z indykatorami. Z warstw tych wybiera się lub losuje reprezentacje pojedyncze, bądź też najlepiej w proporcjach takich, jak ich udział w populacji. Innym kryterium wyłaniania jest wybór lub losowanie jedynie wewnątrz tych grup, których listy są dostępne, a pomiar fizycznie jest możliwy. Jeśli łączne cechy grup odpowiadają charakterystyce populacji, to możliwe jest uogólnianie wyników, w przeciwnym wypadku zawęzić należy przestrzeń definicyjną badanej populacji. To samo uwarunkowanie dotyczy oddolnego wyłaniania prób reprezentatywnych poprzez odwzorowanie struktur potencjalnie najlepiej odpowiadających właściwościom i proporcjom ogółu.

Obok reprezentatywności istotna jest także liczebność prób badawczych. Mniej lub bardziej obostrzone rygory doboru minimalnej liczby respondentów wynikają przede wszystkim z zaplanowanych metod analizy statystycznej oraz z celów ewaluacji. Przy zamiarze uogólniania wniosków na populację stosuje się metody szacowania błędów lub szacowania ufności co do wyniku. Wraz ze zwiększeniem liczby próbek maleje prawdopodobieństwo popełnienia błędu i wzrasta pewność co do dokładności parametrów. Można wprowadzić z góry przyjąć pewien poziom istotności i określić wymagane liczebności, lecz zależą one od rozkładów empirycznych, a te różnią się znacznie od idei teoretycznych. Z tego powodu zasadnicze ustalanie jakości pomiarowej oraz wiarygodności

wyników przeprowadza się dopiero po zebraniu danych. Generalną zasadą jest pobieranie maksymalnie dostępnej liczby próbek, ograniczonej jedynie względami czysto technicznymi i ekonomicznymi. Jakikolwiek zalecenia co do minimalnych liczebności są jedynie sugestiami odnoszonymi do konkretnej sytuacji badawczo-interpretacyjnej. I tak, ufność do wyników zależy od właściwości konkretnej skali pomiarowej, a konkretnie od jej rozdzielczości (zobacz tabele na str. 54). Większość testów istotności można użyć wyłącznie przy próbach liczących kilkaset przypadków, a tylko nieliczne przy kilkudziesięciu. Jednak nawet duża liczebność nic nie pomoże, jeśli nie wystąpi zróżnicowanie opinii respondentów, a zaplanuje się analizę rozkładową. Uwarunkowania te powodują, że trudno zalecić jakiś jeden optymalny schemat doboru prób.

Zasadniczym celem ewaluacji jest jej praktyczna i lokalna użyteczność, a w takim wypadku najważniejsze jest posługiwanie się wieloma próbami, nie tylko celowymi, lecz także okazjonalnymi. Z prób tych nie można wnioskować o cechach populacji, lecz przy dużej liczności i różnorodności środowisk można na ich podstawie szacować ogólnospołeczne standardy ewaluacyjne. Kolejne próby okazjonalne uaktualniają normy oraz wzbogacają wiedzę o tendencjach dzięki porównywaniu z tym, co zmierzono wcześniej. Wartość każdego rodzaju prób ma swe podłoże w opisie charakteryzującym respondentów, zależy jednak głównie od możliwości grupowania bądź rozwarstwiania ze względu na cechy badawczo istotne. O ile więc w badaniach podstawowych najistotniejsza jest reprezentatywność próby względem populacji, o tyle w ewaluacji praktycznej cenniejsza jest różnorodność prób i wielość grup. Osiąga się to w dwojaki sposób: przeprowadzając pomiary w wielu środowiskach oraz rozwarstwiając ogół badanych na mniejsze podzbiory statystyczne, zwane *przekrojami*.

➤ **Przekroje statystyczne** tworzyć można jako rozwarstwienie ogółu badanych na podstawie z góry przyjętych założeń, bądź też na podstawie kryteriów ustalonych dopiero dzięki pomiarowi. W pierwszym przypadku podział wynika ze znanych wcześniej takich charakterystyk różnicujących, jak np.: teren badań, środowisko, profile szkół, specjalizacje grup lub klas, odmiana realizowanego programu, formy organizacji zajęć, osoba wykładowcy itp. Są to więc wszelkie kryteria grupujące, niezbędne do prawidłowej interpretacji rezultatów. W drugim przypadku wykorzystuje się kwantyfikatory filtrujące, które służą do pobierania ze zbioru danych tylko tych frakcji, które spełniają określone kryterium co do zmierzonej wartości zmiennych. Kwantyfikatorem filtrującym może być dowolna zmienna cząstkowa, której przypisana zostanie funkcja odsiewania lub grupowania przypadków. Co więcej – funkcję tę pełnić mogą zarówno dowolne pozycje skali (np. wskaźnik poziomu motywacji), jak też zmienne zawarte w metryczce narzędzia (np. płeć lub wiek respondentów). Tego rodzaju dynamiczny podział na frakcje wzbogaca proces wnioskowania.

Na potrzeby praktyki ewaluacyjnej proponuję umowną konwencję podziału na przekroje statystyczne oraz nazewnictwo prób badawczych w zależności od zasięgu próbkowania. Podział taki jest strukturą hierarchiczną, w której ogół respondentów rozwarstwiamy kolejno na próby mniejsze: grupy i klasy. W każdym przekroju zalecane są pewne optymalne i minimalne liczebności, a ponadto przyjmuje się jednolite, zwarte sposoby kodowania zbiorów:

➤ **Ogół** – to zbiór wszystkich dwupomiarowych prób pobranych z badanej populacji (np. studentów pierwszego roku), a także z pomiarów wieloletnich (np. kolejnych pierwszych roczników). Optymalna liczebność to kilkaset osób, a minimalna 90, koniecznie z podziałem na co najmniej 3 grupy. Próby mniej liczne należy scalać i analizować razem z danymi innych badaczy. W ramach próby zwanej *ogółem* nie należy mieszać różnych populacji, gdyż właśnie ten przekrój jest podstawą wyznaczania standardów ewaluacyjnych.

➤ **Grupa** – to próba pobrana w pełnym cyklu dwupomiarowym ze środowiska w miarę jednorodnego (np. szkoła, profil, kierunek, specjalność, rocznik itp.). Optimum liczebności wynosi 50÷80, minimum 30 osób, w tym koniecznie muszą być zawarte przynajmniej 2 klasy. Każdą z grup kodujemy jedną dużą literą (A, B, ... Z) lub jedną cyfrą (1, 2, ... 9). Kodowanie potrzebne jest do zachowania dla osób trzecich niejawności źródeł pochodzenia próby, a także do zautomatyzowanego wyodrębniania przekrojów podczas analizy.

➤ **Klasa** – to próba homogeniczna, pochodząca od respondentów poddanych identycznemu procesowi oddziaływania (np. wspólne zajęcia, ten sam program i wykładowca). Optimum liczebności wynosi 25÷40, minimum 15 osób. Każdą z klas oznaczamy dużą literą lub cyfrą tej grupy, do której dana klasa należy, oraz małą literą przypisaną klasie (Aa, Ab, Ac, Ba, Bb, ... Zz).

W konkretnym przypadku badania jakości procesów edukacyjnych, dotyczących informatyzacji, podmiotami ewaluacji muszą być uczestnicy zajęć komputerowych. Pomiary *InfoKultury* można przeprowadzać już od drugiego poziomu nauczania, począwszy od czwartych klas szkół podstawowych, poprzez gimnazja i licea, aż po studia wyższe i podyplomowe. Jednakże muszą to być kształcenia ogólnoinformatyczne, a nie kursy specjalistyczne.

<i>Ślady</i>		<i>Wzorce</i>
operacjonalizacja zadań	.....	cele i problemy ewaluacji
obiekty i próby badawcze	.....	przedmiot i podmioty ewaluacji
teren, okres i cykl badań	.....	zasięg przestrzenny i zakres czasowy
instrumenty i plan badań	.....	narzędzia i strategia pomiaru
eksploracja i interpretacja	.....	metody analizy i syntezy
dobór respondentów	.....	reprezentatywność i liczebność prób
przekroje statystyczne	.....	populacja, ogół, grupy, klasy

## **Jak utworzyć narzędzie – budowa kwestionariusza**

Opracowanie narzędzia pomiarowego o odpowiedniej jakości jest trudnym zadaniem. Nie ma żadnych uniwersalnie skutecznych recept na tę dolegliwość. Narzędzie musi być spoiwem obranej metodologii badawczej, zintegrowanej z celami i problemami merytorycznymi, dlatego opis sposobu konstruowania narzędzia ograniczam do studium konkretnej implementacji *Kwestionariusza Kultury Informatycznej*, zbudowanego na wzorcu *Skali dwuważonych ocen*. Niezależnie od wzorca – każde narzędzie badawcze musi trafnie próbować indykatum, a poruszane kwestie powinny być percepcyjnie i predykcyjnie dopasowane do szerokiego spektrum potencjalnych podmiotów ewaluacji. Z tego względu zdania tworzące pozycje kwestionariusza należy redagować pod kątem wysokiej komunikatywności, a co za tym idzie – w formie prostych językowo wyrażen. Pod tą pozornie błahą osnową kryją się jednak istotne wątki treściowe. Każde ze zdań pełni funkcję pobudzania respondentów tezą ważną eksploracyjnie, z jednoczesnym próbkowaniem reakcji na ten bodziec. Klucz do sukcesu badawczego tkwi właśnie w optymalnym doborze i redakcji owych tez, formułowanych w postaci bodźców-stwierdzeń. Prześledźmy proces tworzenia narzędzia skalowanego, budowanego na założeniach metodologicznych, przedstawionych szczegółowo w części teoretycznej.

➤ **Definiowanie i ekstrakcja indykatum** jest fazą koncepcyjnego opracowania struktur pojęciowych, precyzujących komponenty tego, co ma być mierzone, lub precyzyjniej – co będzie wskazywane. Przypomnę, że indykatum wyraża zmienną ogólną, mierzoną za pomocą skali (np. cechę zwaną *InfoKulturą*), natomiast przedmiot ewaluacji może być całkiem inny, wyznaczany pośrednio (np. *jakość zajęć informatycznych*). W pierwszym rzędzie należy więc nadać zwartą nazwę indykatum oraz zdefiniować sens pojęciowy i/lub operacyjny. W uproszczeniu – *InfoKulturę* zdefiniowałem jako splot cech mentalnych: stanów uświadomienia i postaw wobec informatyzacji, co oznacza, że są to właściwości osobnicze wymagające budowy teoretycznego modelu struktury i pomiaru za pomocą zoperacjonalizowanych wskaźników pośrednich. W różnych psychologicznych koncepcjach umysłu w miarę zgodna jest tendencja do wyodrębniania struktur afektywnych i kognitywnych. Stanowiło to podstawę do ekstrakcji dwóch głównych komponentów indykatum, nazwanych przeze mnie *Doznaniem* i *Poznaniem*. Te dwie zmienne globalne, możliwe do interpretacji jako wyniki połówkowe lub jako czynniki główne, wymagały zdefiniowania zawężonego tylko do kontekstu *InfoKultury*. Na dalszym etapie ekstrakcji wybiera się czynniki mające najbardziej istotny wpływ na poziom indykatum. Ze względu na bardzo ograniczoną liczbę czynników, w dużym stopniu jest to wybór arbitralny, zależny od twórcy skali, chociaż podstawą są również teorie wywodzące się z nauk społecznych. Jako czynniki wytyczające komponenty

afektywne zdecydowałem się przyjąć *Emocje*, *Motywacje* i *Opinie*. Dwie pierwsze nazwy własne odzwierciedlają pełnione role czynników wyznaczających cechy mentalne i jako istotne nie wymagają uzasadnienia wyboru. Natomiast nazwę *Opinie* niejako z konieczności nadałem równie ważnym komponentom postaw, dotyczącym emocjonalnego oceniania. Do czynników kognitywnych desygnowałem zmienne o nazwach: *Poglądy*, *Wprawa* i *Rozwaga*. Pierwsza z nich dotyczy wyrażania wnikliwych ocen, opartych nie na emocjach, lecz na racjonalnych przesłankach. Druga reprezentuje poziom doświadczeń, a trzecia wyraża roztropność w odniesieniu do zachowań przy komputerze.

Kolejnym etapem ekstrakcji jest rozwarstwienie czynników na składniki, przy czym także i w tym wypadku konieczność redukcji liczby komponentów zmusza do wyboru kwestii jedynie najbardziej istotnych. Tym razem większe znaczenie mają cele praktyczne ewaluacji i zadania badawcze, chociaż cechy wybrane jako składowe reprezentatywne muszą taksonomicznie mieścić się w strukturach kategorii nadrzędnych. Przykładowo składnik o nazwie *Ocena*, próbujący jakość zajęć informatycznych, jest komponentem czynnika *Opinie* oraz elementem połówkowego zbioru wyznaczników afektywnych pod nazwą *Doznanie*. Zwięzłe nazewnictwo zmiennych jest wysoce umowne, pełni jednak ważną rolę w ewaluacji. Oczywiście każda ze zmiennych w procesie ekstrakcji musi być zdefiniowana znacznie precyzyjniej, również z opisem wzajemnych relacji. Zagadnienie to przedstawiłem w kolejnej części książki, dlatego tutaj każdy z elementów ekstrakcji objaśniam jedynie ogólnikowo. Pragnę jednak w tym miejscu zwrócić uwagę na trudności z nadawaniem nazw. Otóż często brakuje odpowiednich słów na wyrażenie funkcji wszystkich zmiennych cząstkowych. Na przykład bardzo ważną strukturą postaw jest funkcja oceniania, jeśli jednak nazwę *Ocena* nada się jednemu ze składników, to już czynnikowi nadrzędnemu trzeba nadać inną nazwę (*Opinie*), choć może ona nie być optymalna. Podobnie jest z parami składników komplementarnych, tworzących czynniki. Może budzić zdziwienie nazwa *Aplauz* występująca w parze z *Oceną*. Słowo to jednak dość trafnie oddaje nadzwyczaj afektywny sposób wyrażania zachwyty wobec zjawiska, jakim powszechnie obdarzana jest komputeryzacja. Podobnie jak aplauz na widowni wybrzmiewa po pewnym czasie, tak samo wskaźniki nadmiernie wysokiej aprobaty komputerów spadają po zgłębieniu tajników sztuki informatycznej, przechodząc w bardziej racjonalne *Poglądy* oraz w krytyczny *Osąd*. Podczas ekstrakcji szukajmy więc optymalnych nazw dla zmiennych, nie bojąc się przenośni i daleko idących uproszczeń.

Zauważmy w powyższym przykładzie, że między komponentami indykatum zachodzą pewne zależności strukturalne (sieć powiązań) a także dynamiczne, kiedy to wartość jednego z czynników lub składników maleje, a innego rośnie. Oprócz par komplementarnych, w procesie ekstrakcji poszukuje się również



par relacyjnych. Mogą one sprzęgać ze sobą dwa komponenty przynależne do różnych połówek skali, choćby czynniki doznaniowe z poznawczymi. Oprócz schematu ekstrakcji, opisującego strukturę rozwarstwienia zmiennej ogólnej, już na etapie tworzenia narzędzia niezbędne jest konstruowanie modelu sieci potencjalnych powiązań, umożliwiającego oszacowanie trafności i rzetelności pomiaru. Tego typu model stanowi rozwinięcie schematu interpretacji.

Wróćmy do ekstrakcji zmiennych. W konstruowaniu kwestionariusza założyłem, że przeciw wagę dla ocen afektywnych powinny stanowić komponenty kognitywne o nazwach *Osąd* i *Wgląd*. Pierwszy z tych składników odnosi się do poglądów w sprawie przydatności edukacyjnej i użyteczności komputera w preferowanych dziedzinach. Drugi składnik weryfikuje poziom świadomości co do możliwych zastosowań. Jako reprezentację motywacji postanowiłem przyjmując *Ambicje*, oznaczające gotowość do uczenia się oraz *Intencje*, wyrażające zamiar przeznaczania czasu na pracę z komputerem. Poziom motywacji zależy od przestrzeni, jaka dzieli aktualny stopień biegłości od wyznaczonego sobie indywidualnie pułapu osiągnięć w tym zakresie. Dlatego ważne jest ustalenie poziomu składników kognitywnych w ujęciu retro- i futurospekcyjnym. Stan dotychczasowych doświadczeń wyraża zmienna o nazwie *Obycie*, natomiast przeświadczenie o możliwości nauczenia się obsługi komputera w przyszłości weryfikuje składnik komplementarny o nazwie *Pewność*. W pomiarach postaw istotne jest badanie komponentów emocjonalnych. Parę składników, którym powierzyłem tę rolę, wybrałem na podstawie wcześniejszych obserwacji zjawisk towarzyszących komputeryzacji. U jednych osób chodzi o stres, u innych o uzależnienia. Są to oznaki negatywne, natomiast w ewaluacji poszukuje się wartości, dlatego nazwy przyjęte dla składników wyrażają pozytywność, osiąganą poprzez odwrócenie polaryzacji. Tak więc *Spokój* jest zmienną, która oznacza bezstresowość lub niski poziom niepokoju, natomiast *Odpór* to zmienna, która wyraża niepodatność na uzależnienie i odporność na zniewolenie przez komputer. Ostatnią parę składników stanowią ważne komponenty kognitywne. Również i one wybrane zostały pod wpływem zjawisk zaobserwowanych podczas prowadzenia zajęć komputerowych. Wśród słuchaczy dorosłych trafia się na dość liczne przypadki braku wiary we własne możliwości samokształcenia w dziedzinie technologii informacyjnych, natomiast u najmłodszych adeptów komputera ujawniają się działania impulsywne i nierozważne, prowadzące do uszkodzeń systemów bądź utraty ważnych plików. Wprowadziłem zatem do skali składnik o nazwie *Zdolność*, wyrażający samoocenę własnego potencjału, oraz składnik *Ogląda*, sprawdzający umiejętność pohamowania nadmiernego pobudzenia i swoistego wchłaniania przez komputer. Są to typowe symptomy świadomości informatycznej. Wszystkie zmienne cząstkowe, oprócz nazwania i zdefiniowania, muszą dać się wyrazić za pomocą empirycznych wskaźników.

➤ **Dobór i weryfikacja wskaźników** jest fazą operacjonalizacji zmiennych elementarnych, polegającą na przejściu od definicji pojęciowych do reprezentacji werbalnych, wyrażających dyspozycje do zachowań odzwierciedlających badane cechy mentalne. W fazie tej poszukuje się specyficznych pośredników pomiędzy właściwościami niemierzalnymi wprost a symptomami wyrażanymi deklaratywnie. W narzędziu skalowanym zarówno reprezentacje werbalne są ściśle skodyfikowane, jak też potencjalne deklaracje są skwantyzowane, co oznacza, że treści wypowiedzi respondentów są z góry narzucone, a ponadto poziomy akceptacji bądź odrzucenia mogą być wybierane jedynie ze stałego zbioru jednolitych opcji. W konkretnym przypadku budowy *Kwestionariusza Kultury Informatycznej* przyjąłem schemat poziomów ekstrakcji 2-3-2-2, co daje łączną liczbę 24 wskaźników. Pozycje kwestionariusza stanowią celowo i mozolnie dobierane zdania, które – poprzez swą treść i ładunek stymulujący – pobudzają respondentów do wypowiedzi z określonym natężeniem akceptacji. Progi natężeń wyznaczone są kategoriami możliwych wypowiedzi (zwanymi kwantyfikatorami kwantyzacyjnymi), przy czym w preferowanym tutaj wzorcu *Skali dwuważonych ocen* zalecam stosowanie siedmiu następujących opcji:

absolutnie nie	nie	raczej nie	brak zdania	raczej tak	tak	absolutnie tak
----------------	-----	------------	-------------	------------	-----	----------------

Przypomnę, że przyjęliśmy szeroko rozumiane pojęcie wskaźników w funkcji zarówno próbników, jak i wyznaczników. W tej części opisu skoncentrujemy się na aspektach próbkowania. Dobór wskaźników w praktyce polega na ułożeniu treści takich bodźców-stwierdzeń, które: 1) wyrażają tezę lub antytezę istotną w indeksacji zmiennej ogólnej; 2) próbują zmienne cząstkowe zdefiniowane w procesie ekstrakcji; 3) testują aspekty czasowe i przestrzenne symptomów; 4) umożliwiają kontrolę par zmiennych powiązanych wzajemnymi relacjami.

Studium działań nad doбором wskaźników do *Kwestionariusza Kultury Informatycznej* przedstawię w układzie zgrupowanym wokół sześciu czynników. Opiszę skrótowo funkcje, jakie były oczekiwane od każdego ze wskaźników, oraz tło, stanowiące przesłanki wyboru takiej, a nie innej materii próbkowania. Podam też treść pierwotnych bodźce-stwierdzeń, które nie sprawdziły się empirycznie, dlatego zamieszczam je w nawiasach kątowych  $\langle \rangle$ , natomiast strzałką  $\Rightarrow$  oznaczam stwierdzenia ostatecznie przyjęte do kwestionariusza.

Pierwszy czynnik, jaki zamierzałem zbadać, dotyczył afektywnych opinii wyrażanych wobec zjawiska informatyzacji. Zasadniczym celem zastosowania narzędzia miała być ewaluacja kursów komputerowych, toteż do skali należało wmontować wskaźnik próbujący jakość zajęć. Okazało się, że trudno jest sformułować stwierdzenie, które poruszałoby tę delikatną kwestię w sposób pośredni. Chodziło o to, ażeby nie pytać wprost o jakość kursu. Żaden jednak z kolejnych pomysłów nie sprawdzał się empirycznie. Kardynalnym błędem

dwóch pierwszych sformułowań było to, że odnosiły się one tylko do nielicznej grupy respondentów negatywnie nastawionych do komputerów. Tymczasem bodźce-stwierdzenia powinny sytuacyjnie przystawać do większości badanych.

⟨Ten kurs pomoże mi pokonać moje uprzedzenie do komputerów.⟩

⟨Ten kurs pomoże mi poprawić moje poglądy wobec komputerów.⟩

⟨Kurs pomoże mi w innych zajęciach wymagających posługiwania się komputerem.⟩

Trzecie zdanie spełniałoby ten warunek, gdyby nie fakt, że wiele osób wciąż nie dostrzega roli komputera na innych zajęciach poza informatyką. Jedynym dobrym rozwiązaniem okazało się indagowanie w miarę bezpośrednie, z jawną sugestią poddającą w wątpliwość szeroko rozumianą przydatność kursu:

⇒ Wątpię w przydatność takich kursów komputerowych.

Przyjęty bodziec jest introspekcją spolaryzowaną ujemnie, co oznacza, że respondent wyraża swe własne odczucia, a wartością jest zaprzeczenie tezy.

Do kontrolnej pary z powyższym bodźcem dołączyłem stwierdzenie dotyczące oceny tego, czy dla zaawansowanych słuchaczy kurs był też korzystny:

⟨Nawet studenci obcy z komputerem wyniosą z tego kursu coś pożytecznego.⟩

⇒ Nawet obcy z komputerem odniosą z tego kursu pożytek.

Tak sformułowany bodziec jest ekstraspekcją spolaryzowaną dodatnio, co znaczy, że respondent wypowiada się o innych, a wartością jest potwierdzenie tezy. Po usunięciu słowa ograniczającego krąg badanych tylko do studentów i po drobnej korekcie redakcyjnej, wskaźnik ten w pełni sprawdził się empirycznie. Parę próbkującą jakość kursu celowo umieściłem w środkowej części kwestionariusza, ażeby respondent nie rozpoczął od oceniania tak bezpośredniego.

Na pierwszą pozycję w kwestionariuszu zaplanowałem bodziec o dużym ładunku afektywnym, umożliwiającym respondentom pełne uzewnętrznienie ewentualnych negatywnych postaw. Ma to duże znaczenie dla zjednania tych, którzy są nieprzychylni komputeryzacji i przez to mogliby być też niechętni samej ewaluacji. Dwa pierwotne sformułowania nie sprawdziły się w praktyce. Prawdopodobną przyczyną złego rozkładu empirycznego były niuanse semantyczne, a mianowicie – uprzedzenie i nastawienie są ocenami wyprzedzającymi, z góry zakładającymi negację, nawet jeszcze przed poznaniem zjawiska.

⟨Jestem uprzedzony(a) do komputerów.⟩

⟨Jestem niechętnie nastawiony(a) do komputerów.⟩

Znacznie lepszym okazał się wskaźnik eksplikacyjny, próbkujący stan bieżący, przy czym zmieniłem obiekt oceny: z komputerów na proces komputeryzacji. Korzystne było też wyeliminowanie alternatywnych końcówek rodzajowych.

⇒ Odczuwam raczej niechęć do komputeryzacji.

Do pary kontrolnej włączyłem stwierdzenie testujące aspekt globalny, jednak pierwsza z wcześniejszych wersji była zbyt zawężona, a druga zbyt nieprecyzyjna znaczeniowo. Dopiero trzecia wersja okazała się treściowo optymalna.

⟨Tylko ludzie bardzo inteligentni posługują się komputerami.⟩

⟨Tylko bystrzy ludzie posługują się komputerem.⟩

⇒ Tylko uzdolnieni do komputera skorzystają z jego zalet.

Obydwa wskaźniki tworzące parę są spolaryzowane ujemnie, co jest w pełni prawidłowe, ponieważ nie powinno być żadnego schematyzmu w polaryzacji bodźców-stwierdzeń, a więc nie może też być automatyzmu typu polaryzacji naprzemiennej. Wszystkie cztery przedstawione dotychczas wskaźniki tworzą czynnik o nazwie *Opinie*, będący afektywnym komponentem postaw.

Drugim czynnikiem doznaniowym, istotnym w procesach edukacyjnych, są *Motywacje*. Z teorii znane jest zjawisko cykliczności pobudzenia i nasilenia tych dyspozycji, dlatego poziom motywacji jest niezwykle ważnym kryterium oceniania wartościującego. Jednak ze względu na dużą zmienność konieczne jest każdorazowe odnoszenie poziomu chwilowego dyspozycji szczegółowej do poziomu cechy ogólniejszej, o wiele bardziej stałej. Tak powstał zamysł, ażeby jeden ze wskaźników motywacji był jednocześnie wskaźnikiem uniwersalnym, buforowym dla innych skal i potrzebnym do badania homogeniczności prób. Bodziec-stwierdzenie o takiej funkcji nie może dotyczyć indykatum, toteż nie ma w nim mowy o komputerach ani o informatyce. Pierwotna treść stwierdzenia była zbyt łatwo akceptowana i dopiero bardziej rygorystyczne sformułowanie przyniosło wzorcowy rozkład, wymagany od wskaźników buforowych.

⟨Jestem zawsze gotowy(a) uczyć się nowych rzeczy.⟩

⇒ Zawsze wyrażam gotowość uczenia się trudnych rzeczy.

Po weryfikacji na bardzo licznych próbach badawczych okazało się, że w ten sposób zredagowane stwierdzenie jest znakomitym próbnikiem różnicującym przypadki, natomiast poziom tej cechy jest niemal stały, mimo upływu czasu.

Paralelnym wskaźnikiem, tym razem ściśle związanym z indykatum, jest pozycja kwestionariusza testująca motywacje do uczenia się o komputerach. Kolejne wersje bodźca-stwierdzenia różniły się tylko nieznacznie, przy czym pierwsze zdanie było powszechnie akceptowane na zbyt wysokim poziomie.

⟨Chciał(a)bym nauczyć się czegoś więcej na temat komputerów.⟩

⟨Chciał(a)bym nauczyć się znacznie więcej na temat komputerów.⟩

⇒ Chcę nauczyć się znacznie więcej o komputerach.

Przykład ten jest ilustracją, jak nawet niewielkie zmiany redakcyjne – choćby nieprecyzyjne treściowo stopniowanie – służą tzw. *justowaniu estymatorów*. W odróżnieniu od globalnego wskaźnika buforowego, lokalny wskaźnik motywacji wykazuje znaczne wahania, zależne od środowisk i od sezonu szkolnego. Wiąże się to bardziej z fazami zaspokajania ambicji, a mniej z jakością zajęć.

Kolejne wskaźniki motywacji to dyspozycje dotyczące przeznaczania czasu na pracę przy komputerze. Mimo że weryfikacja bodźca-stwierdzenia wypadła pomyślnie, to jednak podejrzewam możliwość występowania błędu

nieliniowej zależności pomiędzy natężeniem wypowiedzi a natężeniem cechy. Przeanalizujemy poniższe stwierdzenie pod kątem potencjalnych sytuacji:

⇒ Chcę więcej czasu przeznaczyć na pracę z komputerem.

Otóż z roku na rok przybywa osób, które spędzają przy komputerze coraz więcej godzin. Może to być czas wykorzystany z pożytkiem, lecz często jest po prostu marnotrawiony. Wśród respondentów mogą być takie osoby, które już i tak zbyt długo przesiadują przy komputerze. W ich sytuacji wartością byłoby zaprzeczenie. Jednakże dla większości respondentów nadal wartością jest chęć intensywniejszego poznania komputera, więc nie można zmieniać polaryzacji. Nie ma pewności, czy wskaźnik jest monotoniczny, jednak postanowiłem go pozostawić w skali, gdyż zmiany motywacji mają również podobny przebieg – po zaspokojeniu aspiracji gotowość do działań w naturalny sposób maleje. Częściowym wyjaśnieniem stanu faktycznego jest wnioskowanie z wzajemnej relacji pomiędzy wypowiedziami wobec pary kontrolnej, ponieważ wskaźnik paralelny do poprzedniego pozostaje z nim w swoistym dysonansie.

⟨Myślę, że komputery nie oszczędzają czasu.⟩

⇒ W przeszłości komputery pomogą mi zaoszczędzić czas.

Nawet jeśli aktualnie respondenci wyrażają potrzebę pewnej „straty czasu”, to jednak przewidują pozytywne tego efekty w przyszłości. Jest to znamieny przykład wskaźnika futurospekcyjnego, oznaczającego prognozowanie. Tego typu trudniejsze do oceny tezy warto umieszczać przy końcu kwestionariusza.

Jako trzeci istotny czynnik doznaniowy obrałem *Emocje*. Komputer jest medium silnie pobudzającym, zazwyczaj dostarczającym odczucia pozytywne. Jeśli jednak pojawiają się doznania negatywne, to wymagają one rozpoznania i przeciwdziałania, dlatego w kwestionariuszu umieściłem wyłącznie wskaźniki zjawisk niekorzystnych. Pierwotnym moim zamiarem było sprawdzenie, czy respondenci potwierdzają wartość tezy o lepszych efektach uczenia się dwóch osób przy jednym komputerze. Okazało się jednak, że stwierdzenie:

⟨Uczyć się komputera milej parami, lecz używać go trzeba samotnie.⟩

pobudziło frywolne skojarzenia, wyrażane dopiskiem '*niestety tak*' i emotikoną roześmianej buzi. Zmieniłem zatem całkowicie treść bodźca-stwierdzenia:

⇒ Bał(a)bym się zostać sam na sam z komputerem.

To stwierdzenie również może rozbawić, lecz jest niezwykle czułym próbnikiem takich stanów emocjonalnych, kiedy to użytkownik komputera odczuwa silny stres, jeśli tylko nie ma w pobliżu osoby mogącej pomóc w razie trudności. Co ciekawe, stres ten może występować nawet wtedy, gdy nie ma żadnej awarii. W wielu przypadkach obecność osoby wspierającej jest pożądana, jednak są też takie osoby, które czują się zażenowane brakiem wystarczającej wiedzy. Konieczne było dodanie wskaźnika kontrolującego relacje wewnątrzgrupowe.

⟨Czuje się nieswojo, gdy widzę, że inni wiedzą więcej niż ja na temat komputerów.⟩

⟨Czuje się nieswojo, gdy widzę, że inni o komputerach wiedzą więcej ode mnie.⟩

⇒ Jest mi źle w grupie, gdy inni lepiej znają komputery.

Zmiana treści wynikała jedynie z poszukiwania bardziej zwartej formy zdania, przy zachowaniu tych samych właściwości rozkładu empirycznego.

W dobie nasilających się uzależnień od komputera, a zwłaszcza od gier i od Internetu, bardzo ważne jest wykrywanie tych niepokojących symptomów. Wprowadzone pierwotnie zdanie (z porównaniem do narkotyków) okazało się zbyt różnicujące w zależności od środowiska. Wbrew zamiarom wskaźnik taki testował raczej wrażliwość na zjawisko narkomanii niż na komputeromanię. Wykryto to w badaniach przeprowadzonych w krajach, w których występuje zupełnie inna skala zjawisk. Nietrafność potwierdziła się, gdy usunięto odniesienie do narkotyków, kiedy to różnicowanie środowiskowe zanikło. Jednak nadal z oczywistej tezy, że można się uzależnić, nic poznawczo nie wynikało.

⟨Można się uzależnić od komputerów tak samo jak od narkotyków.⟩

⟨Można się uzależnić od komputerów.⟩

⇒ Obawiam się uzależnienia od komputera.

Dopiero wprowadzenie ostatniej wersji dało pożądaną rezultat, dzięki zastosowaniu emocjonalnej formy introspekcyjnej, gdzie każdy wypowiada się o sobie. Warto zatem zwrócić uwagę na doniosłą rolę introspekcji, a także na to, że porównywanie do innego obiektu, niebędącego przedmiotem ewaluacji, może prowadzić do nietrafnych wniosków. Naprawa treści tegoż bodźca-stwierdzenia odbyła się jednak kosztem wskaźnika bliźniaczego. W parze kontrolnej chciałem umieścić wskaźnik ekstrasppekcyjny, kiedy to respondenci wypowiadają się o innych. Aby uniknąć powtarzania formuły '*obawiam się*', musiałem zmienić treść wskaźnika występującego w parze, pierwotnie zredagowanego tak:

⟨Obawiam się, że pewnego dnia komputery przejmą władzę i zniewolą ludzi.⟩

⟨Obawiam się, że pewnego dnia komputery zniewolą ludzi.⟩

na postać ⇒ Pewnego dnia komputery zniewolą ludzi.

Ze względu na dalece prognostyczny charakter tej tezy rozkład empiryczny okazał się niekorzystny. W zasadzie kwantyfikatory selekcyjnej konkluzyjności wskazują na potrzebę przeredagowania treści, jednak ze względów merytorycznych postanowiłem pozostawić tak śmiałą tezę. Chodzi o badanie, czy respondenci mają świadomość, że – obok niewątpliwie olbrzymich korzyści – globalna komputeryzacja stanowi także coraz bardziej realne zagrożenia.

Czwarty czynnik w ekstrakcji indykatorum nazwałem *Poglądami*. Otwiera on grupę wskaźników tworzących komponenty kognitywne. Jedną z kwestii, jakie zaplanowałem w tym obszarze próbkować, dotyczy zakresu wykorzystania komputera w procesach edukacji intencjonalnej. Zagadnienie to wynikało z ówczesnej batalii o szersze zastosowanie technologii informacyjnej w szkole.

⟨Używanie komputerów powinno być częścią wszystkich kursów.⟩

⟨Komputer w pewnym zakresie powinien być używany na wszystkich zajęciach.⟩

⟨Komputer winien być używany tylko na kursach Informatyki.⟩

Pomimo zmian dokonywanych w treści wskaźnik ten niezmiennie wykazywał zbyt mało wyrazistą polaryzację. Tezy te nie uzyskiwały kwalifikowanego większościowo poparcia bądź odrzucenia. Z przesłanek metodyki wykorzystania nowoczesnych mediów w kształceniu wynika potrzeba stosowania komputera na różnych przedmiotach w szkole, tymczasem uczący się nie dostrzegają tego. Prawie połowa badanych, zamiast zaprzeczyć, zgadza się ze zdaniem:

⇒ W szkole komputer winien być używany tylko na Informatyce.

Jest to przejawem zinternalizowania stanu faktycznego i przyjęcia za własny, fałszywy osąd wartości w gruncie rzeczy negatywnego zjawiska, jakim jest brak wykorzystania komputera na innych zajęciach, choćby na matematyce. Zdecydowałem o pozostawieniu tego wskaźnika w skali, ponieważ z pewnością z biegiem czasu stan świadomości w tej kwestii będzie się poprawiał.

Do pary kontrolnej z poprzednim wskaźnikiem zaplanowałem pierwotnie mierzyć osąd użyteczności kursów, w których uczestniczyli respondenci:

⟨Ten kurs wywrze wpływ na to, jaką specjalizację wybiorę w przyszłości.⟩

⟨Kurs pozwoli mi docenić możliwości stosowania komputerów w mojej dziedzinie.⟩

Po okresie weryfikacji doszedłem jednak do wniosku, że są to zagadnienia zbyt szczegółowe, a jako takie nie wymagają omawiania na każdym rodzaju kursu, więc nie mogą być uniwersalnym wyznacznikiem jakości zajęć. Z tego powodu zmieniłem funkcję wskaźnika, próbując osąd użyteczności narzędzia:

⇒ Uważam, że komputery są użyteczne w mojej dziedzinie.

Przyjęte w takiej postaci stwierdzenie ma jednak wady. Otóż młodsi uczniowie nie posiadają jeszcze ukształtowanej preferencji co do przyszłych dziedzin, w których zamierzają się specjalizować. Nadto w różnych dziedzinach inaczej przedstawia się użyteczność komputerów. Powoduje to zakłócenie introspekcyjnej funkcji wskaźnika przez czynniki zewnętrzne. W makroskali populacji zakłócenia te niwelują się wprawdzie, lecz w pomiarach lokalnych mogą być wyraźne (np. inny osąd wyrażą uczniowie profilu humanistycznego od profilu informatycznego). Właśnie z tego powodu tak bardzo ważne jest próbkowanie dwukrotne i analiza zmian, jakie zachodzą (a nie jedynie stanów chwilowych). Wprowadzenie do skali nieoptymalnego wskaźnika powoduje, że nie można go już później poprawić bez konieczności rezygnacji z wcześniej pozyskanych zasobów danych. Cykliczny proces optymalizacji i weryfikacji narzędzia trzeba na pewnym etapie zamknąć, podejmując decyzję o zaprzestaniu jakiegokolwiek dalszej modyfikacji bodźców-stwierdzeń. Ma to istotne znaczenie w budowie narzędzi skalowanych, gdyż do ich normalizacji i do wyznaczenia standardów potrzebne jest zgromadzenie licznego zbioru danych empirycznych.

W poszukiwaniu optymalnych wskaźników zdarza się, że kolejne wersje projektowanych bodźców-stwierdzeń różnią się treścią, chociaż w założeniach winny próbować tę samą cechę. Jeśli na przykład badamy stan świadomości związanej z informatyzacją i komputeryzacją, to mamy do wyboru bardzo wiele możliwości próbkowania różnych aspektów składających się na ową cechę.

⟨Znajomość podstaw informatyki powinna być wymagana od wszystkich uczniów szkół średnich.⟩

⟨Myślę, że komputer jest narzędziem, które będę musiał(a) używać.⟩

⇒ Już dziś wiem, w czym komputer jest niezastąpiony.

Każda z wersji musi jednoznacznie przynależć do wydzielonej z góry grupy, w tym wypadku komponentów poznawczych, dotyczących wglądu w zachodzące procesy. Podczas weryfikacji empirycznej najlepiej sprawdzają się takie wskaźniki, które próbują właściwości w sposób najbardziej ogólny. Warto jednak do pary kontrolnej wprowadzić tezę znacznie bardziej szczegółową.

⟨Myślę, że w najbliższych latach rola komputerów w życiu codziennym będzie wzrastać.⟩

⇒ Komputer może radzić, jak rozwiązywać życiowe problemy.

Pierwsze stwierdzenie okazało się dla respondentów tak oczywiste, że nie było wystarczającego zróżnicowania wypowiedzi. Dopiero drugie sprawdziło się pod względem wysokiej konkluzyjności, także u zaawansowanych respondentów, dla których informatyka była specjalizacją. Dostrzegają oni bowiem niuanse rozróżniania pojęcia *komputer*, rozumianego jako urządzenie (*hardware*), od potencjalnej funkcji oprogramowania (*software*) w roli systemów eksperckich.

Piątym czynnikiem ewaluacji *InfoKultury* jest zmienna o nazwie *Wprawa*, która wyraża samoocenę uprzednich doświadczeń i biegłości w posługiwaniu się komputerem. Wszystkie wskaźniki, składające się na ten introspekcyjny czynnik, powinny zatem z założenia próbować cechy własne respondenta.

⟨Miałem(am) już do czynienia z komputerem.⟩

⟨Mam negatywne doświadczenia w posługiwaniu się komputerami.⟩

⇒ Mam pozytywne doświadczenia w pracy z komputerem.

Warto zwrócić uwagę, że bez potrzeby używania formy czasu przeszłego dwa ostatnie bodźce-stwierdzenia mają zdatność próbkowania retrospekcyjnego. Z tego względu dobrany do pary wskaźnik kontrolny, pierwotnie także retrospekcyjny, postanowiłem zmienić na silniej testujący stan aktualny.

⟨W przeszłości potrzebowałem(am) użyć komputera.⟩

⟨W przeszłości potrzebowałem(am) komputera, ale nie wiedziałem(am) jak go użyć.⟩

⟨Umiał(a)bym poradzić sobie z komputerem.⟩

⇒ Umieć poradzić sobie z komputerem.

Przyjęte do kwestionariusza, szerokie znaczeniowo sformułowanie obejmuje swoim zasięgiem także sytuacje radzenia sobie z występującymi awariami.



Oprócz mierzenia poziomu obycia z komputerem, uznałem też za istotne ustalanie stopnia pewności siebie podczas posługiwania się tym medium.

⟨Czuje się niepewnie każdorazowo, gdy zaczynam pracować na komputerze.⟩

⟨Czuję się pewnie, gdy siadam do pracy z komputerem.⟩

⇒ Czuję się niepewnie, gdy siadam przy komputerze.

Sposób redagowania tegoż bodźca-stwierdzenia może być przykładem, jak należy poszukiwać optymalnie zwartej formy zdania, z jednoczesnym doprecyzowaniem czasowego aspektu wskaźnikowania. W tym wypadku chodziło o wykrycie symptomów ewentualnej trwałości stanu niepokojącego, który przecież w miarę nabywania doświadczeń powinien ustępować. Najbardziej zakorzenione braki perspektyw na pozytywne zmiany wykrywał bodziec następujący:

⟨Myślę, że nigdy nie będę w stanie posługiwać się komputerem.⟩

Formy zaprzeczenia w stwierdzeniu nie powinno się jednak stosować, dlatego odwróciłem tezę, zmieniając polaryzację wskaźnika na spolaryzowaną dodatnio:

⟨Myślę, że będę w stanie posługiwać się komputerem.⟩

⇒ Z pewnością będę w stanie posługiwać się komputerem.

Podczas weryfikacji empirycznej okazało się, że dodanie zwrotu 'z pewnością' daje znacznie lepszy rozkład i bardziej konkluzyjne parametry statystyczne. Jest to typowy wskaźnik futurospekcyjny, próbujący stany przyszłe. Łącznie czynnik samooceny co do umiejętności i wprawy testowany jest w całej czasoprzestrzeni z użyciem wskaźników diagnostycznych oraz prognostycznych.

Ostatniemu, szóstemu czynnikowi nadałem nazwę *Rozwaga*, ponieważ z jednej strony wyraża on roztropną powściągliwość w zachowaniach podczas pracy z komputerem, a z drugiej – stanowi konglomerat wskaźników kognitywnych, które podczas wypowiedziania się respondentów zmuszają ich do głębszego namysłu. Próbkowanie oglądy zachowań powierzyłem stwierdzeniu:

⇒ Praca z komputerem wymaga dużej koncentracji uwagi.

Zagadnieniem badawczym jest to, że komputer silnie pobudza użytkownika, nie tylko w sferze emocjonalnej, lecz także behawioralnej. Użytkownik ulega nawykowi działania pochopnego, prowadzącego do wielu strat, których można byłoby uniknąć, gdyby czynności były przemyślane. Aspektem paralelnym do nadmiernej szybkości działania jest długotrwałość przebywania, połączona z charakterystycznym zatraćaniem się psychicznym, zdrowotnym i fizycznym. Braki świadomości tego typu zagrożeń wykrywane są poprzez sformułowanie:

⟨Długotrwała praca przy komputerze jest szkodliwa.⟩

⇒ Zbyt długotrwała praca przy komputerze jest szkodliwa.

Dodanie pozornie błędnego słówka 'zbyt' zmieniło korzystnie jakość wskaźnika. Przykład pary kontrolnej wskaźników próbujących oglądę zachowań ilustruje prawidłowy sposób badania aspektów czasowych w sytuacjach, gdy zarówno nadmierna szybkość działań, jak i zbytnia rozwlekłość są antywartościami.

W dobrej skali pomiarowej umieszcza się zarówno wskaźniki próbkujące komponenty afektywne, jak i kognitywne. W sposób naturalny wyłonić można także inne grupy bodźców-stwierdzeń, zakładając pewne kryterium podziału. Najbardziej typowym jest podział ze względu na trudność kwestii, a więc na te pozycje skali, które respondentom sprawiają mniejsze lub większe problemy podczas wypowiedzania się. We wzorcu *Skali dwuważonych ocen* zalecam podobne grupowanie, polegające na wyłonieniu wypowiedzi udzielanych pod wpływem gotowych skryptów reakcji na bodziec-stwierdzenie oraz wypowiedzi wymagających refleksji i dłuższego zastanowienia nad wyborem stanowiska. Do grupy bodźców pobudzających do głębszego przemyślenia należy teza:

⟨Używanie komputera jest możliwe już po minimalnym poinstruowaniu.⟩

⇒ Można używać komputera już po krótkim poinstruowaniu.

Testuje ona szacunkowy poziom zdolności informatycznych. Wprawdzie sformułowana jest w konwencji tezy globalnej, odnoszącej się bardziej do ogółu, lecz jednak próbkuje pośrednio osobisty poziom świadomości w tej kwestii. Chodziło mi głównie o wmontowanie do skali wskaźnika, który by wyznaczał próg pomiędzy prowadzeniem słuchaczy kursu „za rączkę” przez instruktora a pracą samodzielną z wykorzystaniem instrukcji elektronicznych (*Pomocy*). Do pary kontrolnej z powyższym wskaźnikiem pierwotnie desygnowałem tezę, która nie dała odpowiednich rezultatów w fazie weryfikacji empirycznej:

⟨Opanowanie komputera jest niemożliwe w pojedynkę.⟩

Dopiero zmiana treści przyniosła korzystny rozkład i wyższą konkluzyjność:

⇒ Umiejętności komputerowe można zdobyć samemu.

W stwierdzeniu tym celowo ukryty jest pewien niuans formalno-definicyjny, który respondentów o wyższym poziomie ścisłego rozumowania w percepcji przekazów werbalnych powinien od razu naprowadzić na logiczną wypowiedź. Otóż umiejętności są właściwością przyswajaną osobiście, internalnie, a więc nikt inny nie wyręczy nas z ich nabywania. Brak tej świadomości powoduje, że miliony osób na świecie wydają bająnskie sumy na nieefektywne kursy.

Łącznie *Kwestionariusz InfoKultury* tworzą 24 bodźce-stwierdzenia, w tym dokładnie po 12 afektywnych i kognitywnych, choć nie są one ułożone kolejno. Jedynym wymogiem jest to, aby obok siebie umieszczone były pary kontrolne. Formularz powinna rozpocząć prezentacja instytucji prowadzącej badania, np.:

*Uniwersytet im. Adama Mickiewicza, Zakład Technologii Kształcenia*

Dalej pozostawiamy miejsce na dane respondenta, datę oraz lokalizację kursu:

*Imię i nazwisko lub kod ..... data .....*

*Szkoła ..... profil, kierunek .....*

Przed pozycjami próbkującymi warto umieścić apel do respondentów o rzetelne i samodzielne wypełnianie kwestionariusza, a na końcu należy podziękować. Każde stwierdzenie uzupełniają jednolite opcje wypowiedzi do wyboru:

*Prosimy o rzetelne wyrażanie własnych odczuć wobec zagadnień!*

1. Odczuwam raczej niechęć do komputeryzacji.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
2. Tylko uzdolnieni do komputera skorzystają z jego zalet.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
3. Mam pozytywne doświadczenia w pracy z komputerem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
4. Umiejętności sobie z komputerem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
5. Czuję się niepewnie, gdy siadam przy komputerze.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
6. Z pewnością będę w stanie posługiwać się komputerem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
7. Można używać komputera już po krótkim poinstruowaniu.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
8. Umiejętności komputerowe można zdobyć samemu.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
9. Bał(a)bym się zostać sam na sam z komputerem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
10. Jest mi źle w grupie, gdy inni lepiej znają komputery.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
11. Zawsze wyrażam gotowość uczenia się trudnych rzeczy.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
12. Chcę nauczyć się znacznie więcej o komputerach.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
13. Wątpię w przydatność takich kursów komputerowych.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
14. Nawet obcy z komputerem odniosą z kursu pożytek.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
15. W szkole komputer winien być używany tylko na Informatyce.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
16. Uważam, że komputery są użyteczne w mojej dziedzinie.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
17. Już dziś wiem, w czym komputer jest niezastąpiony.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
18. Komputer może radzić, jak rozwiązywać życiowe problemy.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
19. Praca z komputerem wymaga dużej koncentracji uwagi.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
20. Zbyt długotrwała praca przy komputerze jest szkodliwa.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
21. Chcę więcej czasu przeznaczyć na pracę z komputerem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
22. W przyszłości komputery pomogą mi zaoszczędzić czas.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
23. Obawiam się uzależnienia od komputera.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
24. Pewnego dnia komputery zniewolą ludzi.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak

*Dziękujemy za udział w badaniach poziomu Kultury informatycznej!*

### **Jak zbierać wypowiedzi – technika pomiaru**

Sposób przeprowadzenia pomiaru wynika bezpośrednio z opracowanego wcześniej projektu ewaluacji. W zależności od przyjętej strategii (zob. str. 105): eksperymentu, badań relacyjnych, przekrojowych bądź panelowych, a także od planowanych zakresów przestrzennych i czasowych próbkowania (str. 103), inne są metody selekcji prób badawczych i pobierania danych empirycznych. Zasadnicze różnice dotyczą reprezentacji i odnoszą się do ról przypisanych poszczególnym grupom. Zróżnicowane są też momenty pobierania próbek i odstępy czasu pomiędzy pomiarami dystansowymi. Na przykład, w eksperymencie, polegającym na jednorazowym, krótkotrwałym oddziaływaniu jakimś silnie pobudzającym przekazem medialnym, drugi pomiar wykonywany jest tuż po pobudzeniu. Jeśli więc przedmiotem badań jest reklama, to skutki jej oddziaływań mogą być zmierzone w ciągu paru minut. Jeśli jednak ewaluacja dotyczy procesów edukacyjnych, a zależy nam na badaniach długofalowych, to pomiary w pełnych powtarzanych cyklach mogą trwać kilka lat.

Chociaż plany badawcze są różne, to jednak sama czynność ankietowania jest standaryzowana. Polega na rozdaniu kwestionariusza i zebraniu wypełnionych formularzy. Czynność tę w praktyce może wykonać każda osoba, chociaż najlepiej, jeśli ankieter jest znany respondentom. W pomiarach jakości zajęć z prośbą o wypełnienie formularzy winien zwracać się wykładowca. Do ewaluacji procesu edukacyjnego niezbędne są zawsze dwa pomiary identycznym kwestionariuszem – już na samym początku oraz blisko końca danego okresu kształcenia (cały kurs, semestr, rok). Wykładowca powinien zapewnić słuchaczy, że pomiary te nie służą do oceniania uczniów, lecz wyłącznie do weryfikacji efektów pracy dydaktycznej. Warto przy tym wystosować ustny apel o pełną samodzielność w zakreślaniu wypowiedzi, z uzasadnieniem, że przecież mierzone postawy są u każdego indywidualną cechą umysłu.

Na wnikliwe wypełnienie druku wystarcza około 10 minut. W tym czasie osoba przeprowadzająca pomiar nie powinna podejmować jakichkolwiek prób objaśniania ewentualnych zapytań. Wyjątkiem może być wcześniejsze wytłumaczenie najmłodszym uczniom, że w ich sytuacji zwrot: *'w mojej dziedzinie'* oznacza to, co ich interesuje i czym w przyszłości zamierzają się zajmować. Przed zebraniem kwestionariuszy warto poprosić respondentów o sprawdzenie, czy podczas wypełniania formularza nie przeoczyli którejś pozycji ze skali. Natychmiast po zebraniu druków należy posegregować i oznakować każdą wyodrębnioną próbę badawczą oraz dokładnie opisać zbiory, ażeby uchronić się przed zapomnieniem istotnych informacji: skąd pochodzą dane, z jakiego okresu, z którego pomiaru, jakie to były próby badawcze (grupy oraz klasy), o jakich obserwowalnych cechach, a ponadto w przypadku ewaluacji zajęć – kto prowadził zajęcia, czym charakteryzował się proces dydaktyczny itp.

Warto zwrócić uwagę na to, iż ja opowiadam się za podawaniem nazwisk respondentów. Uważam, że anonimowość nie jest tutaj konieczna, ponieważ stawiane w *Kwestionariuszu* tezy nie są ani drażliwe, ani nie naruszają dóbr osobistych. Natomiast sygnowane imiennie formularze wypełniane są bardziej skrupulatnie, zwłaszcza przez uczniów młodszych. W uzasadnionych sytuacjach można oczywiście podjąć decyzję o całkowitej anonimowości wszystkich respondentów, lecz należy wtedy zadbać o zakodowanie prac w taki sposób, ażeby możliwe było ułożenie do pary danych uzyskanych od tej samej osoby. Jest to niezbędne do przeprowadzenia analizy indeksowej, której podstawę stanowi odniesienie do fluktuacji wypowiedzi jednostkowych.

Istota ewaluacji różnicowo-dystansowej polega na dwukrotnym badaniu takich cech mentalnych, które uznawane są za względnie trwałe. Jeśli więc w drugim pomiarze ujawnią się wyraźne zmiany postaw i uświadomienia, to prawdopodobnie musiały zadziałać jakieś zewnętrzne czynniki modyfikujące. Oczekuje się prawidłowości przemian, a na tle owych założeń pomiar ujawnia trendy rzeczywiste. Warto przeprowadzić wstępną analizę wypowiedzi już po pierwszym pomiarze, jednakże niedopuszczalne jest wyciąganie wniosków w rodzaju: *Ta osoba źle się wypowiedziała...* Byłoby to działanie niedorzeczne interpretacyjnie, sprzeczne ze strategią ewaluacji demokratycznej, a ponadto groźniejsze łamanie kardynalnych zasad etyki badawczej<sup>38</sup>, zgodnie z którą nie wolno dezawuować indywidualnych postaw, choćby skrajnie odmiennych.

Zalecane tutaj wstępne zapoznanie się z wypowiedziami pochodzącymi z pierwszego pomiaru ma zupełnie inny – pożyteczny cel. Z kwestionariusza można wydobyć informacje o aktualnym poziomie kompetencji informatycznych, co umożliwi dopasowanie stopnia trudności zajęć do potencjalnych możliwościach percepcyjnych danej grupy. Ponadto, narzędziem tym daje się wykryć stany prawie niedostrzegalne na lekcji, takie jak: ukrywany stres przed komputerem lub wobec grupy, błędne przeświadczenia o niemożliwej do pokonania trudności obsługi, wrodzona niechęć do technicyzacji, *a priori* brak akceptacji przedmiotu, metod lub treści programowych. Jeśli nauczyciel wykryje na samym początku tego typu niekorzystne zjawiska, wówczas może je w ciągu roku szkolnego skutecznie niwelować. Uruchamiane są bowiem niejako samoistnie załączkowe mechanizmy ewaluacji formatywnej. Niektórzy zwolennicy ewaluacji zalecają ciągle współuczestnictwo ewaluatora w badanym procesie oraz tzw. *monitoring*, czyli systematyczne diagnozowanie w toku, jednakże nie wolno przesadzać z wielokrotnością pomiarów lub z nieustanną kontrolą tych samych uczestników zajęć. Ciągłe monitorowanie prowadzi do „samospelniających się” celów ewaluacji, a nie do poprawy efektów samego procesu. Jedną grupę badawczą wystarczy zdiagnozować dwukrotnie, przed i po kursie, warto natomiast przeprowadzać pomiary na większej liczbie grup.

Pod względem jakości metody badawczej najlepszym rozwiązaniem jest ewaluacja wieloletnia, z kilkoma pełnymi cyklami pomiarów zwieńczonych eksplikacją – zwłaszcza w takim wypadku, gdy na kolejnych rocznikach w tej samej szkole bada się jakość realizacji prowadzonego przez siebie przedmiotu nauczania. Diagnoza pochodząca z danego roku staje się podstawą do optymalizacji procesu w kolejnym roku. Ponadto, możliwość konfrontacji tego, jak zdaniem wykładowcy udały się zajęcia, ze wskaźnikami empirycznymi opinii słuchaczy, jest niezwykle cenna w samodoskonaleniu zawodowym.

### **Jak kodować wypowiedzi – strukturyzacja danych**

Uporządkowanie danych surowych według ściśle zdefiniowanego formatu jest niezwykle ważne, gdyż umożliwia wzajemną wymianę zbiorów pomiędzy różnymi badaczami, łatwe przesyłanie pocztą elektroniczną i gromadzenie reprezentatywnych zasobów empirycznych, z których wyznacza się standardy ewaluacyjne, niezbędne do wnioskowania uogólniającego. Unifikacja formatu potrzebna jest też przed wprowadzeniem danych do programu przetwarzającego, jednak w zależności od producenta wymagane formaty mogą być różne. Do obróbki statystycznej za pomocą programu InfoCult™ Analyser potrzebne są pliki kodowane w najbardziej tradycyjnym i łatwo odczytywalnym formacie ASCII. Chodzi o tzw. czysty kod tekstowy, reprezentujący jedynie znaki alfanumeryczne, bez atrybutów typograficznych. W tej postaci każdemu znakowi odpowiada dokładnie jeden bajt kodujący. Do przygotowania danych można wykorzystać systemowy edytor tekstu (*Notatnik* lub *WordPad*), zapisując dane do pliku w formacie tekstowym. Utworzone odrębnie dla każdej próby małe pliki \*.txt przed analizą scala się w jeden większy plik i zmienia rozszerzenie nazwy na \*.dxt. Rozszerzenie to jest rozpoznawane przez Analizator.

Strukturyzacja danych polega na rangowaniu wypowiedzi i wprowadzaniu wskaźników empirycznych do macierzy w ściśle ustalonej kolejności. Do analizy niezbędne są komplementarne dane, pochodzące z obydwu pomiarów, dlatego trzeba dobrać parami druki obu ankiet tych samych osób, z początku i z końca zajęć. Dla każdego respondenta przeznaczona jest tylko jedna linia macierzy, co odpowiada jednemu wierszowi w edytorze tekstu. Wpisywanie rozpoczyna się od kodu grupy i klasy oraz znaku równości. Dalej umieszcza się rozdzielone kreską dwa ciągi liczb z pomiarów 'przed' i 'po'. Ciągi powstają poprzez nadanie rang od 1 do 7 kolejno stopniowanym opcjom wypowiedzi, przy czym zawsze opcji 'absolutnie nie' przypisuje się 1, za 'brak zdania' 4, a za 'absolutnie tak' 7, bez względu na polaryzację wskaźnika. Kompletny wiersz pary danych jednego respondenta przyjmuje postać następującą:

```
Aa=125526662246161652446713-116616561176161764367711
```

W takiej postaci dane są całkowicie anonimowe. Z tego względu w pomocniczym pliku tekstowym należy zapisać i przechowywać szczegółowe opisy kodów: skąd są poszczególne grupy, jakie to klasy, o jakim profilu, kto uczył, kiedy przeprowadzono pomiar, ile godzin trwał kurs itp. Z uwagi na poufność tych opisów nie są one włączane do zasadniczego pliku z danymi.

Wyjaśnienia wymaga sprawa depolaryzacji i nadawania wag wskaźnikom. Czynności te wykonuje automatycznie Analizator. Jeśli jednak badacz używa innego programu do obliczeń statystycznych, to konieczne jest odwrócenie polaryzacji tych wypowiedzi, które odnoszą się do bodźców-stwierdzeń wyrażających antytezę wartości, a więc wymagających zaprzeczenia. Taką cechą mają m.in. dwie pierwsze pozycje *Kwestionariusza InfoKultury* i wszystkie te, które w schemacie polaryzacji oznaczono znakiem minusa. Ponadto, obliczenia należy przeprowadzać nie na rangach 1÷7, lecz na prawidłowo nadanych wagach z zakresu (-3...+3). Po depolaryzacji każda wypowiedź, która akceptuje tezę niosącą wartość lub przeczy antytezie, ma znak dodatni, natomiast wypowiedzi zaprzeczające wartości lub akceptujące antytezę mają znak ujemny. Neutralny jest jedynie '*brak zdania*', któremu przypisuje się wagę zerową (0). Sposób przetwarzania rang na wagi przedstawiam w następnym rozdziale.

Bardzo ważną procedurą strukturyzacji jest grupowanie danych. Obok siebie, w kolejnych wierszach należy umieścić wszystkie przypadki należące do danej klasy. Kolejność klas wewnątrz grupy powinna być tak dobrana, ażeby obok siebie znalazły się próby, które zamierzamy porównywać. Naturalnym układem jest sąsiedztwo klas prowadzonych przez jednego wykładowcę. Dobrym rozwiązaniem jest uporządkowanie według profilu, np. odrębnie nauki ścisłe i humanistyczne. Podobnie jest z układem grup – od ich układu zależy możliwość umieszczania rezultatów na wspólnym wykresie. Analizator wyodrębnia dowolne przekroje na podstawie przypisanych kodów, lecz nie zmienia kolejności zbioru danych. Operację taką w każdej chwili wykonać można za pomocą edytora, przenosząc całe bloki tekstowe. Warto również przygotować kilka plików z różnymi wersjami pogrupowania tych samych danych. Dotyczy to zwłaszcza strategii dalszego rozwarstwiania prób badawczych ze względu na wartości celowo dobranych zmiennych grupujących bądź filtrujących.

<i>Ślady</i>		<i>Wzorce</i>
rozwarstwienie zmiennych	.....	definiowanie i ekstrakcja indykatum
operacjonalizacja zmiennych	.....	dobór i weryfikacja wskaźników
optymalizacja próbników	.....	redakcja bodźców-stwierdzeń
strategie badawcze	.....	pomiary dystansowe i cykliczne
pozyskiwanie danych	.....	próbkiwanie reakcji-wypowiedzi
strukturyzacja danych	.....	kodowanie i grupowanie wypowiedzi

## Jak przetwarzać dane – obliczenia statystyczne

Z proponowaną tu metodologią ewaluacji za pomocą skal dwuważonych wiąże się ściśle pewien, zalecany przeze mnie jako optymalny, zespół metod statystycznych. Zagadnienia te przedstawiam w odrębnym dziale, w którym czytelnik znajdzie nie tylko definicje, wzory i objaśnienia parametrów, ale także przykłady ich obliczania. Wiele z opisanych tam miar ma innowacyjny charakter, stąd procedury przetwarzania nie są dostępne w funkcjonujących na rynku pakietach statystycznych. Czytelnik może do obliczeń wykorzystać uniwersalne narzędzie, jakim jest arkusz kalkulacyjny, bądź też dedykowany InfoCult™ Analyser, umieszczony na płycie CD-R dołączonej do książki. Opis posługiwania się Analizatorem znajduje się w dziale *Eksploatacja zasobów empirycznych*, dlatego tutaj przedstawię jedynie ogólne zasady obróbki danych.

Przypomnę, że w procesie kodowania wypowiedzi zalecam wpisywanie rang 1÷7 jako wygodniejszy sposób wprowadzania danych. Nie wolno jednak sumować rang ani liczyć średnich z danych surowych, gdyż część bodźców-stwierdzeń jest spolaryzowana ujemnie (tezy fałszywe). Dopiero depolaryzacja poprzez nadanie odpowiednich wag wskaźnikom oraz agregacja zmiennych częściowych według schematu interpretacji umożliwia prawidłowe przetwarzanie danych i poprawne wnioskowanie. Analizator wykonuje te czynności automatycznie, natomiast w każdym innym programie trzeba je odpowiednio zaimplementować. Jest prosty sposób na usprawnienie czynności depolaryzacji i nadawania wag podczas przetwarzania za pomocą arkusza kalkulacyjnego.

**Depolaryzacja.** Każda pozycja *Kwestionariusza InfoKultury* jest numerowana, zatem kolejnym wskaźnikom nadaje się tę samą numerację porządkową 1÷24. Wskaźniki nie posiadają nazw własnych, dlatego numeracja stanowi jedyny identyfikator, na podstawie którego wiąże się daną wypowiedź z bodźcem-stwierdzeniem. W zależności od dodatniej/ujemnej polaryzacji stwierdzenia, przypisuje się tej pozycji odpowiednio znak plusa/minusa. Powstaje w ten sposób *schemat polaryzacji* 24 wskaźników. Do arkusza kalkulacyjnego wprowadza się ów schemat w nieco zmienionej postaci, jako zbiór liczb +1 lub -1. Jeśli na pierwszej karcie arkusza znajduje się macierz danych wyrażonych za pomocą rang [R], to na drugiej karcie można utworzyć tego samego rozmiaru macierz wag [W] poprzez powielenie formuł opartych na wzorze  $W = P \cdot (R - 4)$ . W uproszczeniu do pojedynczego przypadku rezultaty wyglądają następująco:

Numer wskaźnika	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Polaryzacja [P]	-1	-1	1	1	-1	1	1	1	-1	-1	1	1	-1	1	-1	1	1	-1	1	1	1	1	-1	-1
Rangi [R]	1	2	5	5	2	6	6	6	2	2	4	6	1	6	1	6	5	2	4	4	6	7	1	3
Wagi [W]	3	2	1	1	2	2	2	2	2	2	0	2	3	2	3	2	1	2	0	0	2	3	3	1

Identyczna procedura dotyczy danych z obu pomiarów 'przed' i 'po', przy czym powstała macierz warto skopiować i wkleić tylko jako *wartości* na trzecią kartę.



Zaleca się, ażeby wszystkie obliczenia statystyczne przeprowadzać na macierzy wag [W] utworzonej z elementarnych wskaźników, a nie na skumulowanych szeregach rozdzielczych. Na szeregach nie da się przeprowadzić np. analizy indeksowej. Warto jednak wspomnieć też, że z macierzy wag nie można obliczać np. średniej geometrycznej. Obliczone parametry statystyczne, jeśli nie są współczynnikami bezwzględnymi, to kalibrowane są na *skali wag* albo na *skali ocen*. Konwersji na skalę ocen dokonuje się dopiero po obliczeniach, a polega ona na dodawaniu wartości równej 2,5 do parametrów pozycyjnych, reprezentujących poziomy cech. Dzięki temu bardzo prostemu przekształceniu rezultaty liczbowe ewaluacji przybliżone są do poziomów odczytywanych jako utrwalony skrypt stopniowania wartości ocen szkolnych.

Podczas prowadzenia obliczeń statystycznych w arkuszu kalkulacyjnym największą trudność sprawia wyłanianie odpowiednich wskaźników tworzących prawidłowe przekroje analityczne. Analiza odbywa się na czterech poziomach rozwarstwienia indykatum. Poziom elementarny tworzą wskaźniki, a ich pary kontrolne po złączeniu stają się składnikami. Z połączenia czterech z góry ustalonych wskaźników powstają czynniki, z dwunastu rezultaty połówkowe, a ze wszystkich 24 – wynik ogólny. Jakakolwiek pomyłka przy scalaniu wag, ułożonych w kolumnach macierzy i reprezentujących wskaźniki, powoduje całkowite zafałszowanie parametrów. Proces scalania, odwrotny do ekstrakcji, nazwijmy *agregacją*. Przyjrzyjmy się sposobom postępowania przy scalaniu.

*Agregacja.* Każda ze zmiennych cząstkowych powinna mieć jednoznacznie wyznaczoną kolumnę w arkuszu. Pierwszy wiersz przeznaczony jest na symboliczny opis zawartości kolumn (tzw. główka). Dla wskaźników warto przyjąć taką samą strukturę, jak dla danych surowych, lecz wartościami są tutaj już zdepolaryzowane wagi. W kolumnach A i B umieszcza się kody grup i klas. Wskaźniki z pierwszego pomiaru zajmują kolumny od D do AA, a z drugiego – od AC do AZ. Pomiedzy każdą grupą zmiennych warto pozostawić kolumnę separującą. Dla składników rezerwuje się grupy kolumn BB÷BM oraz BO÷BZ, a dla czynników kolumny CB÷CG oraz CI÷CN. Rezultaty połówkowe tworzy się w kolumnach CP÷CQ i CS÷CT, a wyniki ogólne 'przed' i 'po' oblicza w CV i CW. Numery wskaźników i symbole zmiennych z pierwszego pomiaru oznaczymy 'prim', a drugiego 'bis'. W przydzielaniu kolumn dla zmiennych i w procesie scalania danych posługiwać się trzeba strukturalnym schematem agregacji:

Numery kontrolnych par wskaźników, odpowiadające pozycjom w Kwestionariuszu											
13×14	1×2	11×12	21×22	9×10	23×24	15×16	17×18	3×4	5×6	7×8	19×20
Umowne nazwy zmiennych – składników, czynników, rezultatów połówkowych i wyników											
<i>Ocena</i>	<i>Aplauz</i>	<i>Ambicje</i>	<i>Intencje</i>	<i>Spokój</i>	<i>Odpór</i>	<i>Osąd</i>	<i>Wgląd</i>	<i>Obycie</i>	<i>Pewność</i>	<i>Zdolność</i>	<i>Ogląda</i>
<i>Opinie</i>		<i>Motywacje</i>		<i>Emocje</i>		<i>Poglądy</i>		<i>Wprawa</i>		<i>Rozwaga</i>	
<i>Doznanie</i>						<i>Poznanie</i>					
<i>InfoKultura</i>											

Zajmijmy się sposobem agregacji wskaźników na przykładzie wybranych dwóch składników: *Ambicji* i *Intencji*, tworzących czynnik *Motywacje*. Technicznie agregacja polega na wyborze odpowiednich wskaźników, natomiast obliczeniowo jest to łączenie poprzez wyznaczenie ich średniej arytmetycznej:

Formuła = ( $\$kolumna\ wiersz + \dots + \$kolumna\ wiersz$ )/liczba wskaźników  
symbol \$ przed nazwą kolumny oznacza, że adres kolumny jest stały

Z przyjętego dla *Kwestionariusza InfoKultury* schematu wynika, że składnik *Ambicje* jest wypadkową wskaźników nr 11 i 12. Umieszczono je w arkuszu kalkulacyjnym w kolumnach N i O oraz AM i AN. Składnik *Ambicje* 'przed' oblicza się w komórce BD2 poprzez formułę  $=(\$N2+\$O2)/2$  z założeniem, że dane pierwszego respondenta wprowadzono właśnie do wiersza nr 2. Formułę tę powiela się automatycznie, przeciągając w dół przez komórki obejmujące wszystkich respondentów (*przypadki*). W adresach samoczynnie zmieniają się numery wierszy. Analogicznie wprowadza się formułę obliczającą *Ambicje* 'po'  $=(\$AM2+\$AN2)/2$  do komórki BQ2, po czym przeciąga ją w dół kolumny. Dla składnika *Intencje* splot stanowią wskaźniki 21 i 22, stąd formuły 'przed' BE  $=(\$X2+\$Y2)/2$  oraz 'po' BR  $=(\$AW2+\$AX2)/2$ . Czynnik *Motywacje* składany jest z czterech wskaźników: 11, 12, 21 i 22, dlatego formuły są następujące: 'przed' CC  $=(\$N2+\$O2+\$X2+\$Y2)/4$ ; 'po' CJ  $=(\$AM2+\$AN2+\$AW2+\$AX2)/4$ . Także i te formuły powielić trzeba na wszystkie przypadki. Fragment arkusza z oznaczeniami, wagami i rezultatami obliczeń powinien mieć poniższą postać:

A	B	C	...	N	O	...	X	Y	...	AM	AN	...	AW	AX	...	BD	BE	...	BQ	BR	...	CC	...	CJ
gr.	kl.	nr	...	11'	12'	...	21'	22'	...	11''	12''	...	21''	22''	...	Am'	In'	...	Am''	In''	...	Mo'	...	Mo''
A	a	=	...	0	2	...	2	3	...	3	2	...	3	3	...	1	2,5	...	2,5	3	...	1,75	...	2,75

Po wypełnieniu formułami całego obszaru komponentów indykatoru, otrzymuje się niezbędne do dalszych obliczeń przekroje analityczne przez zmienne. W tych przekrojach przede wszystkim zestawiane i porównywane są wartości pochodzące z obu pomiarów oraz nieliczne miary dotyczące spójności wypowiedzi. Większość dalszych obliczeń odbywa się w przekrojach przez przypadki.

Agregacja wskaźników poprzez obliczanie średnich arytmetycznych z wag jest pierwszym etapem obliczania statystyk z próby. Kolejne etapy zależą od celu ewaluacji i zadań badawczych. Najprostszym zamiarem może być dążenie do ustalenia wyłącznie średnich poziomów zmiennych w danej grupie/klasie. W arkuszu kalkulacyjnym wmontowane są podstawowe funkcje statystyczne. Często wystarczy jedynie wybrać odpowiednią funkcję i podać zakres komórek, z których komputer ma dokonać obliczeń. Przykładowo, średnie arytmetyczne *Motywacji* w grupie A oblicza się z kolumn CC i CJ w zakresie 151 wierszy, gdyż tylu respondentów było w tej grupie. Łatwo też obliczać współczynniki korelacji pomiędzy zmiennymi, wskazując zakres w obrębie dwóch kolumn.

Jednak badanie rozkładów realnych wymaga własnego definiowania procedur, gdyż większość wbudowanych funkcji oparta jest na rozkładzie normalnym. To samo dotyczy testów istotności oraz estymacji, choć te najbardziej zaawansowane strategie nie są raczej potrzebne w praktyce ewaluacyjnej. Bardziej przydatne są metody badania zmienności oraz techniki porównawcze, dlatego – oprócz chwilowych miar pozycyjnych – do analizy rozkładowej potrzebne są miary dyspersji, do analizy różnicowej miary tendencji, a do analizy indeksowej miary dynamiki. Minimalny zestaw parametrów, jaki należy obliczać, to: poziomy średnie, odchylenia realne, trendy zmian oraz efekty fluktuacji. Jeśli natomiast opracowuje się i weryfikuje nowe narzędzie skalowane, to konieczne staje się też obliczanie kwantyfikatorów i wyznaczników jakości pomiarowej.

**Obliczenia.** Każdy parametr statystyczny wyrażony wzorem można obliczyć w arkuszu, zamieniając wzór na elementarne formuły. Zawile równania stają się przejrzyste, jeśli rozdzieli się je na małe fragmenty, które przetwarzane są odrębnie. Zresztą, większość miar wymaga obliczania etapowego. Na przykład do wyznaczenia odchyłeń najpierw należy ustalić wartość średnią. Tak więc rozłożenie procedur statystycznych na działania podstawowe jest zasadniczą regułą postępowania. Zgodnie z zasadami matematycznymi najpierw należy ustalić wartości wyrażeń wewnętrznych, np. ujętych w nawiasy lub zawartych w obrębie operatora sumy  $\Sigma$ . Istotnym zagadnieniem podczas przetwarzania jest wybór właściwych składników, które są zliczane. We wzorach występują bowiem sumy zupełne, cząstkowe, warunkowe lub funkcje selekcyjne. Należy zwracać szczególną uwagę na indeksy sterujące obliczaniem. Jeśli we wzorze pod znakiem  $\Sigma$  występuje wyłącznie indeks iteracyjny, to należy zsumować kolejno wszystkie przypadki z danej próby. Jeśli jednak występuje tam indeks selekcyjny, to sumuje się wyłącznie przypadki spełniające podany warunek. W arkuszu kalkulacyjnym dostępnych jest wiele przydatnych funkcji sterujących i formuł, które ułatwiają proces selekcji i obliczania. Całość koniecznych obliczeń w pełni automatycznie wykonuje program InfoCult™ Analyser.

### **Jak konstruować tabele – strukturyzacja oznak**

W ewaluacji splotowej, wskutek rozwarstwienia indykatum, występuje wielość zmiennych cząstkowych. Z obliczeń statystycznych uzyskuje się wiele różnych parametrów. Ogólnym pojęciem wiążącym zmienne z parametrami jest zbiór kategorii, które nazwałem *oznakami*. Termin ten oznacza zespół relacji w rodzaju: właściwość → wartość, cecha → miara bądź obiekt → atrybut. Tego typu zależności dwóch elementów najlepiej jest przedstawiać w postaci tabelarycznej, zwłaszcza wtedy, gdy dysponujemy wielością oznak. Wydaje się celowe przyjęcie w miarę jednolitych, a jednocześnie optymalnych prawideł budowy tabel, które mieściłyby zmienne i rezultaty obliczeń statystycznych.

Poprawnie ustrukturyzowany format tabelaryczny umożliwia całościowe ujmowanie oznak, odczytywanie prawidłowości statystycznych i wykrywanie ewentualnych zjawisk niekorzystnych lub nietypowych. W tabelach porządkuje się komplet parametrów niezbędnych do analiz i do tworzenia wykresów. Strukturę tabel tworzą: komórki, kolumny i wiersze oraz tzw. główki i boczki. W główce tabeli powinien znaleźć się opis zawartości kolumn, dlatego komórki przynajmniej pierwszego wiersza przeznaczają się nie na dane, lecz na opisy. Analogicznie w boczku tabeli, najlepiej z lewej strony, opisuje się zawartość wierszy. Kolumny tworzą układ pionowy, a rzędy poziome stanowią wiersze. Elementarnym obiektem jest komórka, która zawiera analizowane wartości. Ważnym uzupełnieniem jest ogólny opis zawartości, umieszczony nad tabelą lub w nagłówku, często połączony z numeracją kolejnych tabel. Opis tytułowy precyzuje, jakie cechy lub obiekty znajdują się w macierzy komórek.

Przyjmijmy jako zasadniczą konwencję takiej struktury, w której zmienne cząstkowe usytuowane są w kolumnach tabel. Przy takim założeniu główkę tabeli konstruuje się zgodnie ze schematem interpretacji bądź ekstrakcji:

Ocena	Aplauz	Ambicje	Intencje	Spokój	Odpór	Osąd	Wgląd	Obycie	Pewność	Zdolność	Ogłoda
-------	--------	---------	----------	--------	-------	------	-------	--------	---------	----------	--------

Przykład ten ilustruje opis główki tabeli zawierającej wyłącznie *Składniki*. Na wyższym poziomie agregacji zmiennych korzystne jest konstruowanie tabeli zbiorczej, przedstawiającej *Czynniki* w zestawieniu z rezultatami połówkowymi i wynikami ogólnymi. Struktura główki jest wówczas bardziej rozbudowana:

Czynniki						Rezultaty połówkowe		Wyniki
Opinie	Poglądy	Motywacje	Wprawa	Emocje	Rozwaga	Doznanie	Poznanie	Kultura

Zestawienie to ułatwia wykrywanie czynników, które wpłynęły na wyższe lub niższe rezultaty połówkowe, wyrażające komponenty afektywne i kognitywne. Tabele zawierające wyłącznie wskaźniki mają w główce tylko numery pozycji. W wyjątkowych sytuacjach, gdy tabela w całości dotyczy wyników ogólnych, zamiast zmiennych można w kolumnach umieścić kolejne próby badawcze:

kl.	1a	1b	1c	2a	2b	2d	2e	3a	4c	4e	4f	4g	5a	5b	6c	7d	8f	9g
-----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Generalnie jednak rezultaty poszczególnych prób badawczych najlepiej jest sytuować w grupach wierszy, podobnie jak przypadki w macierzach danych.

Grupowanie wierszy w tabelach niezbędne jest z tego powodu, że dla każdej próby oblicza się wiele miar statystycznych. Parametry kolejnych grup lub klas powinny być skupione w trzech kategoriach: 1) miary chwilowe 'przed', 2) miary chwilowe 'po' oraz 3) miary dystansowe. Sąsiedztwo miar dobiera się tak, ażeby blisko siebie znalazły się te parametry, które należy analizować wspólnie, tj.: poziomy i odchylenia, różnice i trendy, wektory i efekt fluktuacji. Jeśli potrzebne jest porównanie miar położonych w dużej odległości od siebie, to warto wykonać kwerendę tylko tych wierszy, które są potrzebne do analizy.

Aby jednak nie pogubić się w interpretacji wydobytych wartości, konieczne jest wycinanie wierszy razem z symbolicznymi oznaczeniami parametrów.

M'	Każdy symbol literowy, łącznie ze znakami lub indeksami, ma jednoznacznie przypisane odrębne znaczenie. Najczęściej są to pierwsze litery od angielskich nazw parametrów. Indeksy 'prim' i 'bis' oznaczają odpowiednio rezultaty pierwszego i drugiego pomiaru. Znaki plus, minus lub plus-minus przed literami mają zróżnicowane znaczenie, opisane dokładnie w dziale poświęconym statystyce. Oprócz przedstawionego tu sposobu oznakowania parametrów warto w boczku tabeli zamieszczać także nazwy parametrów. Właśnie w tym celu należy przestrzegać zasady stosowania zwięzłych, krótkich nazw przydzielanych parametrom.
d'	
M''	
d''	
±M	
±r	
±t	
+e	
-e	
±e	

Program InfoCult™ Analyser automatycznie generuje wszystkie niezbędne do analizy tabele statystyk, uzupełnione kompletnymi główkami i boczkami.

W strukturyzacji oznak pożądanymi są dwa dalsze zabiegi: *kondensowanie* i *abstrahowanie*, prowadzące do większej wartości i przejrzystości zasobów.

**Kondensowanie.** Czynność ta dotyczy grupowania i scalania zasobów empirycznych, pochodzących z wielu prób i przekrojów przez przypadki. Działanie to jest ściśle powiązane ze strategią doboru prób. Jeśli zasoby danych pochodzą z tej samej populacji, to można łączyć kolejne rezultaty w większą całość, o ile nie ma ku temu metodologicznych przeciwwskazań. Próbkę, pochodzącą z klas o różnych oddziaływaniach, w pierwszym rzędzie analizuje się lokalnie, lecz warto gromadzić dane z wielu środowisk, scalać je i analizować globalnie. Im zbiór jest liczniejszy, tym bardziej rezultaty zbliżone są do populacyjnych. W początkowym etapie kondensowanie polega na scalaniu jedynie własnych zasobów badacza. Kolejność łączonych prób musi być ustalona już podczas strukturyzacji danych. Zwieńczeniem procesu kondensowania jest scalanie zasobów pochodzących z różnych kręgów, uzyskanych od innych badaczy.

**Abstrahowanie.** Działanie to jest przeciwieństwem kondensowania. Polega na wydobywaniu z zasobów tylko tych oznak, które są istotne eksploracyjnie. Redukcja liczby prób typowych, które nie wpływają na ustalone już standardy, zwiększa przejrzystość podczas analizy. Podobnie abstrahowanie od nieistotnych parametrów ułatwia interpretację. Ma to szczególne znaczenie podczas automatycznego generowania statystyk przez Analizator. Użytkownik powinien celowo ustawiać opcje wyboru tylko niezbędnych kwerend. Parametry istotne dla wskaźników – konkluzyjność, rozziw i bilans, dla składników – spójność i zbieżność, a dla czynników i wyników – selektywność, zawilość i trafność nie są potrzebne na innych poziomach analizy zmiennych. Im bardziej miary są uogólnione, tym bardziej są one wiarygodne w budowaniu teorii ogólnych. lecz wnioskowanie na potrzeby praktyki i podejmowanie decyzji dotyczących oddziaływań lokalnych oprzeć trzeba na parametrach bardziej szczegółowych.

## **Jak interpretować parametry – analiza ilościowa**

Obszerne objaśnienia każdej z osobna miary statystycznej, stosowanej w ewaluacji za pomocą skal dwuważonych, przedstawiłem w odrębnym dziale. Tutaj skoncentrujemy się na ogólnych zasadach odczytywania istoty parametrów w połączeniu z charakterystyką różnych metod analizy ilościowej. Każdy z obliczonych parametrów pełni bowiem określoną rolę w odmiennych strategiach analitycznych. W jednych metodach może mieć znaczenie kluczowe, a w innych marginalne. Co więcej – pełna interpretacja parametru wymaga odniesienia do zmiennej, jaką dana miara reprezentuje. Zagadnienie łącznej interpretacji oznak w kontekście objaśniania zjawisk empirycznych wymaga jednak wstępnego przygotowania do percepcji sensu czysto statystycznego. Trzeba starać się zrozumieć, jakiego typu jest dany parametr i co wyraża:

- czy poziomy, informujące o chwilowym potencjale lub natężeniu cechy;
- czy dyspersje, określające wzajemną rozbieżność opinii respondentów;
- czy wektory, wskazujące na siłę i trend zmian zachodzących w czasie;
- czy indeksy, wyznaczające względną dynamikę i efektywność procesu;
- czy wyznaczniki, charakteryzujące dobroć pomiaru bądź atrybuty prób.

Samo obliczenie wartości liczbowej parametru jest tylko wstępem do analizy statystycznej. Występują bowiem takie miary, które świadczą o pozytywnym zjawisku wtedy, gdy ich wartość nie wzrasta, lecz dąży do zera. Ponadto inny jest sens miar chwilowych, a inny dystansowych. Są parametry mianowane, lecz są też współczynniki bez jednostek. Pewne miary można interpretować geometrycznie, lecz niektórych nie da się poprawnie przedstawić na wykresie. Wszystko to sprawia, że interpretacja miar jest jednocześnie formą analizy.

*Analiza pozycyjna* miar chwilowych polega na szacowaniu wartości poziomów kalibrowanych umownie, jako że ewaluacja dotyczy cech niewymiernych. Na kontinuum natężenia cech ustala się progi, którym nadawane są wagi. Najczęściej obliczanym parametrem pozycyjnym jest średnia arytmetyczna z wag. Ze względu na swoistą niewymierność średnie poziomy muszą być odnoszone do miar normujących (standardów ewaluatywnych) lub do skryptów oceniania (np. skali stopni szkolnych). W interpretacji procesów lokalnych wystarcza porównywanie poziomów uzyskanych przez poszczególne próby (ryc. 9 i 10).

*Analiza rozkładowa* polega na ocenie zjawiska formowania się standardów wartości na podstawie zróżnicowania wypowiedzi. Bazowymi parametrami są odchylenia od średniej. Im większe odchylenia, tym bardziej respondenci różnią się w opiniach. Bez zróżnicowania opinii problem nie byłby wart badania, a nadmiernie duży rozrzut świadczyłby o nieuformowanej jeszcze wartości. Rozkłady często analizuje się na wykresach liczebności (np. histogramach). Interpretacja polega wówczas na ocenie charakterystyki spektrum wypowiedzi na podstawie graficznego układu skupisk wokół opcji wyboru (ryc. 2, 5 i 18).

*Analiza gradientowa* jest graficznym sposobem określania kierunku zmian. Wszystkie wskaźniki odwzorowane są w układzie biegunowym, gdzie dodatkowo nadaje się im zwrot i kierunki zależne od cechy, jaką próbują. Na tym samym wykresie zestawione są poziomy średnie i wysokie z obu pomiarów. Analiza polega na porównywaniu stanów 'przed' i 'po', przy czym ocenia się równomierność i nasilenie zmian oraz określa zwrot wektora wypadkowego. Gradient wyznacza kierunek największego przyrostu, zwróconego w stronę dominujących w procesie cech ogólnych, mianowanych na wykresie (ryc. 34).

*Analiza różnicowa* polega na interpretacji parametrów zmienności z oceną intensywności zmian. Kluczowymi statystykami są miary tendencji, przy czym najłatwiejsza do odczytania znaczeń jest trójka różnic na poziomach HML. Porównuje się miary pozycyjne z pierwszego i drugiego pomiaru tych samych prób badawczych, objaśniając zwłaszcza trendy niekorzystne, niewspółbieżne. Chodzi tu o ocenę harmonii zmian cech u respondentów o wyższym i niższym poziomie indykatum, co ułatwiają miary korelacji i zawilości. Brak harmonii może pośrednio świadczyć o niedopasowanym oddziaływaniu (ryc. 36 i 38).

*Analiza indeksowa* jest procedurą umożliwiającą porównywanie rezultatów uzyskiwanych przez grupy silnie zróżnicowane. Podstawowymi parametrami są tu względne indeksy przyrostu i spadku. Ich wypadkowa jest miarą efektów wszelkich oddziaływań. Indeksy statystyczne interpretuje się jako niezależne od stanu początkowego, toteż bez względu na pierwotne cechy respondentów każda z prób badawczych ma szansę na uzyskanie porównywalnych wzrostów, możliwych do wyrażenia procentowego. Strategia indeksowa oparta jest na analizie fluktuacji elementarnych wypowiedzi (zob. ryc. 39, 40 i 41).

Specyficzną kategorię miar stanowią parametry kwantyfikacyjne i wyznaczniki jakości pomiarowej. Nie służą one do wartościowania indykatum, lecz do weryfikacji i optymalizacji narzędzia pomiarowego, a także do szacowania trafności i rzetelności badawczej. Każda z tych miar ma zupełnie inną interpretację statystyczną i zawężone, wysoce specjalistyczne zastosowanie, dlatego ich sens szczegółowo objaśniany jest wraz z opisami strategii metodologicznych oraz w wyodrębnionym dziale zamykającym część statystyczną.

<i>Ślady</i>		<i>Wzorce</i>
nadawanie rang i wag	.....	depolaryzacja wskaźników
łączenie wskaźników	.....	agregacja zmiennych w oznaki
strukturyzacja oznak	.....	budowa zasobów tabelarycznych
kumulacja klas i grup	.....	kondensowanie prób do ogółu
selekcja i redukcja miar	.....	abstrahowanie od nieistotności
interpretacja miar chwilowych	.....	analiza pozycyjna i rozkładowa
interpretacja miar zmienności	.....	analiza różnicowa i indeksowa

## **Jak odczytywać wykresy – wizualizacja oznak**

Przedstawianie parametrów statystycznych w postaci graficznej stanowi ważną formę nie tylko prezentacji rezultatów, lecz powinno być również integralną procedurą podczas eksploracji i analizy zasobów empirycznych. Chodzi o to, że poszukiwanie znamiennych oznak, które są podstawą formułowania wniosków, znacznie łatwiej jest przeprowadzać na wykresach niż w tabelach. Zasada ta dotyczy większości miar, z wyjątkiem tych, których nie da się poprawnie przedstawić w formie graficznej (np. wariancji). Są jednak też takie rodzaje wykresów (np. rozkłady), bez analizy których w ogóle trudno się obyć. Z tego względu konieczna jest umiejętność właściwego odczytywania znaczeń wyrażonych wizualnie. Każdy wykres jest komunikatem niosącym umowne kody, które muszą być jednoznacznie rozumiane. Ma to znaczenie zwłaszcza w sytuacji automatycznego generowania wykresów przez programy komputerowe. Od osoby obsługującej komputer zależy: czy rodzaj wykresu pasuje do typu parametrów, czy osie są poprawnie wymiarowane i optymalnie skalowane, czy wygenerowany komunikat odczytywany będzie zgodnie z intencją.

Korzyści graficznej analizy parametrów mają swe źródło zarówno w ogólnych zaletach wizualizacji, jak też w szczególnych, wysoce specjalistycznych formach obrazowania zjawisk statystycznych. Ważna jest zwłaszcza szybkość recepcji wzrokowej. Z układów przestrzennych w jednej chwili odczytywane są tendencje i kierunki zmian pozytywnych i negatywnych, a także wzajemne rozkłady opinii zgodnych i niezgodnych. Istotną rolę odgrywa również pamięć wzrokowa, która ułatwia wykrycie różnic na wykresach aktualnie oglądanych i wcześniej widzianych. Wszelkie porównania, tak ważne w analizie zasobów empirycznych, możliwe są dzięki składaniu wielu parametrów na wspólnym wykresie. W ten sposób uwidaczniane są m.in.: pary kontrolne wskaźników, relacje stanu początkowego i końcowego, różnice międzygrupowe, szeregi lub cykle czasowe. Wszystko to pomaga dostrzec wielorakie oznaki, stanowiące niezwykle cenne przesłanki do formułowania wniosków z badań. Niezbędna jest jednak recepcja i percepcja wykresów, a więc umiejętność odczytywania kodów obrazowych, jak też zdolność do interpretowania sensu merytorycznego.

➤ **Recepcja wykresu** to odbiór warstwy syntaktycznej zawartej w obiektach tworzących dany rysunek. Jako pierwsze odczytuje się elementy należące do takich kanonów składniowych, jak: podpis pod wykresem i opisy wewnątrz, legenda, rodzaj układu współrzędnych oraz wymiarowanie i orientacja osi. Następnie rozpoznać należy to, co stanowi istotę odwzorowania parametru: położenie punktów, długości odcinków lub słupków, usytuowanie pasemek, skupienie obiektów, kształty krzywych, kąty nachylenia, pola powierzchni, zwroty i kierunki wektorów bądź jeszcze inne specyficzne dla wykresów formy. Na tym etapie postrzega się przede wszystkim oznaki kwantytatywne.



➤ **Percepcja wykresu** to interpretacja warstwy semantycznej niesionej przez ów swoisty komunikat wizualny. Z dobrze opracowanych wykresów odczytać można symptomy kwalitatywne, służące do wnioskowania wartościującego. Jednak odczytanie komunikatu i wnioskowanie zależą już w znacznej mierze od odbiorców. Dla jednych widoczne na wykresie poziomy, przyrosty i efekty wydawać się będą odpowiednio wysokie, a dla innych zbyt niskie. Co gorsza – trudność interpretacji polega na tym, że w pewnych sytuacjach spadki wcale nie muszą być symptomami negatywnymi (np. zjawisko tonowania wypowiedzi pierwotnie zbyt impulsywnych w miarę uświadamiania wartości rzeczywistej). Z tego powodu w percepcji wykresów pomocne są zwięzłe opisy zewnętrzne oraz tabele normujące skalę pomiarową. Istotą interpretacji wartościującej jest odnośnienie do standardów ewaluacyjnych, a odczytywanie sensu jakościowego na wykresie rozpoczyna się od identyfikacji zmiennych. Konieczne jest rozpoznanie: co jest argumentem, a co wartością statystyki; na których osiach mianowane są zmienne lub przypadki; czy miana osi są ciągłe, czy dyskretne; czy wartości przyrastają w poziomie, czy w pionie; czy jest to miara chwilowa, czy dystansowa; co przedstawia przebieg funkcji, jaka jest dynamika, jakie są trendy i efekty, a co najważniejsze – jakie są wzajemne układy elementów. To wszystko musi być interpretowane łącznie z definicyjnym sensem zmiennych.

Zagadnienia interpretacji przedstawiam szczegółowo w odrębnym dziale, tutaj scharakteryzuję rodzaje wykresów związanych ze skalą dwuwazną:

*Wykres punktowy* ilustruje średnie poziomy cech kalibrowane na skali ocen. Obok siebie umieszczone są dwa parametry chwilowe z pomiarów 'przed' i 'po'. Umożliwia to porównanie stanu początkowego z końcowym, a ponadto na tym samym rysunku zestawione są zmienne częściowe lub rezultaty różnych grup. Punkt usytuowany wyżej oznacza odpowiednio wyższy poziom wartości cechy, a nachylenie linii łączącej punkty obrazuje dynamikę zmian (ryc. 9, 66, 72).

*Wykres odcinkowy* jest wizualizacją trzech poziomów HML, oznaczających progi podziału respondentów na frakcje o wysokim, średnim i niskim poziomie cech. Odcinki rozmieszczone są pionowo na skali ocen w parach 'przed' i 'po', przy czym długość odcinka wyznacza prawdopodobny rozrzut ok. 68% przypadków. Wykres ten umożliwia analizę harmonii zmian, np. czy oddziaływania edukacyjne dopasowane były do słuchaczy mniej i bardziej zdolnych (ryc. 36 i 67).

*Wykres pasmowy* ukazuje realny rozrzut kwalifikowanej większości i rozziw wypowiedzi. Długość pasemek rozpostartych poziomo na skali wag określa zróżnicowanie opinii, natomiast ich usytuowanie w strefach zgodności bądź niezgodności ze standardem ewaluacyjnym ilustruje natężenie akceptacji tez zawartych w bodźcach-stwierdzeniach. Umożliwia to graficzne rozpoznawanie konkluzyjności poszczególnych wskaźników. Dodatkowe czarne pole wyznacza różnicę pomiędzy średnimi jakości i intensywności wypowiedzi (ryc. 75, 68).

*Wykres rozkładu* ilustruje najczęściej liczebności grup respondentów, którzy wybrali te same opcje wypowiedzi (histogram, ryc. 18). W uogólnieniach można doprowadzić do krzywej ciągłej, której kształt jest najistotniejszym przesłaniem wizualizacji (rozkład, ryc. 19). Na podstawie charakterystyki rozkładu podejmuje się ważne decyzje podczas tworzenia skali pomiarowej, a podczas analizy – testy istotności opiera się na prawdopodobieństwie wynikającym z rozkładu.

*Wykres biegunowy* jest zbiorczą reprezentacją wszystkich wskaźników skali. W zależności od jakościowo-definicyjnych cech wskaźników nadaje się im odpowiedni zwrot i odrębny kierunek na współrzędnych biegunowych. Długości wektorów wiodących wyrażają średnie i wysokie poziomy cech elementarnych, podczas gdy osie i bieguny mianowane są za pomocą cech ogólnych. Graficznie analizuje się rozkład warstw i ustala wektor wypadkowy (ryc. 34).

*Wykres słupkowy* jest uogólnionym odwzorowaniem zachodzących przemian: tendencji rozwojowych i dochodzenia do konsensusu, toteż obok siebie znajdują się dwie statystyki: *Trendu* i *Bilansu*. Poziome słupki zorientowane są w prawo lub w lewo, zależnie od trendów dodatnich lub ujemnych. O intensywności zmian świadczą długości słupków, jednak pamiętać należy o tym, że *Bilans* zazwyczaj jest pozytywny także przy wartości zerowej (ryc. 70, 82).

*Wykres wektorowy* ukazuje cztery parametry fluktuacji wypowiedzi, niezbędne do analizy indeksowej. Każdą zmienną reprezentują dwa wektory, skierowane w górę i w dół, liczone odpowiednio dla przyrostów i spadków. Suma długości wektorów ilustruje intensywność oddziaływań prowadzących do zmian cech mentalnych u respondentów, natomiast różnica wyznacza dynamikę (ryc. 41). Indeksy te umożliwiają porównywanie prób niehomogenicznych (ryc. 83).

### **Jak interpretować symptomy – analiza jakościowa**

Swoistą właściwością dobrej, standaryzowanej skali pomiarowej jest to, że wszystkie jej bodźce-stwierdzenia skupione są wokół próbkowania jednej, najbardziej uogólnionej zmiennej jakościowej. Wyniki pomiaru zmian postaw i świadomości (a pośrednio także jakości zachodzących procesów), wyrażone na poziomie zmiennej ogólnej, są najbardziej wiarygodne. Jednak rozpoznanie wyłącznie poziomu ogólnego byłoby stwierdzeniem faktu: *Jak jest?* a nie: *Dlaczego tak jest?* Prawidłowa interpretacja rezultatów ewaluacji splotowej wymaga analizy jakościowej na każdym poziomie oznak i dopiero na tej podstawie dokonuje się generalizującej syntezy symptomów wartości indykatum.

W procesie oceniania wartościującego niezbędne jest stosowanie strategii relatywistycznych. Wysokie/niskie wartości parametrów statystycznych same w sobie nie są jakością, podobnie jak przyrosty/spadki wcale nie muszą być automatycznie pozytywne/negatywne. Oznaki empiryczne mogą jednak stać się symptomami wartości zjawisk, jeśli odczytuje się je w ujęciu względnym.

*Względność poziomów* jest typową właściwością skali kalibrowanej umownie. Poziom zmiennej może być uznawany za wartość tylko wtedy, jeśli przekracza próg ustalony jako norma dla homogenicznej kategorii respondentów. Każda elementarna teza skali ma swój graniczny poziom akceptacji, do którego dążą średnie z licznych prób. Nadmierna zatem, skrajnie bezkrytyczna aprobata może być też zinterpretowana jako przejaw niepełnej świadomości problemu.

*Zróżnicowanie opinii* stanowi wartość jedynie w ściśle ustalonych warunkach. Po pierwsze – tezę zawartą w bodźcu-stwierdzeniu akceptować musi kwalifikowana większość respondentów, lecz – po drugie – ze względu na doniosłość problemu nie powinna to być aprobata zdecydowana i jednomyślna. Istnieje więc pewien jakościowo optymalny rozkład wypowiedzi, świadczący o badaniu ważkich zagadnień, wokół których konstytuują się standardy ewaluatywne.

*Spójność wypowiedzi* odnosi się do wewnętrznej zgodności przypadków. Jawne przypadki unikania wypowiedzi (większość opcji 'brak zdania') dyskwalifikują respondentów. Jednak potencjalna niespójność kontrolnych par wskaźników jest zjawiskiem naturalnym, wynikającym z różnych perspektyw próbkowania czasoprzestrzennego. O jakości wypowiedzi świadczy raczej miara dystansowa zbieżności, która powinna wskazywać na proces spajania własnych opinii.

*Zmienność wypowiedzi* w drugim pomiarze jest wartością świadczącą o dynamice procesów, a zatem stagnacja może być uznana za zjawisko negatywne. Ogólnie oczekuje się przyrostów, lecz w pewnych wyjątkach spadek poziomu pierwotnie nadmiernego może być oznaką pozytywnego dojrzewania. Wartość zmian musi być ustalana w odniesieniu do norm początkowych i końcowych.

*Współbieżność przemian* dotyczy harmonijnej transpozycji w trzech frakcjach prób rozwarstwionych na osoby o wysokim, średnim i niskim poziomie badanych cech mentalnych. Wartością procesów oddziaływania jest takie dopasowanie metod i treści, ażeby każda z frakcji odniosła korzyści, niezależnie od pierwotnego poziomu. Przejawami niedopasowania są tendencje rozbieżne.

*Fluktuacja wypowiedzi* sama w sobie jest wartością oznaczającą, że zadziałały czynniki modyfikujące postawy i świadomość. Ponadto wahania wypowiedzi potwierdzają, że drugi pomiar identycznym narzędziem jest w pełni poprawny metodologicznie. Także i tu oczekuje się efektów wzrostowych, jednak zbędne jest odnoszenie do norm, gdyż indeksy są parametrami samonormującymi.

*Sezonowość zjawisk* sprawia, że poprawne wartościowanie zakłócać jest przez chwiejne stany takich cech, które z natury mają cykliczne, niemonotoniczne przebiegi (jak np. poziomy motywacji). Do oceny wartości tych komponentów indykatorum należy podchodzić ze szczególną ostrożnością – ewentualne spadki motywacji mogą być bowiem normalnym następstwem zaspokojenia aspiracji, a nie symptomem wadliwego procesu oddziaływań. Poprawną interpretację ułatwia wskaźnik buforowy, który powinien mieć wartość w miarę stałą.

Relatywizm wyznaczanych wartości powoduje, że podczas przechodzenia od analizy oznak empirycznych do wartościującej interpretacji symptomów, ewaluator musi podjąć ostateczne decyzje w kilku zasadniczych kwestiach:

*Co konstytuuje normy wartości?* Strategia zakłada oparcie się na standardach ustalonych empirycznie, o ile są one zgodne z oceną sędziów kompetentnych. Wykorzystać można normy rzeczywiste, wyznaczone na podstawie ograniczonych liczebnie prób, bądź estymatory teoretyczne, oczekiwane dla populacji. Pamiętać jednak należy, że standardy te ulegają powolnej dezaktualizacji.

*W czym przejawia się wartość?* O ile jakość to takie kategorie nieliczbowe, jak np.: zgodność/zróźnicowanie, spójność/niespójność, zbieżność/rozbieżność, o tyle wartość jest nieskrajnym optimum na osi przeciwieństw. Niezbędne jest zróźnicowanie opinii, lecz zgodność co do standardów. Wypowiedzi byłyby spójne, jednak aspekty czasoprzestrzenne tworzą niespójność. Składniki są zbieżne, choć czynniki rozbieżne. Skrajności świadczą zatem raczej o braku wartości.

*Jakie cechy tworzą wartość?* Częściowo wynika to z rozwarstwienia indykatum. Jeśli ewaluacja dotyczy cech mentalnych, to najczęściej poszukuje się jakości w sferach: afektywnej, kognitywnej i behawioralnej. Pożądane są przychylne postawy, świadomość i dyspozycyjność. Dodatkowymi wartościami mogą być: wnikliwość w treść bodźców-stwierdzeń, logika wypowiedzi i trafność opinii.

*Jaką dynamikę winny mieć cechy?* Komponenty indykatum mają zróźnicowaną podatność na zmiany. Najszybciej przyrastać powinny wskaźniki poznawcze, natomiast dyspozycje do działań mogą pozostawać na stałym poziomie. Oceny afektywne, pierwotnie zawyżane, w drugim pomiarze mogą obniżyć się do norm, co nie jest negatywne, podobnie jak nieznaczny spadek poziomu motywacji. Każdą ze zmian wartościować należy w odniesieniu do miar normujących.

*Co było inspiracją wypowiedzi?* Opcje wybierane przez respondentów zależą od treści niesionej przez bodziec-stwierdzenie, a w szczególności od tego, czy zostanie uruchomiony gotowy skrypt reakcji, czy raczej tworzenie wypowiedzi. Wskaźniki próbujące uświadomienie mają niższy poziom akceptacji, lecz są bardziej wartościowe poznawczo niż te, które wynikają z reakcji impulsywnej.

*Co stymulowało dyspozycje?* Ocenę wartości komponentów wolicjonalnych uzależnić należy od tego, czy dany pomiar dokonywany był w fazie potencjalnego pobudzenia, czy raczej w fazie zaspokojenia aspiracji. Utrzymanie np. wysokiej gotowości do uczenia się pod koniec kursu rodzi wątpliwość, czy oddziaływanie było aż tak stymulujące, czy może nie zostały spełnione oczekiwania.

*Czy relacje są zgodne z modelem?* Dwukrotny pomiar umożliwia analizę zmian komponentów pod kątem wzajemnych zależności. Zachodzące relacje pomiędzy parami składników, a zwłaszcza czynników, stanowią najistotniejszą przesłankę wartościowania. Jeśli model teoretyczny dopuszcza niewspółbieżność, to potwierdzenie empiryczne jest mimo wszystko symptomem pozytywnym.

## **Jak formułować wnioski – synteza rezultatów**

Opracowanie wyników ewaluacji wiąże się ściśle z udzieleniem wyczerpujących odpowiedzi na pytania zawarte w projekcie badawczym (str. 96÷106). Najlepszą formą sprawozdania jest przedstawienie opisu symptomów wartości łącznie z tabelami parametrów i wykresami oznak. W zależności od przeznaczenia sprawozdanie powinno być mniej lub bardziej zwięzłą syntezą jedynie najistotniejszych rezultatów empirycznych. W skrajnym uogólnieniu badania hipotez rezultatem może być sformułowanie twierdzeń (np. „Metoda czynnościowego kształtowania pojęć daje wysokie efekty”) lub zaleceń (np. „W procesie edukacji niezbędna jest harmonia struktur poznawczych i emocjonalnych”). Rezultaty pośrednie zazwyczaj potrzebne są jedynie ewaluatorowi, chyba że mają ilustrować metodologię bądź stanowią dowód potwierdzający hipotezę. Błędem byłoby odrębne omawianie wskaźników, gdyż służą one nade wszystko do kumulatywnego próbkowania przez skalę wielopozycyjną. Ich istotna rola ujawnia się głównie w fazie konstruowania narzędzia i w procesie weryfikacji jakości pomiarowej. Analiza spójności par kontrolnych jest jedną z metod, lecz sama w sobie nie stanowi celu ani problemu badawczego, a właśnie osiągnięte cele i rozwiązane problemy powinny stanowić treść sprawozdania z badań.

Ewaluacja splotowa ma jednak tę specyfikę, że dotyczy badań obiektów i zjawisk o złożonej strukturze. Z tego powodu rezultatami eksploracji są nie tylko wyniki ogólne, lecz także ich komponenty. W raporcie z badań powinny zatem zostać umieszczone wszelkie istotne oznaki, ujawnione na poziomach składników, czynników, rezultatów połówkowych i ogólnych. Ważne jest to, jakie komponenty wpłynęły na wyższe lub niższe poziomy indykatum, jakie były tendencje poszczególnych zmiennych i jak układały się wzajemne relacje. Na poziomie składników według modelu oczekiwana jest swoista zbieżność, podczas gdy na poziomie czynników dopuszczalna jest wycinkowa niewspółbieżność, choć i tak wzorcem pożądanym są przyrosty wartości zmiennych. Jakikolwiek przypadki niezgodnych z oczekiwaniami zjawisk empirycznych wymagają ustalenia przyczyn i objaśnienia ich w sprawozdaniu, natomiast symptomy typowe wystarczy zsyntezować i przedstawić jako fakty badawcze. Ażeby jednak nie były to jedynie artefakty, wnioskowanie oprócz trzeba na przesłankach walidacyjnych, warunkujących uogólnianie i upowszechnianie.

*Powtarzalność rezultatów* stanowi najcenniejszą przesłankę potwierdzania, że obserwowane zjawiska nie są przypadkowe. Najbardziej wartościowe jest potwierdzenie zewnętrzne przez innego badacza, który uzyska zbliżone relacje pomiędzy strukturami modelu. W praktyce ewaluacyjnej duże znaczenie mają pomiary cykliczne, które umożliwiają porównywanie efektów oddziaływań tego samego wykładowcy na kolejne roczniki słuchaczy. Podobną moc walidacyjną ma powtarzalność uzyskiwana w równoległych grupach homogenicznych.

*Wiarygodność wniosków* ma swe źródło w prawidłowości metodologicznej oraz w odpowiedniości zastosowanych metod i technik. Poprawność wnioskowania trudno jest zweryfikować, gdyż statystyczne wyznaczniki jakości pomiarowej są jedynie przesłankami pośrednimi, same w sobie wymagającymi poprawnej interpretacji. Zdarza się, że autorzy raportów podają miary rzetelności bądź trafności zupełnie automatycznie, nie zdając sobie sprawy z całkowitej bezzasadności w konkretnej sytuacji badawczej<sup>7</sup>, a przecież odpowiedniość miar ma dużo większe znaczenie niż ich popularność (zob. str. 200). W ewaluacji skalą dwuwazoną podstawy wiarygodności buduje się głównie na specjalnie opracowanych wyznacznikach *konkluzyjności* i *trafności*. Pierwszy wyznacza ważność kwestii i moc wnioskowania, a drugi weryfikuje parametry poprzez porównanie dwóch różnych dróg dochodzenia do wyników. Oznacza to rodzaj kontroli powtarzalności, podobnej w istocie do pomiarów dwiema metodami.

*Zasięg uogólniania* odnosi się do możliwości generalizacji wniosków na obszar szerszy niż teren badań oraz do formułowania standardów oczekiwanych od populacji. Dopuszczalność takich działań uzależniona jest od reprezentatywności i liczności prób badawczych. Naczelną zasadą jest to, że nie ocenia się rezultatów pojedynczych, lecz przypadki kumuluje się w klasy, grupy i ogół. Takie poziomy uogólnień są wystarczające w ewaluacji lokalnej, gdy rezultaty odnosi się wyłącznie do badanych prób. Uogólnienie wniosków na populację i budowanie teorii na rezultatach empirycznych wymaga opisanie w raporcie metodologicznych podstaw generalizacji i dowiedzenia, że spełniono wszystkie jej uwarunkowania. Samo podanie liczebności prób jest niewystarczające.

*Zasięg upowszechniania* określa krąg potencjalnych odbiorców raportu z badań. Najbardziej przejrzystą sytuacją jest ewaluacja na potrzeby lokalne, ponieważ sformułowanie trafnych wniosków z własnych doświadczeń jest łatwiejsze niż interpretacja wyników cudzych. Wnioski deskryptywne i zalecenia formatywne mogą być jednak przydatne także dla innych. Wymiana doświadczeń i kumulowanie danych to działania szczególnie pożądane w ewaluacji za pomocą skal standaryzowanych, których sposób interpretacji jest udostępniony. W takiej sytuacji raport może mieć formę sygnałną. Upowszechnianie badań innowacyjnych obliguje natomiast do publikacji w postaci kompletnej monografii.

<i>Ślady</i>		<i>Wzorce</i>
odczytywanie wykresów	.....	konwencja wizualizacji statystyk
interpretacja wykresów	.....	analiza graficzna oznak
interpretacja zjawisk	.....	analiza jakościowa symptomów
kwalifikacja zjawisk	.....	synteza wartościująca indykatum
weryfikowanie wyników	.....	analiza jakości pomiarowej
formułowanie wniosków	.....	uogólnianie i upowszechnianie

## ZMIENNE EWALUACJI INFOKULTURY

W procesie ewaluacji podejmuje się próbę wartościowania czegoś, co nie tylko wydaje się niemierzalne, lecz jest także trudne do nazwania. Bierze się to z dalekiej drogi pomiędzy tym, co chcemy zbadać, a tym, co daje się mierzyć. Kliniknym przykładem jest choćby *ewaluacja jakości pracy szkoły*. Na takim poziomie sformułowane zadanie jest w pełni przejrzyste i zasadne, jednak trafne wykonanie tegoż zadania w polskich szkołach graniczy z cudem. Ministerstwo podało bowiem jedynie ogólne cele, mylnie nazywając je „wskaźnikami”, natomiast nie podało żadnego wzorca praktycznego wskaźnikowania. Gdyby sprawa była prosta dla fachowców, dostarczono by gotowe narzędzia, tymczasem wykonanie zadania zrzucano na nauczycieli, doszkalanych przez quasi-fachowców. Nikłym tego efektem jest... opasła sprawozdawczość.

Zasadniczy problem poprawności ewaluacji tkwi w definiowaniu, operacjonalizacji i wskaźnikowaniu wyznaczników indykatum, zwanych *zmiennymi*:

**Zmienna** to każda interesująca badacza właściwość struktury indykatum, której podczas próbkowania podmioty ewaluacji nadają określoną wartość. Struktura zmiennych wyprowadzana jest z modelu teoretycznego, a badane właściwości jej komponentów muszą mieć zdefiniowany sens empiryczny.

Prześledźmy na przykładzie *InfoKultury* drogę przechodzenia od zadania badawczego do zmiennych. Wśród bardzo wielu czynników wpływ na ocenę pracy szkoły ma *jakość zajęć* (w tym m.in. informatycznych). Jednym z wielu sposobów oceny zajęć może być badanie zmian cech mentalnych, jakie zaszły u uczniów. Oznacza to, że przedmiot ewaluacji jest mierzony pośrednio przez indykatum związane z procesami umysłowymi, a te wymagają zdefiniowania operacyjno-strukturalnego<sup>39</sup>. Wyodrębnia się i doprecyzowuje komponenty tego, co ma być szukane, buduje się model relacyjny i opracowuje wskaźniki. Następnie – dzięki pomiarom – zmienne „szukane”, będące pierwotnie tylko pojęciami teoretycznymi, stają się zmiennymi „wskazywanymi”, posiadającymi wartość empiryczną. Wskaźniki elementarne stanowią liczbowe argumenty zmiennej losowej, lecz służą także do syntezy wyznaczników jakościowych.

### Wskaźniki – argumenty zmiennej losowej

Zmienne bazowe, tworzące najniższy poziom ekstrakcji, mają specyficzne właściwości. Po pierwsze – odzwierciedlają bezpośrednio wypowiedzi respondentów, stając się w obranej metodzie wskaźnikami pomiarowymi. Po drugie – są podstawowymi argumentami funkcji w analizie statystycznej. Po trzecie – tworzą pary kontrolne, służące do analizy aspektów czasoprzestrzennych. Z właściwości pierwszej wynika możliwość odczytywania sensu jakościowego każdego wskaźnika, poprzez analizę treści merytorycznej bodźca-stwierdzenia, w połączeniu z natężeniem społecznej akceptacji tejże elementarnej tezy. Na poziomie analizy statystycznej większość parametrów liczbowych powiązana jest z jakością komponentów indykatum, lecz niektóre miary tracą ów sens merytoryczny, wyznaczając jakość wyłącznie statystyczną (np. rzetelność). Ponadto, na poziomie wskaźników możliwa jest analiza bez odczytywania sensu treściowego, dlatego nie ma potrzeby definiowania każdej pozycji skali co do jej interpretacji jako części indykatum. Niezbędne jest natomiast podanie schematu interpretacji w kategoriach aspektów pobudzenia i czasoprzestrzeni.

Schemat interpretacji wskaźników Kwestionariusza Kultury Informatycznej Tab. I

Numer i polaryzacja	Wskaźnik	Rodzaj pobudzenia	Aspekt czasowy	Aspekt trwałości	Aspekt osobowy	
1	-	afektywny	impuls	retrospekcja	chwilowość	introspekcja
2	-	afektywny	impuls	futurospekcja	stałość	ekstraspekcja
3	+	kognitywny	namysł	retrospekcja	chwilowość	introspekcja
4	+	kognitywny	namysł	retrospekcja	stałość	introspekcja
5	-	kognitywny	impuls	retrospekcja	chwilowość	introspekcja
6	+	kognitywny	impuls	futurospekcja	stałość	introspekcja
7	+	kognitywny	namysł	futurospekcja	chwilowość	ekstraspekcja
8	+	kognitywny	namysł	futurospekcja	stałość	ekstraspekcja
9	-	afektywny	namysł	futurospekcja	chwilowość	introspekcja
10	-	afektywny	namysł	retrospekcja	stałość	introspekcja
11	+	afektywny	namysł	retrospekcja	stałość	introspekcja
12	+	afektywny	namysł	futurospekcja	chwilowość	introspekcja
13	-	afektywny	impuls	retrospekcja	chwilowość	introspekcja
14	+	afektywny	impuls	futurospekcja	stałość	ekstraspekcja
15	-	kognitywny	impuls	futurospekcja	stałość	ekstraspekcja
16	+	kognitywny	impuls	retrospekcja	chwilowość	introspekcja
17	+	kognitywny	namysł	retrospekcja	chwilowość	introspekcja
18	-	kognitywny	namysł	futurospekcja	stałość	ekstraspekcja
19	+	kognitywny	impuls	retrospekcja	chwilowość	ekstraspekcja
20	+	kognitywny	impuls	retrospekcja	stałość	ekstraspekcja
21	+	afektywny	impuls	futurospekcja	chwilowość	introspekcja
22	+	afektywny	impuls	futurospekcja	stałość	introspekcja
23	-	afektywny	namysł	retrospekcja	chwilowość	introspekcja
24	-	afektywny	namysł	futurospekcja	stałość	ekstraspekcja



**Wypowiedzi – wartościowanie stwierdzeń**

Na potrzeby analizy statystycznej wykorzystuje się wagi liczbowe przypisane poziomom akceptacji też zawartych w bodźcach-stwierdzeniach. Na tej płaszczyźnie wystarczy pamiętać o depolaryzacji wskaźników, które wyrażają antytezę. Poziomy akceptacji mogą być jednak także podstawą analizy jakościowej, jeśli weźmie się pod uwagę sens merytoryczny danego stwierdzenia. Zasadniczo najkorzystniej jest interpretować i jednocześnie porównywać pary kontrolne wskaźników. Daje to pełniejsze przesłanki wartościowania aspektów osobowych i czasoprzestrzennych. Oto przykłady interpretacji wypowiedzi:

1. Odczuwam raczej niechęć do komputeryzacji. *nie*
2. Tylko uzdolnieni do komputera skorzystają z jego zalet. *nie*

Ze względu na polaryzację ujemną zaprzeczenie stanowi wartość pozytywną. Gdyby jednak respondenci wybierali opcję '*tak*' przy pierwszym stwierdzeniu, byłoby to mniej niepokojące niż przy stwierdzeniu drugim, gdyż wskaźnik nr 1 ujawnia zjawisko chwilowe, a nr 2 uzewnętrznia cechę bardziej trwałą, która ponadto w aspekcie osobowym mogłaby wyrażać poczucie wyalienowania.

5. Czuję się niepewnie, gdy siadam przy komputerze. *raczej tak*
6. Z pewnością będę w stanie posługiwać się komputerem. *tak*

Jeśli nawet wskaźnik nr 5 wypada niekorzystnie, to jest on jedynie objawem stanu tymczasowego, który w przyszłości powinien ulec radykalnej zmianie, co wynika z deklaracji ujawnionej przez futurospekcyjny wskaźnik nr 6.

9. Bał(a)bym się zostać sam na sam z komputerem. *absolutnie nie*
10. Jest mi źle w grupie, gdy inni lepiej znają komputery. *raczej tak*

Taki układ wypowiedzi świadczyłby o preferencji uczenia się samodzielnego, co jednak przeczyłoby korzystnej strategii rozpoznawania obsługi komputera przez dwóch uczniów przy jednym urządzeniu, z zachowaniem równomiernej rotacji i pełnieniu funkcji *inquirera* (osoby naprowadzającej przez pytania). Oczywiście zaawansowane umiejętności lepiej jest zdobywać indywidualnie.

13. Wątpię w przydatność takich kursów komputerowych. *absolutnie nie*
14. Nawet obcy z komputerem odniosą z kursu pożytek. *tak*

Para ta stanowi znamieny przykład kontrolowania aspektów osobowych. Obie wypowiedzi, choć mają różne wagi (+3 i +2), to w ujęciu jakościowym odzwierciedlają podobną wartość. Łatwiej bowiem zdecydowanie wypowiedzieć się o sobie (nr 13) niż o innych (nr 14). Gdyby jednak to drugie stwierdzenie było negowane, to świadczyłoby o zbyt niskim poziomie merytorycznym kursu.

17. Już dziś wiem, w czym komputer jest niezastąpiony. *tak*
18. Komputer może radzić, jak rozwiązywać życiowe problemy. *tak*

Akceptacja obydwu powyższych stwierdzeń jest przejawem niespójności wypowiedzi. Deklaracja, że się coś wie, jest sprzeczna ze wskaźnikiem testującym wiedzę. Komputer jako maszyna (*hardware*) nie może przecież niczego radzić<sup>40</sup>.

23. Obawiam się uzależnienia od komputera. *nie*  
 24. Pewnego dnia komputery zniewolą ludzi. *raczej tak*

Brak obaw o uzależnienie się jest wyrazem własnej odporności. Jednak do pary kontrolnej desygnowałem kontrowersyjne stwierdzenie, które nie ma wyraźnej polaryzacji wartości. Bierze się to stąd, że wskaźnik nr 24 jest prognostykiem stanu przyszłego, a ponadto bada kwestię dotyczącą innych. Taki rodzaj stwierdzenia jest najtrudniejszy do formułowania wypowiedzi. Oczekuje się zaprzeczenia, lecz świadomość coraz liczniejszych uzależnień od komputera skłania do akceptacji przepowiedni z gatunku *science fiction*.

### **Składniki – komponenty zmiennych cząstkowych**

Wartościowanie pojedynczych wskaźników dostarcza cenne przesłanki do wnioskowania, jednak nie jest wystarczająco wiarygodne. Z tego powodu zasadniczą procedurą eksploracyjną jest agregacja wskaźników. Pierwszym poziomem agregacji jest łączenie par kontrolnych w zmienne cząstkowe zwane *składnikami*. To właśnie one stanowią podstawowe komponenty indykatum. Obydwa wskaźniki tworzące składnik muszą próbować podobną właściwość, aby po ich złączeniu możliwe było zdefiniowanie wspólnego sensu merytorycznego, niezbędnego do interpretacji danych empirycznych. Definicje mogą być mniej lub bardziej rozbudowane, gdyż i tak sens przedmiotowy zawarty jest w treści stwierdzeń. Są one jednak niezbędne po to, aby udostępnić innym badaczom koncepcję twórcy skali, dotyczącą modelu ekstrakcji indykatum. Oprócz definicji każdemu składnikowi należy nadać zwięzłą nazwę.

Schemat interpretacji składników Kwestionariusza Kultury Informatycznej

Tab. II

Pozycje skali i polaryzacja	Składnik	Nazwa	Sens merytoryczny
13- α 14+	afektywny	<i>Ocena</i>	stopień akceptacji zajęć z informatyki
1- α 2-	afektywny	<i>Aplauz</i>	wyraz impulsywnej aprobaty komputerów
11+ α 12+	afektywny	<i>Ambicje</i>	aktualna chęć uczenia się o komputerach
21+ α 22+	afektywny	<i>Intencje</i>	zamiar używania komputerów w przyszłości
9- α 10-	afektywny	<i>Spokój</i>	brak stresu przed komputerem i wobec klasy
23- α 24-	afektywny	<i>Odpór</i>	odporność na uzależnienie się od komputera
15- α 16+	kognitywny	<i>Osąd</i>	uznanie przydatności i możliwości zastosowań
17+ α 18-	kognitywny	<i>Wgląd</i>	wnikliwość i trafność oceny zastosowań
3+ α 4+	kognitywny	<i>Obycie</i>	poziom własnych doświadczeń i zaradności
5- α 6+	kognitywny	<i>Pewność</i>	siła przekonania o swych umiejętnościach
7+ α 8+	kognitywny	<i>Zdolność</i>	ocena możliwości opanowania komputera
19+ α 20+	kognitywny	<i>Ogłada</i>	roztropność decyzji i działań wobec komputera

W ewaluacji splotowej model badawczy zawiera także relacje pomiędzy komponentami indykatum, przy czym owe zależności są istotnymi obiektami analizy jakościowej. Opis zachodzących relacji pomiędzy komplementarnymi parami składników ma na celu przybliżenie sposobu interpretacji zjawisk

empirycznych, a przy okazji jest egzemplifikacją tego, jak można redagować wnioski wartościujące. Stanowi cenne rozwinięcie schematu interpretacji.

#### **Ocena | Aplauz – tak uczą, bo trzeba**

Para ta ustala zależności pomiędzy postawami wobec instytucjonalnej edukacji informatycznej a komputeryzacji jako syndromu ogólnospołecznego. Potrzeba rozwoju technologicznego jest akcentowana bardzo silnie, toteż na tym tle ujawnia się uzależniona od faktycznej jakości ocena kursu, w którym badani uczestniczyli. Z kursami 'a priori' związane są nad wyraz duże nadzieje.

#### **Ambicje | Intencje – chcę się uczyć i używać**

Para ta określa związki między dyspozycjami do poznawania i działania. Ujawnia ona proporcje w dążeniu do pożytecznego wykorzystania komputera na tle użytkowania przyjemnościowego. Efektywne zastosowanie technologii informatycznych wymaga nabycia wiedzy, podczas gdy do zabawy wystarczy dysponowanie czasem. Spadek ambicji dotyczy głównie młodszych populacji.

#### **Spokój | Odpór – bez stresu i nałogu**

Relacja ta dotyczy równowagi emocjonalnej pomiędzy ewentualnym dyskomfortem, wynikającym ze strachu przy pracy z komputerem, a niepokojem przed możliwym uzależnieniem się. Jedni walczą ze stresem, aby usiąść przy maszynie, inni walczą z nałogiem i samym sobą, aby od niej wreszcie odejść. Stres ujawnia się w populacjach osób starszych, a nałóg..., gdzie popadnie.

#### **Osąd | Wgląd – wiem co i dlaczego**

Relacja ta wyznacza stan świadomości faktograficznej względem zdolności eksplikacyjnej. Wydawałoby się, że zasoby wiedzy i umiejętności objaśniania wzrastają wraz z wiekiem, lecz zasada ta wcale nie jest tak jednoznaczna przy kompetencjach z zakresu technologii informacyjnej. Wciąż jeszcze dorośli, czasem bezradni wobec komputera, korzystają z pomocy dużo młodszych.

#### **Obycie | Pewność – poćwiczę, to potrafię**

Ta para składników określa proporcje pomiędzy posiadanym poziomem doświadczeń informatycznych a samooceną potencjalnych dyspozycji do nabywania tego typu umiejętności. Chwilowe niedostatki umiejętności łatwiej można uzupełnić, jeśli podmiot procesu edukacyjnego jest przekonany o tym, że będzie w stanie osiągnąć cel i nie ma oporów przed podjęciem ćwiczeń.

#### **Zdolność | Ogłada – co mogę, a co wolno**

Ostatnia para komplementarna służy do ustalania równowagi pomiędzy oceną zdolności do szybkiego i samodzielnego kształtowania kompetencji informatycznych a roztropnością decyzji i działań podejmowanych podczas pracy z komputerem. O kulturze technicznej świadczy pewien rodzaj ogłady, wynikający ze świadomości zagrożeń powodowanych działaniami pochopnymi.

## Czynniki – dominanty zmiennych splotowych

Kolejnym poziomem agregacji jest łączenie wskaźników w zmienne zwane *czynnikami*. Każdy czynnik jest splotem czterech wskaźników, a także stanowi wypadkową wartość pary składników komplementarnych. Elementy tworzące czynniki zostały już zdefiniowane, dlatego ich sens merytoryczny – związany treściowo z konkretnym indykatem – odczytać można z meritum zmiennych niższego poziomu. Na poziomie czynników ważniejsze jest podanie schematu interpretacji, który definiuje struktury w ogólniejszych kategoriach taksonomicznych, istotnych badawczo. Świadczą o tym nazwy nadane czynnikom, z których część funkcjonuje w naukach społecznych jako pojęcia kluczowe.

Schemat interpretacji czynników Kwestionariusza Kultury Informatycznej

Tab. III

Pozycje skali i polaryzacja wskaźników	Czynnik	Nazwa	Kategoria taksonomiczna
1- 2- 13- 14+	afektywny	<i>Opinie</i>	praoceny wynikające z nastawienia
15- 16+ 17+ 18-	kognitywny	<i>Poglądy</i>	oceny wynikające z uświadomienia
11+ 12+ 21+ 22+	afektywny	<i>Motywacje</i>	predyspozycje do zaspokajania aspiracji
3+ 4+ 5- 6+	kognitywny	<i>Wprawa</i>	dyspozycje do działań optymalnych
9- 10- 23- 24-	afektywny	<i>Emocje</i>	predyspozycje do reagowania uczuciowego
7+ 8+ 19+ 20+	kognitywny	<i>Rozwaga</i>	dyspozycje do podejścia rozumnego

Czynniki wartościuje się poprzez odniesienie do standardów, a także poprzez analizę relacji splotowej pary komponentów afektywnego i kognitywnego.

### **Opinie | Poglądy – intuicja czy wiedza**

Komputeryzacji wciąż jeszcze towarzyszy podekscytowanie nowością, toteż w większości oceny tego zjawiska są przesycone emocjami, a przez to niezbyt adekwatne do stanu faktycznego. Opinie przede wszystkim odzwierciedlają nastawienia, a wynikają bardziej z internalizacji „nowomodnych” trendów (*Wszyscy wychwalają komputer, więc zapewne... musi on być wspaniały?!*) niżli z wiedzy o stanie rzeczy (*Wiem, w czym jest on nieodzowny!*). Podobnie jest z zajęciami informatycznymi – uczniowie z góry obdarzają je znacznym kredytem aprobowania, a w miarę przechodzenia przez proces kształcenia ów kredyt zaufania do nauczyciela zaczyna być rozliczany. Kiedy na lekcjach – i to tych najrzetelniej prowadzonych – komputer przestaje być zabawką, a staje się narzędziem pracy intelektualnej, zrozumiałe jest pewne osłabienie fascynacji. Ewentualny spadek poziomu ocen intuicyjnych podczas przejścia na oceny wynikające z uświadomienia nie może być jednak nadmierny. Zbyt optymistyczne opinie powinny raczej przejść w bardziej racjonalne poglądy<sup>41</sup>, choć osąd komputeryzacji staje się trafny dopiero po wnikliwym wglądzie w to zjawisko. Pomoc w kształtowaniu trafnych przekonań co do rzeczywistej jakości i przydatności procesów informatycznych, a nie tylko co do wyimaginowanych możliwości komputera, jest ważnym zadaniem dla nauczyciela.

**Motywacje | Wprawa – aspiracje i zaspokojenie**

Spłot związków oraz tok wzajemny oddziaływań obu tych zmiennych ma szczególnie silny, cykliczny i nieliniowy charakter. Pobudzenie do działania raczej nigdy dłużej nie utrzymuje się na stałym poziomie, a co istotniejsze – w perspektywie ewaluacji długofalowej wcale nie musi narastać. Oczywiście poziom motywacji silnie uzależniony jest od emocji związanych z obiektem, zjawiskiem lub procesem, którego dana osoba doświadcza, przy czym relacje afektywno-motywacyjne są dość klarowne: im wyższa aprobatą cechy indywidualnie uznawanej za wartość, tym większa chęć do działania. Bardziej zawile są relacje motywacyjno-kognitywne oraz motywacyjno-behawioralne, dlatego w żaden sposób nie należy domniemywać, że czynniki te, z punktu widzenia efektywności procesu edukacyjnego, powinny równocześnie osiągać coraz wyższe pułapy. Zazwyczaj są to zjawiska cyklicznie niewspółbieżne.

Dopuszczalny jest pewien spadek motywacji, gdy osiągnię się już pułap osobistych dążeń w określonym, choćby najbardziej ekscytującym obszarze zainteresowań. Po chwilowym zaspokojeniu ambicji związanych z komputerem ważne stają się intencje na przyszłość. Chodzi tu o chęć dalszego samodoształcania, gdyż pułap pełnych kompetencji informatycznych zdaje się nie mieć kresu. Jednakże w danym cyklu rozwojowym, w czasie trwania konkretnego kursu, wytyczony przez nauczyciela cel zajęć musi być w pełnej symbiozie z poziomem indywidualnych aspiracji i zdolności uczniów, a nie wyłącznie w zgodzie ze standardami wymagań i wyznaczonym programem nauczania. W przeciwnym wypadku u części uczniów ukształtuje się trwała niechęć do informatyki, podobnie jak to dzieje się z matematyką i z fizyką.

**Emocje | Rozwaga – impuls czy refleksja**

Łagodny stres może być katalizatorem przemian, lecz zbyt stresogenne zajęcia wywołują trwałe zniechęcenie do procesu komputeryzacji. Występuje blokada procesów poznawczych oraz wzbudzenie syndromu wyuczonej bezradności. A przecież odwrotnie – celem edukatorów oprócz nauczania winno być także stymulowanie optymistycznego stylu wyjaśniania. Cóż, mimo empirycznych dowodów nauczyciele ignorują fakt, że wyciskają trwałe piętno na całym życiu i zdrowiu uczniów. Na takie skrajności jak fobia komputerowa czy nałogowe uzależnienie się jedyną receptą jest kształtowanie roztropności w zachowaniach swoich podopiecznych. Trzeba – szczególnie na zajęciach przy komputerach – równoważyć zachwiane proporcje pomiędzy podejściem impulsywnym a refleksyjnym. Coraz więcej uczniów, od wczesnych lat posiadających w domu komputer, podchodzi beztrudno do zajęć informatycznych. Skutki są fatalne: od utraty plików aż po uszkodzenia systemów. Na tym tle ujawnia się onieśmienie i zagubienie tych, którzy komputera nie posiadają. Nie można dopuszczać, ażeby pozostawali oni w cieniu samozwańczych guru.

Jakość relacji wewnątrz grup i typy zachowań na zajęciach przy komputerze można oceniać bezpośrednio poprzez obserwację. Jednak stany negatywnych emocji są zazwyczaj skrzętnie ukrywane. Po prostu nie wypada ujawniać, że się nie lubi zajęć z informatyki. Co gorsza – utrwała się przekonanie, że lepiej jeśli z dwóch uczniów siedzących przy jednym komputerze zadanie wykonuje ten bardziej wprawiony. Wówczas ten drugi jest też zadowolony. Potencjalne tego typu nieprawidłowości wykryć można już za pomocą pierwszego pomiaru, a jakiegokolwiek tego objawy należy eliminować od samego początku kursu.

### **Wyniki – rezultaty zmiennych globalnych**

Końcowymi fazami agregacji wskaźników jest utworzenie rezultatów półówkowych i scalenie wszystkich komponentów indykatum w wynik ogólny. Jedna z półówkowych zmiennych globalnych kumuluje wszystkie wskaźniki afektywne (*Doznanie*), a druga kognitywne (*Poznanie*), według schematu:

Schemat agregowania rezultatów półówkowych *InfoKultury*

Tab. IV

Pozycje skali i polaryzacja wskaźników												Rezultat	Nazwa
1-	2-	9-	10-	11+	12+	13-	14+	21+	22+	23-	24-	afektywny	<i>Doznanie</i>
3+	4+	5-	6+	7+	8+	15-	16+	17+	18-	19+	20+	kognitywny	<i>Poznanie</i>

*Doznanie* – to stan emocjonalny oznaczający splot odczuć inicjowanych zarówno przez bodźce zewnętrzne, jak i generowanych wewnątrz samoistnie. Ów komponent afektywny reagowania w pierwszym rzędzie wyzwala gotowy, intuicyjny skrypt oceny sytuacji, regulujący zachowanie. Ze względu na swą impulsywność doznanie bywa często oparte na pozornie nieuzasadnionych, subiektywnych przesłankach. Okazuje się jednak, że w razie braku gotowego skryptu, pod wpływem impulsu, umysł może wygenerować zaskakująco nowatorskie, znakomite rozwiązania. Jest to niezwykle cenne w heurystycznej podbudowie procesów twórczych. Z tego względu doznania, zwłaszcza emocje i motywacje, stają się kluczowe w podmiotowych koncepcjach edukacji.

*Poznanie* – to stan uświadomienia, oznaczający splot zinternalizowanych struktur wiedzy ogólnej. Reagowanie pobudzające komponenty kognitywne oparte jest zazwyczaj na głębszym i dłuższym namyśle, co sugeruje pewną wyższość w procesie edukacyjnym. Informacja niesiona przez bodziec ulega przetworzeniu w całym kontekście obszernych zasobów, śladów i wzorców pamięciowych, w logice bardzo wielu aspektów wartościowania. W rezultacie internalizacji wzbogacane są zarówno zasoby wiedzy, jak też modyfikowane osobiste standardy ewaluatywne. Świadomość jest bardziej ceniona niż pamięć faktograficzna. Poznanie bywa podłożem przede wszystkim logicznego, algorytmicznego podejścia do rozwiązywania problemów, jednak w procesach edukacyjnych niestety zbyt często staje się przejawem odtwórczej rutyny.

**Doznanie | Poznanie – afektywność i kognytywność**

Metaforą połówkowego rozwarstwienia indykatum jest odwieczny dylemat: *Kierować się sercem czy rozumem?* Bez rozstrzygnięcia o wyższości któregoś z podejść trzeba próbować mierzyć i oceniać zmiany zachodzące w obu sferach oddziaływań edukacyjnych – afektywnej i kognitywnej. Obie te sfery są jednakowo ważne, dlatego w procesie kształcenia należy starać się je równoważyć. Zjawiska empiryczne, dotyczące relacji stanów emocjonalnych i poznawczych, ujawniają bardzo ciekawe związki pomiędzy odczuciami a przemyśleniami<sup>42</sup>. Pierwotne nastawienia mogą wspierać lub hamować rozwój. To, czego pragną uczniowie, nie idzie w parze z celami edukacji (np. niezwykle silna presja na nauczyciela, aby na lekcjach informatyki zezwalał na uruchamianie gier). Ich brak świadomości prowadzi do kształtowania antywartości (np. brutalność jako nagradzany cel, szafowanie odnawialnym życiem, osiąganie zysków bez aspektów etycznych, zatracenie się w bytach wirtualnych). To, czego wymaga nauczyciel, może też nie mieć wartości (np. nauka w większości niespójnych języków programowania z nieprzyjaznym interfejsem, obsługa coraz bardziej skomplikowanych pakietów, rozbudowywanych wyłącznie po to, aby zarabiać na nieustannych szkoleniach). To, co oferują wytwórcy, jest wątpliwej jakości (zawodność systemów, błędy w aplikacjach, luki w zabezpieczeniach robione celowo, by przejąć kontrolę nad cudzym komputerem). Wszystko to sprawia, że ocena informatyzacji, w miarę uświadamiania aspektów wartościujących i odkrywania mrocznych stron tego procesu, staje się coraz bardziej trafna. W narzędziu pomiarowym nie ma bezpośrednich próbników owych zjawisk, lecz poziomy ogólnej akceptacji procesu pośrednio zależą od splotu wszystkich odczuwanych bądź uświadomionych ocen afektywnych i kognitywnych.

**Kultura – uogólnienie symptomów**

W poszukiwaniu nazwy dla zmiennej ogólnej, powstałej dzięki agregacji kompletu wskaźników, sięgnąłem do pojęcia '*kultura*'. Sens definicyjny tego pojęcia jest na tyle szeroki, iż pomieścić może treść każdej ze zmiennych cząstkowych, tj: wartości, oceny, normy, postawy, dyspozycje, motywy, odczucia, wypowiedzi, uczenie się, internalizację itp. W uogólnionym znaczeniu *Kultura* to sprzyjający stosunek wobec korzystnych zjawisk społecznych, pozytywne odczucia temu towarzyszące, pożądane aspiracje, motywacje i osady wartościujące, przy czym istotna jest tu również świadomość przedmiotowa danej materii – ujawniana poprzez refleksyjność wypowiedzi, mająca swe odbicie w zgłębianiu zagadnień, w roztropności oraz w trafności przy poszukiwaniu ideału i dążeniu do prawdy obiektywnej. W takim ujęciu desygnaty *Kultury* sytuują się jako pośrednie pomiędzy wzniosłym określeniem cywilizacyjnego dorobku ludzkości, a znacznie ponad prostym synonimem obycia i ogłady. Mogą one oznaczać atrybuty grup społecznych lub osobnicze stany psychiczne.

Szeroki sens pojęcia przypisanego indykatum umożliwia konstruowanie według podanego wzorca także innych narzędzi do ewaluacji aspektów kulturowych, niezwiązanych z informatyzacją. W każdym konkretnym przypadku definicja i opis sensu merytorycznego indykatum wymaga doprecyzowania:

➤ *Kultura informatyczna* – to wyższa od alfabetyzacji forma uświadomienia, uzewnętrzniana w pożądanym stylu zachowań przy komputerze, w sensownym korzystaniu z możliwości użytkowych, w roztropnym i trafnym wyborze optymalnych technologii informacyjnych – co wynika bardziej z wiedzy ogólnej i zasad normatywno-etycznych aniżeli z biegłego znanstwa informatyki. Jest to w zasadzie nowe jakościowo lub przynajmniej zhumanizowane kontinuum funkcjonującego od dawna pojęcia *kultury technicznej* oraz nabierająca coraz bardziej istotną społecznie rangę częśćka *kultury osobistej*.

W konwersacji z komputerem istotą winna być wysoka koncentracja na interakcji, a więc skupienie na tym, co w krótkich komunikatach lub w kontekstowych pomocach starają się przekazać twórcy programu bądź systemu. Wszakże w dobrych aplikacjach zawarty jest olbrzymi ładunek dorobku kompetentnych zespołów twórczych. Komputer „nie myśli”, dlatego użytkownik, poza sprawnością czynności manualnych, musi wykazywać się adekwatnym do problemu informatycznego poziomem logiki rozumowania i postępowania. Przydatna jest kompetencja opracowania optymalnych strategii po to, ażeby korzystać nie tylko z gotowych procedur użytkowych (z ofert menu), lecz aby tworzyć także nowe, indywidualnie zoptymalizowane sposoby użytkowania. Niezbędna jest orientacja w szerokim zakresie przydatności pewnych kategorii oprogramowania do takich praktycznych zastosowań, kiedy to wykorzystanie komputera przynosi rzeczywiście więcej pożytku, niż marnotrawienia czasu. I wreszcie – konieczna jest etyka użytkowania, ogląda zachowań i trafność podejmowanych decyzji po to, by nie dokonywać spustoszeń na dyskach, w sieciach i w całych systemach. Tylko przy takim podejściu komputeryzacja będzie procesem usprawniającym, a nie przytłaczającym. Właśnie w tym tkwi głębszy sens kształtowania pozytywnych, silnie poświadanych cech mentalnych, których splecione symptomy proponuję nazywać *Kulturą informatyczną* bądź skrótowo *InfoKulturą*. Ewaluacja tego indykatum jest dziś dziejową potrzebą.

<i>Ślady</i>		<i>Wzorce</i>
desygnowanie zmiennych	.....	definicje strukturalno-operacyjne
wskaźnikowanie zmiennych	.....	argumenty a wartość wypowiedzi
nazwy, kategorie i treści	.....	taksonomia a sens merytoryczny
zależności zmiennych	.....	relacje par komplementarnych
rezultaty połówkowe	.....	bilansowanie <i>Doznania</i> i <i>Poznania</i>
uogólnienie symptomów	.....	meritum <i>Kultury informatycznej</i>



## PARAMETRY SKALI DWUWAŻONYCH OCEN

Tak się składa, że hasło *matematyka*, a jeszcze bardziej *statystyka* wielu rodaków o humanistycznej duszy niesłusznie przyprawia o ból głowy<sup>43</sup>. Zdaje więc sobie sprawę, że do tego działu książki zajrzą tylko nieliczni. Jednak dla zapaleńców, którzy postanowili przełamać nieuzasadnione opory, postaram się konsekwentnie przeprowadzić propedeutyczne objaśnienie, z podaniem elementarnych konwencji obowiązujących w tej dziedzinie wiedzy<sup>44</sup>. Z oczywistych względów do opisu włączyć muszę statystyki charakterystyczne dla *Skali dwuważonych ocen*. Są one bowiem nieodłącznym elementem strategii przeprowadzania ewaluacji w fazach eksplorowania i interpretowania zjawisk.

Podczas badania prawidłowości statystycznie oczekiwanych, równolegle występuje odkrywanie zjawisk niespodziewanych. I – co ciekawe z punktu widzenia ewaluatora – przejawy te są cenniejsze do analizy i interpretacji niż zjawiska typowe. Wymiary wartości poszczególnych składowych indykatum określa się na podstawie obserwowalnego spektrum wszelakich, zarówno swoistych, jak i nieszablonowych charakterystyk, zwanych *oznakami*.

**Oznaka** – to dostrzegana podczas interpretowania właściwość zmiennych odnoszona do konkretnego poziomu ekstrakcji komponentów (np. *spójność składników*). Właściwość ta na wyższym poziomie agregacji może zanikać lub przechodzić w stan przeciwny (np. *niespójność czynników*). Z tego powodu oznaki bada się na wszystkich poziomach rozwarstwienia indykatum.

W procesie kwantyfikacji oznaki stanowią łącznik pomiędzy jakościowymi symptomami zmiennych a ich statystycznymi parametrami ilościowymi. Ów proces przebiega następująco: przyjmujemy, iż indykatum tworzą symptomy *Poznania* i *Doznania*, a znamiennej oznaką jest m.in. *niespójność czynników*. Ową niespójność można zaobserwować przez porównywanie miar tendencji do zmian. Jedną ze statystyk opisujących tendencje jest parametr o nazwie *Trend* oznaczony symbolem  $[\pm t]$ . Jeśli po obliczeniu wartości tego parametru czynnik poznawczy ma trend dodatni, a czynnik doznaniowy trend ujemny, to uzyskujemy statystyczne potwierdzenie zjawiska niespójności czynników.

Przedstawię w tym miejscu szczegółowe objaśnienia podstawowych pojęć ogólnych, występujących w obszarach kwantyfikacji i analizy statystycznej.

**Parametr statystyczny** – to ujęta definicyjnie matematyczna regularność lub nieregularność mogąca występować w elementach zbiorów, połączona z ilościowym określeniem wymiaru tej właściwości odkrytej w rzeczywistych danych empirycznych. Parametr statystyczny jest integralnym splotem rodzaju miary z opisem pojęcia, wypełnionym symboliką, wzorami i liczbami.

Warto dodać, że w żadnym razie nie można pomijać i oddzielać definicyjnego sensu parametru od jego wartości. Jeśli otrzymamy statystyki jakiegoś eksperymentu, a nie rozumiemy ich znaczenia, to są one dla nas bezwartościowe. Również dane empiryczne mogą być bezużyteczne, o ile nie potrafimy obliczyć choćby fundamentalnych parametrów. Przedstawianie w pracach dyplomowych jedynie histogramów liczebności w żadnym wypadku nie można uznać jako prezentację wyników badań, lecz wyłącznie jako wizualizację danych. Na szczęście coraz więcej osób potrafi posługiwać się arkuszem kalkulacyjnym, który jest uniwersalnym narzędziem pomocnym w opracowywaniu wyników, dzięki wbudowanym funkcjom statystycznym oraz mechanizmom wizualizacji. A jeśli nawet nie ma potrzebnej funkcji, to wzór każdego parametru można przekształcić w odpowiednie formuły liczące. Właśnie do tego potrzebna jest znajomość znaczenia miar, symboli, znaków i operatorów we wzorach.

➤ **Miary statystyczne** – to uogólnione kategorie parametrów, abstrahujące od zmiennych i od ich wartości liczbowych. W precyzyjnym posługiwaniu się pojęciami warto nazwę *miara* używać w odniesieniu do ogólniejszego obszaru definicyjno-abstrakcyjnego, a nazwę *parametr* w ukonkretnionym obszarze desygnująco-ilościowym. Przykładowo miarami *dyspersji* (rozproszenia) mogą być m.in. parametry: *rozrzutu* opartego na odchyleniach, *rozstępu całkowitego* wyników skrajnych lub *rozstępu ćwiartkowego* bazującego na kwartylach.

➤ **Nazwy i symbole parametrów** – to oznaczenia przypisane do formalnych definicji konkretnych statystyk, występujące najczęściej jako połączenie nazwy parametru ze znakiem literowym. Zwykle podanie samej nazwy nie wystarcza do precyzyjnego określenia, o jaki parametr chodzi, stąd dla jednoznaczności przyjęto zwyczaj podawania również nazwiska autora danej statystyki, np.: współczynnik korelacji *r*- Pearsona, wyznacznik rzetelności *alfa* Cronbacha. Ostatecznie najlepszym opisem symbolu jest podanie rozwijającego go wzoru.

W *Skali dwuważonych ocen* proponuję następujące konwencje:

- nazwy parametrów są zwarte, ograniczone do minimum, dwujęzyczne;
- symbole pochodzą od pierwszych liter angielskich nazw parametrów;
- symbole dotyczące pomiaru pierwszego mają znak *prim* ', a drugiego *bis* '';
- symbole uzupełniane są charakterystycznymi znakami +, -, ±, ˇ lub ^.

Dodatkowe oznakowywanie symboli wyniku nie tylko z braku odpowiedniej liczby liter adekwatnych do nazwy, lecz przede wszystkim z potrzeby nadania konkretnym parametrom symboliki znamionującej właściwości danej miary.

Zasady uzupełniającego oznakowywania symboli są następujące:

- Znak wartości dodatniej  $[+p]$  lub znak wartości ujemnej  $[-p]$  jest integralną składową symbolicznego oznaczania tych parametrów  $[p]$ , które zawsze są wyłącznie ujemne bądź wyłącznie dodatnie (z wyjątkami, gdy są zerowe), na przykład:  $-b, -c, +c, -d, +d, -e, +e$ .
- Znak plus-minus  $[\pm p]$  jest integralną składową symbolu parametru  $[p]$  powstałego z porównania statystyk z dwóch pomiarów  $[p'' - p']$ , kiedy to wynik może być dodatni lub ujemny, na przykład:  $\pm X, \pm H, \pm M, \pm L, \pm b, \pm c$ .
- Znak plus-minus  $[\pm p]$  występuje również jako integralna część symbolu parametru  $[p]$  powstałego jako efekt zestawienia indeksowego bądź relacji pomiędzy zbiorami danych lub ich statystykami, kiedy to wynik może być dodatni lub ujemny, na przykład:  $\pm e, \pm i, \pm r, \pm t$ .
- Znaki uzupełnienia diakrytycznego  $[\sim i \sim]$  występują nad parametrami  $[p]$  będącymi wyznacznikami jakości pomiarowej i pełniącymi funkcje kwantyfikatorów selekcyjnych:  $\check{c}, \check{s}$ , kumulacyjnych:  $\hat{a}$  oraz buforowych:  $\check{r}$ .

Szczególną uwagę należy zwrócić na istotną różnicę pomiędzy oznakowaniem parametrów  $[+p]$  i  $[-p]$  a operatorami sumowania  $[+]$  i odejmowania  $[-]$ .

➤ **Operatory działań** – to umowne symbole w postaci znaków graficznych, którym definicyjnie przyporządkowano określoną funkcję matematyczną lub logiczną na ściśle wytyczonych elementach. Stosowanymi tutaj operatorami są nie tylko znaki pomiędzy dwoma argumentami działań arytmetycznych: dodawania, odejmowania, mnożenia i dzielenia, lecz także symbole działań wieloargumentowych: kreska ułamkowa, znak potęgowania i pierwiastkowania, znak modułu, nawiasy ustalające priorytet, a szczególnie znak sumowania  $n$ -krotnego lub warunkowego z indeksami iteracyjnymi i selekcyjnymi.

Konwencje stosowania operatorów działań są następujące:

- Znak sumy  $[+]$  i znak różnicy  $[-]$  rozpoznaje się jako operatory dodawania i odejmowania wtedy, gdy z obu stron otoczone są spacjami, na przykład:  $a - b + c, X'' - X', 2,5 + Q$ .
- Ze względu na potrzebę jednoznacznego odróżniania operatorów od oznaczeń (sumy od znaku plus, a różnicy od znaku minus) zaleca się stosowanie różnych nawiasów w miejscach mogących budzić wątpliwości, na przykład:  $l(-d)^2 + h(+d)^2, [\pm H] + [\pm M] + [\pm L], 1 - |(\pm t) - (\pm e)|$ .
- Znak modułu  $|p|$  jako operator wartości bezwzględnej używany jest także w tych wszystkich miejscach, w których mogłyby powstać wątpliwości co do priorytetu działań (co najpierw – odejmowanie czy ustalenie znaku?), na przykład:  $|+d| - |-d|, |-b'| - |-b''| - |s'| |s''|, |+e| - |-e|$ .

- Znak sumy  $\sum$  jest operatorem  $n$ -krotnego dodawania kolejnych składników zbioru  $n$ -elementowego. Nad znakiem  $\sum$  znajduje się wskaźnik krotności  $n$ . Pod znakiem  $\sum$  standardowo znajduje się indeks iteracyjny  $i = 1$  oznaczający, że należy dodać wszystkie elementy zbioru, począwszy od pierwszego.
- Szczególną odmianą sumowania krotnego jest tu sumowanie warunkowe. Pod znakiem  $\sum$  sumy warunkowej znajduje się indeks selekcyjny w postaci definicji podzbioru, na przykład  $i \in \{ i: X_i < 0 \}$ , co w tym wypadku oznacza, że należy zsumować wyłącznie elementy mniejsze od zera, czyli ujemne.
- We wzorach występuje ponadto funkcja selekcyjna  $\text{Min}(a | b)$  oznaczająca konieczność wyboru tylko jednego, mniejszego z dwóch argumentów  $a$  i  $b$ .

Działania matematyczne i logiczne, prowadzące do wyznaczenia parametrów, wykonuje się zarówno na pierwotnych elementach zbiorów danych empirycznych, jak też na innych, wcześniej obliczonych tzw. parametrach bazowych.

➤ **Zbiory danych empirycznych** – to ściśle uporządkowane w macierzach wagi liczbowe, pochodzące z pomiaru skalą i powstałe z przypisania umownych wartości każdej z pojedynczych wypowiedzi respondentów. Są one budulcem do statystycznej analizy stanu zjawisk lub obiektów poddawanych ewaluacji. Wzory statystyczne buduje się na elementach abstrakcyjnych, natomiast obliczenia prowadzi na wartościach liczbowych pozyskanych z doświadczenia. W tej konwencji szereg podany jako  $n$ -kolejnych elementów  $\{ X_1; X_2; \dots X_n \}$  podczas obliczeń przyjmuje konkretną postać liczbową, np.  $\{ 3; 1; \dots 2 \}$ . Znaki indeksu dolnego oznaczają porządek usytuowania  $i$ -tego elementu  $X_i$ . I wprawdzie podczas sumowania kolejność składników może być dowolna, jednak szczególną uwagę należy zwrócić na to, co znajduje się w wierszach, a co w kolumnach macierzy i jaki podzbiór (przekrój) bierze się do obliczeń.

Przyjrzyjmy się strukturze wzorów wyznaczających parametry bazowe:

$$\begin{array}{l} \text{średnia arytmetyczna} \\ \text{moment zwykły} \end{array} \quad \hat{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n (X_i) \quad [1]$$

$$\begin{array}{l} \text{średnia kwadratowa} \\ \text{moment centralny} \end{array} \quad \tilde{Y} = \sqrt{\frac{1}{n} (Y_1^2 + Y_2^2 + \dots + Y_n^2)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i)^2}$$

gdzie:  $X_1, X_2 \dots$  – wartości kolejnych przypadków,  $n$  – liczba przypadków

$Y_1, Y_2 \dots$  – wartości różnic kolejnych przypadków od ich średniej arytmetycznej

$$Y_1 = X_1 - \hat{X}, \quad Y_2 = X_2 - \hat{X} \quad \dots \quad Y_n = X_n - \hat{X}$$

Istota „bazowości” polega na tym, że średnia arytmetyczna argumentów często wchodzi w skład wzorów służących do obliczania innych parametrów, w tym zwłaszcza momentów centralnych. Stąd, ażeby obliczyć miarę rozproszenia  $\tilde{Y}$  jakichś wyników (np. odchylenie standardowe punktów uzyskanych z testu), należy uprzednio wyznaczyć wartość przeciętną  $\hat{X}$  z tych samych wyników.

Z kolei miara rozproszenia  $\tilde{Y}$  może być składową jeszcze innego parametru. I to zapewne między innymi także z tej właściwości agregowania miar bierze się przekonanie, iż wzory statystyczne są nie do rozszyfrowania. Najbardziej polecanym sposobem obliczania lub interpretacji wzorów jest rozpoczynanie od działań elementarnych i postępowanie sukcesywne małymi krokami.

**Przykład 1. Średnie kwadratowe**

Kolejnych 30 studentów uzyskało następujące liczby ( $x$ ) punktów na teście z matematyki:

9	12	11	5	12	8	14	13	2	10	14	10	5	11	6	8	10	3	4	9	6	11	7	10	12	9	11	7	13	8
---	----	----	---	----	---	----	----	---	----	----	----	---	----	---	---	----	---	---	---	---	----	---	----	----	---	----	---	----	---

Obliczyć odchylenie standardowe, które jest średnią kwadratową z różnic wobec średniej.

Kroki liczenia: 1° Suma  $\Sigma x = 270$  2° Średnia  $\Sigma x/n = 270/30 = 9$  3° Różnice  $(x - \Sigma x/n)$ :

0	3	2	-4	3	-1	5	4	-7	1	5	1	-4	2	-3	-1	1	-6	-5	0	-3	2	-2	1	3	0	2	-2	4	-1
---	---	---	----	---	----	---	---	----	---	---	---	----	---	----	----	---	----	----	---	----	---	----	---	---	---	---	----	---	----

4° Każdą z różnic  $Y$ , powstałą z odjęcia  $(x - 9)$ , należy podnieść do kwadratu, otrzymując:

0	9	4	16	9	1	25	16	49	1	25	1	16	4	9	1	1	36	25	0	9	4	4	1	9	0	4	4	16	1
---	---	---	----	---	---	----	----	----	---	----	---	----	---	---	---	---	----	----	---	---	---	---	---	---	---	---	---	----	---

5° Suma  $\Sigma Y^2 = 300$  6° Wariancja  $\Sigma Y^2/n = 300/30 = 10$  7° Odchylenie stand. =  $\sqrt{10} \approx 3,16$

Jak wynika z powyższego przykładu, każdy elementarny krok przy obliczaniu parametrów statystycznych jest łatwym działaniem matematycznym. Jednak przy licznych zbiorach kroków jest dużo, stąd tak bardzo przydatne jest użycie automatycznych procedur arkusza kalkulacyjnego lub pakietu statystycznego.

W większości statystyki mogą być liczone zarówno dla zmiennej ogólnej, jak i dla dowolnego splotu zmiennych tworzących indykatum. Oznacza to, że do wyznaczenia wartości statystyk konkretnego składnika jako argument  $X_i$  bierze się średnią z wag tej kontrolnej pary wskaźników, która tworzy dany składnik. Podobnie podczas obliczania statystyk czynników – argumentami podstawianymi do wzorów są średnie z wag 4 wskaźników tworzących dany czynnik. Analogicznie do obliczenia wyników połówkowych bierze się średnie z 12 wag, a do wyniku ogólnego – średnią z 24 wag nadanych wskaźnikom. Ta generalna zasada *syntezy oznak* wyjątkowo nie ma zastosowania podczas analizy spójności i zbieżności (dotyczy wówczas tylko poziomu składników), jak również przy wyznaczaniu rzetelności skali (gdyż jest to miara zbiorcza). Wszystkie wzory dotyczą parametrów uogólnionych na całą próbę badawczą jako osiągnięcia danej klasy, grupy bądź ogółu, a nie indywidualnych osób.

<i>Ślady</i>		<i>Wzorce</i>
oznaki w ewaluacji	.....	kwantytatywne właściwości zmiennej
miary statystyczne	.....	ogólne kategorie parametrów
parametr statystyczny	.....	splot definicji, symbolu, wzoru i liczby
oznaczenia i działania	.....	symbole i operatory
zbiory danych empirycznych	.....	elementy zbiorów, indeksy
parametry bazowe	.....	średnia arytmetyczna i kwadratowa

### Miary chwilowe z pomiaru jednokrotnego

Jakkolwiek w badaniu procesów nieodzowne są dwa pomiary, to jednak wyznaczanie parametrów statystycznych konieczne jest dla każdego pomiaru z osobna – odrębnie dla stanu początkowego i końcowego. Miary chwilowe są bowiem podstawą konstruowania dystansowych miar zmienności w czasie. Ponadto statystyki 'przed' i 'po' potrzebne są również do pełnej analizy zjawisk – służą do oceny stanów cech mentalnych na wejściu i na wyjściu ewaluacji. Wzory definiujące miary chwilowe są identyczne dla pomiarów 'przed' i 'po', przy czym rezultaty obliczeń różni się odpowiednio znakami *prim* lub *bis*.

Generalnie zalecam stosowanie siedmiostopniowej gradacji próbkowania i nadawanie opcjom wypowiedzi respondentów kolejnych wag całkowitych  $\{-3; -2; -1; 0; 1; 2; 3\}$ . Formalnie nie ma jednak przeszkód, aby stosować wagi inne. Uogólniając, dla wag  $[w]$  nadanych wypowiedziom oraz dla gradacji  $[g]$ , określającej liczbę progów próbkowania zmiennej, muszą być przestrzegane konstytutywne w *Skali ważonych ocen* zasady kwantyzacji i zależności:

zbiór wag:	$\{w_1; w_2; w_3; \dots w_g\}$	liczba nieparzysta $g = 5$ lub $7$ (optimum)
kres dolny skali:	$\{w_1\} = -X_{\max}$	kres górny skali: $\{w_g\} = X_{\max}$
tożsame moduły:	$\{ w_1 ; \dots  w_{(g+1)/2-1} \} \equiv \{ w_{(g+1)/2+1} ; \dots  w_g \}$	
waga zerowa:	$\{w_{(g+1)/2}\} = 0$	

Przypomnę, że argumentami we wzorach mogą być wprost wskaźniki  $X_i$  powstałe po nadaniu wypowiedziom odpowiednich wag  $\{w_1; w_2; w_3; \dots w_g\}$ , bądź też oznaki skumulowane  $X_{ij}$  powstałe z uśrednienia 2, 4, 12 lub 24 wag. Przykładowo dla składników są to średnie arytmetyczne z wag wskaźników nieparzystych i parzystych  $\{w_{11} \oplus w_{12}; w_{21} \oplus w_{22}; \dots w_{g1} \oplus w_{g2}\}$ . Jeśli chcemy obliczyć miary tego składnika, który tworzony jest z dwóch wypowiedzi wobec bodźców-stwierdzeń nr 1 i nr 2 w skali, to najpierw musimy obliczyć średnie arytmetyczne z kolejnych, odpowiednich par kontrolnych wypowiedzi nr 1 i 2 wszystkich respondentów w danej próbie badawczej.

#### Przykład 2. Agregacja wskaźników

30 studentów uzyskało poniższe wagi ( $w_{i1}$ ) za wypowiedzi wobec pierwszej pozycji skali:

1	2	1	3	2	3	1	3	2	0	3	-1	3	1	3	2	0	3	2	3	3	-1	2	1	2	3	3	1	3	2
---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---

Ci sami studenci uzyskali poniższe wagi ( $w_{i2}$ ) za wypowiedzi wobec drugiej pozycji skali:

2	3	1	3	1	2	1	2	2	-2	3	1	3	1	2	2	1	3	1	3	2	1	2	1	0	3	1	-3	1	2
---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---

Oznaki skumulowane  $X_{ij}$ , potrzebne do obliczeń miar składników, są średnimi z powyższych wag.

1,5	2,5	1	3	1,5	2,5	1	2,5	2	-1	3	0	3	1	2,5	2	0,5	3	1,5	3	2,5	0	2	1	1	3	2	-1	2	2
-----	-----	---	---	-----	-----	---	-----	---	----	---	---	---	---	-----	---	-----	---	-----	---	-----	---	---	---	---	---	---	----	---	---

1° Ze zbioru wag ( $w_{i1}$ ) obliczmy wartość średnią pierwszego wskaźnika  $[i:1]$   $\Sigma/n = 56/30 \approx 1,87$

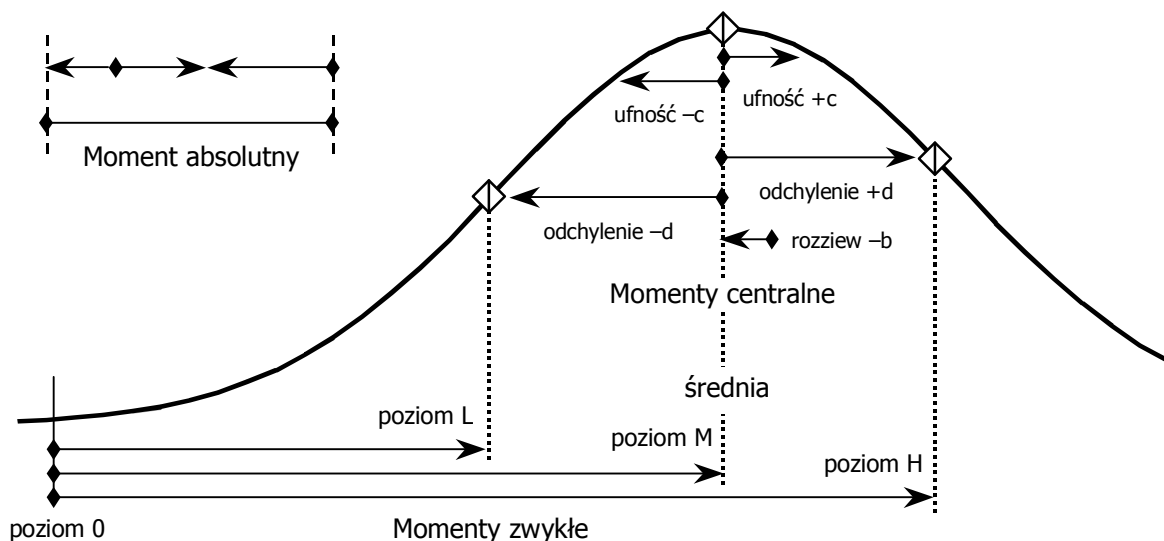
2° Ze zbioru wag ( $w_{i2}$ ) obliczmy wartość średnią drugiego wskaźnika  $[i:2]$   $\Sigma/n = 45/30 = 1,5$

3° Ze zbioru oznak ( $X_{ij}$ ) można obliczyć wartość średnią składnika  $[i:\{1;2\}]$   $\Sigma/n = 50,5/30 \approx 1,68$

Istotne znaczenie w prawidłowej interpretacji parametrów statystycznych ma rozróżnianie pewnych ogólnych kategorii miar, jakimi są tzw. *momenty*. W zasadzie chodzi tu o rozpoznawanie trzech kategorii: momentów zwykłych, centralnych i absolutnych o prostszej denotacji niż ujęcia stricte matematyczne.

➤ **Momenty zwykłe** – to parametry zmiennej losowej odnoszone wprost do liczbowych wartości wag przypisanych wskaźnikom. Z założenia punktem odniesienia powinien być poziom zerowy, jednakże w wypadku skal z wagami umownymi poziom zerowy jest też jedynie umowny. Co więcej, poziom ten może być przesuwany zgodnie z przyjętą konwencją, tym niemniej w danym zbiorze statystyk należy mieć świadomość tego, co było punktem odniesienia. Przy wyskalowaniu osi miarami wag  $\langle -3; +3 \rangle$  poziomem zerowym jest waga = 0 przypisana neutralnej wypowiedzi 'brak zdania'. Jeśli natomiast przeskaluje się wymiarowanie osi na skalę ocen szkolnych, to poziom neutralny znajdzie się w środku pomiędzy poziomem *niedostatecznym* a *dostatecznym*, natomiast „wirtualny” poziom zerowy znajdzie się o 2,5 symboliczne jednostki poniżej. Owa wirtualność wynika z faktu, że nie istnieje ocena szkolna równa zero. W skalach dwuważonych przyjmijmy za momenty zwykłe uznawać wszystkie te parametry pozycyjne, które wyznaczają pewien odstęp od umownego poziomu zerowego. Są to średnie [Q], [V], [B'] i [N''] oraz poziomy [L], [M] i [H].

➤ **Moment centralne** – to parametry zmiennej losowej wytyczone względem poziomu przyjętego jako główna pozycja środkowa. Najczęściej owym centrum jest wartość średnia i w odniesieniu do niej zarówno oblicza się, jak również wokół niej kreśli parametr centralny. Przyjmijmy uproszczenie, że momentem centralnym jest każdy parametr, który wyznacza zakres odstępu od średniej. W tej kategorii mieszczą się odchylenia  $[-d]$  i  $[+d]$ , przedziały ufności  $[-c]$  i  $[+c]$  oraz rozziw  $[-b]$ . Momenty centralne są miarami zróżnicowania wypowiedzi.



Ryc. 24. Graficzna interpretacja momentów zwykłych, centralnych i absolutnych

➤ **Momenty absolutne** – to parametry powstałe bezpośrednio z wag lub ich różnic, a także z różnic pomiędzy innymi parametrami, przy czym z każdej ze składowych różnic bierze się wyłącznie wartość bezwzględna. Oznacza to, że nawet jeśli występują argumenty o wartościach ujemnych, to i tak sumuje się wyłącznie ich moduły bez uwzględniania znaku. W interpretacji geometrycznej odpowiada to dodawaniu samych długości wektorów równoległych. Momentami absolutnymi są tutaj średnie [I] i [W], rozrzut [d] oraz spójność [c].

### **Jakość | Intensywność – średnie oznak**

Para połączonych ze sobą parametrów uśredniających wartości oznak ma kluczowe znaczenie w konstruowaniu *Skali dwuważonych ocen*. Wynika to z opisaną wcześniej techniki porównywania dwóch wymiarów wypowiedzi. Powszechnie znany klasyczny moment zwykły, będący średnią arytmetyczną z wag, wyznacza *Jakość* 'Quality [Q], natomiast mniej znany i rzadko stosowany moment absolutny wyznacza jej średnią *Intensywność* 'Intensity [I]:

$$Q = \frac{1}{n} \sum_{i=1}^n (X_i) \quad (... +3) \quad I = \frac{1}{n} \sum_{i=1}^n |X_i| \quad (0; +3) \quad [2]$$

gdzie  $X_i$  – waga nadana wypowiedzi  $i$ -tego respondenta lub średnia z wag  
 $n$  – liczebność próby (klasy, grupy lub ogółu badanej zbiorowości)

Z porównania wzorów wynika, że przy obliczaniu *Intensywności* jedyną odmianą jest działanie na wartościach bezwzględnych z wag. Trzeba jednak pamiętać, że miara intensywności może być stosowana wyłącznie w sytuacji, gdy zmienna losowa  $X$  opisana jest wagami z zakresu  $\langle -X_{\max} \dots +X_{\max} \rangle$ , przy czym moduły każdej z wag opcji przeciwstawnych muszą być równoważne. Oznacza to, że dla przyjętego tutaj zbioru 7 wag symetrycznie przypisanych wypowiedziom  $\{-3; -2; -1; 0; 1; 2; 3\}$  musi zachodzić relacja tożsamości podzbiorów  $\{|-3|; |-2|; |-1|\} \equiv \{|1|; |2|; |3|\}$ , chociaż w danych empirycznych nie każdy element musi wystąpić. W skali znormalizowanej obowiązuje  $[Q] > 0$ .

### **Przykład 3. Jakość, Intensywność**

30 studentów uzyskało następujące wagi (x) za wypowiedzi wobec jednej z pozycji skali:

1	2	1	3	2	3	1	3	2	0	3	-1	3	1	3	2	0	3	2	3	3	-1	2	1	2	3	3	1	3	2
---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---

Obliczyć Jakość i Intensywność wypowiedzi wg wzoru [2].

Kroki liczenia: 1° Suma  $\Sigma x = 56$  2° Średnia  $\Sigma x/n = 56/30 \approx 1,87$  3° Moduły z wag  $|x|$ :

1	2	1	3	2	3	1	3	2	0	3	1	3	1	3	2	0	3	2	3	3	1	2	1	2	3	3	1	3	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

4° Suma  $\Sigma |x| = 60$  5° Średnia  $\Sigma |x|/n = 60/30 = 2$  stąd  $[Q] \approx 1,87$   $[I] = 2$

Tak zdefiniowana *Intensywność* wypowiedzi nigdy nie jest niższa od *Jakości*. Na przyrost miary intensywności składają się bowiem wszystkie wypowiedzi o wagach różnych od zera, a więc także te niezgodne z opinią większości.



**Rozziew – dysjunkcja oznak**

Strategicznym parametrem skali dwuważonej jest niezgodność pomiędzy mniejszością respondentów wypowiadających się przeciwnie niż kwalifikowana większość. Miara ta jest składową selekcyjnego kwantyfikatora konkluzyjności. Z definicji *Rozziew* 'Bias'  $[-b]$  jest różnicą jakości i intensywności wypowiedzi:

$$-b = Q - I \quad \text{lub} \quad -b = \frac{2}{n} \sum_{i \in \{i: X_i < 0\}} (X_i) \quad \langle -3; 0 \rangle \quad [3]$$

gdzie  $Q$  – średnia jakość wypowiedzi,  $I$  – średnia intensywność ze wzoru [2]  
 $n$  – liczebność próby,  $m$  – liczba przypadków, dla których waga  $X_i < 0$

Drugi wzór powstał po rozwinięciu i przekształceniu  $Q - I$ . Formuła indeksu selekcyjnego  $i \in \{i: X_i < 0\}$  określa dodawanie wyłącznie składników ujemnych. Wynika stąd, że jeśli rozziew występuje ( $-b \neq 0$ ), to ma zawsze wartość ujemną.

**Przykład 4. Rozziew**

Przyjmijmy wagi ( $x$ ) takie same jak w przykładzie 3, uzyskane przez 30 studentów:

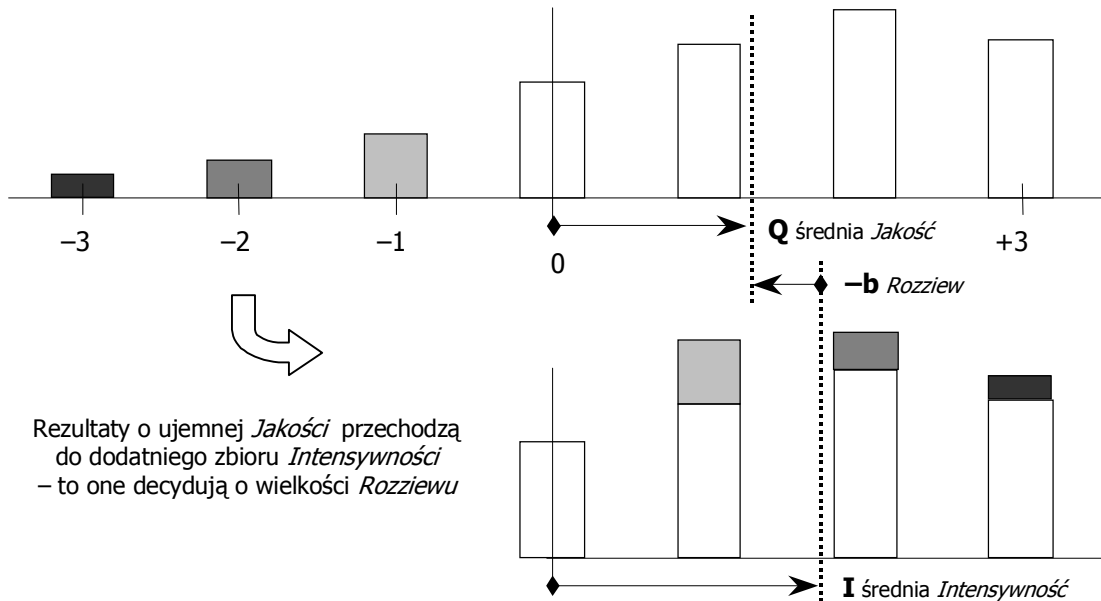
1	2	1	3	2	3	1	3	2	0	3	-1	3	1	3	2	0	3	2	3	3	-1	2	1	2	3	3	1	3	2
---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---

Obliczyć Rozziew wypowiedzi według obydwu sposobów ze wzoru [3].

Kontynuacja przykładu 3:  $[Q] \approx 1,87$   $[I] = 2$   $6^\circ$  Rozziew  $[-b] = Q - I \approx 1,87 - 2 \approx -0,13$

1 $^\circ$  Dwa przypadki  $x_i < 0$     2 $^\circ$  Suma  $\sum x_i = -2$     3 $^\circ$  Rozziew  $[-b] = 2 \cdot \sum x_i / n = 2 \cdot (-2) / 30 \approx -0,13$

Zwróćmy uwagę, że obliczenie według wzoru graficznie prostszego jest tu dłuższe niż przy użyciu pozornie bardziej skomplikowanego wzoru z sumą  $\sum$ . W pierwszym sposobie trzeba było bowiem wcześniej obliczyć  $[Q]$  oraz  $[I]$ . Natomiast znak sumy  $\sum$  z zawilim formalnym zapisem indeksu selekcyjnego w praktyce okazał się dodawaniem tylko dwóch przypadków wag ujemnych -1.



Ryc. 25. Graficzna interpretacja miary rozziewu określającej niezgodność w próbie

Parametr *Rozziewu* ma duże znaczenie głównie na poziomie wskaźników, jako że pośrednio służy do weryfikowania bodźców-stwierdzeń i skalowania narzędzia pomiarowego. W miarę przechodzenia na wyższe poziomy agregacji oznak, zwłaszcza na poziom wyników, rozziew zanika, czyli zmierza do zera. Ponadto niezgodność wypowiedzi zasadniczo maleje w drugim pomiarze.

### Odchylenia realne – dewiacja oznak

W skalach dwuważonych do analizy dewiacji używa się bardziej precyzyjnych miar niż miary standardowe. Są to *Odchylenia realne* 'Real Deviations', zdefiniowane przez dwa parametry, wyznaczające osobno dewiacje ujemne oraz dodatnie. Oblicza się je odrębnie dla osiągnięć poniżej przeciętnych [*l*] i ponadprzeciętnych [*h*], dodatkowo z proporcjonalnym rozdysponowaniem przypadków zgodnych ze średnią { $X_i = Q$ }. *Odchylenie ujemne* [-*d*] obejmuje przypadki gorsze od średniej, a *odchylenie dodatnie* [+*d*] lepsze od średniej:

$$-d = -\sqrt{\frac{1 + \frac{h}{l}}{n} \sum_{i \in \{i: X_i < Q\}}^l (X_i - Q)^2} \quad +d = +\sqrt{\frac{1 + \frac{l}{h}}{n} \sum_{j \in \{j: X_j > Q\}}^h (X_j - Q)^2} \quad [4]$$

gdzie  $X_i, X_j$  – wagi lub średnie z wag wypowiedzi kolejnych respondentów  
 $Q$  – średnia jakość wypowiedzi z próby *n*-elementowej (wzór [2])  
 $l, h$  – liczebności podzbiorów {*i*} oraz {*j*} w próbie *n*-elementowej

W obu wzorach występują sumy warunkowe z indeksami selekcyjnymi. Formalny zapis  $i \in \{i: X_i < Q\}$  nakazuje wybieranie tylko tych składników  $X_i$ , które są mniejsze od średniej  $Q$ . Podobnie jest z zapisem  $j \in \{j: X_j > Q\}$ , lecz wytycza on składniki o wartości wyższej od  $Q$ . Podczas obliczania odchylenia należy zwrócić szczególną uwagę na to, że współczynniki przed znakiem sumy różnią się. W pierwszym wzorze występuje iloraz  $h/l$  a w drugim  $l/h$ .

#### Przykład 5. Odchylenia realne

30 studentów uzyskało następujące wagi (*x*) za wypowiedzi wobec jednej z pozycji skali:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

Obliczyć odchylenia realne wg wzoru [4].

Kroki liczenia: 1° Suma  $\Sigma x = 30$  2° Średnia jakość  $Q = \Sigma x/n = 30/30 = 1$  3° Różnice ( $x - Q$ ):

0	-3	1	2	1	-1	1	1	-2	0	1	0	-1	2	1	-4	-2	-1	1	2	-3	1	-2	0	2	-1	0	1	2	1
---	----	---	---	---	----	---	---	----	---	---	---	----	---	---	----	----	----	---	---	----	---	----	---	---	----	---	---	---	---

4° Różnic ujemnych  $l = 10$ , dodatnich  $h = 15$  5° Przypadki ujemne  $Y_i$  podnosimy do kwadratu:

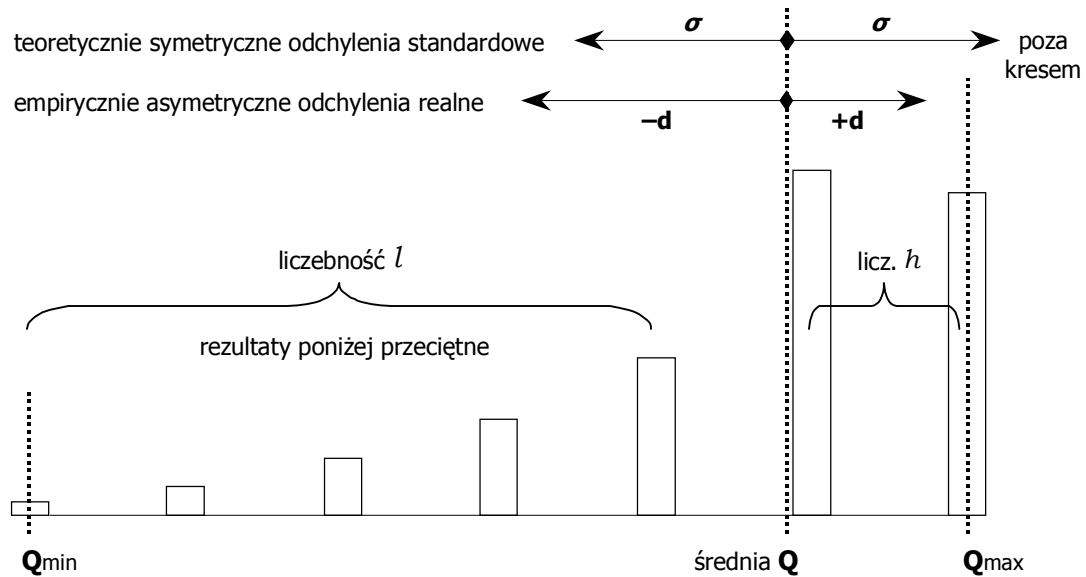
	9				1			4				1			16	4	1			9		4			1				
--	---	--	--	--	---	--	--	---	--	--	--	---	--	--	----	---	---	--	--	---	--	---	--	--	---	--	--	--	--

6° Suma  $\Sigma Y_i^2 = 50$  7° Odchylenie [-*d*] =  $-\sqrt{\{50 \cdot (1+15/10)/30\}} \approx -\sqrt{4,17} \approx -2,04$

8° Przypadki dodatnie  $Y_j$  podnosimy do kwadratu:

		1	4	1		1	1			1			4	1				1	4		1			4			1	4	1
--	--	---	---	---	--	---	---	--	--	---	--	--	---	---	--	--	--	---	---	--	---	--	--	---	--	--	---	---	---

9° Suma  $\Sigma Y_j^2 = 30$  10° Odchylenie [+*d*] =  $\sqrt{\{30 \cdot (1+10/15)/30\}} \approx \sqrt{1,67} \approx +1,29$



Ryc. 26. Odchylenia realne wierniej odwzorowują dewiacje rozkładu

Empiryczne rozkłady z próby, wynikające z pomiarów skalami postaw, mają specyficzną asymetrię. Oto skutek naturalnego w tej metodzie dążenia wypowiedzi w stronę wyników pozytywnych, najczęściej mamy do czynienia z rozkładem lewoskośnym. Wszak z lewej strony, poniżej średniej, respondenci mają większą swobodę wyboru. Gdyby obliczano klasyczne odchylenia standardowe i dodawano je do średniej, wówczas dochodziłoby do paradoksalnej sugestii, jakoby rozproszenie wyników mogło sięgać poza kres górny skali. Z tego względu w skalach dwuważonych istota analizy rozkładowej oparta jest na odchyleniach realnych, wierniej odzwierciedlających rozkłady empiryczne. Są one znacznie trafniejszą miarą cenionego w statystyce wyznacznika dewiacji.

Przypomnę klasyczny wzór na odchylenie standardowe i podam matematyczną zależność pomiędzy popularną miarą (poprawną tylko dla rozkładów symetrycznych) a ogólniejszymi miarami dwuskładowych odchylen realnych:

$$\sigma = \sqrt{\frac{1}{n} \sum_i^n (X_i - Q)^2} = \sqrt{\frac{l(-d)^2 + h(+d)^2}{l+h}} \quad \sigma \neq 0 \quad [5]$$

gdzie  $X_i$  – wagi lub średnie z wag wypowiedzi kolejnych respondentów  
 $Q$  – średnia jakość wypowiedzi z próby o liczebności  $n$ - przypadków  
 $l, h$  – liczebności podzbioru przypadków poniżej- i ponadprzeciętnych  
 $|-d|, |+d|$  – wartości bezwzględne odchylen realnych ze wzoru [4]

Odchylenie standardowe jest pierwiastkiem z wariancji całkowitej, natomiast odchylenia realne to pierwiastki z wariancji cząstkowych, liczonych osobno z lewej i prawej strony względem średniej. Parametry te są identyczne tylko dla rozkładów symetrycznych, zaś w rozkładach asymetrycznych odchylenie standardowe jest średnią kwadratowo-ważoną z obu odchylen realnych.

**Przykład 6. Odchylenie standardowe**

Przyjmijmy wagi ( $x$ ) takie same jak w przykładzie 5, uzyskane przez 30 studentów:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

Obliczyć odchylenie standardowe według klasycznego wzoru [5].

Kroki liczenia: 1° Suma  $\Sigma x = 30$  2° Średnia jakość  $Q = \Sigma x/n = 30/30 = 1$  3° Różnice ( $x - Q$ ):

0	-3	1	2	1	-1	1	1	-2	0	1	0	-1	2	1	-4	-2	-1	1	2	-3	1	-2	0	2	-1	0	1	2	1
---	----	---	---	---	----	---	---	----	---	---	---	----	---	---	----	----	----	---	---	----	---	----	---	---	----	---	---	---	---

4° Każdą z różnic  $Y$ , powstałą z odjęcia ( $x - 1$ ), należy podnieść do kwadratu, otrzymując:

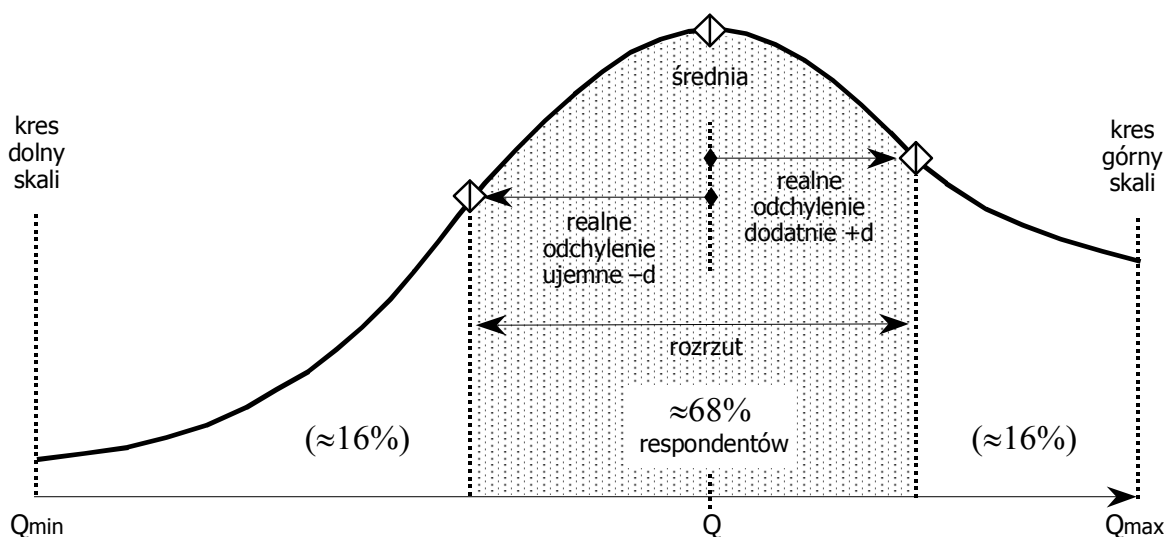
0	9	1	4	1	1	1	1	4	0	1	0	1	4	1	16	4	1	1	4	9	1	4	0	4	1	0	1	4	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	---

5° Suma  $\Sigma Y^2 = 80$  6° Wariancja  $\Sigma Y^2/n = 80/30 \approx 2,67$  7° Odchylenie  $\sigma \approx \sqrt{2,67} \approx 1,63$

Warto zwrócić uwagę na fakt, że wprowadzie przewija się tu pojęcie *wariancji*, lecz statystyka ta nie ma większego zastosowania w skalach dwuważonych, gdyż przy jej liczeniu traci się informacje dotyczące asymetrii rozkładu.

**Rozrzut – dyspersja oznak**

Często zamiennie ze słowem *odchylenie* używa się nazwy *rozrzut*. Jest to niepoprawne, gdyż odchylenie (dewiacja) to średniokwadratowa odległość od środka, natomiast rozrzut (dyspersja) to standardowy zakres rozpostarcia. Wyznaczanie dewiacji względem średniej jest istotą momentów centralnych, a wytyczanie dyspersji jest domeną niezwiązanych momentów absolutnych.



Ryc. 27. Realne miary dewiacji i dyspersji rozkładu empirycznego

W skalach dwuważonych trzymajmy się interpretacji następującej:

- *Odchylenia realne* – to dwie statystyki odstępstwa przypadków od średniej, tworzone ze średnich kwadratowych odległości od pozycji centralnej.
- *Rozrzut* – to statystyka zakresu rozproszenia przypadków, utworzona z obu odchyżeń i wytyczająca większościami kwalifikowany obszar dyspersji.
- *Rozstęp* – to całkowita rozpiętości pomiędzy dwoma skrajnymi przypadkami.

Zgodnie z przyjętą konwencją rozróżniania dewiacji od dyspersji, zdefiniujemy *Rozrzut* 'Dispersion' [d] jako sumę modułów obu odchyłeń realnych:

$$d = |-d| + |+d| \quad d \neq 0 \quad [6]$$

gdzie  $|-d|, |+d|$  – wartości bezwzględne odchyłeń realnych ze wzoru [4]

Parametr ten stanowi odniesienie standaryzacyjne dla innych miar, występuje wówczas w mianowniku wzorów. Przyjmuje się zatem założenie, że *Rozrzut* musi być większy od zera. Z tego powodu trzeba sprawdzać to, czy w jakiegokolwiek badanej próbie, wobec dowolnego bodźca-stwierdzenia nie ma przypadkiem całkowitej jednomyślności wypowiedzi (zły wskaźnik). Optymalny rozrzut wespół z minimalnym rozziwem są budulcem miary konkluzyjności.

Z cech rozkładów quasi-normalnych wynika oszacowanie, że w zakresie *Rozrzutu* powinny mieścić się rezultaty kwalifikowanej większości respondentów  $\approx 68\%$ , a reszta teoretycznie w równej części powyżej i poniżej. Jednakże przy dużej asymetrii zdarza się, że powyżej zakresu rozrzutu brakuje miejsca.

#### Przykład 7. Rozrzut

Przyjmijmy wagi (x) te same jak w przykładzie 5, uzyskane przez 30 studentów:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

Obliczyć *Rozrzut* wypowiedzi wg wzoru [6] oraz ustalić w tym zakresie liczebności wg ryc. 27.

Kontynuacja przykładu 5:  $[-d] \approx -2,04$   $[+d] = 1,29$   $6^\circ$  *Rozrzut*  $[d] \approx |-2,04| + |1,29| \approx 3,33$

1° Średnia Q = 1    2° Zakres rozrzutu  $\langle Q - |-d|; Q + |+d| \rangle = \langle 1 - 2,04; 1 + 1,29 \rangle = \langle -1,04; 2,29 \rangle$

3° W przedziale  $\langle -1,04; 2,29 \rangle$  mieści się 22 respondentów, poniżej 3, a powyżej 5 respondentów.

4° Odpowiada to procentowemu rozkładowi liczebności:  $\approx 73\%$ , 10% i  $\approx 17\%$ .

Powyższy przykład ilustruje inne zjawisko. Otóż kres dolny *Rozrzutu*  $\langle -1,04 \rangle$  znajduje się bardzo blisko wagi równej -1. Z tego powodu – uwzględniając błąd oszacowania wartości średniej oczekiwanej – w zasadzie te trzy przypadki -1 na równi mogłyby znaleźć się w zakresie dolnym, razem z rezultatami -2 i -3. Wówczas rozkład procentowy wynosiłby:  $\approx 63\%$ , 20% i  $\approx 17\%$ . W statystyce „prawda zwykle leży gdzieś pośrodku”, co potwierdza teorię, że na podstawie odchyłeń realnych można szacować liczebności przypadków mieszczących się w obszarze rozrzutu i poza nim. Warto sprawdzić to choćby na przykładzie 1.

Ślady	Wzorce
wskaźniki i ich agregacja	opcje wypowiedzi, wagi i ich średnie
kategorie miar statystycznych	momenty zwykłe, centralne i absolutne
przeciętne, uśrednienie oznak	jakość a intensywność wypowiedzi
niezgodność, dysjunkcja oznak	rozziw wypowiedzi przeciwstawnych
nieregularność, dewiacja oznak	odchylenia realne i standardowe
rozproszenie, dyspersja oznak	rozrzut, zróżnicowanie wypowiedzi

### Spójność – koniunkcja oznak

Narzędzia pomiarowe budowane zgodnie ze strategią ewaluacji splotowej powinny zawierać kontrolne pary wskaźników, próbujących dokładnie ten sam składnik indykatum, lecz z innej perspektywy czasoprzestrzennej. Zalecanym układem w kwestionariuszu jest umieszczanie par kontrolnych obok siebie. Przyjmujemy zatem, że kolejne pozycje kwestionariusza o numerach nieparzystych i parzystych {1 i 2; 3 i 4 ... 23 i 24} tworzą koniunkcję wskaźników tej samej zmiennej cząstkowej. W takim razie powinny one posiadać swoiste podobieństwa właściwości statystycznych, zwane koherentnością. Jako najprostszą miarę koherentności zdefiniujemy *Spójność* 'Consistency' [c]:

$$c = 1 - \frac{1}{nX_{\max}} \sum_{i=1}^n |X_{i;p-1} - X_{i;p}| \quad (\dots 1) \quad [7]$$

gdzie  $p = 2; 4; 6 \dots$

$X_{i;p-1}, X_{i;p}$  – nieparzyste i parzyste wskaźniki kontrolnych par wypowiedzi

$X_{\max}$  – najwyższa waga przypisana do kresu górnego skali, tutaj  $X_{\max} = +3$

$n$  – liczebność próby (klasy, grupy lub ogółu badanej zbiorowości)

Parametr ten zbudowany jest na różnicach pomiędzy wskaźnikami tworzącymi parę kontrolną, a ponieważ większe różnice byłyby miarą niespójności, stąd dla odwrócenia moment absolutny występuje we wzorze jako odjemnik. Dzielenie przez wartość kresowej wagi  $X_{\max}$  jest zabiegiem standaryzującym, umożliwiającym porównywanie spójności skal opartych na różnych wagach. Potencjalnie najwyższą wartością tego parametru jest 1, a realnie dobrą co najmniej 0,5. Do niższej spójności należy jednak podchodzić z ostrożnością.

Podczas tworzenia nowej skali parametr ten co prawda może służyć jako kwantyfikikator pomocniczy, lecz nie jako selekcyjny. Może być więc przydatny jedynie do optymalizacji, lecz nie do odrzucania bodźców-stwierdzeń. Wynika to z celowo zalecanej strategii różnicowania obu wskaźników pary kontrolnej, dlatego podczas interpretacji *Spójności* należy przede wszystkim wnikać w ich relacje czasoprzestrzenne – czy to owe zróżnicowania były przyczyną niespójności, czy może respondenci byli niekonsekwentni w swych wypowiedziach.

#### Przykład 8. Spójność

30 studentów uzyskało wagi ( $x_{p-1}$ ) oraz ( $x_p$ ) odpowiednio dla nieparzystej i parzystej pozycji skali:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
1	-1	3	3	2	-1	2	2	0	2	2	2	1	3	1	-1	-1	0	3	2	-1	2	-1	2	3	1	2	2	1	2

Obliczyć Spójność kontrolnej pary wskaźników według wzoru [7].

Kroki liczenia: 1° Ustala się wartości bezwzględne (Y) z różnic pomiędzy parami wskaźników:

0	1	1	0	0	1	0	0	1	1	0	1	1	0	1	2	0	0	1	1	1	0	0	1	0	1	1	0	2	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

2° Suma  $\Sigma Y = 18$  3° Niespójność  $\Sigma Y/n/X_{\max} = 18/30/3 = 0,2$  4° Spójność [c] =  $1 - 0,2 = 0,8$

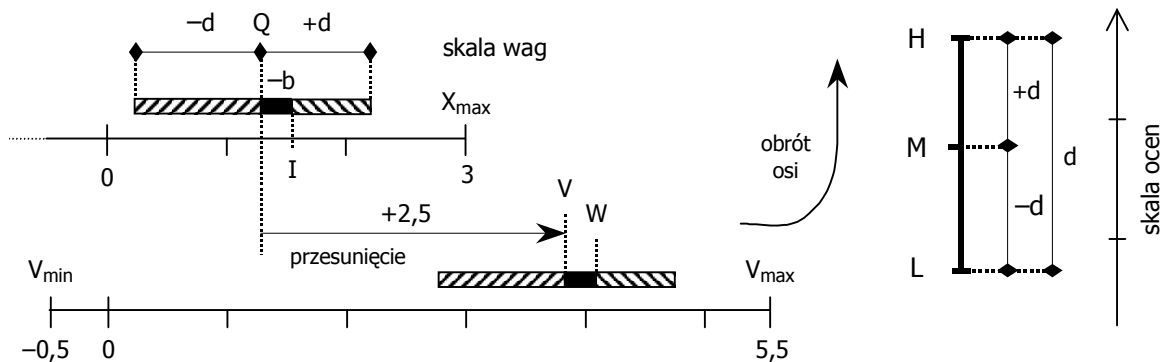
**Wartość | Ważkość – średnie przeskalowane**

Przypomnę, że w *Skali dwuważonych ocen* zaproponowałem ewaluatorom jednolitą konwencję przeskalowywania poziomów, mającą na celu ułatwienie interpretacji i szersze uprzystępnianie wyników. Chodzi o przesunięcie wartości średnich ze skali wymiarowanej za pomocą wag, na skalę wyrażaną za pomocą ocen szkolnych (akademickich). W celu zapewnienia całkowitej jednoznaczności nadałem odmienne nazewnictwo i oznaczenia dla parametrów przed i po konwersji na skalę ocen. Jakość i intensywność po przesunięciu przyjmują odpowiednio nazwy *Wartość* 'Value [V] oraz *Ważkość* 'Weight [W]:

$$V = 2,5 + Q \qquad W = 2,5 + I \quad (\dots 5,5) \quad [8]$$

gdzie  $Q$  – średnia jakość wypowiedzi,  $I$  – średnia intensywność wypowiedzi, obliczone ze wzoru [2]

Przeskalowanie poziomów jest dodaniem liczby 2,5 do uprzednio wyliczonych wartości średnich. Zwracam uwagę na to, że o ile *Wartość* można by obliczać także z wag przeskalowanych, o tyle *Ważkość* musi być wyznaczana zawsze z wag pierwotnych  $\langle -3\dots+3 \rangle$ . Z tego powodu konwersję ze skali wag na skalę ocen należy przeprowadzić na końcu, dopiero po obliczeniu danego parametru.



Ryc. 28. Konwersja ze skali wag na skalę ocen

Teoretycznie wartości wypowiedzi na skali ocen mogłyby wypełniać cały zakres  $\langle -0,5\dots+5,5 \rangle$ , lecz w praktyce pomiarowej wszystkie uśrednione oznaki cząstkowe z reguły mieszczą się w zakresie  $\langle 2\dots5 \rangle$ . Tak więc do wizualizacji wyników na wykresach umieszcza się cztery górne przedziały ocen 2, 3, 4, 5, dodatkowo z rozpiętością  $\pm 0,5$  podziałki. Natomiast do opisu jakościowego wykorzystuje się nazwy czterech poziomów w rozumieniu osiągnięć akademickich (*niedostateczny*, *dostateczny*, *dobry*, *bardzo dobry*). W takiej skali ocen *Wartość*  $V=2$  oznacza poziom niedostateczny i nie tam tutaj miejsca na poziom *mierny* czy *dopuszczający*, dla którego każda z proponowanych nazw okazuje się nietrafna. Parametr [V] jest bazową miarą dla takich estymatorów standardu ewaluatywnego, jak *Baza* i *Norma* oraz *Poziomy HML*.

**Baza | Norma – średnie oczekiwane**

Empiryczną miarą najbardziej zbliżoną do teoretycznego estymatora jest *Norma*. Wbrew pejoratywnym skojarzeniom, nie jest to jakiś narzucony pułap do osiągnięcia, lecz specyficzna podstawa normalizacji skali, umożliwiająca porównywanie wyników. *Bazę* 'Base [B'] i *Normę* 'Norm [N']' oblicza się jako średnie arytmetyczne ze średnich wartości osiągniętych odpowiednio w pomiarach początkowych i końcowych przez wszystkie podgrupy respondentów:

$$B' = \frac{1}{k} \sum_{i=1}^k (V_i') \quad N'' = \frac{1}{k} \sum_{i=1}^k (V_i'') \quad \langle 2,5; 5,5 \rangle \quad [9]$$

gdzie

$V_i'$  – średnie wartości w i-tych klasach (grupach) w pomiarze początkowym

$V_i''$  – średnie wartości w i-tych klasach (grupach) w pomiarze końcowym

$k$  – liczba klas lub grup wyodrębnionych z ogółu badanej zbiorowości

Istotną cechą *Normy* w porównaniu ze średnią ogólną dla całej próby badawczej jest to, że niweluje ona nieco wpływ zróżnicowanych liczebności w poszczególnych grupach. Wyjaśnię to na przykładzie posiadanych przeze mnie danych empirycznych. Otóż dysponuję wynikami kilkuset studentów pedagogiki, a z innych kierunków studiów (w tym technicznych) otrzymuję zwykle reprezentacje mniejsze niż stuosobowe. Rozdzielczy sposób obliczania najpierw średnich grupowych i dopiero z nich obliczanie *Normy* powoduje, że nie jest ona zdominowana przez osiągnięcia tej najliczniejszej grupy.

Z drugiej strony – włączenie do normy rezultatów nawet nielicznej, lecz wyjątkowo nietypowej grupy, w jakimś stopniu mogłoby wypaczyć standardy. Z tego względu dane empiryczne współtworzące *Normę* i określające standardy winny pochodzić z grup możliwie wiernie reprezentujących badaną populację.

**Ufność realna – tolerancja średnich**

Jeśli badacz chce uogólnić wnioski na populację, to nie może poprzestać jedynie na wyznaczeniu statystyk z próby. Na ich podstawie dodatkowo musi oszacować, jak dalece są one wiarygodne i na ile mogłyby ulec zmianie, gdyby zbadał całą populację. Jedną z metod określania tolerancji średnich jest *estymacja przedziałowa*, oparta na charakterystyce rozkładu normalnego, który dla poziomu istotności np.  $\alpha = 5\%$  ma wartość 1,96. Zakres tolerancji, przy której z błędem  $\leq 5\%$  hipotetyczna średnia populacji mieści się w pobliżu średniej obliczonej z próby, nazywa się *przedziałem ufności*. Do wyznaczenia przedziału ufności potrzebna jest jeszcze znajomość odchylenia i liczebności. Ze względu na to, że statystyka skal dwuważonych opiera się na założeniu potencjalnej asymetrii rozkładu, także i tu jako bazowe wyznaczniki dewiacji wykorzystajmy obydwie odchylenia realne, osobno ujemne [-d] i dodatnie [+d].



Analogicznie do dewiacji, również *Ufność realna* 'Real Confidence' definiowana jest przez dwa parametry – przedział ufności ujemny  $[-c]$  i dodatni  $[+c]$ :

$$-c = U(\alpha) \cdot \frac{-d}{\sqrt{n}} \quad +c = U(\alpha) \cdot \frac{+d}{\sqrt{n}} \quad U(5\%) = 1,96 \quad [10]$$

gdzie 1,96 – wartość z tabel rozkładu normalnego dla istotności  $\alpha = 5\%$   
 $-d, +d$  – odchylenie realne ujemne i dodatnie obliczone ze wzoru [4]  
 $n$  – liczebność próby (klasy, grupy lub ogółu badanej zbiorowości)

Przedziały ufności  $\alpha=5\%$  jako momenty centralne wyznaczają zakres tolerancji wokół wartości średniej z próby. W razie potrzeby zmiany poziomu istotności, odpowiednią wartość współczynnika  $U(\alpha)$  odczytuje się z tabel lub oblicza za pomocą funkcji arkusza kalkulacyjnego<sup>45</sup>. Najpewniejszą sytuację zaufania co do wartości średniej mamy wówczas, gdy przedziały ufności są najmniejsze. Oznacza to konieczność racjonalnego zwiększania liczebności prób badawczych, o ile celem ewaluacji jest uogólnienie wniosków na populację. Poniższy przykład dobitnie ilustruje, że na podstawie tak nielicznej próby  $n=30$  osób nie można precyzyjnie oszacować wartości średniej oczekiwanej dla populacji.

#### Przykład 9. Ufność realna

Przyjmijmy wagi ( $x$ ) takie same jak w przykładzie 5, uzyskane przez 30 studentów:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

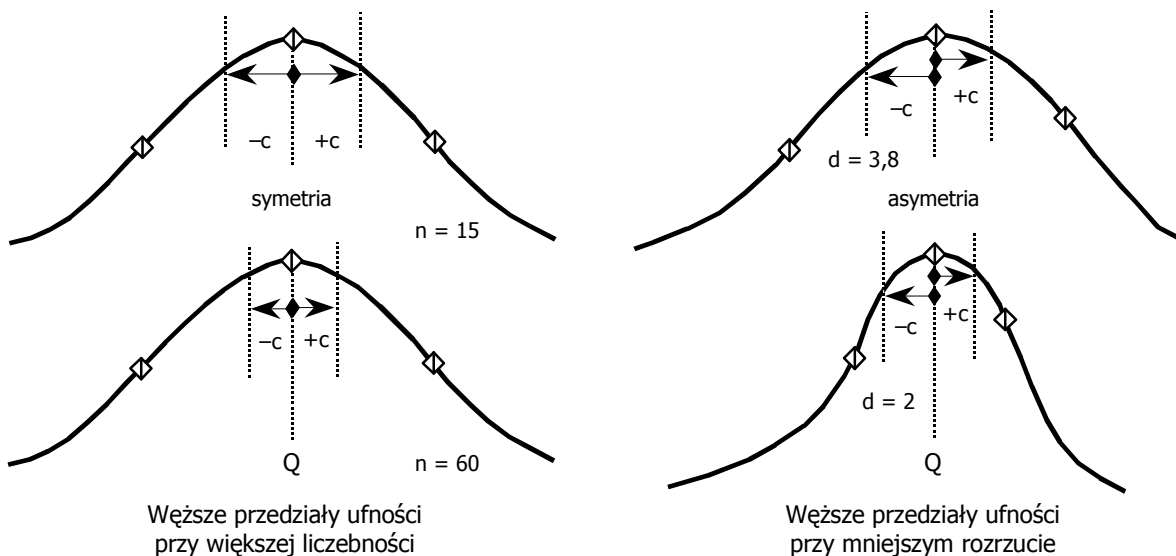
Obliczyć Ufność realną wg wzoru [10] oraz ustalić tolerancję średniej wg ryc. 29.

Wyliczenia z przykładu 5:  $[-d] \approx -2,04$   $[+d] \approx 1,29$   $Q = 1$

1° Przedział ufności ujemny  $[-c] = 1,96 \cdot (-2,04) / \sqrt{30} \approx -0,73$  Poziom istotności  $\alpha = 5\%$

2° Przedział ufności dodatni  $[+c] = 1,96 \cdot 1,29 / \sqrt{30} \approx 0,46$  3° Zakres ufności  $|-0,73| + 0,46 \approx 1,19$

4° Tolerancja średniej  $\langle Q - |-c|; Q + |+c| \rangle \approx \langle 1 - 0,73; 1 + 0,46 \rangle \approx \langle 0,27; 1,46 \rangle$



Ryc. 29. Przedziały ufności realnej zależne od liczebności, rozrzutu i skośności

### Poziomy HML – estymacja pozycji

Najczęściej celem ewaluacji nie jest uogólnienie na populację, lecz wyszczególnienie lokalne, polegające na oszacowaniu pozycji, jaką na skali ocen zajmuje wybrana grupa badawcza. Do tego wystarczy estymacja trójpunktowa. Średnia arytmetyczna oraz punkty po obu jej stronach w odległości odchyłeń uznawane są za najlepsze estymatory pozycyjne. Wyznaczmy więc trzy łatwo interpretowalne *Poziomy*: *wysoki* 'high' [H], *średni* 'mid' [M] i *niski* 'low' [L], wykorzystując zdefiniowane wcześniej odchylenia realne oraz wartość średnią:

$$H = M + |d| \quad M = V \quad L = M - |d| \quad (\dots 5,5) \quad [11]$$

gdzie  $+d, -d$  – odchylenia realne,  $V$  – średnia wartość ze wzoru [8]

Poziomy HML osadzone są zawsze na skali ocen, a ponieważ podziałka tej skali jest umowna, więc nie jest konieczne określanie przedziałów ufności. Poziomy te nabierają jednak wiarygodności w ujęciu różnicowo-dystansowym, określającym zmiany, jakie zachodzą pomiędzy pierwszym i drugim pomiarem. Stanowią zatem ważne miary bazowe dla dystansowych parametrów trendu.

#### Przykład 10. Poziomy HML

Przyjmijmy wagi (x) takie same jak w przykładzie 5, uzyskane przez 30 studentów:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

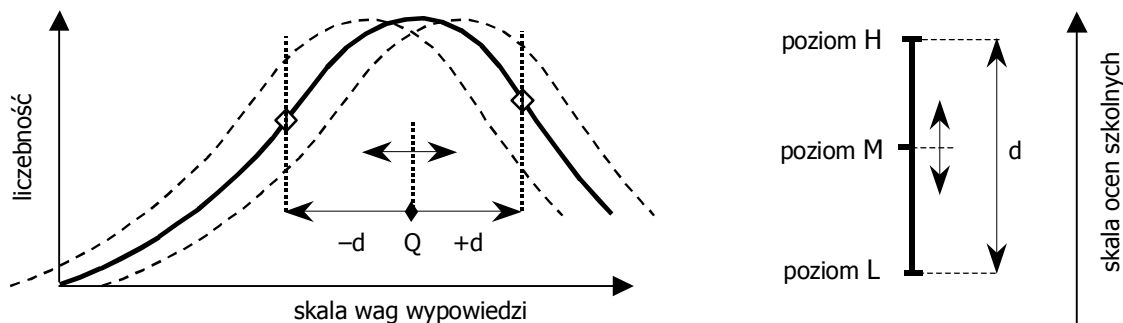
Obliczyć Poziomy HML wg wzoru [11] z konwersją na skalę ocen szkolnych wg wzoru [8].

Wyliczenia z przykładu 5:  $[-d] \approx -2,04$   $[+d] \approx 1,29$   $Q = 1$  Założono przesunięcie  $+2,5$

1° Poziom średni  $M = Q + 2,5 = 1 + 2,5 = 3,5$     2° Poziom niski  $L = M - |d| \approx 3,5 - 2,04 \approx 1,46$

3° Poziom wysoki  $H = M + |d| \approx 3,5 + 1,29 \approx 4,79$     4° Rozrzut  $[d] = 3,33$

Struktura liczebności na skali ocen rozkłada się tak, jak na skali wag. W zakresie rozrzutu  $[d]$  plasuje się nieco ponad  $2/3$  respondentów, a poza nim po ok.  $1/6$  z każdej strony. Jest wysoce prawdopodobne, że w klasie 30-osobowej wynik w zakresie HL uzyska 20÷21 uczniów, a poniżej i powyżej po ok. 5-ciu. Ewentualne przesunięcie średnich  $Q$  i  $M$  nie zmienia tych proporcji. Poziomy HML mogą być kresami podziałów na jakościowo różne 3 podgrupy.



Ryc. 30. Wyższa wiarygodność liczebności w przedziale HL niż ufność do poziomu M

**Skośność – asymetria rozkładu**

Asymetria rozkładu empirycznego jest powszechnym zjawiskiem, warto więc wyznaczać współczynnik charakteryzujący stopień tegoż odkształcenia. Zamiast klasycznego momentu centralnego rzędu trzeciego, zastosujemy tu łatwiejszy do obliczenia, standaryzowany parametr *Skośności* 'Skewness [s], będący stosunkiem różnicy do sumy modułów odchyłeń realnych:

$$s = \frac{|+d| - |-d|}{|+d| + |-d|} \quad \langle -1; +1 \rangle \quad [12]$$

gdzie  $|+d|$ ,  $|-d|$  – wartości bezwzględne odchyłeń realnych ze wzoru [4]

Zerowa wartość współczynnika skośności świadczy o symetrii rozkładu empirycznego, dodatnia o asymetrii prawostronnej, a ujemna – o asymetrii lewostronnej. Warto jednak pamiętać, że nawet znaczna skośność rozkładu nie jest wyznacznikiem jakości procesu, a jedynie oznaką, na ile wiarygodny dla konkretnej próby jest współczynnik autokorelacji. Ocena skośności jest natomiast konieczna w testach istotności, przy podejmowaniu decyzji, czy uprawnione jest oparcie modelu na rozkładzie normalnym. Asymetria rozkładu wpływa w niewielkim stopniu na obniżenie konkluzyjności wskaźników.

**Przykład 11. Skośność**

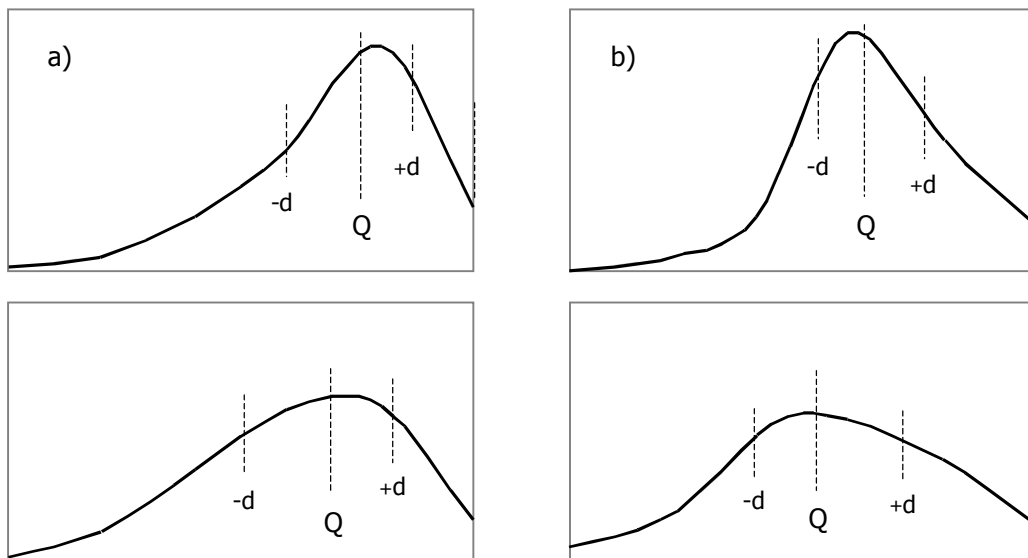
Przyjmijmy wagi (x) takie same jak w przykładzie 5, uzyskane przez 30 studentów:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
---	----	---	---	---	---	---	---	----	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	---	---	---	---	---

Obliczyć Skośność rozkładu według wzoru [12].

Wyliczenia z przykładu 5:  $[-d] \approx -2,04$   $[+d] \approx 1,29$  Asymetria lewostronna

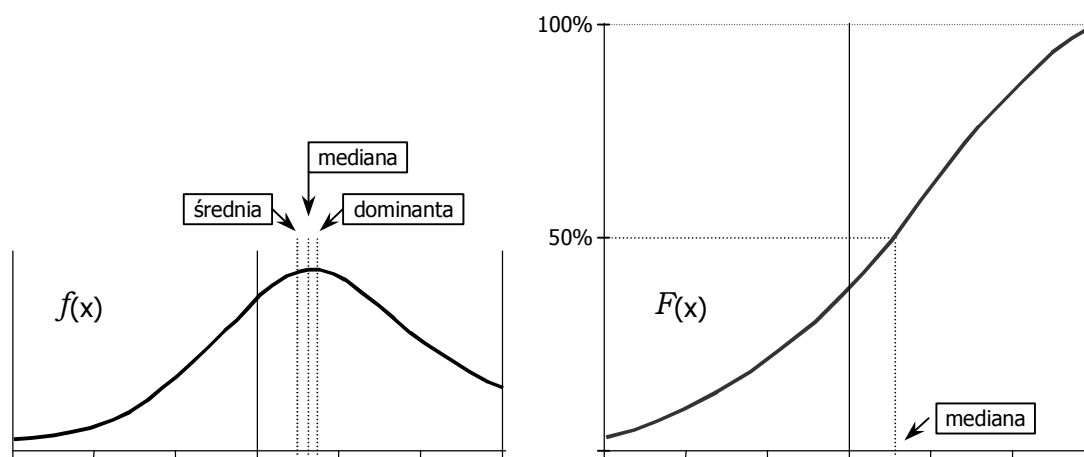
1° Skośność  $s \approx (|1,29| - |-2,04|) / (|1,29| + |-2,04|) \approx -0,75 / 3,33 \approx -0,225$



Ryc. 31. Skośność – przykłady asymetrii a) lewostronnej, b) prawostronnej

### Częstość realna – estymacja rozkładu

Skośność rozkładów empirycznych powoduje, że opieranie statystyk zaawansowanych (np. testów istotności) na symetrycznym modelu rozkładu normalnego nie jest poprawne. Dzieje się tak m.in. dlatego, że we wzorze na krzywą Gaussa występuje tylko jeden parametr odchylenia, choć odchylenia z obu stron są różne. Ponadto parametr przesunięcia wyznacza dominantę, a nie średnią, która w rozkładach asymetrycznych jest w innym miejscu.



Ryc. 32. Gęstość prawdopodobieństwa i dystrybuanta rozkładu realnego – parametry pozycyjne przy asymetrii lewostronnej

Z tego względu do analiz warto wykorzystywać zmodyfikowany wzór na funkcję gęstości lub częstości (w zależności od skalowania) zmiennej losowej oczekiwanej dla populacji, a szacowanej na podstawie rozkładu z próby empirycznej:

$$f(x) = \begin{cases} \frac{e^{-\frac{(x-Q_{\text{mod}})^2}{2|d|}} \cdot \frac{|+d|^2}{|d|^2}}{\xi \sigma \sqrt{2\pi}} & x < Q_{\text{mod}} \\ \frac{e^{-\frac{(x-Q_{\text{mod}})^2}{2|+d|}} \cdot \frac{|-d|^2}{|d|^2}}{\xi \sigma \sqrt{2\pi}} & x \geq Q_{\text{mod}} \end{cases} \quad [13]$$

gdzie  $|d|$ ,  $|+d|$  – wartości bezwzględne odchyłeń realnych ze wzoru [4]  
 $\sigma$  – odchylenie standardowe ze wzoru [5], oczekiwane przy symetrii  
 $x$  – zmienna ciągła,  $e$  – liczba 2,71828...  $\pi$  – liczba 3,14159...  
 $Q_{\text{mod}}$  – estymator wartości modalnej, dla której  $f(x)$  osiąga maksimum (dominanta)  
 $\xi$  – estymator standaryzujący, przy którym pole pod krzywą  $f(x)$  jest równe 1

Funkcja  $f(x)$  – nazwana *Częstością realną* 'Real Frequency' – dla rozkładów symetrycznych jest równoważna gaussowskiej, a przy asymetrii jest korygowana proporcjami odchyłeń realnych. Do wykreślenia krzywej niezbędne jest oszacowanie wartości modalnej  $Q_{\text{mod}}$ , czyli estymacja położenia maksimum funkcji zwanego dominantą. Ponadto na potrzeby skalowania osi Y w jednostkach gęstości prawdopodobieństwa, liczebności w przedziałach lub częstości występowania niezbędne jest oszacowanie parametru standaryzującego  $\xi$  (ksi).

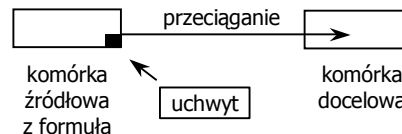
Oszacowanie parametrów rozkładu realnego w dość łatwy sposób można przeprowadzić za pomocą arkusza kalkulacyjnego, przy czym takie działanie ma uzasadnienie tylko na wyższych poziomach agregacji zmiennych, a więc dla czynników, rezultatów połówkowych bądź dla wyniku ogólnego. Chodzi o to, że aproksymacja krzywej wymaga większej liczby progów kwantyzacji niż zapewnia to próbkowanie wskaźników (7) czy składników (13 poziomów). Wzrastającą rozdzielczość skali, wyrażoną w liczbie rozróżnialnych poziomów badanej cechy, uzyskuje się dzięki agregacji wskaźników i oblicza ze wzoru:

$$\text{rozdzielczość} = \text{liczba zagregowanych wskaźników} \cdot (\text{gradacja próbkowania} - 1) + 1$$

Przy zalecanej gradacji próbkowania równej siedmiu progom (7 opcji wypowiedzi) oznacza to liczbę zróżnicowanych poziomów czynników  $4 \cdot (7-1)+1=25$ , dla rezultatów połówkowych  $12 \cdot (7-1)+1=73$  a dla wyników aż  $24 \cdot (7-1)+1=145$ . Rozdzielczość określa maksymalną możliwą liczbę słupków na histogramach. O ile więc dla zmiennych zagregowanych ma sens aproksymowanie wielu słupków do postaci krzywej ciągłej, zdefiniowanej za pomocą funkcji, o tyle na poziomie wskaźników lepiej pozostać przy analizie histogramów.

### Przykład 12. Częstość realna

Przeprowadźmy estymację rozkładu za pomocą arkusza kalkulacyjnego\*. Do tego celu niezbędna jest umiejętność posługiwania się myszą przy zautomatyzowanym powielaniu formuł i ciągów liczbowych. Objasnię ową czynność. Jeśli do jednej komórki wpiszemy określony wzór formuły liczącej, to można automatycznie powielić ten wzór na sąsiednie komórki, chwytając myszą za tzw. uchwyt i przeciągając go (z wciśniętym przyciskiem myszy) wzdłuż wiersza lub kolumny.



W kopiach formuł automatycznie modyfikowane są adresy, do których odnosi się wzór. W ten sam sposób można utworzyć ciąg liczb, jeśli zaznaczy się dwie komórki zawierające dwa pierwsze elementy ciągu, po czym przeciągnie uchwyt. Do estymacji czynnika potrzebny jest ciąg 25 poziomów  $x \in \langle -3; 3 \rangle$  o skoku 0,25.

1° Wpiszmy do komórki A1 pierwszy element = -3 a do komórki B1 drugi = -2,75

2° Zaznaczymy myszą komórki A1 i B1 tak, żeby obie były aktywne (otoczone grubszą ramką).

	A	B	...	M	N	O	P	Q	R	...	X	Y
1	-3	-2,75	...	0	0,25	0,5	0,75	1	1,25	...	2,75	3

3° Chwyćmy myszą za uchwyt w dolnym rogu komórki B1 i przeciągnijmy go aż do komórki Y1.

W efekcie przeciągania automatycznie zostaje wygenerowany zbiór 25 argumentów [x] funkcji  $f(x)$ .

\* Objasnienia dotyczą arkusza MS Excel. Aby na ekranie widzieć cały obszar komórek, na których działamy, należy ustawić wyższą rozdzielczość ekranu, minimum 1152×864 pikseli, a ponadto zaznaczyć kolumny od A do Y, po czym zmniejszyć ich format na szerokość 4,9. Kolumna Z będzie zawierać wyniki obliczeń, warto więc ustawić jej czcionki na pogrubione.

## Wprowadzanie danych

4° Wpiszmy dane skumulowanych liczebności\* w trzecim wierszu do komórek o adresach A3÷Y3:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
0	0	0	0	1	1	1	2	9	13	16	46	70	114	135	151	163	128	119	84	58	29	7	1	2

Największa liczebność występuje w kolumnie Q, czyli dla empirycznej wartości modalnej  $[Q_x] = 1$

5° Zliczmy: ile osób uzyskało wyniki poniżej  $[l]$ , a ile co najmniej przeciętne  $[h]$

6° Do komórki Z1 wpisujemy formułę obliczania liczebności  $low = \text{SUMA}(A3:P3)$

7° Do komórki Z2 wpisujemy formułę obliczania liczebności  $high = \text{SUMA}(Q3:Y3)$

8° Do Z3 wpisujemy formułę obliczania łącznej liczby badanych  $= \text{SUMA}(A3:Y3)$

	Z	AA
1	<b>559</b>	$[l]$
2	<b>591</b>	$[h]$
3	<b>1150</b>	$[n]$

## Obliczanie średniej

9° W komórce A5 wpisujemy formułę  $=A\$1*A3$  Symbol \$1 oznacza stałość adresu wiersza 1.

10° Chwyćmy myszą uchwyt komórki A5 i przeciągnijmy formułę poziomo aż do komórki Y5.

0	0	0	0	-2	-1,75	-1,5	-2,5	-9	-9,75	-8	-11,5	0	28,5	67,5	113	163	160	179	147	116	65,3	17,5	2,75	6
---	---	---	---	----	-------	------	------	----	-------	----	-------	---	------	------	-----	-----	-----	-----	-----	-----	------	------	------	---

Uzyskany zbiór liczb stanowi składniki potrzebne do obliczenia empirycznej wartości średniej  $[Q]$ .

11° W komórce Z5 wpisujemy formułę sumy i dzielenia przez  $[n] = \text{SUMA}(A5:Y5)/Z3$

12° W komórce AA5 wpisujemy symbol średniej  $[Q]$ . Sprawdźmy wynik:

	Z	AA
5	<b>0,886304</b>	$[Q]$

Wartość średniej  $[Q]=0,886304$  usytuowana jest poniżej empirycznej wartości modalnej  $[Q_x]=1$

## Obliczanie odchylenia realnych

13° W komórce A8 wpisujemy formułę  $=A3*(A\$1-\$Z5)^2$  Adres \$Z oznacza stałą kolumnę Z.

14° Chwyćmy myszą uchwyt komórki A8 i przeciągnijmy formułę poziomo aż do komórki Y8.

0	0	0	0	8,33	6,95	5,69	9,13	32	34,8	30,7	59,4	55	46,2	20,1	2,81	2,11	16,9	44,8	62,7	71,9	53,9	18,2	3,47	8,94
---	---	---	---	------	------	------	------	----	------	------	------	----	------	------	------	------	------	------	------	------	------	------	------	------

Uzyskany zbiór liczb stanowi składniki wariancji potrzebne do obliczenia odchylenia realnych.

Odchylenie ujemne składa się z przypadków  $[x]$  poniżej średniej  $[Q]$ , czyli z kolumn od A do P.

15° W komórce Z7 wpisujemy formułę  $=\text{PIERWIASTEK}((1+Z2/Z1)/Z3*\text{SUMA}(A8:P8))$

16° W komórce AA7 wpisujemy symbol odchylenia ujemnego  $[-d]$ . Wynik:

	Z	AA
7	<b>0,746097</b>	$[-d]$

Odchylenie dodatnie składa się z przypadków  $[x]$  powyżej średniej  $[Q]$ , czyli z kolumn od Q do Y.

17° W komórce Z8 wpisujemy formułę  $=\text{PIERWIASTEK}((1+Z1/Z2)/Z3*\text{SUMA}(Q8:Y8))$

18° W komórce AA8 wpisujemy symbol odchylenia dodatniego  $[+d]$ . Wynik:

	Z	AA
8	<b>0,692018</b>	$[+d]$

## Obliczanie odchylenia standardowego

19° W komórce Z9 wpisujemy formułę  $=\text{PIERWIASTEK}((Z1*Z7^2+Z2*Z8^2)/(Z1+Z2))$

20° W komórce AA9 oznaczmy skrótem odchylenie standardowe  $[sd]$ . Wynik:

	Z	AA
9	<b>0,718814</b>	$[sd]$

## Wstępne podstawienie wartości modalnej

21° W komórce Z11 wpisujemy wartość modalną taką, jak dominanta empiryczna  $= 1$

22° W komórce AA11 wpisujemy skrót symbolu wartości modalnej  $[Q_{mod}]$ .

	Z	AA
11	<b>1</b>	$[Q_{mod}]$

## Wstępne podstawienie parametru standaryzującego

23° W komórce Z12 wpisujemy liczbę wskaźników tworzących czynnik  $= 4$

24° W komórce AA12 wpisujemy nazwę parametru standaryzującego  $[ksi]$ .

	Z	AA
12	<b>4</b>	$[ksi]$

Liczba 4 oznacza, że podziałki jednostkowe skalowania osi X są podzielone na ćwiartki (skok 0,25).

\* Dane empiryczne, zapisane w formie skumulowanych liczebności, to tzw. szereg rozdzielczy. Kolejne elementy szeregu informują o tym, ilu respondentów uzyskało w pomiarze rezultat zgodny z wartością określoną dla każdego z poziomów zróżnicowania argumentu  $[x] \in \langle -3; 3 \rangle$ .

*Estymacja rozkładu Częstości realnej*

25° W komórce A10 wpiszmy formułę zgodną ze wzorem na *Częstość realną*  $f(x)$  dla  $[x] < Q_{\text{mod}}$   
 $=1/\$Z12/\$Z9/PIERWIASTEK(2*PI())*EXP(-((A1-\$Z11)^2)/2*\$Z8^2/\$Z7^4)$

26° Chwyćmy myszą uchwyt komórki A10 i przeciągnijmy formułę poziomo tylko do komórki R10.

27° W komórce R10 zmieńmy końcowy fragment formuły dla  $[x] \geq Q_{\text{mod}}$  ...  $*\$Z7^2/\$Z8^4)$

28° Chwyćmy uchwyt komórki R10 i przeciągnijmy tę zmienioną formułę poziomo do komórki Y10.

0	0	0	0	0	0	0	0	0,01	0,01	0,02	0,04	0,06	0,09	0,11	0,13	0,14	0,13	0,1	0,07	0,04	0,02	0,01	0	0
---	---	---	---	---	---	---	---	------	------	------	------	------	------	------	------	------	------	-----	------	------	------	------	---	---

Obliczone współczynniki częstości są pamiętane w komórkach bardziej dokładnie niż wyświetlane.

*Weryfikacja częstości skumulowanej (dystrybuanty)*

29° W komórce Z10 wprowadźmy formułę sumowania częstości =SUMA(A10:Y10) 

Z	AA
---	----

30° W komórce AA10 wpiszmy symbol dystrybuanty [Fx]. Sprawdźmy wynik: 

10	<b>1,00562</b>	[Fx]
----	----------------	------

Wartość ta docelowo ma być równa jedności, czyli współczynnik standaryzujący wymaga korekty.

*Estymacja rozkładu Liczebności realnej*

31° W komórce A4 wpiszmy formułę =\\$Z3\*A10 Przeciągnijmy tę formułę do komórki Y4.

32° W zaznaczonym obszarze od A4 do Y4 zmieńmy format komórek na liczbowe całkowite:

0	0	0	0	0	0	1	3	7	15	28	48	74	103	132	152	160	148	118	81	47	24	10	4	1
---	---	---	---	---	---	---	---	---	----	----	----	----	-----	-----	-----	-----	-----	-----	----	----	----	----	---	---

Taki format jest bardziej przejrzysty, a mimo to wartości są pamiętane dokładniej niż wyświetlane.

*Weryfikacja liczebności skumulowanej*

33° W komórce Z4 wprowadźmy formułę sumowania liczebności =SUMA(A4:Y4) 

Z	AA
---	----

34° W komórce AA4 wpiszmy symbol liczebności estymowanej [nE] Wynik: 

4	<b>1156,463</b>	[nE]
---	-----------------	------

Wartość ta docelowo ma być równa 1150, czyli współczynnik standaryzujący wymaga korekty.

*Weryfikacja estymatora średniej*

35° W komórce A6 wpiszmy formułę =A\\$1\*A4 Przeciągnijmy tę formułę do komórki Y6.

-0	-0,01	-0,03	-0,1	-0,3	-0,81	-1,91	-3,99	-7,25	-11,2	-14	-11,9	0	25,8	65,8	114	160	185	177	141	94,8	53,9	26	10,7	3,73
----	-------	-------	------	------	-------	-------	-------	-------	-------	-----	-------	---	------	------	-----	-----	-----	-----	-----	------	------	----	------	------

36° W komórce Z6 wpiszmy formułę sumy i dzielenia przez [nE] =SUMA(A6:Y6)/Z4 

Z	AA
---	----

37° W komórce AA4 wpiszmy symbol średniej estymowanej [QE] Wynik: 

6	<b>0,869362</b>	[QE]
---	-----------------	------

Wartość ta docelowo ma być identyczna z  $Q=0,886304$ , czyli należy skorygować wartość modalną.

Przerwijmy realizację przykładu, abym mógł skomentować aktualny stan. Dotychczasowe czynności potrzebne były po to, ażeby utworzyć formularz obliczeniowy. Po zapisaniu do pliku, formularz ten może służyć wielokrotnie do wyznaczania rozkładu wszystkich innych czynników empirycznych, a po modyfikacji formularza – także do analizy rezultatów połówkowych i wyników. Pierwszy wiersz arkusza/formularza zawiera narastające wartości zmiennej  $[x]$  i jest to ciąg wartości skalujących osi odciętych  $X$  w jednostkach wagowych jakości cechy. Rozkłady różnych funkcji  $f(x)$  znajdują się w wierszach 3÷10. Oś rzędnych  $Y$  może być skalowana w odmiennych jednostkach: jeśli chcemy uzyskać wykres liczebności skalowanej bezwzględna liczbą respondentów  $[n]$ , to pobieramy dane empiryczne z 3. wiersza, a rozkłady oczekiwane z wiersza 4. Jeśli natomiast potrzebny jest estymowany dla populacji rozkład gęstości prawdopodobieństwa lub częstości względnej, bądź frekwencji procentowej, to pobieramy współczynniki z 10. wiersza opracowanego powyżej formularza.

W każdym dalszym wykorzystaniu formularza użytkownik wprowadza dane tylko w wierszu trzecim, do komórek o adresach A3÷Y3. Ponadto musi podać dwa parametry: wartość modalną  $Q_{\text{mod}}$  w komórce Z11 oraz współczynnik standaryzujący  $\xi$  [ksi] w komórce Z12. W pierwszym podejściu przyjmuje się wartość modalną taką, jaka wynika z pomiaru, czyli jest to poziom cechy, jaki wystąpił najczęściej (najwyższy słupek histogramu liczebności). Natomiast początkową wartość współczynnika standaryzującego przyjmuje się jako liczbę całkowitą, która odpowiada liczbie wskaźników tworzących dany poziom zmiennej (czynnik, rezultat połówkowy lub wynik). W procesie standaryzacji rozkładu wstępne wartości  $Q_{\text{mod}}$  i  $\xi$  muszą zostać doprecyzowane. Wyjaśnię istotę oszacowywania tych dwóch ważnych parametrów estymacji.

**Standaryzacja rozkładu.** W fazie weryfikacji estymowanego w powyższym przykładzie rozkładu stwierdziliśmy, że charakterystyki rozkładu oczekiwanego nie są w pełni zgodne z pierwotnymi statystykami danych empirycznych. Chodzi o liczebność ( $1156,463 \neq 1150$ ) oraz o średnią ( $0,869362 \neq 0,886304$ ). Wprawdzie różnice te nie są duże, lecz m.in. po to tworzyliśmy formularz, aby je całkowicie wyeliminować. Oprócz tego, po skumulowaniu współczynników częstości, dla rozkładu standaryzowanego powinniśmy uzyskać dokładnie 1, co oznacza 100% przypadków, a nie 100,562%. Także i tę nieprawidłowość można usunąć. Czynność standaryzacji polega na stopniowym korygowaniu wartości parametrów  $Q_{\text{mod}}$  i  $\xi$  tak długo, aż uzyska się najlepsze przybliżenie par wartości empirycznych i estymowanych  $\{[n] = [nE]; [Q] = [QE]\}$ , a ponadto osiągnie się wartość graniczną dystrybuanty równą jedności  $\{[F_x] = 1\}$ . Kolejne fazy standaryzacji to wpisywanie najpierw do komórki Z11, a później do Z12 liczb o coraz większej dokładności w części ułamkowej, zmienianych małymi krokami iteracji tak długo, aż uzyska się wystarczającą zgodność zawartości komórek Z3 = Z4 i Z5 = Z6, a w komórce Z10 pojawi się dokładnie 1. Przybliżenia o podanej precyzji uzyskiwane są przy tzw. ogólnym formacie komórek.

### Kontynuacja przykładu 12. Standaryzacja rozkładu

#### Estymacja wartości modalnej (dominanty)

Średnia estymowana  $[QE]$  jest poniżej średniej  $[Q]$  z próby, należy więc dominantę przesunąć w górę.

38° W komórce Z13 wpiszmy formułę oszacowania przesunięcia (Q-QE) = Z5-Z6

Z	AA
---	----

39° W komórce AA13 wpiszmy symbol przesunięcia  $[\pm m]$ . Wynik: 

13	<b>0,016942</b>	$[\pm m]$
----	-----------------	-----------

Jest to szacunkowa wielkość, o jaką należy zmienić dominantę, zatem  $Q_{\text{mod}} \approx 1 + 0,016942$

40° W komórce Z11 zmieńmy wstępną wartość  $Q_{\text{mod}}$  z równej 1 na dokładniejszą 1,016942

41° Porównajmy nową wartość średniej estymowanej ze średnią Q z próby:

5	<b>0,886304</b>	[Q]
---	-----------------	-----

Wartości te są już bardzo bliskie, lecz dominantą wymaga jeszcze korekty.

6	<b>0,886212</b>	[QE]
---	-----------------	------

42° W komórce Z11 zmieniajmy małymi porcjami wartość  $Q_{\text{mod}}$  tak długo, aż średnie będą równe.

Poszukiwana wartość modalna  $Q_{\text{mod}}$  wynosi  $\approx 1,017033172$  co daje średnią: 

6	<b>0,886304</b>	[QE]
---	-----------------	------

Dominanta jest powyżej poziomu =1, czyli kolumnę Q prawidłowo zakwalifikowano jako  $x \leq Q_{\text{mod}}$



## Estymacja parametru standaryzującego (ksi)

Dystrybuanta estymowana [Fx] przekracza wartość graniczną 1, należy więc zwiększyć liczbę [ksi].

43° W komórce Z14 wpiszmy formułę oszacowania parametru  $\xi(4) = 4 * Z10$ 

Z	AA
---	----

44° W komórce AA14 wpiszmy symbol parametru [ksi4]. Wynik: 

14	<b>4,022184</b>	[ksi4]
----	-----------------	--------

Jest to przybliżona, poszukiwana wartość parametru standaryzującego estymowany tutaj rozkład.

45° W komórce Z12 zmieńmy wstępną wartość [ksi] z równej 4 na dokładniejszą 4,022183726

46° Porównajmy liczebność estymowaną z liczebnością rzeczywistą próby [n]: 

3	<b>1150</b>	[n]
---	-------------	-----

Wyświetlane wartości w komórkach Z3 i Z4 są zgodne z próbą badawczą.

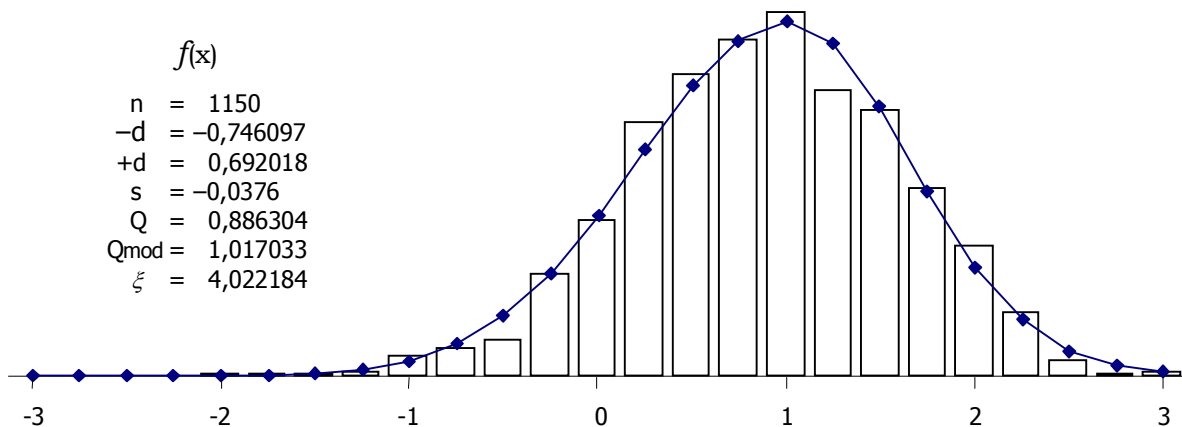
Dystrybuanta obliczana w komórce Z10 przyjęła standardową wartość =1.

4	<b>1150</b>	[nE]
10	<b>1</b>	[Fx]

47° Porównajmy rozkład empiryczny (komórki A3÷Y3) z rozkładem estymowanym (komórki A4÷Y4):

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
0	0	0	0	1	1	1	2	9	13	16	46	70	114	135	151	163	128	119	84	58	29	7	1	2
0	0	0	0	0	0	1	3	7	14	27	46	71	101	129	150	159	149	120	83	49	25	11	4	1

48° Wykonajmy wykres estymowanej funkcji częstości nałożony na histogram liczebności empirycznej.



Ryc. 33. Histogram rozkładu empirycznego i krzywa rozkładu estymowanego – oszacowanie czynnika *Rozwaga* dla populacji w wieku 16÷20 lat

Z porównania obu wykresów wynika, że estymowana za pomocą wzoru krzywa częstości realnej w miarę wiernie odwzorowuje rozkład empiryczny. Standardowy błąd średniokwadratowy różnic wynosi zaledwie 0,5%. Lokalne niedopasowania próbek dyskretnej do wygładzonej krzywej rozkładu oczekiwanego są zjawiskiem typowym, nawet przy bardzo licznych próbach.

<i>Ślady</i>		<i>Wzorce</i>
koniunkcja oznak	.....	spójność kontrolnych par wskaźników
średnie przeskalowane	.....	wartość i ważkość na skali ocen
średnie oczekiwane, standardy	.....	baza i norma jako estymatory
różnice i tolerancja średnich	.....	przedziały ufności realnej
estymacja pozycji	.....	oczekiwane poziomy HML
asymetria rozkładu z próby	.....	skośność lewo- i prawostronna
estymacja rozkładu	.....	funkcja częstości realnej

### Miary dystansowe z pomiaru dwukrotnego

Ruchomość parametrów powoduje, że najbardziej wartościowe są miary różnicowo-dystansowe. Wymusza to konieczność dwukrotnego pomiaru identycznym narzędziem. Przyjęliśmy wcześniej konwencję oznaczania parametrów obu pomiarów indeksami *prim* i *bis* [ $M'$ ;  $M''$ ], natomiast ich różnice znakiem plus/minus [ $\pm M$ ]. Oprócz miar różnicowych, znak  $\pm$  zintegrowany z symbolem posiadają także te miary indeksowe, które mogą przyjmować obie wartości walencyjne: dodatnie lub ujemne. Zdefiniujemy kategorie miar dystansowych:

➤ **Miary różnicowe tendencji** – to parametry zmiennej losowej wyznaczające kierunki i nasilenie zachodzących przemian, powstałe z porównania stanów końcowego i początkowego, jako różnica statystyk z obu pomiarów [ $p'' - p'$ ]. Parametrem trójwektorowym tendencji jest *Zmiana* [ $\vec{A}$ ] reprezentowana przez trzy wektory składowe  $\{\pm H, \pm M, \pm L\}$ , a parametrem skumulowanym jest *Trend* [ $\pm t$ ] standaryzowany względem kresu zmian potencjalnie możliwych.

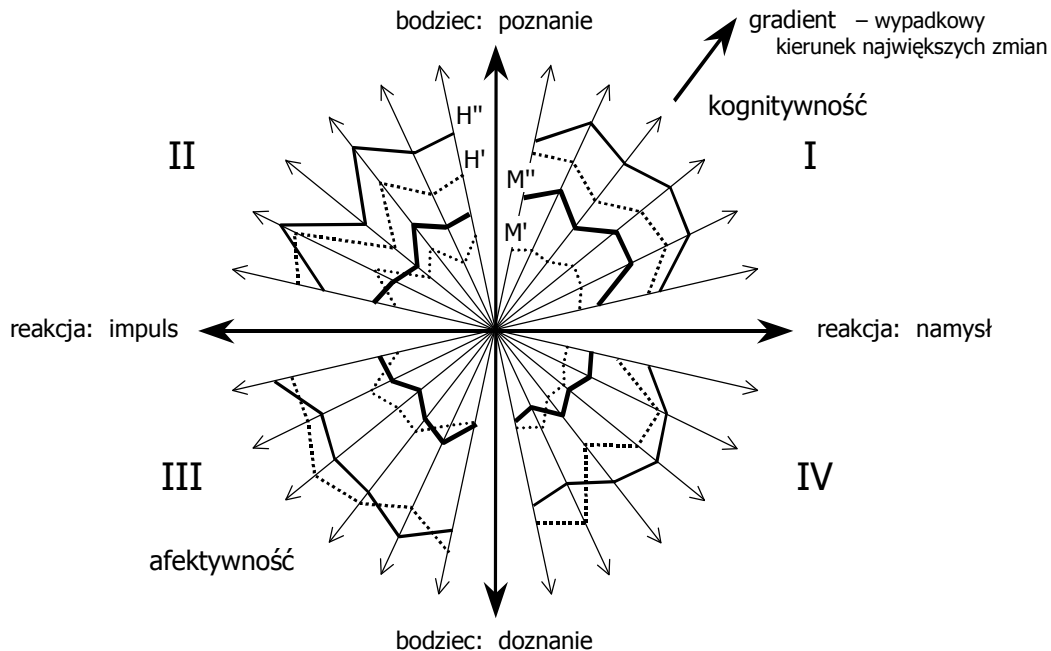
➤ **Miary różnicowe konwergencji** – to parametry zmiennej losowej dotyczące tych oznak, których charakterystyka świadcząca o pozytywnych przemianach powinna zmierzać do niwelowania różnic [ $p'' - p'$ ]  $\rightarrow 0$ . Do tej kategorii miar zalicza się parametry *Zawiłości* [ $\pm i$ ], *Bilansu* [ $\pm b$ ] i *Zbieżności* [ $\pm c$ ].

➤ **Miary indeksowe korelacji** – to parametry zmiennej losowej określające siłę i znak zależności między dwoma paralelnymi zbiorami danych, np.  $X'' \leftrightarrow X'$ , a przez to pośrednio wyznaczające relacje pomiędzy cechami wskaźnikowanymi przez te dane. Z tej kategorii wykorzystamy miarę autokorelacji [ $\pm r$ ], definicyjnie zgodną z popularnym współczynnikiem korelacji *r*-Pearsona. Indeksowość miary oznacza tu liczenie względne celem uzyskania skalarnego parametru z zakresu  $\langle -1; 1 \rangle$ , niezależnego od poziomów wartości składowych.

➤ **Miary indeksowe dynamiki** – to parametry zmiennej losowej wyznaczające zwroty i amplitudy fluktuacji przemian, powstałe z porównania wahań indywidualnych wypowiedzi każdego z respondentów [ $X_i'' - X_i'$ ] względem zmian potencjalnie możliwych jako maksymalne. Parametrem *n*-wielowektorowym dynamiki jest *Fluktuacja* [ $\vec{F}$ ], wektorami składowymi *Przyrost* [ $+e$ ] i *Spadek* [ $-e$ ], a parametrami skumulowanymi *Zasięg* [ $e$ ] i *Efekt* [ $\pm e$ ]. Także i tutaj indeksacja ma na celu uniezależnienie od poziomu absolutnego, oderwanie parametrów od wymiarowania pozycyjnego i od skalowania jednostkami umownymi.

Uogólniając – istotą tworzenia miar różnicowych jest wykorzystanie różnic pomiędzy statystykami pochodzącymi z obu pomiarów, przy czym w wypadku określania tendencji pozytywną oznaką jest przyrost wartości, a przy konwergencji zerowanie. W tworzeniu miar indeksowych do ilorazowej budowy złożonych parametrów standaryzowanych wykorzystuje się zarówno bazowe, jak też wcześniej już zagregowane statystyki, przy czym pozytywnymi oznakami empirycznymi są względnie wysokie wartości dodatnie tych parametrów.

Proces przemian edukacyjnych jest mocno uwikłanym splotem oddziaływań i efektów. Jakość wpływu i rezultatów poszczególnych grup poddanych ewaluacji trudno jest wyrazić pojedynczym parametrem statystycznym, stąd warto próbować różnych metod analizy. Większość miar można przedstawić na wykresach, toteż jedną z alternatyw analizy parametrycznej jest analiza graficzna. Opiszę jej przebieg. Na współrzędnych biegunowych należy nanieść kierunki o równomiernie rozłożonych kątach obrotu (dla skali o 24 pozycjach kąt =  $360^\circ / (24+4) \approx 12^\circ 51'$ ), na nich umieszcza się *Poziomy* M', M'', H' i H'' wskaźników odpowiednio w takich ćwiartkach układu, które wymiarowane są adekwatnie do próbkowanych cech mentalnych. Dwanaście wskaźników *Poznania* znajduje się w ćwiartkach I i II, a dwanaście wskaźników *Doznania* w ćwiartkach III i IV. Ponadto wskaźniki wypowiedzi, które wymagają dłuższego namysłu zgrupowane są w ćwiartkach I i IV, a te, na które reagujemy gotowym skryptem wypowiedzi – w ćwiartkach II i III. Punkty wyznaczone przez te same *Poziomy* łączymy liniami łamanymi (ciągłymi i kropkowanymi). Dla większej przejrzystości wykresu nie nanosi się poziomów dolnych L' i L''.



Ryc. 34. Wykres biegunowy – obraz zmian wskaźników

Interpretacja obejmuje grupy wskaźników: w grupie III oznak afektywnych widoczne są zarówno zmiany przyrostowe, jak i spadkowe; w ćwiartce I wszystkie wskaźniki oznak kognitywnych mają tendencje wzrostu; w dwóch pozostałych grupach II i IV więcej jest zmian pozytywnych niż negatywnych; w ćwiartce II jest największe zróżnicowanie poziomów. Stan na wejściu był najniższy w grupie oznak poznawczo-refleksyjnych, a najwyższy w afektywnych. Proces edukacyjny powinien te zróżnicowania niwelować. Wypadkowy gradient największych zmian można oszacować jako skierowany na sferę poznawczą.

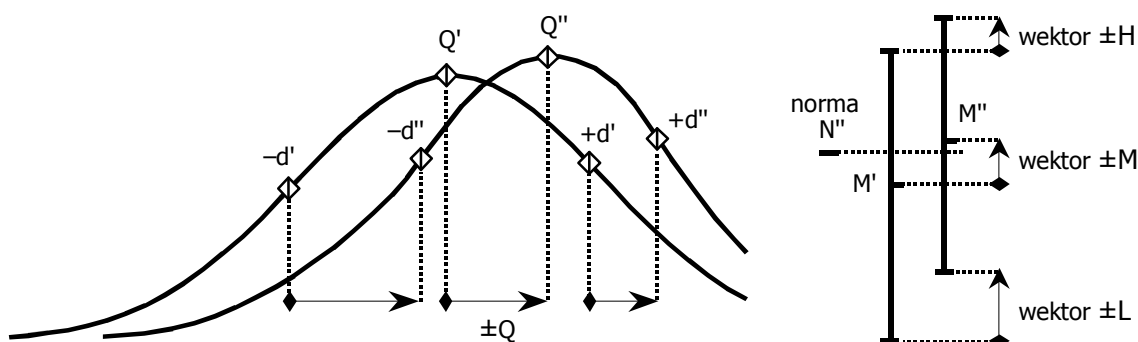
### Zmiana – przesunięcie rozkładu

Spróbujmy matematycznie zdefiniować parametr zmian bardziej zwarty niż jej graficzna interpretacja. *Zmiana* 'Alteration' [ $\bar{A}$ ] w ujęciu abstrakcyjnym jest zbiorem estymatorów pozycyjnych, reprezentujących tendencje do dys-tansowych przemian zachodzących na trzech kluczowych poziomach wartości:

$$\bar{A} = \{\pm H, \pm M, \pm L\} \quad [14]$$

przy czym  $\pm H = H'' - H'$        $\pm M = M'' - M'$        $\pm L = L'' - L'$

W ujęciu praktycznym *Zmiana* jest zbiorem różnic pomiędzy poziomami HML w pomiarze końcowym wobec pomiaru początkowego, obligatoryjnie odnoszonych podczas wnioskowania do któregoś z normujących momentów zwykłych, np. do średniej wartości rezultatów końcowych w grupie kontrolnej [ $M_k''$ ], do *Bazy* [ $B'$ ] lub *Normy* [ $N''$ ], a nade wszystko wzajemnie do siebie  $\pm H \leftrightarrow \pm M \leftrightarrow \pm L$ .



Ryc. 35. Przesunięcia momentów i wektory zmian poziomów HML

Przesunięcia nie zawsze są dodatnie i współbieżne. Zmiany na poziomie niskim L zazwyczaj są największe, gdyż w dolnym obszarze jest duża przestrzeń do przemieszczeń, natomiast ewentualne przyrosty na poziomie wysokim H są najmniejsze, ponieważ tam blisko już jest kres górny skali ocen ( $H_{\max} = 5,5$ ).

#### Przykład 13. Zmiana

30 studentów uzyskało następujące wagi ( $x'$ ) w pierwszym pomiarze oraz ( $x''$ ) w drugim:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć wektory zmian według wzoru [14] i wcześniejszych przykładów.

Zbiór  $X'$  jest identyczny jak w przykładach 5 i 10 – bierzemy stamtąd  $M' = 3,5$   $L' = 1,46$   $H' = 4,79$

Liczmy HML'': 1° Suma  $\Sigma x'' = 60$  2° Średnia jakość  $Q'' = 60/30 = 2$  3° Różnice ( $x'' - Q''$ ):

0	-2	1	1	0	0	1	1	0	0	0	1	0	1	0	-2	-1	0	1	0	-1	0	-1	0	0	-1	0	1	1	-1
---	----	---	---	---	---	---	---	---	---	---	---	---	---	---	----	----	---	---	---	----	---	----	---	---	----	---	---	---	----

4° Różnic ujemnych  $l = 7$ , dodatnich  $h = 9$  5° Przypadki ujemne  $Y''_i$  podnosimy do kwadratu:

	4														4	1														1
--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	---

6° Suma  $\Sigma Y''_i^2 = 13$  7° Odchylenie  $[-d''] = -\sqrt{\{13 \cdot (1+9/7)/30\}} \approx -\sqrt{0,99} \approx -1,0$



**Przykład 14. Zawilość**

Przyjmijmy próbę badawczą i zbiory danych ( $x'$ ) oraz ( $x''$ ) identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć zawilość według wzoru [15], na podstawie zmian HML obliczonych w przykładzie 13.

$$M' = 3,5 \quad M'' = 4,5 \quad \pm M = 1,0 \quad H' = 4,79 \quad H'' = 5,23 \quad \pm H = 0,44 \quad L' = 1,46 \quad L'' = 3,5 \quad \pm L = 2,04$$

$$1^\circ \text{ Min}(M'|M'') = 3,5 \quad \text{Min}(H'|H'') = 4,79 \quad \text{Min}(L'|L'') = 1,46 \quad V_{\max} = 5,5$$

$$2^\circ \text{ Zawilość } [\pm i] = (0,44 + 1,0)/(2 \cdot 5,5 - 4,79 - 3,5) - (2,04 + 1,0)/(2 \cdot 5,5 - 1,46 - 3,5) \approx 0,03$$

Wyliczony w przykładzie parametr *Zawilości* jest bliski zeru, co oznacza w zasadzie brak zawilości procesów oddziaływania. Wynika to z dodatnich, harmonijnych zmian, odpowiednio większych na poziomie L i mniejszych na poziomie H. Jednakże w praktyce ewaluacji procesów edukacyjnych właśnie tak niewielkie wartości dodatnie rzędu  $0,03 \div 0,06$  są najbardziej oczekiwane, oznaczają bowiem, że oddziaływanie miało optymalnie intensywne natężenie, a poziom trudności był dopasowany i przy tym wcale nie tak łatwy. Wartość  $[\pm i] > +0,1$  oznaczałaby zbyt dużą trudność, a wartość ujemna – zbyt dużą łatwość.

**Bilans – równoważenie rozziwu**

Do wykrywania subtelnych przesunięć z jednej do drugiej strefy skali dwuważonej, a więc do badania zmian wewnętrznej zgodności w grupach, wykorzystajmy charakterystyczną dla skali metodę bilansowania parametrów powstałych z zestawienia momentów zwykłych i absolutnych. *Bilans* 'Balance'  $[\pm b]$  wyznacza różnicę pomiędzy modułami *Rozziwu* w pierwszym i drugim pomiarze, standaryzowanymi metodą podzielenia przez *Rozrzuty* wypowiedzi:

$$\pm b = \frac{|-b'|}{d'} - \frac{|-b''|}{d''} \quad \langle -1; +1 \rangle \rightarrow 0 \quad [16]$$

gdzie  $|-b'|$ ,  $|-b''|$  – moduły rozziwów w pomiarach 'przed' i 'po', ze wzoru [3]  
 $d'$ ,  $d''$  – rozrzuty w pomiarach początkowym i końcowym, ze wzoru [6]

Parametr ten interpretuje się podobnie jak bilans w księgowości. Jego zerowa wartość oznacza *zrównoważenie zdań* i nie ma co oczekiwać, by był dodatni. „Superata” jest możliwa jedynie wówczas, gdy początkowo był rozziw, który udało się zniwelować. Natomiast wartość ujemna świadczyłaby o niekorzystnym zjawisku rozwarstwienia grupy, w której część respondentów przesunęła się do strefy niezgodności z opinią kwalifikowanej większości. W odróżnieniu od parametrów uśrednionych, *Bilans* jest miarą pojedynczych przypadków, a ściślej – wyłącznie przypadków *zdań odrębnych*. Nie jest więc przydatny do uogólnień: ani dotyczących populacji, ani na poziomie indykatorum. Jest natomiast przydatny do oceny jakości lokalnych procesów wewnątrzgrupowych.

**Przykład 15. Bilans**

Przyjmijmy próbę badawczą  $n=30$  i zbiory danych ( $x'$ ) oraz ( $x''$ ) identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć bilans według wzoru [16], wykorzystując wcześniej już obliczone miary dyspersji.

$d' \approx 3,33$     $d'' \approx 1,73$    ponadto Rozziew  $|-b''| = 0$  ponieważ brak wag ujemnych w zbiorze ( $x''$ )

1° Sześć przypadków ( $x'$ )  $< 0$  ich suma = -10   2° Rozziew  $[-b'] = 2 \cdot \sum x_i / n = 2 \cdot (-10) / 30 \approx -0,67$

3° Bilans  $[\pm i] \approx |-0,67| / 3,33 - |0| / 1,73 \approx +0,2$

Parametr bilansu ma taką właściwość, że najbardziej wyrazisty jest na poziomie pojedynczych wskaźników, natomiast na poziomie indykatum dąży do zera, analogicznie jak tworzące go składowe rozziewu. Jest więc to miara w szczególnym sensie konwergencyjna (o statystyce zbieżnościowej). Chociaż oczekuje się zerowania, to pozytywne są każde dodatnie wartości *Bilansu*.

**Zbieżność – niwelacja niespójności**

Konwergencyjną (w pełnym tego słowa znaczeniu) miarą dystansową jest parametr *Zbieżności* 'Convergence  $[\pm c]$ ' utworzony z różnicy między wymiarami *Spójności* obu wypowiedzi odnoszących się do kontrolnych par wskaźników:

$$\pm c = c'' - c' \quad \langle -1; +1 \rangle \rightarrow 0 \quad [17]$$

gdzie  $c''$ ,  $c'$  – spójności odpowiednio w pomiarze końcowym i początkowym

O ile odrębnie każda z bazowych miar spójności wcale nie jest i nie musi być zerowa, o tyle ich różnica powinna dążyć do zera. Niespójność bowiem zależy zarówno od tego, jak bardzo czasoprzestrzennie zróżnicowane są pary kontrolne bodźców-stwierdzeń, jak też od tego, czy (abstrahując od treści) sami respondenci wypowiadali się starannie – wewnątrznie spójnie, czy niespójnie. W ujęciu różnicowym zredukowane są niespójności immanentne kontrolnych stwierdzeń skali (objawiające się tym, że poziom wypowiedzi wobec każdego twierdzenia dąży do pewnej oczekiwanej wartości, różnej nawet dla par kontrolnych). Pozostają więc mierzone jedynie różnice niespójności wypowiedzi. Zadaniem ewaluatora jest interpretacja, czy różna od zera wartość parametru *Zbieżności* jest w konkretnym przypadku wyznacznikiem pozytywnego niwelowania, czy może przejawem generowania wtórnie bezzasadnej niespójności. Te pierwsze przypadki zazwyczaj mają wartość dodatnią, a te drugie ujemną, chociaż interpretacja nie jest tak prosta. Wyobraźmy sobie bowiem sytuację hipotetyczną, w której respondenci w pierwszym pomiarze nadmiernie często wybieraliby opcję '*brak zdania*', a w drugim różnicowali już wypowiedzi zgodnie z treścią stwierdzeń. Zbieżność wówczas byłaby ujemna, nawet gdyby drugie wypowiedzi były w pełni trafne. W praktyce takie sytuacje są jednak rzadkie.

**Korelat – współzależność zmian**

Zależność pomiędzy wynikami obydwu pomiarów określa współczynnik autokorelacji. Przyjmując krótszą nazwę *Korelat* 'Correlate' [ $\pm r$ ], obliczamy go jako iloraz kowariancji wypowiedzi tych samych respondentów w stosunku do średniej geometrycznej z wariancji pomiarów początkowego i końcowego:

$$\pm r = \frac{\sum_{i=1}^n \{(X'_i - Q')(X''_i - Q'')\}}{\sqrt{\sum_{i=1}^n (X'_i - Q')^2 \cdot \sum_{i=1}^n (X''_i - Q'')^2}} \quad \langle -1; +1 \rangle \quad [18]$$

gdzie  $X'_i, X''_i$  – wagi wypowiedzi  $i$ -tych respondentów w obu pomiarach  
 $Q', Q''$  – średnie jakości wypowiedzi w pierwszym i drugim pomiarze  
 $n$  – liczebność próby (klasy, grupy lub ogółu badanej zbiorowości)

Jest to w pełni odwzorowana definicja współczynnika korelacji  $r$ -Pearsona, przy czym w tym konkretnym zastosowaniu interpretacja jest bardziej jednoznaczna. Dodatnie wartości wskazują, w jakim stopniu ci sami respondenci, którzy na wejściu osiągnęli wynik powyżej/poniżej średniej, osiągnęli odpowiednio lepszy/gorszy wynik na wyjściu. Ujemne wartości oznaczają, że role odmieńczyły się, tj. „lepsi i gorsi” zamienili się miejscami w rankingu wyników końcowych. Wartości bliskie zeru świadczą o braku zależności pomiędzy wypowiedziami w obu pomiarach. Wymiar liczbowy i interpretacja są identyczne także po konwersji na skalę ocen. Współczynnik ten zbudowano dla rozkładu symetrycznego, stąd może nie być wiarygodny przy dużej asymetrii.

**Przykład 16. Korelat**

Przyjmijmy próbę badawczą i zbiory danych ( $x'$ ) oraz ( $x''$ ) identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć współczynnik autokorelacji według wzoru [18], wykorzystując średnie już obliczone.

$Q' = 1$     $Q'' = 2$    1° Tworzymy dwa zbiory różnic  $\{Y'_i = X'_i - Q'\}$  oraz  $\{Y''_i = X''_i - Q''\}$ :

0	-3	1	2	1	-1	1	1	-2	0	1	0	-1	2	1	-4	-2	-1	1	2	-3	1	-2	0	2	-1	0	1	2	1
0	-2	1	1	0	0	1	1	0	0	0	1	0	1	0	-2	-1	0	1	0	-1	0	-1	0	0	-1	0	1	1	-1

2° Obliczamy składniki kowariancji, mnożąc elementy zbioru ( $Y'_i$ ) przez elementy zbioru ( $Y''_i$ ):

0	6	1	2	0	0	1	1	0	0	0	0	0	2	0	8	2	0	1	0	3	0	2	0	0	1	0	1	2	-1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----

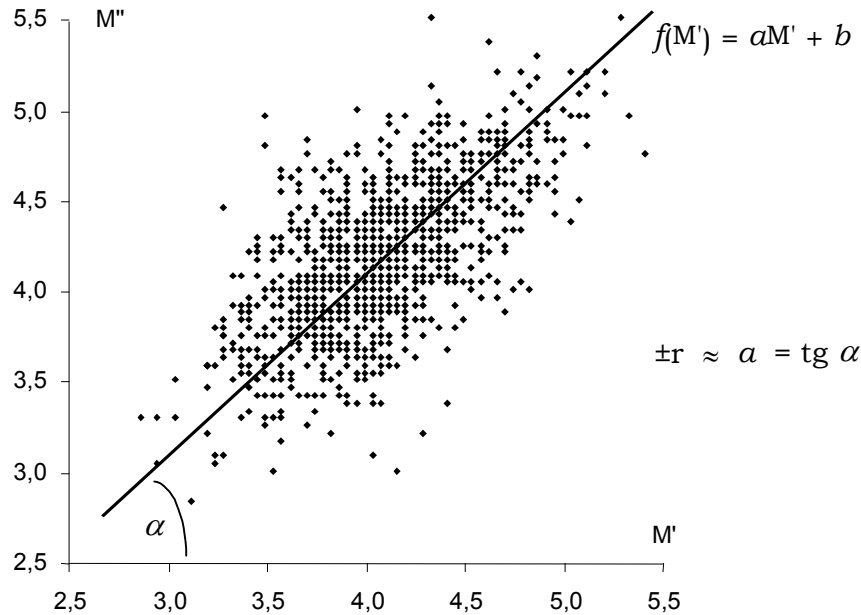
3° Wartość licznika: suma składników = 32   4° Obliczamy kwadraty różnic ( $Y'_i$ ) oraz ( $Y''_i$ ):

0	9	1	4	1	1	1	1	4	0	1	0	1	4	1	16	4	1	1	4	9	1	4	0	4	1	0	1	4	1
0	4	1	1	0	0	1	1	0	0	0	1	0	1	0	4	1	0	1	0	1	0	1	0	0	1	0	1	1	1

5° Suma kwadratów zbioru ( $Y'_i$ ) = 80   6° Suma kwadratów zbioru ( $Y''_i$ ) = 22

7° Wartość mianownika =  $\sqrt{(80 \cdot 22)} \approx 41,95$    6° Korelat [ $\pm r$ ]  $\approx 32/41,95 \approx 0,76$





Ryc. 37. Silna korelacja między wynikami 'przed' i 'po'

Parametr autokorelacji w ograniczonym zakresie może być interpretowany geometrycznie jako współczynnik kierunkowy  $[a]$  funkcji  $f(M') = aM' + b$  określający nachylenie prostej najlepiej dopasowanej do skupiska punktów o współrzędnych  $(M', M'')$ . Punkty te ilustrują zależność średnich 'przed' i 'po'.

### **Trend – intensywność zmian**

Najbardziej skumulowanym, niejako wynikowym parametrem w analizie zmienności jest *Trend*  $Trend [\pm t]$  wyznaczający zwrot i wielkość tendencji do zmian na podstawie sumy transpozycji wszystkich trzech poziomów  $\pm H$ ;  $\pm M$ ;  $\pm L$  standaryzowanych poprzez odniesienie do możliwych zmian maksymalnych:

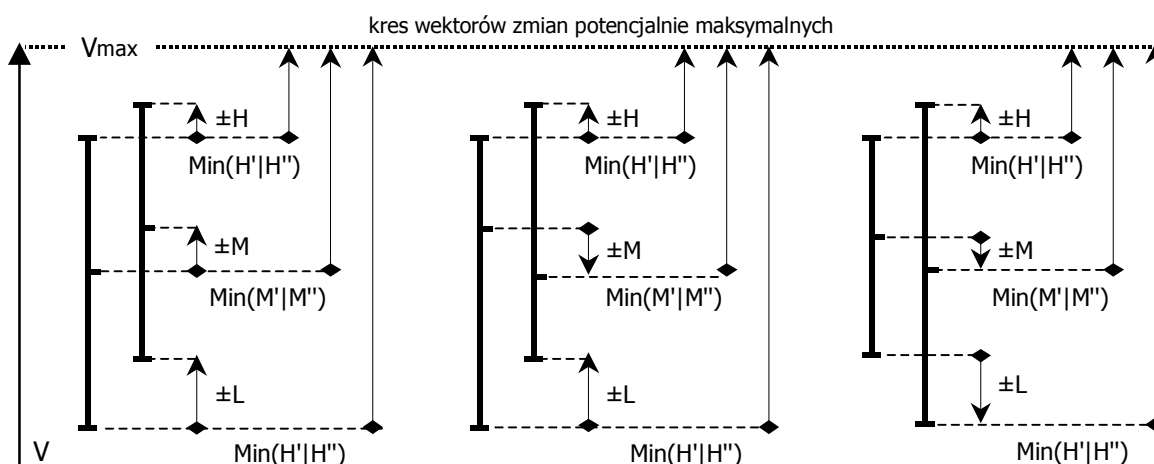
$$\pm t = \frac{[\pm H] + [\pm M] + [\pm L]}{3V_{\max} - \text{Min}(H'|H'') - \text{Min}(M'|M'') - \text{Min}(L'|L'')} \quad \langle -1; +1 \rangle \quad [19]$$

gdzie  $\pm H = H'' - H'$        $\pm M = M'' - M'$        $\pm L = L'' - L'$

$V_{\max}$  – najwyższa wartość górnego kresu poziomu ocen, tutaj  $V_{\max} = 5,5$

Funkcja  $\text{Min}(H'|H'')$  oznacza wybór argumentu o mniejszej wartości

Standaryzacja polega na podzieleniu wartości liczbowej, określającej łączny wymiar zmian, przez wartość liczbową informującą o tym, jak dużą przestrzeń swobodną do osiągnięcia najwyższego poziomu mieli jeszcze respondenci. Tę przestrzeń wyznacza się z różnic pomiędzy górnym kresem skali a poziomami tych parametrów H, M i L, które z pomiarów 'przed' i 'po' mają niższą wartość. W interpretacji graficznej wyznaczanie miary trendu na podstawie przesunięć poziomów HML sprowadza się do składania trzech wektorów tendencji rzeczywistej względem trzech wektorów tendencji potencjalnie możliwej (ryc. 38).



Ryc. 38. Wektory składowe trendu dla trzech różnych sytuacji empirycznych – zmiany rzeczywiste  $\{\pm H, \pm M, \pm L\}$  wobec potencjalnie możliwych

Dopuszczalność sumowania wektorów składowych z różnic poziomów  $\pm HML$  wynika stąd, że dystrybuanty rozkładu w zakresie przedziałów wyznaczonych odchyleniami realnymi  $\langle -d, +d \rangle$  mają charakterystyki zbliżone do liniowych.

#### Przykład 17. Trend

Przyjmijmy próbę badawczą i zbiory danych  $(x')$  oraz  $(x'')$  identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć trend według wzoru [19], na podstawie zmian HML obliczonych w przykładzie 13.

$$M' = 3,5 \quad M'' = 4,5 \quad \pm M = 1,0 \quad H' = 4,79 \quad H'' = 5,23 \quad \pm H = 0,44 \quad L' = 1,46 \quad L'' = 3,5 \quad \pm L = 2,04$$

$$1^\circ \text{ Min}(M'|M'') = 3,5 \quad \text{Min}(H'|H'') = 4,79 \quad \text{Min}(L'|L'') = 1,46 \quad V_{\max} = 5,5$$

$$2^\circ \text{ Trend} [\pm t] = (0,44 + 1,0 + 2,04) / (3 \cdot 5,5 - 4,79 - 3,5 - 1,46) \approx +0,516$$

W praktyce tak bardzo wysoka wartość trendu jak w powyższym przykładzie może być osiągana tylko wyjątkowo na poziomie pojedynczych wskaźników. W miarę przechodzenia na wyższe poziomy agregacji zmiennych, na skutek uśredniania wskaźników następuje urealnianie wartości parametru. Trend jest wyznacznikiem jakości procesów, analizowanym zawsze wspólnie z wektorami zmian i z miarą zawłości. Stanowi też miarę bazową dla wyznaczania trafności.

Ślady	Wzorce
kierunki zmian wskaźników	..... osie, bieguny i gradient przyrostu
przesunięcie rozkładu	..... zmiana jako baza analizy różnicowej
nierównomierność zmian	..... zawłość kontekstu oddziaływań
wartościowość zmian	..... bilans osiągnięcia konsensusu
konwergencja zmian	..... zbieżność, niwelacja niespójności
współzależność zmian	..... korelat, współczynnik autokorelacji
intensywność zmian	..... trend, wynikowy parametr tendencji

**Fluktuacja – wahania wypowiedzi**

Pojęciem kluczowym oderwania się od wszelkich ruchomych poziomów jest fluktuacja. Nie da się jej wyrazić jednym parametrem, lecz można ją zdefiniować zasobem indeksów. W ujęciu abstrakcyjnym *Fluktuacja* 'Fluctuation' [ $\bar{F}$ ] jest zbiorem stosunków wszystkich różnic wyznaczających indywidualne zmiany wypowiedzi w czasie względem możliwych maksymalnych zmian:

$$\bar{F} = \left\{ \frac{X_1'' - X_1'}{X_{\max} - \text{Min}(X_1'' | X_1')}; \frac{X_2'' - X_2'}{X_{\max} - \text{Min}(X_2'' | X_2')}; \dots \frac{X_n'' - X_n'}{X_{\max} - \text{Min}(X_n'' | X_n')} \right\} \quad [20]$$

gdzie

$X_i''$  – waga wypowiedzi (lub średnia) i-tego respondenta w drugim pomiarze  
 $X_i'$  – waga wypowiedzi (lub średnia) i-tego respondenta w pierwszym pomiarze  
 $X_{\max}$  – najwyższa waga przypisana do kresu górnego skali, tutaj  $X_{\max} = +3$   
 Funkcja  $\text{Min}(X' | X'')$  oznacza wybór argumentu o mniejszej wartości;  
 pomija się elementy nieoznaczone przypadków granicznych  $X'' = X' = X_{\max} = 3$

W praktyce oznacza to konieczność obliczania osobno (dla każdego z respondentów) różnic pomiędzy wypowiedziami w pomiarze końcowym i pomiarze początkowym i odnoszenia ich do potencjalnych wektorów fluktuacji.

Porównajmy to ujęcie z definicją *Zmiany*. Otóż przy wyznaczaniu zmiany różnice pomiędzy dwoma pomiarami obliczane są z wcześniej uśrednionych poziomów grupowych, natomiast *Fluktuacja* jest zasobem różnic indywidualnych, z których to pierwotnych elementów dopiero w ujęciu względnym oblicza się parametry standardowe wzrostu i spadku. Istotna odmienność obu strategii polega na tym, że przy analizie zmienności najpierw następuje kumulowanie wszystkich przypadków i dopiero potem różnicowanie, a przy analizie fluktuacji najpierw określa się indeksy zróżnicowania przypadków i dopiero te elementarne zróżnicowania kumuluje, czyli sumuje i uśrednia.

**Przykład 18. Fluktuacja**

Przyjmijmy próbę badawczą i zbiory danych ( $x'$ ) oraz ( $x''$ ) identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć indeksy fluktuacji według wzoru [20], pomijając nieoznaczone przypadki graniczne [ $\sim$ ].

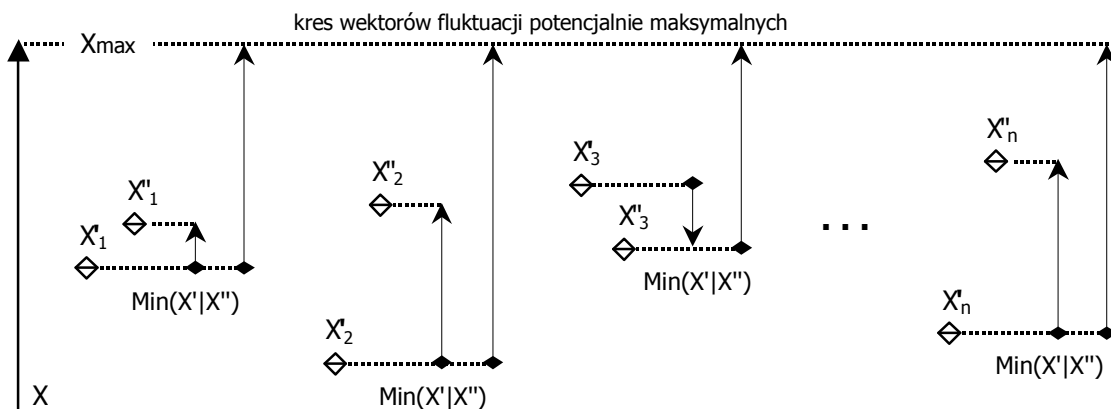
1° Tworzymy dwa zbiory różnic  $\{Y_i = X_i'' - X_i'\}$  oraz  $\{\text{Max}Y_i = X_{\max} - \text{Min}(X_i'' | X_i')\}$  dla  $X_{\max} = 3$

1	2	1	$\sim$	0	2	1	1	3	1	0	2	2	$\sim$	0	3	2	2	1	-1	3	0	2	1	-1	1	1	1	$\sim$	-1
2	5	1	$\sim$	1	3	1	1	4	2	1	2	3	$\sim$	1	6	4	3	1	1	5	1	4	2	1	3	2	1	$\sim$	2

2° Obliczamy indeksowe elementy zbioru opisującego fluktuację, czyli ilorazy  $\{Y_i / \text{Max}Y_i\}$ :

$\frac{1}{2}$	$\frac{2}{5}$	1	$\sim$	0	$\frac{2}{3}$	1	1	$\frac{3}{4}$	$\frac{1}{2}$	0	1	$\frac{2}{3}$	$\sim$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	1	-1	$\frac{3}{5}$	0	$\frac{1}{2}$	$\frac{1}{2}$	-1	$\frac{1}{3}$	$\frac{1}{2}$	1	$\sim$	$-\frac{1}{2}$
---------------	---------------	---	--------	---	---------------	---	---	---------------	---------------	---	---	---------------	--------	---	---------------	---------------	---------------	---	----	---------------	---	---------------	---------------	----	---------------	---------------	---	--------	----------------

Dla większej dokładności obliczeń na tym etapie warto indeksy wyrażać w ułamkach zwykłych.



Ryc. 39. Wektory składowe fluktuacji wypowiedzi kolejnych  $n$ - respondentów

Łączny wektor fluktuacji nie ma zaczepienia do jakiegoś bezwzględnego poziomu ocen. Orientacyjnym punktem odniesienia mogłaby być hipotetyczna oś oscylacji, będąca poziomem osiągniętym prawdopodobnie w połowie okresu badań. Do precyzyjnych obliczeń efektów jakiegoś procesu lub eksperymentu niezbędne jest odniesienie fluktuacji do wektorów absolutnych. Najlepszą bazą okazuje się zbiór możliwych maksymalnych osiągnięć. W zaproponowanej tu metodzie *indeksacji* następuje przejście od różnic do indeksów niezależnych, wolnych od poziomów, mających cechy standardowych amplitud fluktuacji. Indeksacja polega na ilorazowym ujęciu proporcji między sumami wektorów.

### **Przyrost | Spadek – amplitudy fluktuacji**

Fluktuację w najbardziej zagregowanej formie reprezentują dwa wektory: *Przyrost* 'Increase [+e] i *Spadek* 'Decrease [-e], oznaczające przemiany dodatnie i ujemne. Liczy się je osobno dla zysków [i] i strat [j], z uwzględnieniem w obu liczebnościach ( $h$ ,  $l$ ) przypadków braku jakiegokolwiek zmiany, lecz bez przypadków granicznych  $X''=X'=X_{\max}=3$ . Przyrost obejmuje wszystkie rezultaty końcowe nie gorsze, a spadek – wszystkie nie lepsze od stanu początkowego, w proporcji do sumy potencjalnych wektorów fluktuacji maksymalnych:

$$+e = \frac{\sum_{i \in \{i: X''_i \geq X'_i\}} (X''_i - X'_i)}{\sum_{i \in \{i: X''_i \geq X'_i\}} (X_{\max} - X'_i)} \quad -e = \frac{\sum_{j \in \{j: X''_j \leq X'_j\}} (X''_j - X'_j)}{\sum_{j \in \{j: X''_j \leq X'_j\}} (X_{\max} - X''_j)} \quad [21]$$

gdzie

$X''_i$ ,  $X''_j$  – wagi lub średnie z wag wypowiedzi respondentów w drugim pomiarze

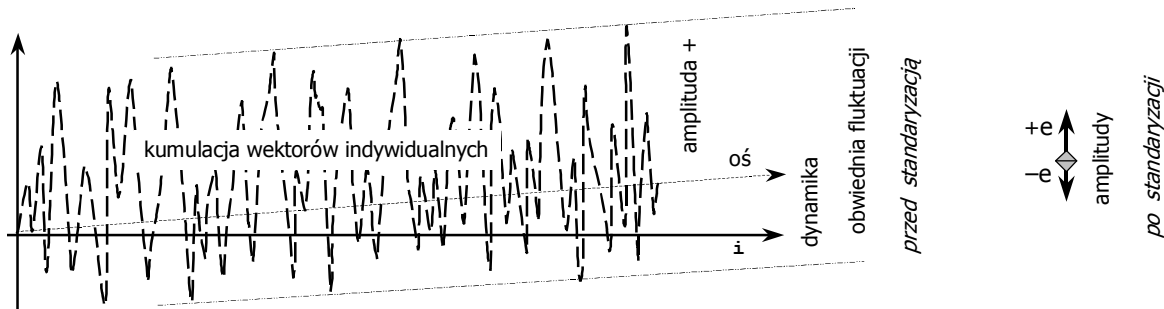
$X'_i$ ,  $X'_j$  – wagi lub średnie z wag wypowiedzi w pierwszym pomiarze

$X_{\max}$  – najwyższa waga przypisana do kresu górnego skali, tutaj  $X_{\max} = +3$

$h$ ,  $l$  – liczebności podzbiorów warunkowych  $\{i\}$   $\{j\}$  w próbie  $n$ -elementowej

Pomija się elementy nieoznaczone przypadków granicznych  $X''=X'=X_{\max}=3$

Oba parametry są standardowymi amplitudami zachodzących przemian, przy czym wzrost jest miarą oddziaływania czynników pozytywnych, zaś spadek miarą reakcji negatywnych. Sposób obliczania tych parametrów powoduje, że są one czułe zwłaszcza na tendencje niekorzystne w górnym poziomie skali. Poniższa rycina jest próbą zilustrowania tego, czym są parametry fluktuacji.



Ryc. 40. Usytuowanie parametrów fluktuacji

### Przykład 19. Przyrost, Spadek

Przyjmijmy próbę badawczą i zbiory danych ( $x'$ ) oraz ( $x''$ ) identyczne jak w przykładzie 13:

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć amplitudy fluktuacji według wzoru [21], pomijając nieoznaczone przypadki graniczne  $\sim$ .

1° Tworzymy zbiory różnic  $\{Y_i = X_i'' - X_i'\}$  oraz  $\{\text{Max}Y_i = X_{\text{max}} - X_i'\}$  dla przypadków  $X_i'' \geq X_i'$

1	2	1	$\sim$	0	2	1	1	3	1	0	2	2	$\sim$	0	3	2	2	1		3	0	2	1		1	1	1	$\sim$	
2	5	1	$\sim$	1	3	1	1	4	2	1	2	3	$\sim$	1	6	4	3	1		5	1	4	2		3	2	1	$\sim$	

2° Obliczamy Przyrost  $[+e] = \sum Y_i / \sum \text{Max}Y_i = 33 / 59 \approx 0,56$

3° Tworzymy zbiory różnic  $\{Y_j = X_j'' - X_j'\}$  oraz  $\{\text{Max}Y_j = X_{\text{max}} - X_j''\}$  dla przypadków  $X_j'' \leq X_j'$

			$\sim$	0					0				$\sim$	0					-1		0								$\sim$	-1
			$\sim$	1					1				$\sim$	1					1		1								$\sim$	2

4° Obliczamy Spadek  $[-e] = \sum Y_j / \sum \text{Max}Y_j = -3 / 8 = -0,375$

Sytuacja z przykładu ułatwia zrozumienie, dlaczego parametry obliczane dla zmiennych zagregowanych są dokładniejsze. Otóż, dla uproszczenia tabel, przyjęte tutaj dane surowe są w postaci liczb całkowitych, więc dotyczą raczej poziomu wskaźników. Z tego powodu wystąpiły trzy przypadki nieoznaczone, gdy wagi 'przed' i 'po' są maksymalne = +3. Dla tych przypadków nie było już wolnej przestrzeni do zmian, więc nie mogły one wpłynąć na dynamikę zmian, choć przecież takie maksymalne wskaźniki są rezultatami nader pozytywnymi. W rzeczywistości pomiarowej na poziomie składników przypadki nieoznaczone występują rzadko, a na poziomie czynników i wyników prawie w ogóle się nie zdarzają. W takiej sytuacji nawet minimalne zmiany na górnym poziomie (np. z 2,75 na 3,0) powodują już znaczny przyrost amplitudy. W praktyce przeciętne amplitudy dla populacji szacuje się na  $[+e] = +0,15$  i  $[-e] = -0,05$ .

**Zasięg – obwiednia fluktuacji**

Para wektorów górnej i dolnej amplitudy fluktuacji służy do budowy miar wynikowych. Suma modułów wzrostu i spadku wyznacza *Zasięg* 'Extent' [e]:

$$e = |+e| + |-e| \quad \langle 0; 2 \rangle \quad [22]$$

gdzie |+e| – wartość bezwzględna przyrostu, |-e| – wartość bezwzględna spadku, obliczone ze wzoru [21]

Parametr ten określa standardową obwiednię fluktuacji, czyli sumę amplitud z wahań wypowiedzi respondentów. Odzwierciedla to głębokość zachodzących przemian i jest w badanym procesie miarą *intensywności oddziaływań* przede wszystkim czynników zewnętrznych. W pomiarze dydaktycznym jakości zajęć wartość liczbowa *Zasięgu* nosi wyraźne piętno siły wpływu nauczyciela. Przy średnim [e] = 0,2 można przyjąć, że [e] > 0,2 jest oddziaływaniem ponadprzeciętnym, [e] > 0,25 silnym, [e] < 0,2 poniżej przeciętnym, a [e] < 0,15 słabym.

**Efekt – dynamika fluktuacji**

Do uogólniającego wnioskowania o rezultatach pomiarów dystansowych wykorzystuje się parametr efektywnościowy. *Efekt* 'Effect' [±e] obliczany jest z różnicy pomiędzy modułem przyrostu [+e] i modułem spadku [-e]:

$$\pm e = |+e| - |-e| \quad \langle -1; +1 \rangle \quad [23]$$

gdzie |+e| – wartość bezwzględna przyrostu, |-e| – wartość bezwzględna spadku, obliczone ze wzoru [21]

Ten najbardziej pożyteczny parametr wyznacza kierunek i skalarną wartość dynamiki przemian zachodzących w badanym procesie. Jest praktyczną miarą *jakości* splotu wszystkich oddziaływań intencjonalnych i okazjonalnych – tak pozytywnych, jak i negatywnych – ze wskazaniem tych, które dominowały. W pomiarze dydaktycznym wskaźnik efektu jest wypadkową akceptacji bądź odrzucania zarówno treści, jak i metod stosowanych przez nauczyciela, lecz silnie modyfikowaną przez interakcję zewnętrznych oddziaływań środowiska. W praktyce na poziomie wyników grup lub klas można przyjąć, że [±e] > 0,15 jest *Efektem* doskonałym, [±e] > 0,10 bardzo dobrym, a [±e] > 0,05 dobrym.

**Przykład 20. Zasięg, Efekt**

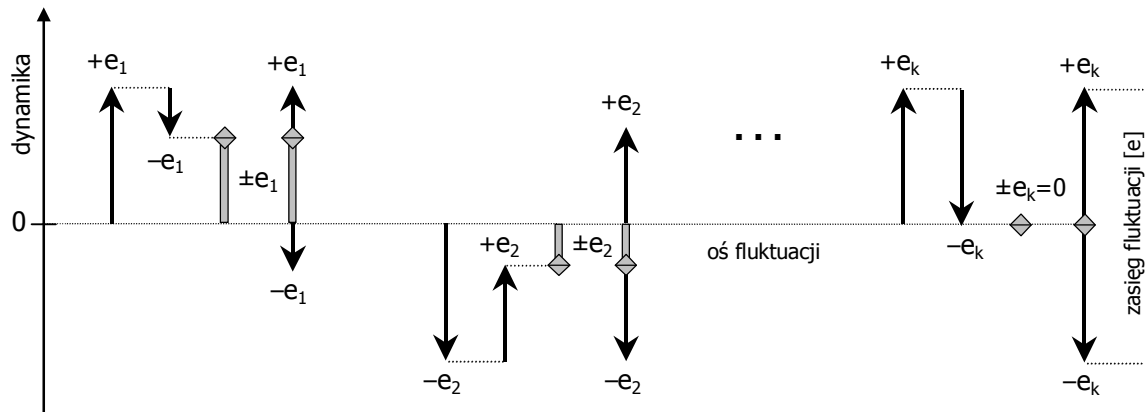
Przyjmijmy próbę badawczą i zbiory danych (x') oraz (x'') identyczne jak w przykładzie 19.

Wyznaczyć obwiednię i dynamikę fluktuacji ze wzorów [22] i [23].

Wykorzystamy statystyki obliczone w przykładzie 19: [+e] ≈ 0,56    [-e] = -0,375

1° Zasięg [e] = |+e| + |-e| ≈ 0,56 + 0,375 ≈ 0,935

2° Efekt [±e] = |+e| - |-e| ≈ 0,56 - 0,375 ≈ +0,185

Ryc. 41. Tworzenie wektorów wynikowych fluktuacji dla  $k$ - grup

Przy wyznaczaniu oraz interpretacji parametrów pochodnych zarówno *Fluktuacji* jak i *Zmiany* warto operować na wektorach, a nie na różnicach. Wprawdzie wyniki różnic dwóch liczb mają także znak plus lub minus, lecz operowanie pojęciem *wektorów* znacznie lepiej ilustruje zjawiska przemieszczania. Przyjąć jednak trzeba jednoznaczne ograniczenie, że są to wektory odwzorowane na jednej osi liczbowej, w pojedynczym wymiarze (w jednowymiarowej przestrzeni euklidesowej, a nie na płaszczyźnie). Nie są to więc wektory równoległe, jak mogłoby wynikać z analizy ryciny 41, ilustrującej mechanizm składania, lecz współosiowe, jak na tej samej rycinie już po złożeniu składowych w wektor wynikowy. W takim założeniu suma/różnica wektorów jest równa sumie/różnicy ich wartości skalarnej, a zwrot przy odejmowaniu skierowany jest zgodnie z wektorem o większej długości.

Zarówno *Efekt* jak i *Trend* są parametrami liczbowymi *niemianowanymi*, co oznacza, że nie przedstawiają one żadnych jednostek (nie są ani wagami, ani poziomami ocen). Ich wartości mieszczą się zawsze w przedziale  $\langle -1; +1 \rangle$  a indeksowy sposób obliczania upoważnia do podawania wartości w postaci ułamków bądź procentów, po wymnożeniu przez 100%. Właśnie ze względu na takie właściwości stanowią one podstawę do porównywania wyników różnych grup respondentów, w tym również prób pochodzących z różnych populacji. Ponadto obydwie te parametry stanowią bazę do skonstruowania wewnętrznego wyznacznika trafności pomiarowej.

<i>Ślady</i>		<i>Wzorce</i>
wahania wypowiedzi	.....	fluktuacja jako baza analizy indeksowej
zmiana a fluktuacja	.....	różnice statystyk, statystyki z różnic
różnice a wektory	.....	składanie wektorów współosiowych
amplitudy fluktuacji	.....	przyrost i spadek, indeksy przemian
obwiednia fluktuacji	.....	zasięg, siła zewnętrznych oddziaływań
dynamika fluktuacji	.....	efekt, wynikowy wyznacznik przemian

## Wyznaczniki jakości pomiarowej

Trudno jest przejrzeć oddzielić rezultaty ewaluacji świadczące o jakości indykator od empirycznych wyznaczników dobroci samego pomiaru. W obu przypadkach wykorzystuje się zazwyczaj te same zbiory danych. Jeśli celem badań było np. ustalenie poziomu natężenia określonych cech mentalnych, to uzyskane rezultaty są zarówno symptomami tych cech, jak też splotem wielu innych czynników, cech i zjawisk. Zawsze aktualne pozostają pytania: Czy wyniki odzwierciedlają to, co było celem ewaluacji; czy raczej są efektem tego, że trafiliśmy na niestandardową grupę badawczą? A może rezultaty są w ogóle skutkiem jakiegoś niezamierzonego błędu w zastosowanej metodologii? W pewnym stopniu wyjaśnia to drugi pomiar tych samych prób badawczych. I choć niektórych kwestii nie udaje się nigdy rozstrzygnąć, to w interpretacji trzeba podejmować decyzje. Ułatwiają to specjalnie skonstruowane wyznaczniki statystyk kwantyfikacyjnych, znamienne w obranym modelu skalowania.

➤ **Wyznaczniki jakości pomiarowej** – to zbiór strategicznych dla danej skali parametrów kwantyfikacyjnych, weryfikujących z jednej strony prawidłowość empirycznego funkcjonowania idei, na której oparto skalowanie narzędzia, a z drugiej strony – weryfikujących aspekty wynikowe procesu badawczego.

W *Skali dwuważonych ocen* do oszacowania jakości badań stosuje się wyznaczniki: konkluzyjności, selektywności, trafności i rzetelności pomiaru. Oczywiście ewaluator, w zależności od potrzeb, może wykorzystać także inne znane statystyki i techniki analizy, bądź ograniczyć się do opisanych poniżej. Każdy z wyznaczników jakości pomiarowej spełnia kluczową rolę w różnych fazach opracowywania narzędzia skalowanego i przeprowadzania ewaluacji:

➤ **Kwantyfikator selekcyjny konkluzyjności wskaźników** – na etapie konstruowania i weryfikowania skali ułatwia podejmowanie decyzji w kwestii doboru do kwestionariusza optymalnych bodźców-stwierdzeń, posiadających zdolność do generowania wypowiedzi zróżnicowanych, lecz walencyjnie zgodnych.

➤ **Kwantyfikator selekcyjny istotności różnic** – w fazie porównywania grup (np. w eksperymencie) służy do testowania, czy różnice średnich są istotne statystycznie. Ponadto pomaga w ustalaniu liczebności prób wystarczających do uzyskania z góry założonej precyzji pomiaru i rozdzielczości wyników.

➤ **Kwantyfikator kumulacyjny trafności wyników** – podczas interpretowania trendów i efektów ewaluacji wspiera formułowanie wniosków uogólniających i orzekanie co do wiarygodności parametrów odzwierciedlających zachodzące przemiany. Służy do porównania wyników obliczonych dwiema metodami.

➤ **Kwantyfikator buforowy rzetelności skali** – po upowszechnieniu wyników umożliwia porównywanie z innymi badaniami, na odmiennych populacjach, w różnych środowiskach, za pomocą innych skal i wobec innego indykatoru. Ponadto może wspierać procesy ekstrakcji zmiennych i optymalizacji skali.



**Konkluzyjność – siła wnioskowania**

Strategicznym wyznacznikiem mocy wnioskotwórczej wskaźników oraz kwantyfikatorem doboru bodźców-stwierdzeń jest standaryzowany parametr *Konkluzyjności* 'Conclusiveness' [č], będący zestawieniem rozrzutu z rozziwem:

$$\check{c} = \frac{\frac{d' + d''}{2} - |-b'| - |-b''| - |s'| |s''|}{X_{\max}} \rightarrow 1 \quad [24]$$

gdzie  $d', d''$  – rozrzuty w pomiarze początkowym i końcowym, ze wzoru [6]  
 $-b', -b''$  – rozziwy w pomiarze początkowym i końcowym, ze wzoru [3]  
 $\pm s', \pm s''$  – skośności w pomiarze początkowym i końcowym, ze wzoru [12]  
 $X_{\max}$  – najwyższa waga przypisana do kresu górnego skali, równa +3

Od średniej dyspersji odejmowane są wartości odzwierciedlające to, jak silna jest opozycja wyrażająca opinie przeciwne poglądom kwalifikowanej większości. Z jednej strony większa zgodność daje solidniejsze podstawy do stwierdzenia, że mamy rzeczywiście do czynienia z uformowanymi już standardami, lecz z drugiej strony nie może to być zgodność całkowita, gdyż wówczas zbędna byłaby ewaluacja. Strategią osiągania konkluzyjności badań za pomocą skal dwuważonych jest uzyskanie optymalnego kompromisu pomiędzy zgodnością (mały rozziw) i jednocześnie konieczną niejednorodnością (duża dyspersja). Moc wnioskotwórcza ma źródło w tym, że respondenci potwierdzają istnienie symptomów wartości tam, gdzie ona wcale nie jest oczywista i jednoznaczna.

Zerowa wartość wyznacznika [č] świadczy o braku podstaw do wnioskowania, a ujemna – o konieczności wyeliminowania bodźca-stwierdzenia przy weryfikacji skali. Wartość [č] > 0,5 oznacza prawidłowość wyjustowania oznak, dopasowanie próbkowania do aktualnej rozpiętości symptomów indykatum i osadzenie semantyki stwierdzeń w strefach zgodnych ze standardami ewaluatywnymi. Konkluzyjność w pewnym stopniu maleje przy dużej skośności rozkładu, jednak spadek jest znaczący tylko wtedy, gdy skośność pojawia się w obu pomiarach. Jeśli natomiast choćby jeden z rozkładów 'przed' lub 'po' jest w miarę symetryczny, to skośność drugiego praktycznie nie ma znaczenia.

**Przykład 21. Konkluzyjność**

Przyjmijmy zbiory danych identyczne jak w przykładzie 13 oraz statystyki obliczone wcześniej.

1	-2	2	3	2	0	2	2	-1	1	2	1	0	3	2	-3	-1	0	2	3	-2	2	-1	1	3	0	1	2	3	2
2	0	3	3	2	2	3	3	2	2	2	3	2	3	2	0	1	2	3	2	1	2	1	2	2	1	2	3	3	1

Obliczyć konkluzyjność wskaźnika wg wzoru [24], wykorzystując wyniki z przykładów 5, 11, 13, 15.

$$[d'] = 3,33 \quad [d''] = 1,73 \quad [-b'] = -0,67 \quad [-b''] = 0 \quad [s'] = -0,225 \quad [-d''] = -1,0 \quad [+d''] = 0,73$$

$$1^\circ \text{ Druga miara skośności: } [s''] = (|+d''| - |-d''|) / d'' \approx (0,73 - 1,0) / 1,73 \approx -0,156$$

$$2^\circ \text{ Konkluzyjność } [\check{c}] \approx \{(3,33+1,73)/2 - |-0,67| - |0| - |-0,225| \cdot |-0,156|\} / 3 \approx 0,61$$

**Selektywność – istotność różnicowania**

Porównywanie średnich rezultatów dwóch prób badawczych daje wprost odpowiedź, która z grup lub klas uzyskała lepszy wynik w ujęciu matematycznym. Jednak w podejściu statystycznym warto sprawdzić, czy różnice są na tyle duże, iż ewentualny błąd pomiaru lub estymacji nie podważy pewności co do zróżnicowania. Mierzmy bowiem tylko wybraną próbę respondentów, a ponadto zakładamy, że to średnia arytmetyczna wyników jednostkowych jest najlepszym oszacowaniem poszukiwanego wyniku ogólnego. A przecież wartość przeciętna ze splotu cech mentalnych wcale nie musi być idealnym wyznacznikiem wypadkowej jakości cech. Z tego powodu zakłada się pewien (w miarę niski) próg prawdopodobieństwa popełnienia błędu nie większego niż  $\alpha$  (np. 1%, 5% lub 10%) i przy takim warunku przeprowadza test selektywności. Dla poziomu istotności  $[\alpha]$  wzór na *Selektywność* 'Selectivity'  $[\check{s}]$  ma postać:

$$\check{s} = \begin{cases} U(\alpha) \cdot \left( \frac{|+d_a|}{\sqrt{n_a}} + \frac{|-d_b|}{\sqrt{n_b}} \right) & \text{dla } Q_a < Q_b \\ U(\alpha) \cdot \left( \frac{|-d_a|}{\sqrt{n_a}} + \frac{|+d_b|}{\sqrt{n_b}} \right) & \text{dla } Q_a > Q_b \end{cases} \quad \begin{matrix} U(1\%) = 2,576 \\ U(5\%) = 1,96 \\ U(10\%) = 1,645 \end{matrix} \quad [25]$$

gdzie  $+d_a, -d_a, +d_b, -d_b$  – realne odchylenia porównywanych prób {a} i {b}  
 $Q_a, Q_b$  – średnie jakości porównywanych prób {a} i {b}, ze wzoru [2]  
 $n_a, n_b$  – liczebności respondentów w porównywanych próbach {a} i {b}

Test polega na porównaniu różnicy obu średnich z parametrem *Selektywności*, równym sumie odpowiednich przedziałów ufności realnej (ryc. 42). Jeśli różnice średnich są większe od  $[\check{s}]$ , to stwierdzamy istotność różnic z pewnością  $100\% - \alpha$ . W przeciwnym wypadku zróżnicowanie nie jest wiarygodne i trzeba przyjąć inny poziom istotności, zmieniając wartość współczynnika  $U(\alpha)$ <sup>45</sup>. Wzór [25] ulega uproszczeniu w sytuacji, gdy porównywane są dwie średnie z obu pomiarów 'przed' i 'po' tej samej grupy. Po przekształceniu można też obliczać  $[n]$  jako liczebności prób konieczne do uzyskania z góry założonej selektywności.

$$\check{s} = \begin{cases} 1,96 \left( \frac{|+d'| + |-d''|}{\sqrt{n}} \right) & \text{dla } Q' < Q'' \\ 1,96 \left( \frac{|-d'| + |+d''|}{\sqrt{n}} \right) & \text{dla } Q' > Q'' \end{cases}$$

**Przykład 22. Selektywność**

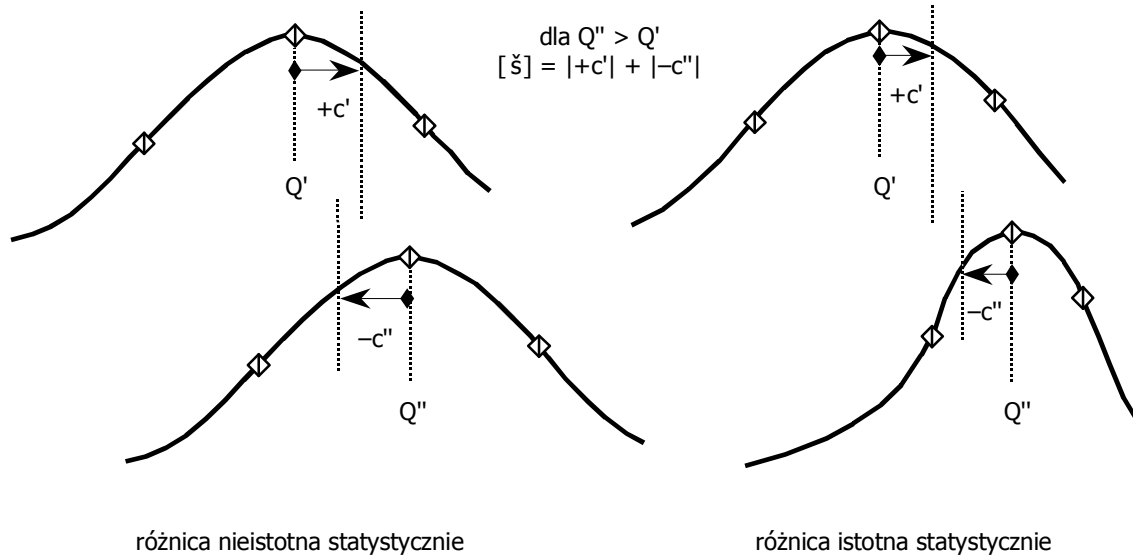
Przyjmijmy próbę badawczą  $n = 30$  i zbiory danych identyczne jak w przykładzie 13.

Zbadać istotność różnic średnich wg wzoru [24], wykorzystując wyniki z przykładów 5 i 13.

$$n' = n'' = 30 \quad Q' = 1 < Q'' = 2 \quad [+d'] = 1,29 \quad [-d''] = -1,0 \quad \text{poziom istotności } \alpha = 5\%$$

$$1^\circ \text{ Selektywność: } [\check{s}] = 1,96 \cdot (|+d'|/\sqrt{30} + |-d''|/\sqrt{30}) \approx 1,96 \cdot (1,29 + 1,0)/\sqrt{30} \approx 0,82$$

$$2^\circ Q'' - Q' = 1 > [\check{s}] \approx 0,82 \quad \text{Wniosek: Średnia } Q'' \text{ jest istotnie większa od } Q' \text{ z pewnością } 95\%.$$



Ryc. 42. Selektywność – oszacowanie istotności różnic pomiędzy średnimi

**Trafność – słuszność wnioskowania**

Wyznacznikiem weryfikacji jakości pomiarowej jest *Trafność* 'Aptness'  $[\hat{a}]$ :

$$\hat{a} = 1 - |(\pm t) - (\pm e)| \rightarrow 1 \quad [26]$$

gdzie  $\pm t$  – trend obliczony ze wzoru [19] jako parametr analizy różnicowej  
 $\pm e$  – efekt obliczony ze wzoru [23] jako parametr analizy indeksowej

Parametr ten jest zestawieniem *Trendu*, wyznaczonego za pomocą różnicowej analizy zmian, z *Efekt*em ustalonym za pomocą indeksowej analizy fluktuacji. Na podstawie zgodności lub niezgodności obu tych miar, obliczonych dwiema różnymi metodami, orzeka się o stopniu wiarygodności wyników odzwierciedlających zachodzące przemiany. Należy przy tym pamiętać, że *Trafność* jest kwantyfikatorem kumulacyjnym, a więc jego wyznaczanie i interpretowanie odnosi się tylko do wyższych poziomów zmiennych zagregowanych, a więc co najmniej do czynników, a głównie do wyników połówkowych i ogólnych. W poniższym przykładzie zrobmy odstępstwo, obliczając trafność wskaźników.

**Przykład 23. Trafność**

Przyjmijmy próbę badawczą i zbiory danych identyczne jak w przykładzie 13.

Wyznaczyć trafność pomiaru wg wzoru [24], wykorzystując obliczenia z przykładów 17 i 20.

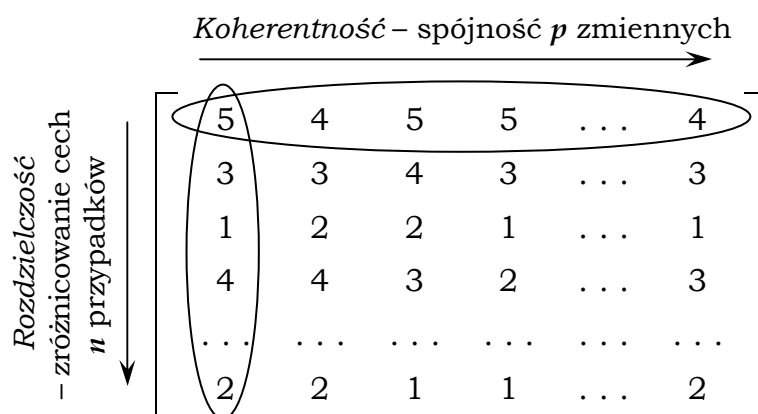
Trend  $[\pm t] \approx +0,516$       Efekt  $[\pm e] \approx +0,185$

Trafność:  $[\hat{a}] = 1 - |(\pm t) - (\pm e)| \approx 1 - |0,516 - 0,185| \approx 0,67$

Zgodnie z przewidywaniem, obliczona trafność wskaźników jest bardzo niska. Potwierdza to konieczność budowania skal pomiarowych złożonych z wielu wskaźników. W ewaluacji *InfoKultury* za pomocą skali 24-pozycyjnej parametry trafności wyników grup osiągały średnio  $>0,96$  a czynników  $>0,95$  (tj.  $\alpha < 5\%$ ).

### Rzetelność – zróżnicowanie wewnątrzspójne

Skala cech mentalnych powinna wyraźnie wskazywać, że badane osoby są na znacznie różniących się poziomach, jednak u konkretnej osoby oznaki mierzonej cechy powinny być spójne i oscylować wokół poziomu charakteryzującego cechę osobniczą. Pierwszą z tak zdefiniowanych właściwości skali winna być *rozdzielczość*, czyli zdolność do różnicowania natężeń indykatum, podobna interpretacyjnie do mocy dyskryminacyjnej w skali Likerta. Drugą – *koherentność*, oznaczająca zdolność do wydobywania spójnych wskaźników i uogólniania koincydentalnych oznak tej cechy, jaką chcemy badać, podobna interpretacyjnie do miar trafności wewnętrznej skal. W przekroju przez przypadki o liczebności [n] „rzetelna” skala ujawnia duże zróżnicowania wypowiedzi respondentów wobec pojedynczego wskaźnika (kolumny macierzy na ryc. 43), natomiast w przekroju przez wskaźniki zmiennych – możliwie największą zgodność wypowiedzi każdego indywidualnie respondenta (wiersz macierzy).



Ryc. 43. Rozdzielczość i koherentność macierzy wskaźników

Obie te właściwości ujmuje parametr o nazwie *Rzetelność* 'Reliability'  $\bar{r}$ , przy czym komponent zróżnicowania cech znajduje się w mianowniku wzoru:

$$\bar{r} = \frac{p}{p-1} \left( 1 - \frac{\sum_{k=1}^p \sum_{j=1}^n \left( X_{jk} - \frac{1}{n} \sum_{j=1}^n X_{jk} \right)^2}{\sum_{j=1}^n \left( \sum_{k=1}^p X_{jk} - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^p X_{jk} \right)^2} \right) \rightarrow 1 \quad [27]$$

gdzie  $X_{jk}$  – waga wypowiedzi  $j$ -tego respondenta wobec  $k$ -tego wskaźnika  
 $n$  – liczebność ogółu respondentów,  $p$  – liczba wskaźników w skali

Jest to wiernie odwzorowany współczynnik rzetelności *alfa* Cronbacha, zalecany jako buforowy wyznacznik jakości, porównawczy dla różnych pomiarów. Świadczyć może o wiarygodności oznak i o prawidłowym doborze wskaźników rzeczywiście tej samej cechy, lecz mierzonej u zróżnicowanych respondentów.

**Przykład 24. Rzetelność**

Jako ilustrację sposobu obliczania współczynnika rzetelności przedstawię zasadę wykorzystania do tego celu arkusza kalkulacyjnego. Dla przejrzystości tabeli przyjmijmy małą liczebnie próbę  $n=5$  i tylko  $p=5$  pozycji skali. Nie ma to wpływu na samą procedurę obliczania, a w arkuszu liczba kolumn i wierszy może być powiększona stosownie do potrzeb. Wykorzystamy dane z ryciny 43.

Zacznijmy od uproszczenia formuły zapisu wzoru, przypisując każdej sumie  $\sum$  dużą literę od A do F. W takiej postaci trzeba jednak pamiętać, że litery przed nawiasem C( B( F( E( nie oznaczają mnożenia.

$$\tilde{r} = \frac{p}{p-1} \left( 1 - \frac{C(B(X-A)^2)}{F(D-E(D))^2} \right)$$

Pozostawmy wolny pierwszy wiersz arkusza na ewentualne opisanie kolumn oraz lewą kolumnę (boczek) na późniejsze opisanie zawartości wierszy:

	A	B	C	D	E	F	G	H
1	$X_{jk}$	k=1	k=2	...	...	p: k=5	sumy D	$(D-E)^2$
2	j=1	5	4	5	5	4	23	81
3	j=2	3	3	4	3	3	16	4
4	...	1	2	2	1	1	7	49
5	...	4	4	3	2	3	16	4
6	n: j=5	2	2	1	1	2	8	36
7	średn. A	3,0	3,0	3,0	2,4	2,6	14	174
8		4	1	4	6,76	1,96	średn. E (z sum D)	suma F
9		0	0	1	0,36	0,16		
10	$(X-A)^2$	4	1	1	1,96	2,56		0,95977
11		1	1	0	0,16	0,16		rzetelność
12		1	1	4	1,96	0,36		
13	sumy B	10,0	4,0	10,0	11,2	5,2	40,4	suma C (z sum B)

- 1° Wypełnijmy 25 komórek o adresach B2 ÷ F6 wartościami przykładowych danych X (jak wyżej).
- 2° W komórce B7 wstawmy formułę obliczania średniej A  $=\text{ŚREDNIA}(B2:B6)$
- 3° Chwyćmy myszą za prawy dolny narożnik komórki B7 i przeciągnijmy formułę poziomo do F7.
- 4° W komórce B8 wprowadźmy formułę  $=(B2-B\$7)^2$  i przeciągnijmy formułę poziomo do F8.
- 5° Zaznaczmy komórkę B8, chwyćmy myszą dolny narożnik i przeciągnijmy formułę w dół do F12.
- 6° Powtórzmy to powielanie formuł dla kolumn C, D, E i F (adresy formuł aktualizują się same).
- 7° W komórce B13 wstawmy formułę obliczania sumy B  $=\text{SUMA}(B8:B12)$
- 8° Chwyćmy myszą narożnik komórki B13 i przeciągnijmy formułę poziomo do F13.
- 9° W komórce G13 wstawmy formułę obliczania sumy C  $=\text{SUMA}(B13:F13)$
- 10° W komórce G2 wstawmy formułę obliczania sumy D  $=\text{SUMA}(B2:F2)$
- 11° Chwyćmy myszą narożnik komórki G2 i przeciągnijmy formułę w dół do G6.
- 12° W komórce G7 wstawmy formułę obliczania średniej E  $=\text{ŚREDNIA}(G2:G6)$
- 13° W komórce H2 wprowadźmy formułę  $=(G2-\$G\$7)^2$  i przeciągnijmy formułę w dół do H6.
- 14° W komórce H7 wstawmy formułę obliczania sumy F  $=\text{SUMA}(H2:H6)$
- 15° W komórce H10 wpiszmy formułę obliczania Rzetelności [r]  $=5/(5-1)*(1-G13/H7)$

Bardzo wysoki wynik  $[\bar{r}] \approx 0,96$  z przykładu 24. jest skutkiem tego, że celowo dobrałem dane tak, aby zilustrować ideę rzetelności. Sprawdźmy zatem, czy uzyskamy oznakę nierzetelności po zamianie wierszy i kolumn w macierzy:

**Przykład 25.** Nierzetelność po odwróceniu macierzy

$X_{jk}$	k=1	k=2	...	...	p: k=5	sumy D	$(D - E)^2$
j=1	5	3	1	4	2	15	1
j=2	4	3	2	4	2	15	1
...	5	4	2	3	1	15	1
...	5	3	1	2	1	12	4
n: j=5	4	3	1	3	2	13	1
śred. A	4,6	3,2	1,4	3,2	1,6	14	8
	0,16	0,04	0,16	0,64	0,16	średn. E (z sum D)	suma F
	0,36	0,04	0,36	0,64	0,16		
$(X - A)^2$	0,16	0,64	0,36	0,04	0,36		
	0,16	0,04	0,16	1,44	0,36		
	0,36	0,04	0,16	0,04	0,16		
sumy B	1,2	0,8	1,2	2,8	1,2	7,2	suma C (z sum B)

Rzetelność  $[\bar{r}] = 5/(5 - 1) \cdot (1 - 7,2/8) \approx 0,13 \rightarrow$  oznaka nierzetelności skali

Wyznacznik rzetelności może być używany poprawnie tylko w wypadku ewaluacji cech mentalnych. Cechy te bowiem u respondentów powinny być zróżnicowane. Jeśli natomiast ewaluacja dotyczy wartościowania obiektów podobnych, to wówczas taki sposób wyznaczania rzetelności nie ma sensu. Po pierwsze – respondenci pełniący rolę jurorów mogą, a nawet powinni być zgodni w wyrażanych ocenach. Po drugie – z samej definicji obiekty podobne nie powinny zbyt różnić się. Ze względu na wewnętrzne właściwości, jakie wyraża *alfa* Cronbacha, do tak bardzo popularnej i powszechnie podawanej w publikacjach miary rzetelności należy podchodzić bardziej rozważnie. Jeśli jednak zastosuje się tę miarę adekwatnie do rodzaju indykatoru, to może ona pomóc w optymalizacji bodźców-stwierdzeń przy zestrzajaniu narzędzia oraz w globalnej legalizacji wersji adaptowanych na inne obszary kulturowe. Fakt, iż jest to kryterium głównie wewnętrzne, nie umniejsza roli miary rzetelności w porównywaniu jakości wyników uzyskiwanych za pomocą różnych skal.

<i>Ślady</i>		<i>Wzorce</i>
ocena jakości pomiarowej	.....	kwantyfikatory selekcyjne i kumulacyjne
miara konkluzyjności	.....	siła wnioskowania ze wskaźników
miara selektywności	.....	test istotności zróżnicowania średnich
miara trafności pomiaru	.....	porównanie trendu zmian z efektem
rozdzielczość skali	.....	zdolność różnicowania przypadków
koherentność skali	.....	spójność kontrolnych par wskaźników
miara rzetelności skali	.....	zróżnicowanie wewnątrzspójne
stosowność miary rzetelności	.....	ewaluacja zróżnicowań mentalnych

## EKSPLORACJA ZASOBÓW EMPIRYCZNYCH

W zasadniczym stadium ewaluacji, już po dokonaniu pomiarów, badacz przystępuje do czynności eksploracyjnych. W odróżnieniu od szerszego pojęcia badań eksploracyjnych, w których kluczową rolę odgrywa strategia pomiaru, tutaj przyjmijmy definicję ograniczającą zakres do operowania na zasobach:

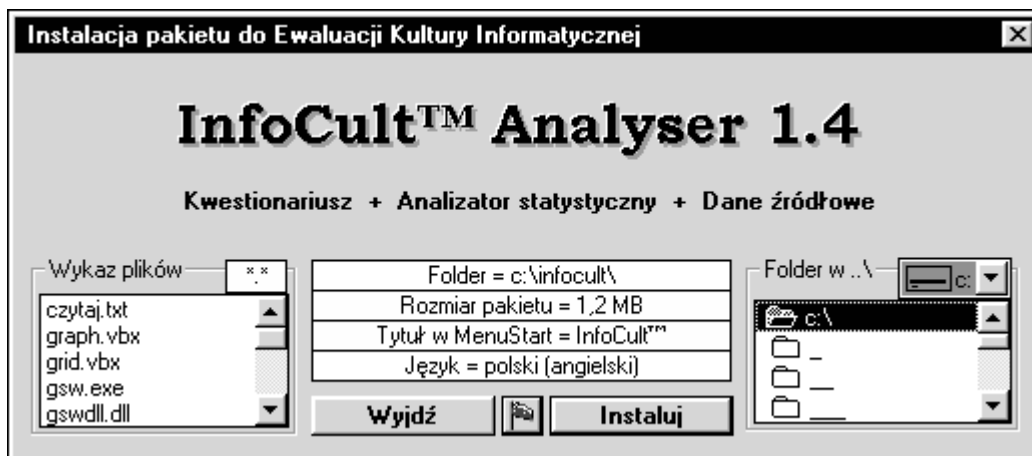
**Eksploracja** zasobów to wielofazowy i wielowątkowy proces poszukiwania i odkrywania znamienych oznak empirycznych, wymagających interpretacji i stanowiących przesłanki formułowania wniosków o stanie rzeczywistości.

W takim ujęciu penetrujemy nie teren w poszukiwaniu próbek, lecz posiadane już próbki empiryczne. Tworzą one zbiór danych, które należy przetworzyć do wartościowszej postaci zasobów statystycznych i graficznych. Materiałem do analiz są parametry i wykresy. Podczas eksploracji z danych wydobywa się to, co miało być *szukane* oraz sprawdza, czy to, co jest realnie *wskazywane*, zgadza się z tym, co było *oczekiwane*. Jest to faza poszukiwania odwzorowań modelu abstrakcyjnego w strukturach wyznaczonych doświadczalnie.

Trudno jest oddzielić fazę eksploracji od interpretacji, gdyż zwykle po odkryciu jakiejś prawidłowości od razu próbuje się znaleźć jej wytłumaczenie. Z tego powodu znaczną część opisu oznak, których należy poszukiwać, przeniosłem do następnej części książki. Tutaj skoncentrujemy się na sposobach generowania zasobów, zawierających w sobie niezbędne do analizy oznaki. Odwzorowanie struktur splotowego modelu indykatum wymaga działań przeciwnych, tj.: ekstrakcji indykatum  $\leftrightarrow$  agregacji zmiennych cząstkowych. Oprócz przekrojów przez zmienne, konieczne są też przekroje przez przypadki, tj.: rozgrupowanie ogółu  $\leftrightarrow$  scalanie prób badawczych. W tworzeniu statystyk potrzebne jest kumulowanie, ale i abstrahowanie. Proces eksploracji musi więc być w pełni sterowalny, choć warto, aby był też maksymalnie zautomatyzowany. Do tego celu najlepszym narzędziem jest komputer z dedykowanym oprogramowaniem. W takim wypadku eksploracja sprowadza się do obsługi Analizatora, który oferuje wszystko to, czego należy poszukiwać i co należy wygenerować, a użytkownik może skupić się przede wszystkim na interpretacji.

## InfoCult™ Analyser – instalacja i uruchomienie

Po włożeniu do napędu komputera płyty CD-R z pakietem samoczynnie powinien wystartować program instalacyjny. Jeśli to nie nastąpi, należy odszukać na płycie plik Install.exe i uruchomić go. Pojawi się następujące okno:



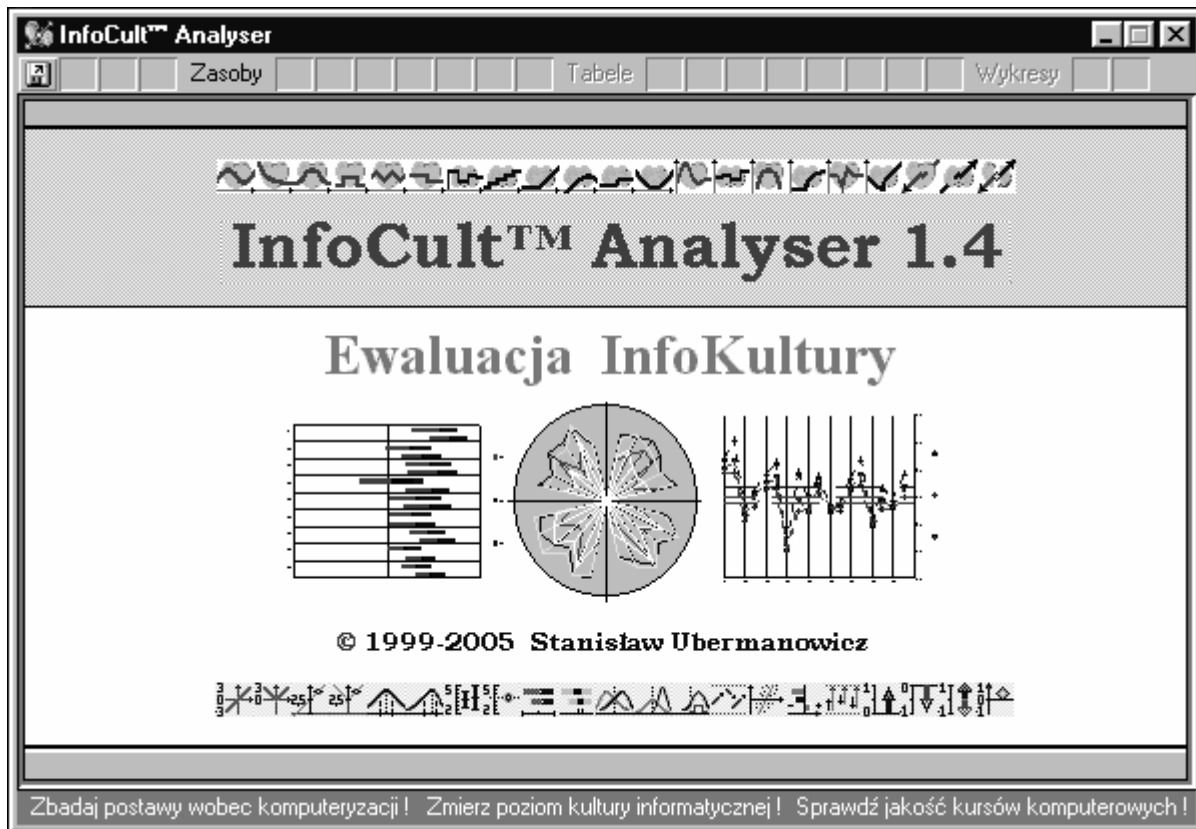
Ryc. 44. Program instalacyjny pakietu

Zachęcam do pozostawienia domyślnie proponowanych lokalizacji pakietu, bez kierowania go do innego katalogu. Przy takiej decyzji wystarczy kliknąć na przycisku [Instaluj], a całość instalowania wykonana zostanie automatycznie. Pliki zostaną przekopiowane na dysk i skupione w folderze o nazwie Infocult. W menu Start / Programy utworzona zostanie grupa InfoCult™ ze skrótami służącymi do uruchamiania Analizatora bądź edytowania Kwestionariusza, a także ze skrótem Uninstall do usuwania pakietu z dysku, bez kasowania zasobów własnych użytkownika. Warto podkreślić to, że pakiet nie umieszcza niczego w rejestrze systemowym, a tym samym nie „zaśmieca” komputera.

Program InfoCult™ Analyser funkcjonuje w różnych wersjach systemu Windows, współpracując ze wspomagającym Serwerem graficznym. Serwer ten jest uruchamiany i zamykany synchronicznie z Analizatorem, nie należy więc go odrębnie wyłączać. Pakiet został utworzony za pomocą narzędzia programowania Microsoft Visual Basic® PE 3.0, którego jestem gorącym zwolennikiem i orędownikiem (zwłaszcza na potrzeby edukacyjne), mimo że wprowadzono już znacznie nowsze wersje, lecz niestety bardzo opasłe i redundantne. Zaletą pakietu VB 3.0 jest to, że może efektywnie działać także na starszych, wolniejszych komputerach o mniejszych pojemnościach pamięci i dysków.

Program jest optymalizowany na małą rozdzielczość ekranu, można więc go używać także przy monitorach VGA. Warto ustawić tryb grafiki komputera zależnie od jakości monitora, lecz na rozdzielczość nie wyższą niż 800×600. Chodzi o to, że eksploracja połączona z analizą zajmuje wiele czasu, podczas którego ewaluator musi uważnie wpatrywać się w zawartość ekranu. Szkoda byłoby niszczyć sobie wzrok, gdyby czcionki na monitorze były zbyt małe.

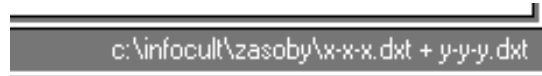




Ryc. 45. Okno startowe Analizatora

Program uruchamia się w typowy dla systemu Windows sposób, poprzez wywołanie polecenia z oferty menu Start / Programy / InfoCult™ / InfoCult™ Analyser. Po uruchomieniu Analizatora aktywność usług jest ograniczona. Dopiero po wprowadzeniu danych i utworzeniu tabel stają się dostępne wszystkie opcje menu oraz ikony poleceń. W dolnej części okna znajduje się *pasek zasobów*, na którym, po wczytaniu a także po dołączeniu danych, wyświetlane są pełne ścieżki dostępu oraz nazwy wszystkich plików załadowanych do Analizatora.

pasek zasobów ze ścieżką i nazwami plików →



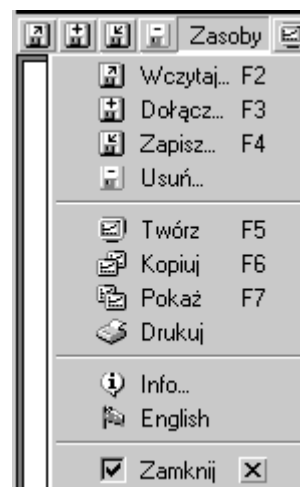
Podczas analizy szczególną uwagę należy zwracać na zmieniające się napisy w *pasku tytułowym* okna. Informują one o bieżącym stanie analizy oraz o tym, czego dotyczy zawartość okna. Ważne jest zwłaszcza to, jakiej klasy, grupy lub ogółu dotyczą parametry wyświetlane w tabelach i na wykresach oraz na jakim poziomie przebiega analiza zmiennych – czy dotyczy wskaźników, składników, czynników, czy wyników. Dodatkowo w wypadku wykresów ważne jest to, czy poziomy i dyspersje ilustrują parametry z obu pomiarów, czy tylko z jednego, w zestawieniu ze średnimi normującymi.



← pasek tytułowy okna  
← menu, ikony poleceń


## Zasoby – menu operowania zbiorami

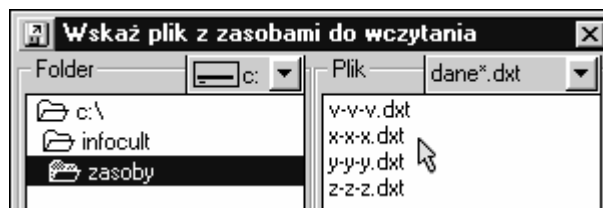
Przez *zasoby* rozumieć należy zarówno otwierane pliki (tj. zbiory danych \*.dxt i wykresy \*.wmf), jak też tabele i wykresy generowane w Analizatorze. Operowanie zbiorami polega na: wyszukiwaniu ich na dysku, wprowadzaniu do pamięci operacyjnej, scalaniu, tworzeniu nowych zasobów, zapisywaniu lub wymazywaniu z dysku, przenoszeniu do innych programów i drukowaniu. Po wybraniu z menu kategorii Zasoby pojawia się lista z ofertą dostępnych poleceń – w początkowej fazie okrojona, a następnie pełna. Zawsze aktywne jest polecenie Wczytaj, służące do ładowania plików, natomiast Dołącz, Zapisz, Usuń, Twórz, Kopiuj, Pokaż oraz Drukuj pojawiają się tylko wówczas, gdy w pamięci znajdują się już wcześniej załadowane dane. Z poleceniami skojarzone są klawisze funkcyjne F2, F3, F4, F5, F6, F7, których wciśnięcie daje taki sam skutek jak wybór z menu. Przydaje się zwłaszcza klawisz F5 do zautomatyzowanego kreowania tabel i wykresów. Obok każdej nazwy polecenia znajdują się takie same ikony, jak na *pasku menu*. Ułatwia to orientację co do tego, jaką funkcję pełni każda z ikon.



Ryc. 46. Menu zasobów

### Wczytaj – ładowanie zasobów

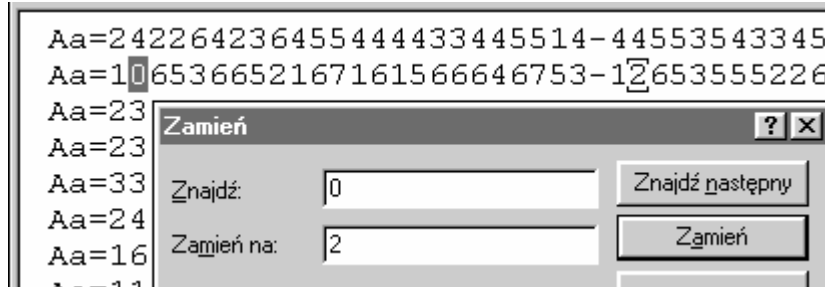
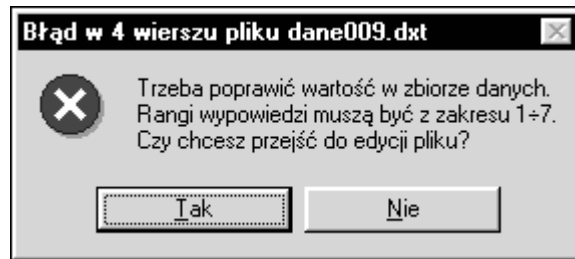
- Kliknięcie ikony  lub wciśnięcie klawisza **F2** odpowiada wywołaniu z menu polecenia Zasoby... Wczytaj, a to uaktywnia okno dostępu do foldera zasobów.
- Wskazanie myszą nazwy wybranej z listy plików o formacie dane\*.dxt powoduje załadowanie zbioru danych do pamięci i wstępne przetworzenie macierzowego zasobu surowych wag przypisanych wypowiedziom.
- Alternatywą wprowadzania zasobów jest przełączenie na format obraz\*.wmf, co umożliwi wybór z listy, wczytanie i podgląd wykresu bądź dowolnego obrazu skalowalnego, bez usuwania rezydujących w pamięci danych.
- Domyślnym miejscem przechowywania zasobów jest dysk twardy z ustaloną przy instalacji ścieżką c:\Infocult\Zasoby. Jeśli dane znajdują się w innym miejscu, to należy wybrać z listy właściwy napęd i wskazać folder zawierający zasoby.
- Dane muszą dotyczyć minimum 60 osób i 3 grup.



Ryc. 47. Wczytywanie

Zasoby danych wprowadzane do programu InfoCult™ Analyser są przetwarzane wyłącznie w pamięci operacyjnej, a zatem treść otwartych plików zawierających surowe dane zawsze pozostaje bez zmian, z wyjątkiem *korekty*.


**Korekta danych.** Jeśli podczas ładowania zasobów program wykryje jakiś błąd w strukturze lub błędną wartość danych, to wyświetlane jest okno informujące o konieczności naprawy zbioru. Po akceptacji przejścia do trybu edycji uruchomiony zostaje systemowy edytor tekstu *WordPad*, za pomocą którego we wskazanym miejscu trzeba wprowadzić poprawkę. Jeśli nie znamy właściwej liczby, wpisujemy taką samą, jak w drugim pomiarze.

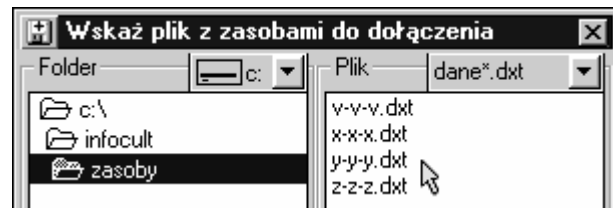


Ryc. 48. Korekta błędu w zbiorze danych

Każde ponowne wczytanie nowego pliku danych powoduje, że wcześniej otwarte zasoby usuwane są z pamięci operacyjnej. W wypadku posiadania zbiorów niezbyt licznych, zamiast wczytywać, lepiej jest dołączać kolejne dane i przetwarzać je razem, co umożliwi pełniejszą interpretację porównawczą.

### **Dołącz – kumulowanie danych**


- Kliknięcie ikony  lub wciśnięcie klawisza **F3** odpowiada wywołaniu z menu polecenia Zasoby... Dołącz, a to uaktywnia okno dostępu do foldera zasobów.
- Wskazanie myszą nazwy wybranej z listy plików o formacie dane\*.dxt spowoduje załadowanie następnych danych do zasobów rezydujących już w pamięci i ich przetworzenie łączne.
- Maksymalna liczebność połączonych danych nie może przekroczyć 1200.

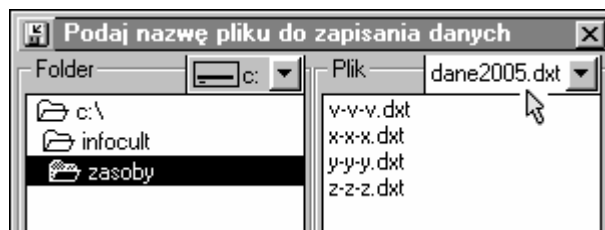


Ryc. 49. Scalanie danych

Warto przemyśleć porządek dołączania danych i wprowadzać je w takiej kolejności, ażeby obok siebie znalazły się rezultaty tych grup, które chcemy bezpośrednio ze sobą porównywać. W celu odniesienia wyników lokalnych do standardów ewaluacyjnych należy dołączyć pliki x-x-x.dxt lub y-y-y.dxt. Jest to konieczne zwłaszcza wtedy, gdy próby z własnych badań są nieliczne. Podczas przetwarzania każdorazowo automatycznie generowany jest zbiorczy plik o nazwie z-z-z.dxt mogący służyć jako kopia zapasowa na wypadek omyłkowego usunięcia pliku. Ponadto dane wczytane do pamięci komputera w optymalnej dla ewaluatora kolejności można w dowolnej chwili zapisać do nowego pliku pod inną, własną nazwą (np. dane2005.dxt).

### Zapisz – archiwizacja zasobów

- Kliknięcie ikony  lub wciśnięcie klawisza **F4** odpowiada wywołaniu z menu polecenia Zasoby... Zapisz, a to uaktywnia okno dostępu do foldera zasobów.
- Wpisanie nowej nazwy dla zasobów o formacie dane\*.dxt oraz wciśnięcie klawisza Enter spowoduje zapisanie do pliku tych danych (np. scalonych), które aktualnie rezydują w pamięci.
- Jeśli na ekranie widoczny był wykres, wówczas nastąpi zapisanie obrazu do pliku o formacie obraz\*.wmf

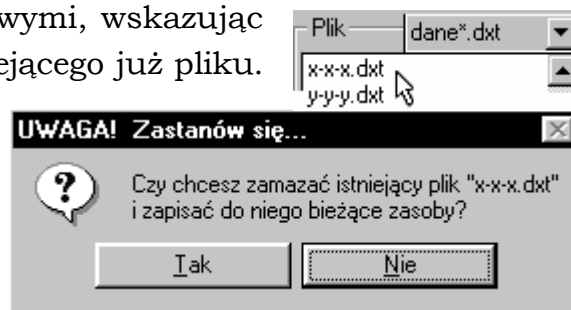


Ryc. 50. Zapisywanie i usuwanie zasobów


W oferowanej wersji programu należy stosować wyłącznie krótkie nazwy plików (do 8 standardowych liter bądź cyfr). Rozszerzenia określające format .dxt, .wmf dodawane są automatycznie. Dobrym nawykiem jest tworzenie nazw opartych na przyrastającej numeracji zasobów (np. dane2005.dxt). Dla bezpieczeństwa danych warto je zapisywać także na drugim nośniku.

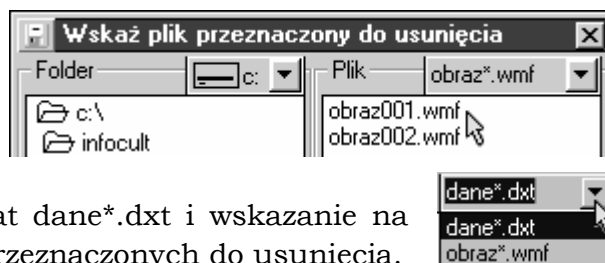
Zasoby zbędne można zastąpić nowymi, wskazując myszą podczas zapisywania nazwę istniejącego już pliku.

Pojawia się wówczas ostrzeżenie oraz zapytanie o akceptację zamiaru tzw. *nadpisania*. Trzeba wówczas uważnie przeczytać komunikat, ażeby nie utracić zbioru, który może być potrzebny.




### Usuń – kasowanie zasobów

- Kliknięcie ikony  odpowiada wywołaniu z menu polecenia Zasoby... Usuń, a to uaktywnia dostęp do foldera zasobów, domyślnie ustawianego na obraz\*.wmf
- Wskazanie myszą nazwy wykresu na liście plików o formacie obraz\*.wmf wywołuje podobne jak powyżej okno dialogowe, ostrzegające przed niezamierzonym wymazaniem pliku z dysku.
- Alternatywą jest przełączenie na format dane\*.dxt i wskazanie na liście nazwy pliku ze zbiorem danych przeznaczonych do usunięcia.

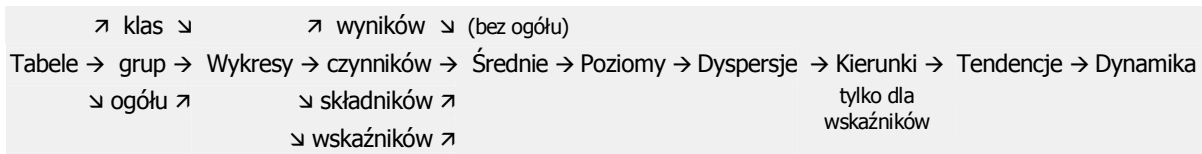


Czynność wymazywania plików z dysku poprzedzana jest zawsze ostrzeżeniem i pytaniem o potwierdzenie zamiaru usunięcia konkretnego zasobu. Polecenie usunięcia można potwierdzić lub je anulować. Usuniętych zasobów nie można już odzyskać, chyba że były to dane przetwarzane ostatnio, a ich kopia znajduje się w zbiorczym pliku z-z-z.dxt. Zalecam zachowanie szczególnej ostrożności podczas usuwania plików lub ich nadpisywania, ażeby omyłkowo bezpowrotnie nie wymazać z dysku ważnych zbiorów danych.

### Twórz – kreowanie sekwencyjne

- Klikanie ikony  lub wciskanie klawisza **F5** bądź wywołanie z menu polecenia Zasoby... Twórz uruchamia kolejne kroki w sekwencyjnym kreowaniu zasobów.
- Zależnie od aktualnego stanu analizy budowana jest nowa tabela z parametrami albo tworzony kolejny wykres, oparty na bieżących przekrojach analitycznych.


**Automatyzacja.** W programie InfoCult™ Analyser 1.4 wprowadziłem bardzo przydatny mechanizm zautomatyzowanego, sekwencyjnego przechodzenia przez kolejne etapy analizy i wizualizacji. Wystarczy konsekwentnie naciskać tylko jeden przycisk ikoniczny [Twórz] lub klawisz funkcyjny F5, aby uzyskać wszystkie możliwe fazy obróbki statystycznej oraz prezentacji graficznej.




Ryc. 51. Schemat sekwencji przy analizie automatycznej

Początkowa, domyślną fazą automatycznej analizy po wczytaniu danych jest utworzenie tabeli z parametrami wyników dla klas. Dalsze fazy to kreowanie kompletu wykresów związanych z wynikami. Następnie tworzona jest tabela miar czynników tych samych klas i związana z nią seria wykresów. W kolejnych cyklach prezentowane są miary składników i wskaźników klas oraz ich wizualizacje, po czym następuje przełączenie przekroju i ukazywane są pełne wyniki dotyczące grup. Po wyświetleniu wszystkich prezentacji grup następuje przejście do analizy rezultatów uogólnionych na populację. Warto zwrócić uwagę na to, że dla ogółu nie są prezentowane wyniki, gdyż byłyby to tabele z jedną tylko kolumną i wykresy zaledwie z pojedynczymi elementami. Ponadto wykres kierunków zmian ukazywany jest wyłącznie w fazie analizy wskaźników, gdyż jedynie ich dotyczy. Oczywiście ewaluator może w każdej chwili zmienić stan autoanalizy, przeskakując do innej, dowolnie wybranej fazy przetwarzania statystycznego lub wizualizacji.

### Kopiuj – ładowanie do schowka

- Kliknięcie ikony  lub wciśnięcie klawisza **F6** odpowiada wywołaniu polecenia Zasoby... Kopiuj, a funkcja ta umieszcza w schowku zasoby widoczne na ekranie.

Mechanizm schowka służy do przenoszenia zasobów pomiędzy różnymi programami: zazwyczaj tabel do arkusza kalkulacyjnego lub edytora tekstu, a wykresów do edytora tekstu lub programu z obróbką grafiki wektorowej. Trzeba jednak pamiętać, że zamknięcie Analizatora za pomocą przycisku  czyści zawartość schowka, jeśli więc chcemy, aby czyszczenie nie następowało, to należy kończyć pracę programu za pomocą polecenia Zamknij w menu.


W razie zamiaru umieszczenia w pamięci podręcznej (zwanej *schowkiem*) miar statystycznych, obliczonych podczas analizy i zawartych w tabeli, należy przed wywołaniem funkcji Kopiuj zaznaczyć obszar komórek przeznaczonych do zapamiętania. Polega to na zakreśleniu myszą wzdłuż wiersza, kolumny lub po przekątnej obszaru z jednoczesnym przytrzymaniem lewego przycisku.

Grupa		1	2	3	4	5	6	7	8	9	Ogół
Próba	n	94	113	55	93	33	15	15	33	59	510
Wartość	M'	3,6	3,6	3,8	3,7	4,1	3,6	4,2	3,7	4,0	3,8
Wartość	M"	3,7	3,7	4,0	3,9	4,3	3,6	4,3	3,7	4,1	3,9
Zmiana	$\pm M$	0,05	0,13	0,13	0,13	0,14	0,08	0,06	-0,02	0,07	0,10

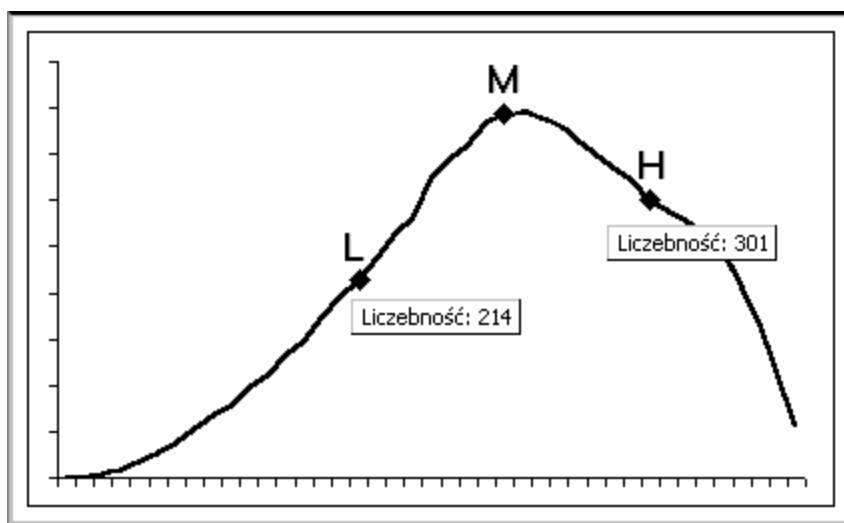
Ryc. 52. Zaznaczanie aktywnego obszaru tabeli

Z kolei przed wysłaniem wykresu do schowka warto sprawdzić, czy ustawiono właściwe opcje tworzenia grafiki (kolorowa, pastelowa, półtonowa, kreskowa), gdyż grafika wysyłana do schowka może mieć inne atrybuty niż wyświetlana.

#### **Pokaż – odtwarzanie ze schowka**


- Kliknięcie ikony  lub wciśnięcie klawisza **F7** odpowiada wywołaniu polecenia Zasoby... Pokaż, a funkcja ta wyświetla na ekranie aktualną zawartość schowka.

Odtwarzanie zawartości schowka to czynność odwrotna do kopiowania. Służy do podglądu tego, co zawiera pamięć podręczna. Na ten okres z ekranu znikają zasoby dotąd widoczne, lecz w pamięci operacyjnej pozostają dane, można więc powrócić do przetwarzania, a co więcej – można w ten sposób obejrzeć różnice np. pomiędzy wykresem ze schowka a wykresem bieżącym. Jeśli informacja przechowywana w schowku ma standardowy format tekstu bądź obrazu, to możliwe jest odtworzenie zawartości skopiowanej z innego programu, na przykład wykresu z arkusza kalkulacyjnego (jak na ryc. 53).

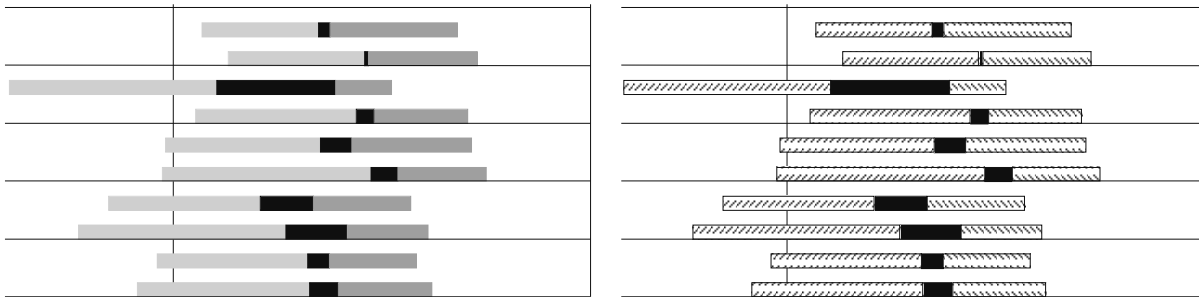


Ryc. 53. Rozkład empiryczny odtworzony ze schowka – kształt charakterystyczny dla skali ocen

### **Drukuj – drukowanie zasobów**


- Kliknięcie ikony  odpowiada wywołaniu z menu polecenia Zasoby... Drukuj, a to wysyła na drukarkę fragment tabeli lub wykres aktualnie widoczny na ekranie.

W razie zamiaru wydrukowania parametrów umieszczonych w tabeli\*, należy przed uruchomieniem funkcji Drukuj zaznaczyć myszą aktywny obszar komórek przeznaczonych do druku w taki sam sposób, jak przy kopiowaniu. Podobnie przed wydrukowaniem wykresu warto sprawdzić opcje tworzenia grafiki, gdyż widoczny na ekranie obraz jest zazwyczaj kolorowy, natomiast drukuje się najczęściej bez barw. Zwłaszcza w artykułach przeznaczonych do publikacji powinno się umieszczać bezbarwne, skalowalne wykresy kreskowe utworzone w grafice wektorowej, dlatego w każdym przypadku najlepiej jest wprowadzić ilustracje do edytora z zaawansowanymi funkcjami przetwarzania rysunków, dopracować je, sformatować, opisać i dopiero stamtąd wydrukować.

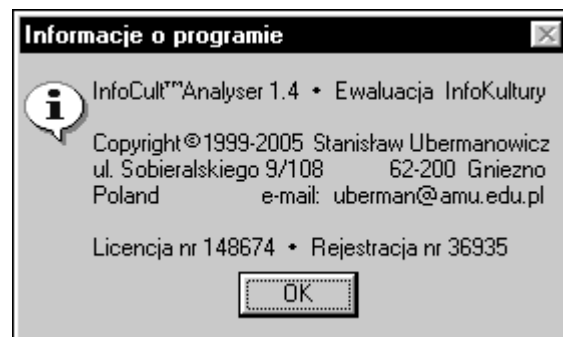


Ryc. 54. Dwa rodzaje wykresów – półtonowy i kreskowy

### **Info... – informacje o programie**

- Wybór z menu oferty Zasoby... Info  wyświetla informacje: o wersji programu, o domowym adresie autora, adresie e-mail oraz o licencji i numerze rejestracyjnym.


Autor posiada wyłączne prawa do pakietu InfoCult™ Analyser, dlatego w sprawach udzielania licencji bądź też konsultacji technicznej konieczny jest kontakt bezpośredni. Natomiast teoria oraz metodologia ewaluacji *InfoKultury* są dorobkiem zawodowym, więc w tych zagadnieniach zachęcam do nawiązania współpracy naukowej poprzez Zakład Technologii Kształcenia na Wydziale Studiów Edukacyjnych Uniwersytetu im. Adama Mickiewicza w Poznaniu.



Ryc. 55. Okienko informacyjne

\* Bezpośrednie drukowanie tabel nie na każdym sprzęcie jest możliwe – należy wówczas za pomocą schowka przenieść treść tabel do edytora lub arkusza i drukować z tamtej platformy.

### **English / Polish – zmiana języka**

- Wybór z menu oferty Zasoby... English  przełącza program na angielską wersję językową. Powrót do wersji polskojęzycznej umożliwia polecenie Zasoby... Polish.
- Pierwotnie, podczas instalowania programu, rozpoznawana jest wersja językowa systemu operacyjnego. Jeśli Windows jest zlokalizowany jako polski, język ten ustawiany zostaje jako domyślny, w przeciwnym razie – domyślnym jest angielski.

Upowszechnianie programu sprawia, że mogą chcieć go używać osoby nie znające języka polskiego. Ponadto funkcja wyboru wersji językowej jest niezwykle przydatna podczas prezentacji wyników badań na konferencjach międzynarodowych. Przełączenie języka jest błyskawiczne, a przy tym zasoby widoczne na ekranie nie znikają, lecz odświeżane są jedynie do nowej wersji.




Równocześnie z podmiianą języka we wszystkich obiektach obsługi programu (m.in. w menu), następuje także zmiana w opisach tabel i na wykresach.

Factor		Opinion	Notion	Motive	Practice	Emotion	Thought	Affection	Cognition	Culture
Norm	N"	4,3	3,7	4,0	3,9	3,8	3,4	4,0	3,7	3,9
Total		n=510								r"=0,83
Value	M'	4,5	3,6	4,2	3,2	3,7	3,3	4,1	3,4	3,8
Value	M"	4,3	3,7	4,0	3,9	3,8	3,4	4,0	3,7	3,9
Alterate	±M	-0,21	0,12	-0,21	0,68	0,09	0,10	-0,11	0,30	0,10

Approval	Ambition	Calmness	Verdict	Contact	Ability
Applause	Intention	Hardiness	Insight	Aplomb	Sagacity
Mean values of 'pre' and 'post' for all					

Ryc. 56. Opisy tabel i wykresów w języku angielskim

### **Zamknij – wyłączenie programu**



- Wywołanie z menu polecenia Zasoby... Zamknij  kończy działanie programu, przy czym zapisywane są opcje użytkownika i pozostawiona zawartość schowka.
- Kliknięcie na systemowy przycisk Zamknij  natychmiast zamyka okno programu, bez zapisywania ustawionych opcji oraz bez zapamiętywania zawartości schowka.

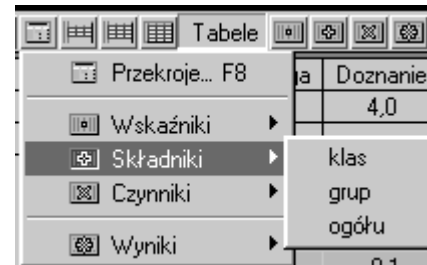
Warto zwrócić uwagę na to, że obydwa sposoby wyłączania programu dają różne skutki, więc powinny być używane w zależności od sytuacji. Jeśli użytkownik chce, ażeby następnym razem program otworzył się z tak samo zmodyfikowanymi opcjami, musi zamknąć aplikację poprzez wybór z menu, a nie poprzez kliknięcie przycisku w prawym górnym rogu okna. Wszystkie moduły współpracujące – w tym także Graphics server służący do tworzenia wykresów – otwierane są i zamykane równocześnie z programem głównym, dlatego nie należy samodzielnie zamykać Serwera podczas pracy Analizatora.



## Tabele – menu analizy statystycznej

Oferta menu służąca do kreowania tabel z parametrami statystycznymi staje się dostępna dopiero po wczytaniu danych surowych o formacie \*.dxt. Natychmiast po wczytaniu lub dołączeniu danych automatycznie pojawia się okno dialogowe z tytułem: Wybierz przekroje statystyczne, które służy do selekcji wyświetlanych miar oraz do ustawiania opcji zawartości tworzonych tabel, takich jak wybór przekroju przez zmienne i przez podgrupy badawcze.



Na pasku menu, obok przycisku Tabele znajdują się przyciski ikoniczne, które umożliwiają szybkie generowanie tabel na podstawie opcji domyślnych, zapamiętanych z poprzedniego otwarcia programu bądź ustawionych w trakcie bieżącej analizy. Kliknięcie na jednej z ikon  spowoduje utworzenie odpowiednio tabeli dla klas, grup lub ogółu, natomiast kliknięcie na którejś z grupy czterech ikon  utworzy odpowiednio tabelę z parametrami wskaźników, składników, czynników lub wyników.



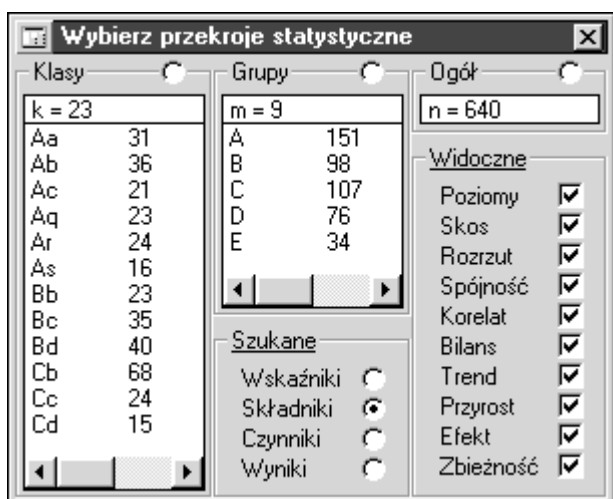
Ryc. 57. Menu tabel

Tabel ze statystyką ewaluacji nie można zapisywać do pliku, lecz można je drukować. Najlepszym jednak sposobem jest skopiowanie do schowka, po czym wklejenie do edytora i wydrukowanie dopiero po odpowiednim opisaniu i sformatowaniu. Przed skopiowaniem do schowka lub przed bezpośrednim wydrukowaniem należy zaznaczyć myszą aktywny obszar komórek w tabeli. Utworzenie tabel jest niezbędne do późniejszego generowania wykresów.

### Przekroje – wybór zmiennych, miar i prób

- Kliknięcie ikony  lub wciśnięcie klawisza **F8** odpowiada wywołaniu polecenia Tabele... Przekroje, a to uaktywnia okno wyboru zmiennych, parametrów i prób.
- W pierwszej kolejności należy w ramce Szukane wybrać przekrój przez zmienne, poprzez kliknięcie jednej z opcji:  Wskaźniki,  Składniki,  Czynniki lub  Wyniki.
- Następnie w ramce Widoczne należy przejrzeć i pozostawić jako zaznaczone parafką tylko te parametry, które chcemy aby znalazły się w tabelach miar statystycznych.
- Ostatnią czynnością jest wskazanie próby badawczej:  Klasy,  Grupy albo  Ogółu, co spowoduje natychmiastowe wygenerowanie tabeli zgodnej z ustawieniami.
- Jeśli chcemy wycofać się bez generowania tabeli, to należy kliknąć na ikonie .

Posługiwanie się oknem wyboru przekrojów (ryc. 58) zastępuje potrzebę wywoływania pozostałych ofert z menu tworzenia tabel. Umożliwia bowiem ustalanie wszystkiego: na jakim poziomie ekstrakcji | agregacji zmiennych ma być prowadzona analiza statystyczna i jakich prób badawczych ma dotyczyć. Wyłącznie w tym oknie można wybrać miary, jakie mają być widoczne w tabelach, a także przejrzeć strukturę prób zakodowanych w pliku danych.



Ryc. 58. Przekroje – próby i miary

**Liczebności.** W oknie dialogowym przekrojów wyświetlane są istotne informacje dotyczące wielkości prób: ogółu [n], liczby klas [k] i grup [m]. Widoczne są też kody literowe przypisane klasom i grupom, wraz z liczbą respondentów w każdej z tych prób. Po wczytaniu pliku z danymi warto odczytać liczebności oraz sprawdzić strukturę zasobów pod kątem tego, czy ustalona kolejność grup i klas jest najwłaściwsza do prowadzenia analiz porównawczych.

### Wskazniki – przetwarzanie argumentów


- Po wywołaniu z menu polecenia Tabele... Wskazniki  pojawia się oferta wyboru przekroju: klas, grup, ogółu. Wybór próby uruchamia generowanie tabeli.

Tabela wskaźników ogółu zawiera najważniejsze wyznaczniki niezbędne do określania poprawności pomiaru i jakości narzędzia. Obowiązkowo należy ją utworzyć i szczegółowo przeanalizować zawarte w niej parametry podczas weryfikacji i optymalizacji nowo tworzonej skali. Dotyczy to zwłaszcza miary *Konkluzyjności* [c], która wyznacza, jak mocną podstawę wnioskowania dają konkretne bodźce-stwierdzenia, pod warunkiem posiadania danych z licznej próby badawczej. Sprawdzić należy, czy któryś z 24 wskaźników nie daje miary konkluzyjności mniejszej od 0,5. Świadczyłoby to bowiem o konieczności usunięcia bądź przeredagowania bodźca-stwierdzenia.

Wskaźnik	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Rozrzut	d"	2,7	2,8	2,3	1,9	2,9	2,2	3,0	3,0	2,8	3,1	2,4	2,5	2,1	2,3	3,4	2,6	2,3	2,9	3,0	2,9	2,6	2,2	2,9	3,2
Rozziew	-b"	-0,2	-0,2	-0,2	-0,1	-0,4	-0,2	-0,6	-0,6	-0,2	-0,6	-0,3	-0,1	-0,1	-0,1	-0,9	-0,2	-0,2	-0,4	-0,5	-0,5	-0,3	-0,1	-0,3	-0,9
Konkluz.	c	0,7	0,7	0,7	0,6	0,7	0,7	0,7	0,7	0,7	0,7	0,8	0,7	0,6	0,6	0,6	0,8	0,7	0,7	0,8	0,7	0,7	0,7	0,6	0,7

Ryc. 59. Selekcyjne kwantyfikatory wskaźników


Drugą miarą świadczącą o jakości kwestionariusza jest *Rozziew* [-b], przy czym należy brać pod uwagę głównie parametr z drugiego pomiaru [-b"]. Wynika to z faktu, że niektórzy uczestnicy w chwili podjęcia kursu mogli nie mieć żadnych uprzednich doświadczeń komputerowych. Jeśli jednak rozziew zarówno w pierwszym, jak i w drugim pomiarze będzie zbyt duży [ $|-b| > |-1|$ ], to wówczas treść tego bodźca-stwierdzenia również należy zmodyfikować. Kolejną miarą pomocną w dopasowywaniu skali do realnego rozproszenia wskaźników jest *Rozrzut*, którego teoretycznym ideałem byłaby wartość  $d=3$  identyczna z zakresem obszaru zgodności, najlepiej z bliską zeru skośnością.

**Normalizacja.** Jeśli posiada się reprezentatywną próbę danych, wówczas najważniejszym parametrem w tabeli wskaźników ogółu staje się *Norma* [N'']. Jest to uśredniony poziom empirycznie wyznaczonych oznak danej cechy, będący najlepszym oszacowaniem wartości oczekiwanej dla całej populacji. Oznacza to, że po zbadaniu np. 600 jednolitych populacyjnie respondentów uzyskamy pewną normę – zwaną *standardem ewaluatywnym* – która teoretycznie powinna być taka sama, jak po zbadaniu innych 600 reprezentantów tej samej populacji, w porównywalnym okresie rozwoju społecznego. Poziomy wartości średnich [M''], uzyskane przez każdą z grup i klas, powinny więc być zawsze porównywane z normami oczekiwanymi [N''] (zob. ryciny 60÷62).

Wskaźnik	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Norma	N''	4,4	4,2	3,9	3,8	4,1	3,9	3,4	3,5	4,7	3,8	3,7	4,5	4,6	4,2	3,5	4,1	4,0	3,8	3,3	3,6	3,9	4,3	4,0	2,7
Grupa	1																								
Wartość	M''	4,3	4,0	2,5	2,5	3,4	3,6	2,8	2,8	4,1	3,7	3,9	4,6	4,7	4,5	3,4	3,7	3,3	3,1	3,9	3,1	3,9	4,0	3,9	2,9
Wartość	M''	4,1	3,9	3,3	3,5	3,6	3,5	2,8	3,0	4,4	3,5	3,4	4,5	4,4	4,2	3,0	3,7	3,9	3,6	3,8	3,2	3,7	4,3	4,0	2,6
Zmiana	±M	-0,2	-0,1	0,8	1,0	0,3	0,0	-0,1	0,2	0,2	-0,2	-0,5	-0,1	-0,4	-0,3	-0,4	-0,1	0,6	0,4	-0,1	0,1	-0,2	0,3	0,1	-0,4
Grupa	2																								
Wartość	M''	4,3	4,4	2,4	2,1	3,0	3,3	2,9	3,0	3,8	3,4	3,9	4,5	4,7	4,4	3,5	3,8	3,5	3,1	3,8	2,9	4,1	4,2	3,8	2,7
Wartość	M''	4,2	4,0	3,7	3,6	3,9	3,4	3,1	3,0	4,2	3,6	3,7	4,1	4,4	4,1	3,4	3,6	3,7	3,7	3,3	3,4	3,5	4,1	4,2	2,9
Zmiana	±M	-0,1	-0,4	1,3	1,5	0,9	0,1	0,2	0,0	0,4	0,2	-0,2	-0,4	-0,3	-0,3	-0,1	-0,3	0,2	0,6	-0,5	0,5	-0,6	-0,1	0,3	0,2

Ryc. 60. Porównywanie wskaźników dwóch grup z poziomami normy

### Składniki – obliczanie komponentów

- Po wywołaniu z menu polecenia Tabele... Składniki  pojawia się oferta wyboru przekroju: klas, grup, ogółu. Wybór próby uruchamia generowanie tabeli.

Składnik		Ocena	Aplauz	Ambicje	Intencje	Spokój	Odpór	Osąd	Wgląd	Obycie	Pewność	Zdolność	Ogłada
Norma	N''	4,4	4,3	4,1	4,1	4,2	3,4	3,8	3,9	3,9	4,0	3,4	3,4
Grupa	1	n=94											
Wartość	M''	4,6	4,2	4,2	4,0	3,9	3,4	3,6	3,2	2,5	3,5	2,8	3,5
Wartość	M''	4,3	4,0	3,9	4,0	3,9	3,3	3,3	3,7	3,4	3,6	2,9	3,5
Zmiana	±M	-0,30	-0,16	-0,31	0,05	0,03	-0,12	-0,22	0,53	0,93	0,12	0,09	0,02
Grupa	2	n=113											
Wartość	M''	4,6	4,3	4,2	4,2	3,6	3,3	3,6	3,3	2,3	3,2	3,0	3,4
Wartość	M''	4,3	4,1	3,9	3,8	3,9	3,5	3,5	3,7	3,7	3,7	3,0	3,4
Zmiana	±M	-0,30	-0,25	-0,31	-0,33	0,27	0,24	-0,17	0,43	1,41	0,49	0,07	0,00

Ryc. 61. Porównywanie składników z poziomami normy


Tabele składników zawierają miary zmiennych cząstkowych, powstałych poprzez połączenie kontrolnych par wskaźników. Argumentami statystyk są średnie arytmetyczne wag dwóch sąsiadujących pozycji w kwestionariuszu [i: { p-1, p } ] o numerach nieparzystych i parzystych według schematu:

Ocena Aplauz Ambicje Intencje Spokój Odpór Osąd Wgląd Obycie Pewność Zdolność Ogłada  
 13×14 1×2 11×12 21×22 9×10 23×24 15×16 17×18 3×4 5×6 7×8 19×20

Z porównania tych par wskaźników obliczane są dwa dodatkowe parametry statystyczne: *Spójność* i *Zbieżność*, które są wyznacznikami weryfikującymi

czasoprzestrzenne aspekty jakości skali pomiarowej. Kontrolna para wskaźników mierzy ten sam symptom, więc powinna być spójna, chyba że niespójność można uzasadnić dużym zróżnicowaniem czasoprzestrzennym. Podczas analizy składników warto też porównywać ze sobą wszystkie pozostałe miary, zwłaszcza z sąsiadujących kolumn, gdyż z tych par tworzone są czynniki.

### Czynniki – obliczanie dominant

- Po wywołaniu z menu polecenia *Tabele... Czynniki*  pojawia się oferta wyboru przekroju: klas, grup, ogółu. Wybór próby uruchamia generowanie tabeli.

Czynnik		Opinie	Poglądy	Motywacje	Wprawa	Emocje	Rozwaga	Doznanie	Poznanie	Kultura
Norma	N"	4,3	3,8	4,1	3,9	3,8	3,4	4,1	3,7	3,9
Grupa	1	n=94								
Wartość	M'	4,4	3,4	4,1	3,0	3,7	3,2	4,0	3,2	3,6
Wartość	M"	4,1	3,5	4,0	3,5	3,6	3,2	3,9	3,4	3,7
Zmiana	±M	-0,23	0,15	-0,13	0,52	-0,04	0,06	-0,14	0,24	0,05
Grupa	2	n=113								
Wartość	M'	4,5	3,5	4,2	2,7	3,5	3,2	4,0	3,1	3,6
Wartość	M"	4,2	3,6	3,9	3,7	3,7	3,2	3,9	3,5	3,7
Zmiana	±M	-0,27	0,13	-0,32	0,95	0,25	0,03	-0,11	0,37	0,13

Ryc. 62. Czynniki, wyniki połówkowe i wynik ogólny dwóch grup

Tabele czynników zawierają parametry zmiennych splotowych, które w interpretacji traktuje się jako połączenie oznak addytywnych par składników:

Opinie	Poglądy	Motywacje	Wprawa	Emocje	Rozwaga
Ocena×Aplauz	Osąd×Wgląd	Ambicje×Intencje	Obycie×Pewność	Spokój×Odpór	Zdolność×Ogłada

Łączenie par składników w czynniki nie jest jednak procesem sumowania statystyk, lecz syntezą dwóch bardziej szczegółowych oznak dominujących w cesze bardziej uogólnionej. Możemy więc zgodzić się, że motywacje wynikają z ambicji i intencji, lecz nie ma sensu np. uśrednianie współczynników korelacji tych składników. Z tego względu w *Skali dwuważonych ocen* do obliczania miar statystycznych na każdym poziomie uogólnienia zmiennych sięga się zawsze po pierwotne argumenty zmiennej losowej, czyli po wskaźniki. W sensie statystycznym są to średnie arytmetyczne z wag nadanych wypowiedziom odnoszącym się do treści bodźców-stwierdzeń tworzących skalę.


Argumentami statystyk czynników I–VI są średnie z wag czterech wskaźników, a w wypadku *Doznania* i *Poznania* średnie z dwunastu wskaźników lub – w ujęciu uogólniania oznak – splot trzech czynników według schematu:

Opinie	Poglądy	Motywacje	Wprawa	Emocje	Rozwaga	Doznanie	Poznanie ...
1×2×13×14	15×16×17×18	11×12×21×22	3×4×5×6	9×10×23×24	7×8×19×20	...	... ř"
I	II	III	IV	V	VI	I×III×V	II×IV×VI ...

Na komponenty doznaniowe, jako zmienne dominujące, składają się *Opinie*, *Motywacje* i *Emocje*, natomiast z obszaru komponentów poznawczych wybrano *Poglądy*, *Wprawę* i *Rozwagę*, jako zmienne charakteryzujące uświadomienie.

Warto tu przypomnieć, że *Doznanie* i *Poznanie* podczas analizy traktuje się jako czynniki globalne, natomiast podczas syntezy jako wyniki połówkowe. W ostatniej kolumnie tabeli podawane są wyniki ogólne *Kultury*, a ponadto właśnie tam wyświetlany jest współczynnik rzetelności  $[\bar{r}] = \alpha$  Cronbacha, jako najczęściej obliczany i podawany w publikacjach buforowy wyznacznik jakości pomiarów skalowanych, który powinien osiągać wartość co najmniej 0,7.

### Wyniki – obliczanie rezultatów

- Po wywołaniu z menu polecenia Tabele... Wyniki  pojawia się oferta wyboru tylko z dwóch prób: klas lub grup. Wybór próby uruchamia wygenerowanie tabeli.

Grupa		A	B	C	D	E	F	G	H	I	Ogół
Próba	n	151	98	107	76	34	34	31	67	42	640
Wartość	M'	3,5	3,3	3,3	3,5	3,7	3,8	4,0	3,9	4,2	3,6
Wartość	M''	4,0	3,5	3,7	3,7	3,8	3,7	4,0	4,0	4,2	3,8
Zmiana	$\pm M$	0,51	0,22	0,35	0,23	0,14	-0,08	0,03	0,05	0,09	0,25
Zawilość	$\pm i$	0,05	0,06	0,02	-0,03	-0,07	0,06	-0,08	-0,02	-0,03	0,01
Trend	$\pm t$	0,24	0,07	0,13	0,11	0,07	-0,02	0,07	0,03	0,09	0,11
Efekt	$\pm e$	0,24	0,10	0,14	0,09	0,05	-0,03	0,00	0,02	0,04	0,11

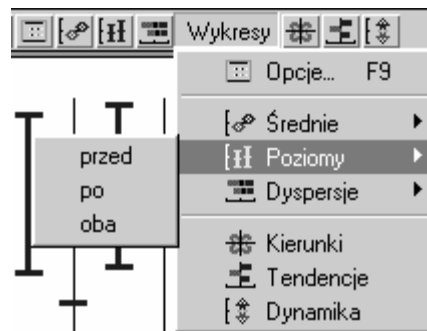
Ryc. 63. Tabela porównawcza wyników dziewięciu grup oraz ogółu

Tabele wyników są zestawieniem parametrów indykatorum *InfoKultury* dla wszystkich klas lub grup, a w ostatniej kolumnie także dla ogółu badanych. W główce tabeli widoczne są kody literowe identyfikujące daną klasę/grupę oraz liczebność tej próby [n]. Taki układ jest najkorzystniejszy do porównań między grupami, zwłaszcza w zakresie średnich poziomów 'przed' i 'po' [M', M'']. Warto ponadto porównywać z sobą inne miary dystansowe tej samej grupy. Jeśli na przykład *Zawilość*  $[\pm i]$  jest ujemna, to należy krytycznie podejść do wartości *Trendu*  $[\pm t]$ , zawyżonej prawdopodobnie wskutek zbyt łatwego programu zajęć, niedopasowanego do wyższych aspiracji i możliwości słuchaczy. Natomiast *Zawilość* powyżej 0,05 może wskazywać na zbyt ambitny program, co daje lepszy *Efekt*  $[\pm e]$ , lecz kosztem zaniedbania najsłabszych w grupie. Najbardziej trafne i jednoznacznie wiarygodne są wyniki tych prób badawczych, w przypadku których *Trend* i *Efekt* osiągają identyczne wartości. Na tej przesłance konstruowany jest wyznacznik *Trafności* pomiarowej.

Ślady		Wzorce
zasoby empiryczne	.....	dane surowe, tabele miar i wykresy
organizacja tabel	.....	selekcja miar wyświetlanych
przekroje przez przypadki	.....	klasy, grupy i ogół respondentów
przekroje przez zmienne	.....	wskaźniki, składniki, czynniki, wyniki
agregacja zmiennych	.....	schematy interpretacji splotów
statystyki ewaluacji, oznaki	.....	miary zjawisk i wyznaczniki jakości

## Wykresy – menu wizualizacji graficznej

Oferty menu służące do tworzenia wykresów stają się dostępne dopiero po wygenerowaniu dowolnej tabeli, a znikają natychmiast po wczytaniu bądź dołączeniu nowych danych. O gotowości do wizualizacji świadczą aktywne przyciski ikoniczne, pojawiające się na pasku menu wokół przycisku Wykresy. Umożliwiają one szybsze niż z poziomu menu tworzenie wykresów, jednak przy założeniu właściwości domyślnych, wcześniej już ustalonych: liczby 1÷3 grup reprezentowanych równocześnie na jednym wykresie średnich oraz opcji określającej to, które pomiary (*przed*, *po*, *oba*) mają być widoczne na wykresach poziomów i dyspersji. Wykresy generowane są zgodnie z ustalonymi opcjami koloru, odcieni, napisów i czcionek.

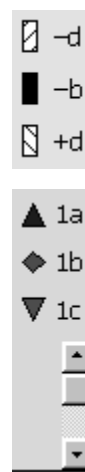


Ryc. 64. Menu wykresów

W celu ustawienia właściwości wykresów wybiera się z menu ofertę Opcje. Trzy kolejne oferty dotyczą wizualizacji miar chwilowych. Wykresy średnich M i poziomów HML tworzone są na wartościach przekonwertowanych na skalę ocen akademickich. Wykresy dyspersji ilustrują rozrzut i rozziw wypowiedzi. Ostatnie trzy oferty wizualizacji oparte są na miarach dystansowych. I tak wykres biegunowy ilustruje kierunki i proporcje zmian wskaźników, wykres pasmowy obrazuje bilans i trend przemian, a wykres wektorowy odzwierciedla dynamikę fluktuacji i tym samym jest ilustracją efektów względnych.









Istotną rolę na ilustracjach odgrywa legenda, nadająca symbolom graficznym oraz kolorom różne znaczenia, w zależności od rodzaju wykresu oraz od wyboru prób badawczych. Na większości wykresów legenda wyznacza, jaki kolor lub kształt reprezentuje dany parametr oznaczony symbolem: -d, +d, -b, M', M'', L', L'', ±b, ±t, +e, -e lub ±e.

W wypadku średnich dany znak graficzny reprezentuje klasę lub grupę, choć to przypisanie zmienia się podczas *przewijania* obrazów dla następnych klas i grup. Przewijanie oznacza tu posługiwanie się obiektem zwanym *suwakiem*, za pomocą którego płynnie lub skokowo przełącza się w przód lub wstecz podgląd obrazów. Jeżeli przewijanie jest dostępne, suwak pojawia się w prawym dolnym narożniku okna.



Wykresy ilustrujące parametry ewaluacji można zapisywać do plików w formacie grafiki wektorowej \*.wmf lub bezpośrednio drukować. Szczególnie jednak w wypadku wykresów wskazana jest obróbka ilustracji, zwłaszcza przed ich przekazaniem do publikacji. Warto więc poprzez schowek lub plik przenieść obraz do edytora z funkcjami obróbki grafiki wektorowej, po czym wyretuszować zbędne elementy, uzupełnić istotnymi opisami, dopracować czcionkę, format, proporcje, tekstury i dopiero z tamtego poziomu drukować.



### Opcje – ustawianie atrybutów grafiki

- Kliknięcie ikony  lub wciśnięcie klawisza **F9** odpowiada wywołaniu polecenia Wykresy... Opcje, a to uaktywnia okno ustawiania właściwości wykresów i opisów.
- W ramce Wyświetlanie wybiera się opcję  kolorowe, pastelowe, półtonowe lub kreskowe, obserwując efekt na ekranie i pozostawiając ustawienie najstosowniejsze do potrzeb.
- Analogicznie należy postępować w ramach: Kopiowanie, Zapisywanie oraz Drukowanie i choć zmiany nie będą widoczne, ujawnią się w schowku, w plikach i na wydruku.
- W ramce Napisy można wybrać inną wersję językową opcjami  polskie, angielskie; zaznaczenie opcji  symboliczne podmienia nazwy zmiennych na symbole ogólne, natomiast zaznaczenie opcji  tytułowe wprowadza do wykresów opisy objaśniające.
- W ramce Czcionki zmienia się wielkość liter opcjami  domyślne, mniejsze, większe, a zaznaczenie opcji  blokowe przełącza czcionki z szeryfowych na bezszeryfowe.
- Napis Alt+PrintScr w ramce Wyświetlanie przypomina o funkcji zrzutu okna do schowka, a w ramce Drukowanie opcja  z ekranu umożliwia wydruk graficzny całego okna.


Szczególną uwagę należy zwrócić na to, że istnieje możliwość wybrania innych opcji dla wykresów wyświetlanych na ekranie, a zupełnie innych dla obrazów kopiowanych do schowka, zapisywanych do pliku bądź wysyłanych na drukarkę. Ustawienia te są niezależne, toteż przed skierowaniem obrazu w inne miejsce warto choć na chwilę ustawić na ekranie taką samą opcję, ażeby podejrzeć, jak wykres będzie wyglądał w pamięci podręcznej, jak na wydruku i jak w pliku. Zwykle na ekranie ogląda się wykresy barwne, podczas gdy do druku najczęściej generuje się ilustracje monochromatyczne. W chwili skierowania obrazu w inne niż ekran miejsce, komputer usuwa cyjanowe tło wykresu, które jest korzystniejsze podczas analizy, gdyż nie razi w oczy tak jak tło białe.



Ryc. 65. Opcje wykresów

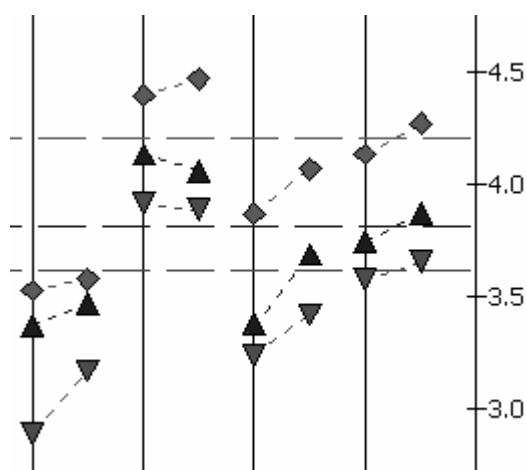
Optymalnie ustawione opcje tworzenia wykresów mogą być zapamiętane przy wychodzeniu z programu i odtworzone przy następnym uruchomieniu, o ile Analizator wyłączy się poleceniem Zamknij  po wywołaniu menu Zasoby, a nie poprzez kliknięcie na systemowym przycisku zamykania okna .

### Średnie – wykres poziomów wartości

- Kliknięcie ikony  powoduje wygenerowanie wykresu punktowego z wartościami średnimi wskaźników, składników, czynników lub wyników – dla klas, grup bądź ogółu – w zależności od tego, jakie wcześniej utworzono tabele.


- Wywołanie z menu polecenia Wykresy... Średnie udostępnia ofertę wyboru liczby 1, 2, 3 klas lub grup, których średnie umieszczone będą na wspólnym wykresie.
- Przy licznej oraz reprezentatywnej próbie respondentów wykres średnich dla ogółu staje się bezpośrednią ilustracją poziomu standardów ewaluacyjnych.

Wykresy *Średnich* zawierają ułożone obok siebie parami wartości M'M" pochodzące z obydwu pomiarów ('przed' i 'po'). Nachylenie linii łączącej średnie ilustruje dynamikę przemian, lecz ma sens tylko w obrębie jednej zmiennej<sup>46</sup>. Oś pionowa skalowana jest za pomocą ocen szkolnych. Na wspólnym wykresie można umieszczać osiągnięcia dwóch lub trzech klas/grup. Kreski poziome wyznaczają średnie oznak, a ponadto na wykresie wyników kreślony jest obszar odchylenia standardowego.

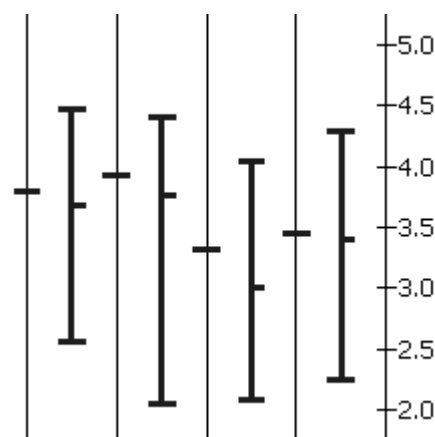


Ryc. 66. Średnie trzech grup

#### Poziomy – wykres rozkładów HML

- Kliknięcie ikony  powoduje wygenerowanie wykresu odcinkowego, ilustrującego poziomy HML wskaźników, składników, czynników lub wyników – dla klas, grup bądź ogółu – w zależności od tego, jakie wcześniej utworzono tabele.
- Wywołanie z menu polecenia Wykresy... Poziomy udostępnia ofertę [przed, po, oba], umożliwiając wybór, z jakiego pomiaru średnie mają być na jednym wykresie.


Wykresy *Poziomów* HML dla ogółu są przybliżeniem rozkładu oczekiwanego dla populacji, natomiast wykresy dla klas oraz grup służą do analizy porównawczej zmian poziomów i rozrzutów oznak w odniesieniu do średnich normujących skalę. W wypadku wybrania opcji [przed] na wykresie widoczne są poziomy H'M'L' uzyskane w pierwszym pomiarze, odniesione do *Bazy B'*. Wybór opcji [po] tworzy wykres z poziomami H"M"L" pochodzącymi z końcowego pomiaru, odniesionymi do *Normy N"*. Wybór opcji [oba] służy do obserwacji przemian, jakie zaszły w danej klasie lub grupie pomiędzy pierwszym a drugim pomiarem, relatywnie na poziomach H'-H", M'-M" oraz L'-L". O jakości procesu świadczy równomierne, dodatnie przesunięcie punktów węzłowych rozkładu. Można przyjąć oszacowanie, że w obszarze rozrzutu pomiędzy poziomami H i L powinny mieścić się wyniki uzyskane przez około 2/3 respondentów.



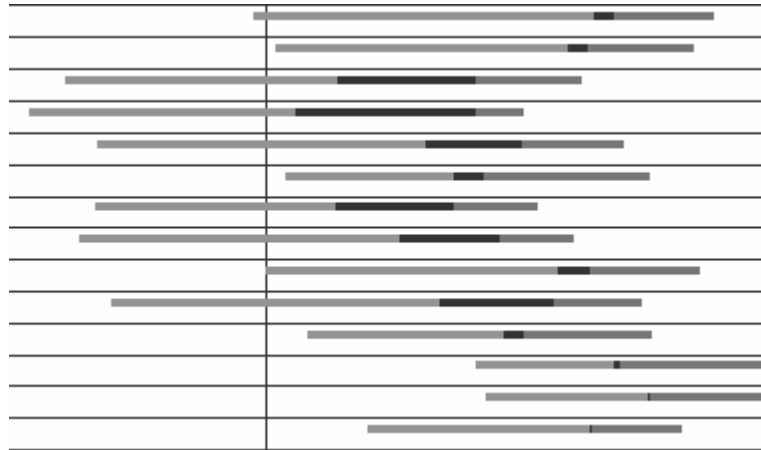
Ryc. 67. Poziomy N" i H"M"L"



### Dyspersje – wykres rozrzutu i rozziewu


- Kliknięcie ikony  powoduje utworzenie wykresu pasmowego, ilustrującego rozrzut i rozziew wskaźników, składników, czynników lub wyników – dla klas, grup bądź ogółu – w zależności od tego, jakie wcześniej utworzono tabele.
- Wywołanie z menu polecenia Wykresy... Dyspersje udostępnia ofertę [przed, po, oba], umożliwiając wybór, z jakiego pomiaru rozrzuty mają być na jednym wykresie.

Wykresy dyspersji służą m.in. do wizualizacji rozkładów empirycznych, lecz głównie do oceny jakości skali. Największą przydatność stwarza wykres rozrzutu i rozziewu wskaźników dla ogółu, umożliwia bowiem oglądanie tego, jak rozkładają się wypowiedzi względem obszaru zgodności skali. Najlepsze są te wskaźniki, które dają mały *Rozziew* [-b] i duży *Rozrzut* [d  $\approx$  3]. Ponieważ jednak mierzy się zmiany, to na skali musi pozostać obszar wolnego miejsca na oczekiwane przemieszczenia rozkładów. Na wykresach pasmowych krańce czarnych pól *Rozziewu* [-b] wskazują z lewej strony na *Jakość* [Q], a z prawej na *Intensywność* [I] wypowiedzi. Całkowita długość pasemek obrazuje *Rozrzut* [d].

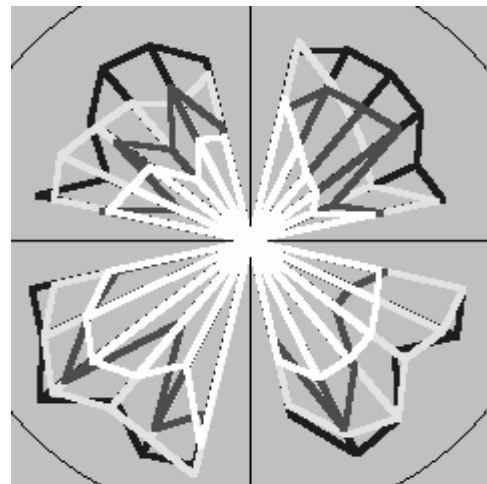


Ryc. 68. Rozrzut i rozziew wskaźników

### Kierunki – wykres gradientu zmian ML


- Kliknięcie ikony  lub wybór z menu polecenia Wykresy... Kierunki tworzy wykres biegunowy zmian wskaźników afektywnych i kognitywnych na poziomach M i L.

Gradient jest to wyznaczany graficznie kierunek największych zmian. Interpretacja wykresu polega na ustaleniu, w jaką stronę od środka nastąpiły maksymalne przyrosty. Z prawej strony skupiono wskaźniki reprezentujące *namysł*, a po lewej *impuls*. U góry ustalono biegun *poznania*, u dołu *doznania*, stąd prawa górna ćwiartka układu to biegun *kognitywności*, a lewa dolna – *afektywności*. Na wykresie obok największe przyrosty pomiędzy jasnymi i ciemnymi „płatkami rozety” widoczne są na kierunkach poznawczych.

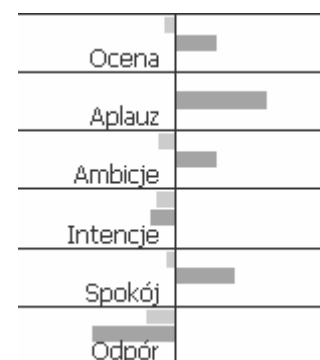


Ryc. 69. Kierunki zmian

### Tendencje – wykres bilansu i trendu zmian


- Kliknięcie ikony  powoduje utworzenie wykresu słupkowego, ilustrującego bilans i trend wskaźników, składników, czynników lub wyników – dla klas, grup bądź ogółu – w zależności od tego, jakie wcześniej utworzono tabele.
- Identyczny skutek daje wywołanie z menu polecenia Wykresy... Tendencje.

Słupki jaśniejsze to *Bilans* a ciemniejsze to *Trend* wypowiedzi. Skierowanie od osi w prawo oznacza tendencję rosnącą, a w lewo malejącą. Długość każdego słupka reprezentuje wartość przemian zachodzących na poziomie konkretnej zmiennej. W reinterpretacji tendencji należy pamiętać, że miara *Bilansu* ma taką charakterystykę, iż jej zerowa wartość jest zjawiskiem oczekiwanym (podobnie jak w bilans w księgowości). Natomiast *Trend* lub *Bilans* o wartościach ujemnych są w każdym przypadku zjawiskiem niepokojącym.

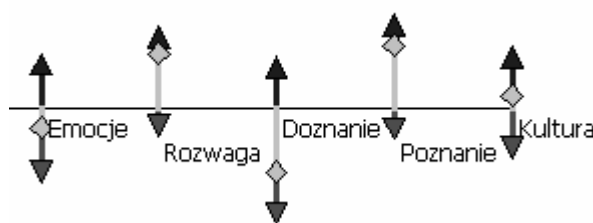


Ryc. 70. Tendencje

### Dynamika – wykres efektu i zasięgu fluktuacji

- Kliknięcie ikony  powoduje utworzenie wykresu wektorowego, ilustrującego efekt i zasięg fluktuacji wskaźników, składników, czynników lub wyników – dla klas, grup bądź ogółu – w zależności od tego, jakie wcześniej utworzono tabele.
- Identyczny skutek daje wywołanie z menu polecenia Wykresy... Dynamika.

Wektory skierowane w górę to *Przyrosty* fluktuacji par wypowiedzi, a zwrócone w dół to *Spadki*. Różnice między wektorami są miarą *Efektu*, ilustrowaną położeniem znacznika w kształcie rombu i długością odcinka – im znacznik wyżej nad osią, tym lepiej. Sumy wektorów to *Zasięg* fluktuacji – im większa rozpiętość, tym silniejsze było oddziaływanie zewnętrzne.



Ryc. 71. Dynamika przemian

Ślady	Wzorce
momenty zwykle pozycyjne	..... punktowy wykres średnich
momenty centralne	..... odcinkowy wykres poziomów HML
wymiary rozproszenia	..... pasmowy wykres rozrzutu i rozziwu
wielkości i kierunki zmian	..... biegunowy wykres gradientów $\pm M \pm L$
wartości i tendencje przemian	..... słupkowy wykres bilansu i trendu
indeksy dynamiki fluktuacji	..... wektorowy wykres efektu i zasięgu

## INTERPRETACJA ZJAWISK EMPIRYCZNYCH

Najtrudniejszą fazą ewaluacji jest wartościowanie zjawisk postrzeganych dzięki pomiarom i formułowanie wniosków o podmiocie badawczym. Wynika to stąd, że badacz dysponuje jedynie niewielką liczbą wskaźników pośrednich, podczas gdy indykatum uwikłane jest w gąszczu wielorakich uzależnień. Trochę przypomina to wykrywanie dowodów śladowych lub wręcz podejmowanie rozstrzygnięć poszlakowych, co jednak stanowi znamię ewaluacji, odróżniając ją od eksperymentów scjentystycznych z izolowaniem prób laboratoryjnych. Pozostawienie podmiotów ewaluacji w naturalnym środowisku oddziaływań utrudnia wyodrębnianie czynników, lecz jest jednocześnie taktyką korzystną. Trudność polega także na tym, że wyższe wartości liczbowe wskaźników wcale nie muszą być przejawem wartości aksjonormatywnych. Proste porównanie poziomów, na przykład wskaźników nr 1 i nr 24, w ogóle nie ma sensu. Konieczne jest stosowanie systemowych strategii analityczno-interpretacyjnych, które umożliwiają przejście od parametrów do rezultatów kwalitatywnych.

Zdefiniujmy pojęcie głównej strategii ewaluacji, jaką jest *interpretacja*:

**Interpretacja** zjawisk empirycznych to proces odtwarzania i wyjaśniania symptomów wartości na podstawie charakterystycznych lub nietypowych oznak, wykrytych poprzez pomiar i statystyczne przetworzenie wskaźników.

Ta ogólna definicja wymaga doprecyzowania czynnościowego, przy czym owe czynności dotyczą przede wszystkim rozumowania. Taktyką interpretacji jest:

- rozdzielenie złożonych problemów badawczych na eksplikacje cząstkowe;
- rozwikłanie tego, co jest surowcem, półproduktem, produktem finalnym;
- objaśnienie przyczyn zwłaszcza nieswoistych właściwości statystycznych;
- orzekanie co do istotności zróżnicowania rezultatów poszczególnych prób;
- wartościowanie zjawisk empirycznych poprzez porównywanie z normami;
- ocenianie zgodności z modelem relacji między komponentami indykatum.

W rezultacie interpretacji formułuje się wykładnie deskryptywne oraz wnioski formatywne. Stawia się diagnozę wartościującą drobny wycinek rzeczywistości, jego stany chwilowe i zmienność oraz prognozuje potencjalne stany przyszłe.

Dowolna interpretacja zjawisk kryje w sobie niebezpieczeństwo zróżnicowanego odczytywania znaczeń przez różne osoby, mimo tych samych oznak. Zawiłość badanych procesów społecznych powoduje, że każdy z ewaluatorów nieco inaczej interpretujących fakty empiryczne może mieć w pewnym sensie rację. Wszystko bowiem zależy od tego, jakie przyjmie się reguły rozumowania i kryteria wartościowania. Twórca narzędzia badawczego, zwłaszcza skalowanego, powinien dostarczyć nie tylko schemat interpretacji sensu zmiennych, lecz także powinien opisać najbardziej charakterystyczne oznaki empiryczne oraz udostępnić normy parametrów oczekiwanych, standaryzujących skalę. Dzięki standaryzacji tworzy się układ odniesienia dla innych ewaluatorów.

Normy standaryzujące Kwestionariusz Kultury Informatycznej

Tab. V

Symbole miar	Oczekiwane wartości liczbowe parametrów z podziałem na przedziały									
	dla populacji i ogółu $n \geq 300$					dla grup badawczych $n \approx 60$				
	bardzo niskie	poniżej niskie	poniżej średnie	powyżej wysokie	bardzo wysokie	bardzo niskie	poniżej niskie	poniżej średnie	powyżej wysokie	bardzo wysokie
Q (wy)	<1,05	<1,2	1,35	>1,5	>1,65	<0,85	<1,1	1,35	>1,6	>1,85
d (wy)	<0,67	<0,84	1,0	>1,16	>1,33	<0,5	<0,75	1,0	>1,25	>1,5
d (cz)	<1,1	<1,3	1,5	>1,7	>1,9	<0,9	<1,2	1,5	>1,8	>2,1
N (wy)	<3,55	<3,7	3,85	>4,0	>4,15	<3,35	<3,6	3,85	>4,1	>4,35
N (cz)	<3,3	<3,55	3,85	>4,15	>4,4	<3,2	<3,5	3,85	>4,2	>4,5
s (wy)	< -0,01	< -0,03	-0,05	> -0,09	> -0,15	< -0,01	< -0,03	-0,05	> -0,11	> -0,2
s (ws)	< -0,05	< -0,12	-0,24	> -0,35	> -0,5	< -0,05	< -0,12	-0,24	> -0,5	> -0,8
-b (ws)	< -0,1	< -0,2	-0,33	> -0,52	> -0,8	< -0,1	< -0,2	-0,35	> -0,62	> -1,0
$\pm b$ (wy)	<-0,04	<-0,02	0	>0,02	>0,04	<-0,05	<-0,025	0	>0,025	>0,05
$\pm b$ (skł)	<-0,16	<-0,08	0	>0,08	>0,16	<-0,2	<-0,1	0	>0,1	>0,2
c (skł)	<0,45	<0,54	0,63	>0,72	>0,81	<0,38	<0,50	0,63	>0,76	>0,88
$\pm c$ (skł)	<-0,02	<0	0,02	>0,06	>0,1	<-0,05	<-0,025	0,02	>0,11	0,2
$\pm i$ (wy)	<0	<0,015	0,03	>0,045	>0,06	<-0,025	<0	0,025	>0,05	>0,08
$\pm i$ (cz)	<-0,03	<0	0,025	>0,055	>0,09	<-0,035	<0	0,025	>0,06	>0,1
$\pm r$ (wy)	<0,5	<0,55	0,6	>0,65	>0,7	<0,4	<0,5	0,6	>0,7	>0,8
$\pm r$ (skł)	<0,35	<0,4	0,45	>0,5	>0,55	<0,2	<0,3	0,42	>0,54	>0,66
$\pm t$ (wy)	<0	<0,03	0,06	>0,1	>0,15	<0	<0,03	0,08	>0,14	>0,2
$\pm t$ (skł)	<-0,12	<-0,04	0,04	>0,12	>0,2	<-0,27	<-0,12	0,03	>0,18	>0,33
$\pm e$ (wy)	<0	<0,03	0,06	>0,1	>0,15	<0	<0,03	0,09	>0,15	>0,2
$\pm e$ (skł)	<-0,07	<0	0,07	>0,145	>0,22	<-0,17	<-0,06	0,05	>0,16	>0,27
č (ws)	<0,6	<0,65	0,7	>0,75	>0,8	<0,5	<0,6	0,7	>0,75	>0,8
š(wy) 5%	<0,07	<0,09	0,11	>0,13	>0,15	<0,13	<0,19	0,25	>0,31	>0,37
š(cz) 5%	<0,13	<0,15	0,17	>0,19	>0,21	<0,23	<0,30	0,38	>0,46	>0,53
â (wy)	<0,95	<0,96	0,97	>0,98	>0,99	<0,93	<0,95	0,97	>0,98	>0,99
â (cz)	<0,93	<0,945	0,96	>0,975	>0,99	<0,9	<0,93	0,95	>0,97	>0,99
ř (ws)	<0,75	<0,775	0,8	>0,825	>0,85	<0,7	<0,75	0,8	>0,825	>0,85

(wy) – wyniki, (cz) – czynniki, (skł) – składniki, (ws) – wskaźniki

Normy standaryzujące skalę tworzy się poprzez estymację parametryczną na podstawie skumulowanych rezultatów badań licznych i reprezentatywnych prób, w tym także różnych populacji. Wartości oczekiwane mają taką cechę, że w miarę zwiększania liczebności prób ulegają zawężeniu przedziały ufności. Z tego powodu w tabeli V podałem przedziały oszacowane dla populacji oraz dla grup o zalecanej liczbie około 60 respondentów. Tabela zawiera jedynie normy na takim poziomie agregacji zmiennych, na jakim dany parametr jest najbardziej istotny. Przykładowo wartość średnia *Jakości* [Q] ma sens jako norma wyniku ogólnego, a *Rozziew* [-b] ma sens głównie przy wskaźnikach.

Każda ze statystyk sama w sobie niesie specyficzną właściwość, jednak w połączeniu z sensem zmiennych i z konkretnym pomiarem powstaje nowa wartość empiryczna, zwana *oznaką*. Interpretacja oznak jest odczytywaniem merytorycznego znaczenia komponentów indykatorum równocześnie z analizą statystycznej istoty parametrów. Okazuje się, że pewne układy zależnościowe i zmiennościowe oznak są w pomiarach na tyle powtarzalne, że można uznać je za zjawiska charakterystyczne dla obranej metodologii badawczej. Stanowią one z jednej strony niezwykle ważne ogniwa modelu teoretycznego, będącego podstawą sprawdzania trafności i rzetelności badań, a z drugiej strony – jako swoiste wzorce wskazują na to, jakich zjawisk należy poszukiwać. Wszystkie te zjawiska empiryczne objaśniam w dalszej części bardzo szczegółowo po to, aby naprowadzić zainteresowanych na strategię poprawnej interpretacji.

### **Względność wypowiedzi – poziomy odniesienia**

Potencjalnym błędem interpretacji może być bezkrytyczne odczytywanie poziomów wypowiedzi i proste przyjmowanie ich jako natężeń badanej cechy. Intensywność wypowiedzi zależy od treści stwierdzeń i w pierwszej kolejności wynika ze specyfiki wskaźników skali. Stwierdzenia łatwiejsze do wyrażenia opinii (np. retrospekcyjne) są akceptowane bądź odrzucane bardziej zdecydowanie niż te trudniejsze (np. futurospekcyjne). Poziomy cech oszacowuje się w ujęciu względnym, jako wyższe bądź niższe od wyznaczonych standardów, lub jako przyrastające bądź opadające w czasie. Średnie poziomy każdego ze wskaźników dążą (po uśrednieniu dużej liczby prób) do wartości granicznych, które stają się bazowymi i normatywnymi kryteriami dla wybranej populacji. Standardy te ulegają przemianom wskutek rozwoju społeczeństwa, jednak te powolne przesunięcia nie przeszkadzają w szacowaniu poziomów chwilowych. Okazuje się bowiem, że o ile różnice pomiędzy pierwszym i drugim pomiarem są znaczące, o tyle w drugim pomiarze następuje zbliżenie poziomów do *Normy* (autonormalizacja). Zmiany poziomów są wręcz pożądane jako wartość, pamiętać zatem należy o potrzebie niewielkiego korygowania norm co kilka lat. Najlepszy sposób to pomiary cykliczne i uaktualnianie danych źródłowych.

Przyjrzyjmy się sposobom wartościowania poziomów poprzez *odniesienie*, posługując się tabelą VI zawierającą średnie oczekiwane dla trzech populacji. Modalne wieku respondentów w 9-miesięcznym cyklu dwupomiarowym zmieniły się odpowiednio dla prób w zakresie:  $X=19\div 20$ ,  $Y=16\div 17$ ,  $Z=13\div 14$  lat. Interpretacja oznak wynikających ze zróżnicowania poziomów ocen polega na:

- porównaniu norm dla różnych populacji, dobranych przykładowo ze względu na wiek (tabela VI – kolumny 1, 10 i 15 objaśnione w punktach a, c i f);
- porównaniu poziomów osiągniętych w dwóch pomiarach dystansowych przez te same grupy osób (kolumny 3, 4 i 19 objaśnione w punktach b i g);
- odniesieniu stanów motywacji do norm neutralnego wskaźnika buforowego (porównaj kolumnę 11 z kolumną 12 objaśnioną w punkcie d);
- porównywaniu ze sobą poziomów wskaźników tworzących pary kontrolne (kolumny nieparzyste z parzystymi, np. 13 i 14 objaśnione w punkcie e);
- porównywaniu rezultatów uzyskanych przez różne grupy, poddanych temu samemu procesowi u jednego wykładowcy w ciągu roku i kolejnych latach;
- porównywaniu grup homogenicznych, np. kontrolnej z eksperymentalną, na którą celowo oddziałuje się kontrolowanym czynnikiem.

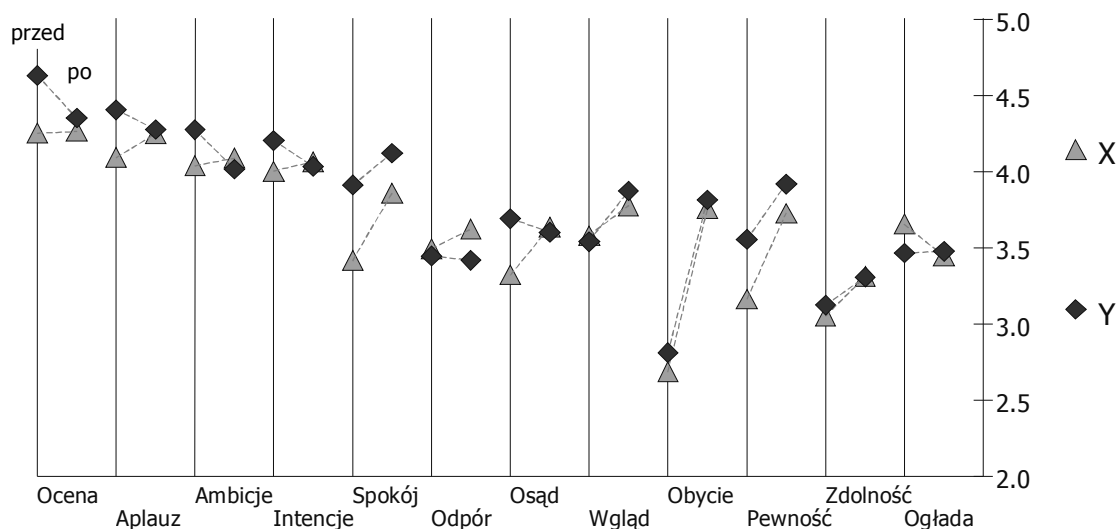
Baza [B'] i Norma [N''] wskaźników u studentów X, licealistów Y i gimnazjalistów Z Tab. VI

i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
$N_X'$	4,1	4,1	3,0	2,4	2,8	3,5	3,0	3,1	3,9	3,0	3,6	4,5	4,5	4,0	2,7	4,0	3,4	3,7	3,7	3,6	3,8	4,2	4,1	2,9
$N_X''$	4,2	4,3	3,8	3,7	3,7	3,8	3,3	3,3	4,3	3,4	3,8	4,4	4,5	4,1	3,0	4,2	3,9	3,7	3,5	3,4	3,9	4,3	4,1	3,1
$N_Y'$	4,5	4,3	2,9	2,7	3,5	3,6	2,9	3,3	4,3	3,6	3,9	4,6	4,8	4,5	3,4	3,9	3,6	3,5	3,8	3,2	4,1	4,3	4,0	2,9
$N_Y''$	4,4	4,2	3,9	3,8	4,1	3,8	3,3	3,4	4,6	3,7	3,7	4,4	4,5	4,2	3,4	3,8	3,9	3,8	3,4	3,6	3,8	4,3	4,1	2,8
$N_Z'$	4,8	4,3	2,7	3,0	3,8	3,8	3,1	3,0	4,3	4,0	3,7	4,7	4,6	4,2	4,5	3,7	3,6	3,0	3,8	3,4	4,1	3,9	3,9	3,2
$N_Z''$	4,7	4,3	3,9	4,0	4,3	4,1	3,4	3,3	4,6	4,2	3,9	4,6	4,6	4,3	4,4	3,9	4,0	3,4	3,3	3,7	4,0	4,1	4,0	3,1
	↑		↑	↑						↑		↑	↑		↑				↑					↑
	a		b	b						c		d	e		f				g					h

- Z tabeli odczytać można i zinterpretować następujące zjawiska globalne:
- a. Zależność emocjonalnego oceniania procesu komputeryzacji od wieku respondentów – wyższy poziom akceptowania przez najmłodszą zbiorowość uczniów Z, a wyraźnie niższa akceptacja przez zbiorowość studentów X.
  - b. Duże przyrosty norm w zakresie nabytych umiejętności, pomimo bardzo niskich (zwykle zaniżanych) poziomów deklarowanych na początku zajęć.
  - c. Niższy poziom bezstresowości u studentów świadczą o silniejszym poczuciu przez nich *informatycznej alienacji* i zażenowaniu z powodu niewiedzy.
  - d. Poziomy motywacji do uczenia się o komputerach (i:12) zdecydowanie wyższe od ambicji uczenia się ogólnego, nie związanego z indykatem (i:11).
  - e. Zawyżone poziomy oszacowania przydatności kursów komputerowych, co wynika z odniesienia do wskaźnika 14, współtworzącego parę kontrolną.
  - f. Niesłuszne poglądy studentów, którzy sytuują używanie komputerów tylko na kursach informatyki i na tym tle wysoce trafne osądy gimnazjalistów.

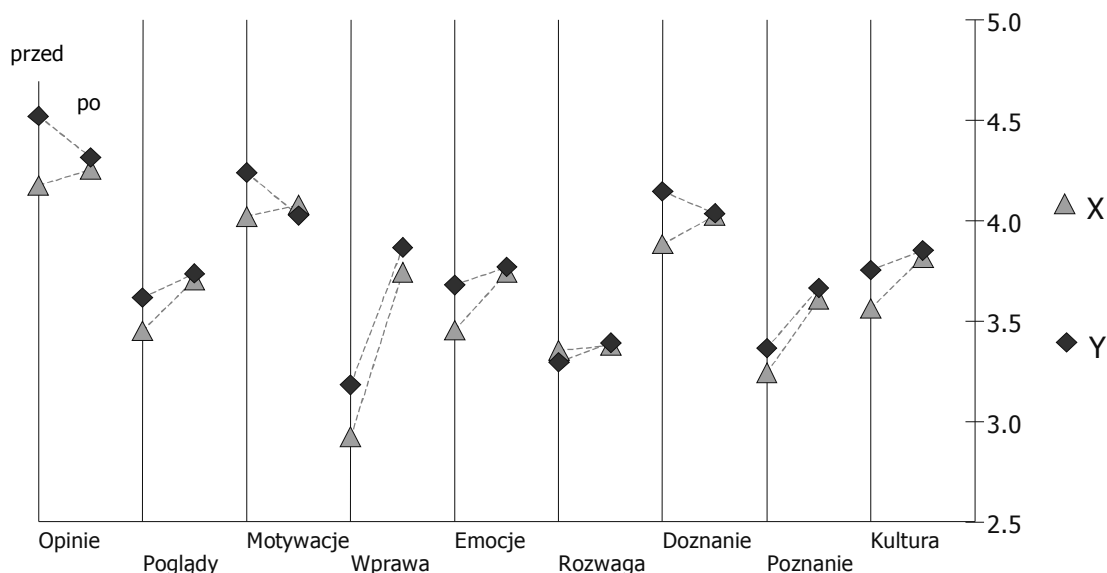
- g. Istotne spadki w drugim pomiarze jednego z ważniejszych komponentów uświadomienia, jakim jest roztropność działań przy komputerze.
- h. Trudność interpretacyjna – czy jest to symptom zbyt niskiej odporności na potencjalne zagrożenia, jakie może stanowić globalna komputeryzacja, czy raczej symptom posiadania świadomości potencjalnie realnych zagrożeń; tego typu wątpliwość może skłonić do modyfikacji bodźca-stwierdzenia.

Bardzo istotne następne zjawisko empiryczne stanowi dowód na to, jak ważną funkcję pełni wizualizacja parametrów. Bez wykresu trudniej byłoby odkryć charakterystyczną tendencję do normowania poziomów (ryc. 72 i 73). *Autonormalizacja* jest empiryczną osobliwością efektu pomiarów za pomocą skal dwuważonych, ujawniająca się jako zbieżność średnich ocen i dążność w drugim pomiarze do samo formujących się poziomów aksjonormatywnych. Mimo odmiennych populacji, właściwości wskaźników powodują, że poziomy wartości komponentów indykatum zbieżają do wypośrodkowanego optimum. Wynika to z procesu świadomościowego dojrzewania respondentów i stanowi potwierdzenie, jak bardzo konieczny w ewaluacji splotowej jest drugi pomiar.



Ryc. 72. Autonormalizacja poziomu składników w trakcie procesu edukacji – tendencja do zbieżności norm dla obu prób X i Y

Na wykresie widoczne jest zjawisko wytyczania norm wypośrodkowanych, wspólnych dla obu populacji. Najbardziej wyrazista jest konwergencja takich zmiennych, jak: *Ocena*, *Aplauz*, *Ambicje*, *Intencje*, *Osąd*, *Zdolność* i *Ogłada*. Niższy poziom *Spokoju* i *Pewności* u studentów bierze się z poczucia, że w ich wieku powinni już potrafić więcej. Z kolei u uczniów obniżający się poziom *Odporu* świadczy o tym, że na zajęciach w szkole nie formuje się odporności na uzależnienia od komputera. Duże nachylenia linii przyrostów wskazują na potencjalną możliwość dalszej zmiany norm danego składnika (np. *Obycia*).



Ryc. 73. Autonormalizacja poziomów czynników i wyników ogólnych – tendencja do zbieżności norm dla obu próbek X i Y

Zjawisko normowania poziomów ujawnia się także na wyższym poziomie agregacji zmiennych. Tu szczególnie wyraźnie widoczna jest tendencja do tonowania nadmiernie rozbudzonych emocji 'przed'. W konsekwencji czynniki doznaniowe zmiernieją do zgodnego z modelem stanu harmonii. Natomiast czynniki poznawcze w obu populacjach zmiernieją w górę, co jest tendencją oczekiwaną z perspektywy celów edukacyjnych. Poziom *Poznania* nie osiągnął jeszcze pułapu, który można byłoby uznać za normę docelową. *Norma* w tej połowie skali ulega szybszej dezaktualizacji. Z tego powodu normy i poziomy cech niemierzalnych są jedynie półproduktami służącymi do dalszej analizy.

### Rozproszenie oznak – analiza rozkładowa

Od jednopunktowej, pozycyjnej analizy poziomów średnich, cenniejszą jest trójpunktowa analiza rozkładowa. Oprócz poziomów uwzględnia się w niej także zróżnicowanie podmiotów tworzących grupy badawcze. Zróżnicowanie wyrażane jest przez miary rozproszenia, a dodatkowo z wartości skumulowanych wyznaczyć można rozkład ilustrujący skupiska cech o tych samych poziomach. Rozkłady różnią się kształtami, jednak w analizie trójpunktowej bierze się pod uwagę jedynie szerokość i skośność. Do tego celu wystarczają parametry dewiacji i asymetrii. Podstawą analizy są *Odchylenia realne* [-d; +d], które wyznaczają wokół poziomu średniego M dwa dodatkowe poziomy H i L. Pomiędzy tymi poziomami mieści się główne skupisko około 2/3 rezultatów, co oznacza kwalifikowaną większość, zatem rezultat przedziałowy jest bardziej wiarygodny, niż sama wartość średnia. Ponadto z różnicy pomiędzy odchyleniami wnioskować można o symetrii bądź asymetrii rozkładu empirycznego.



Odchylenia realne i Skośność  $[\pm s]$  rozkładu wskaźników po scaleniu prób X i Y Tab. VII

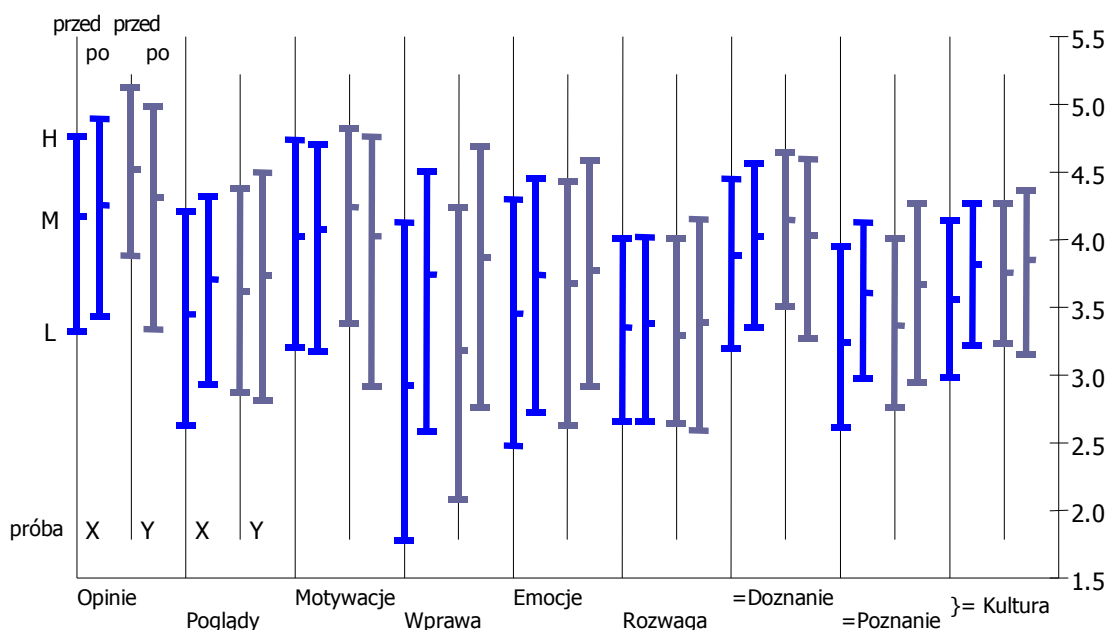
i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
H'	5,1	5,0	4,4	4,0	4,6	4,8	4,1	4,3	5,0	4,5	4,8	5,5	5,5	4,9	4,6	5,0	4,7	4,9	4,7	4,5	4,8	4,9	4,9	4,5
+d'	0,8	0,8	1,5	1,5	1,5	1,2	1,1	1,1	1,0	1,3	1,0	1,0	0,8	0,7	1,6	1,0	1,2	1,3	1,0	1,1	0,8	0,7	0,8	1,6
M'	4,3	4,2	2,9	2,5	3,1	3,6	3,0	3,2	4,0	3,2	3,8	4,5	4,7	4,2	3,0	4,0	3,5	3,6	3,7	3,4	4,0	4,2	4,1	2,9
-d'	-1,9	-1,8	-1,6	-1,5	-1,9	-1,0	-1,6	-1,9	-1,7	-2,0	-1,3	-0,8	-1,0	-1,4	-1,6	-1,5	-1,4	-1,4	-1,2	-1,7	-1,6	-1,3	-1,9	-1,4
L'	2,4	2,4	1,3	1,0	1,2	2,6	1,4	1,3	2,3	1,2	2,5	3,7	3,7	2,8	1,4	2,5	2,1	2,2	2,5	1,7	2,4	2,9	2,2	1,5
$\pm s'$	-0,4	-0,4	-0,1	0,0	-0,1	0,1	-0,1	-0,2	-0,3	-0,2	-0,1	0,1	-0,1	-0,3	0,0	-0,2	-0,1	0,0	-0,1	-0,2	-0,3	-0,3	-0,4	0,1
H''	5,1	5,0	4,8	4,7	4,9	4,9	4,3	4,4	5,1	4,8	4,7	5,0	5,5	4,8	4,6	5,0	4,7	5,0	4,4	4,6	4,8	4,9	4,9	4,5
+d''	0,8	0,8	0,9	1,0	1,0	1,1	1,0	1,0	0,7	1,3	0,9	0,6	1,0	0,7	1,4	0,9	0,8	1,3	1,0	1,1	1,0	0,6	0,8	1,5
M''	4,3	4,2	3,9	3,7	3,9	3,8	3,3	3,4	4,4	3,5	3,8	4,4	4,5	4,1	3,2	4,1	3,9	3,7	3,4	3,5	3,8	4,3	4,1	3,0
-d''	-2,0	-2,0	-1,4	-0,9	-1,9	-1,1	-2,0	-2,0	-2,0	-1,9	-1,5	-1,9	-1,1	-1,6	-1,9	-1,8	-1,5	-1,6	-2,0	-1,9	-1,6	-1,6	-2,1	-1,6
L''	2,3	2,2	2,5	2,8	2,0	2,7	1,3	1,4	2,4	1,6	2,3	2,5	3,4	2,5	1,3	2,3	2,4	2,1	1,4	1,6	2,2	2,7	2,0	1,4
$\pm s''$	-0,4	-0,5	-0,2	0,0	-0,3	0,0	-0,3	-0,3	-0,5	-0,2	-0,2	-0,5	-0,1	-0,4	-0,1	-0,3	-0,3	-0,1	-0,4	-0,3	-0,3	-0,4	-0,4	0,0
	↑	↑		↑		↑						↑	↑						↑				↑	↑
	a	a		b		c						d	e						f				g	h

W tabeli VII warto zwrócić uwagę na następujące właściwości rozkładów:

- Znaczna lewoskośność rozkładów oznak afektywnych, dotyczących oceny komputeryzacji, to efekt skupienia wielu wypowiedzi na poziomie wyraźnie akceptującym i tylko nikłej liczby respondentów dezawuuujących proces.
- Symetria rozkładów oznaki kognitywnej, dotyczącej samooceny umiejętności, świadczy o reprezentatywnym doborze próby, z cechami zdolności zróżnicowanymi *gaussowsko*. Symetrii tej nie zakłóciło nawet bardzo duże przesunięcie rozkładu i zmniejszenie rozrzutu w drugim pomiarze.
- Minimalnie prawoskośny lub symetryczny rozkład znamionuje wskaźniki futurospekcyjne, próbujące stany przyszłe (por. wskaźnik 24).
- Zmiana kształtu rozkładu na silnie lewoskośny świadczy o tym, że tylko nieliczni respondenci w trakcie kursów zaspokoili aspiracje uczenia się o komputerach. Bez analizy rozkładowej trudno byłoby wykryć tego typu niekorzystne zjawisko, kiedy to średnia prawie się nie zmienia, a jednak następuje wyraźne obniżenie motywacji, zwłaszcza na poziomie L.
- Wskaźnik 13 (por. 12) ma taką właściwość, że rozkład sytuje się na zbyt wysokim pułapie, sięgającym kresu górnego. Z analizy wynika jednak, że nie trzeba zmieniać treści bodźca-stwierdzenia. Byłoby to konieczne, gdyby w drugim pomiarze rozkład przesunął się jeszcze bardziej w górę.
- Lewoskośność w drugim pomiarze i przesunięcie rozkładu w dół świadczą o niekorzystnym zjawisku, objawiającym się tym, że w trakcie poznawania słabną odruchy hamujące działania pochopne wobec komputera (zwykle do pierwszego przykrego doświadczenia z utratą ważnych danych).
- Silna lewoskośność charakterystyczna dla wskaźników introspekcyjnych, próbujących emocjonalne stany własne respondentów. Z asymetrii rozkładu wynika, że osób obawiających się uzależnień jest stosunkowo mało.

h. Rozkład lekko prawoskośny bądź symetryczny charakteryzuje wskaźniki ekstrasppekcyjne, próbujące opinie o stanach innych osób. Nadto kwestia ta dotyczy bardzo trudnego prognozowania stanów przyszłych.

Trójpunktową analizę rozkładową korzystnie jest prowadzić na wykresie odcinkowym poziomów HML. Szacuje się wówczas trzy poziomy na skali ocen, wyznaczające frakcje badanych, z podziałem na podgrupy o wysokich, średnich i niskich wartościach cech. Rozpiętość pomiędzy poziomami H-L oznacza rozrzut, jaki dzieli frakcję 'lepszych' od 'gorszych' pod względem *InfoKultury*. Różnice długości odcinków HM względem ML świadczą o skośności rozkładów.



Ryc. 74. Poziomy HML czynników i wyników porównawczo prób X i Y

Również i tu interpretację prowadzi się poprzez różnego rodzaju porównywanie. Zestawienie rezultatów studentów X i licealistów Y umożliwia wykrycie różnic między populacjami; przedstawienie poziomów 'przed' i 'po' ilustruje zmiany, a bliskie ułożenie par czynników komplementarnych ułatwia ocenę wpływu komponentów afektywnych i kognitywnych na ogólny poziom indykatoru.

Nadmiernie ekspresyjne u licealistów *Opinie* 'przed' uległy stonowaniu. *Poglądy* poprawiły się, z wyjątkiem dolnej frakcji L w próbie Y. Zdecydowanie spadły *Motywacje* licealistów, a u studentów nastąpiło jedynie odkształcenie symetrii rozkładu. Bardzo wyraźnie we wszystkich frakcjach HML wzrosły poziomy *Wprawy*, lecz nadal zróżnicowanie umiejętności jest bardzo duże. Poprawiły się stany *Emocji*, głównie dzięki wyzbyciu się stresu. Dość niskie poziomy *Rozwagi* nie uległy poprawie u studentów, natomiast u licealistów udało się jeszcze tę cechę kształtować, z wyjątkiem frakcji L. W sumie próba X osiągnęła harmonijny rozwój *Doznanie* i *Poznanie*, a w próbie Y łączny spadek doznań był jednak zbyt duży, co dało w rezultacie spadek *InfoKultury* frakcji L.

## Miarodajność wskaźników – strefa zgodności

Ocena miarodajności wskaźników jest ważną procedurą na etapie konstruowania nowego narzędzia, jednak wzbogaca także wnioskowanie co do rezultatów ewaluacji. Ową *miarodajność* rozumieć należy jako zdolność skali pomiarowej do pozyskiwania dzięki niej takich wartości parametrów statystycznych, które ujawniają istotne zjawiska empiryczne. Narzędzie badawcze musi różnicować respondentów tak, aby nie wypowiedzieli się identycznie, oraz tak, żeby nie było zbyt dużo wypowiedzi 'brak zdania'. Większość musi być jednak zgodna co do walencyjnej wartości elementarnych tez. Skala wag wypowiedzi zostaje podzielona na dwie strefy: zgodności i niezgodności, po czym analizuje się proporcje usytuowania gęstości wypowiedzi w obu tych strefach.

Do analizy parametrycznej wykorzystuje się miary dyspersji i dysjunkcji oraz wyznaczniki jakości pomiarowej. Parametrem dyspersji jest *Rozrzut* [d] określający zróżnicowanie wypowiedzi, a parametrem dysjunkcji *Rozziew* [-b] wyrażający niezgodność opinii. Na bazie tych miar tworzony jest wyznacznik *Konkluzyjności* [č], świadczący o mocy wnioskowania z każdego wskaźnika. Pomocny jest również parametr *Ważkości* [W], oznaczający intensywność wypowiedzi, a także *Baza* [B'] i *Norma* [N''], wytyczające standardy 'przed' i 'po'. Wartości tych miar odnosi się do norm standaryzujących skalę (str. 220).

Rozrzut [d], Rozziew [-b] i Konkluzyjność [č] wskaźników po scaleniu prób X i Y Tab. VIII

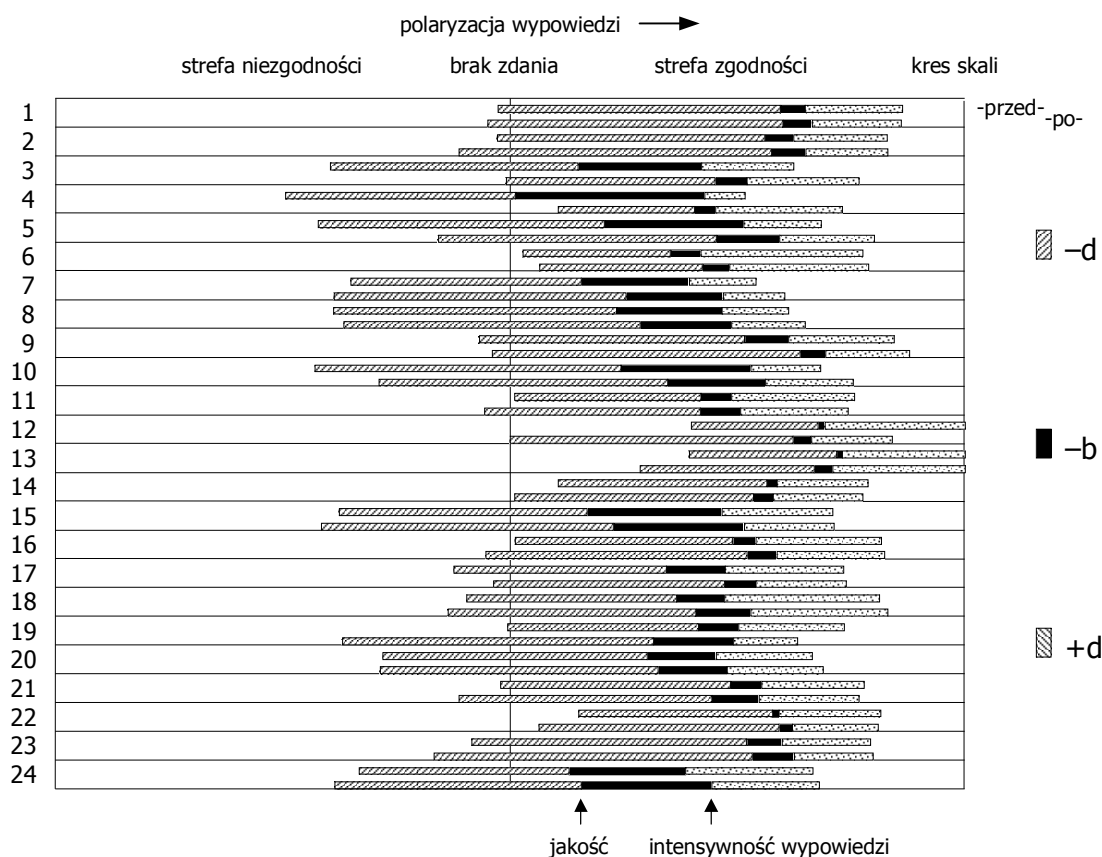
i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
B'	4,3	4,2	2,9	2,5	3,1	3,6	3,0	3,2	4,0	3,2	3,8	4,5	4,7	4,2	3,0	4,0	3,5	3,6	3,7	3,4	4,0	4,2	4,1	2,9	
d'	2,7	2,6	3,1	3,0	3,3	2,2	2,7	3,0	2,7	3,3	2,2	1,8	1,8	2,0	3,3	2,4	2,6	2,7	2,2	2,8	2,4	2,0	2,6	3,0	
-b'	-0,2	-0,2	-0,8	-1,2	-0,9	-0,2	-0,7	-0,7	-0,3	-0,9	-0,2	0,0	0,0	-0,1	-0,9	-0,2	-0,4	-0,3	-0,3	-0,5	-0,2	0,0	-0,2	-0,8	
N''	4,3	4,2	3,9	3,7	3,9	3,8	3,3	3,4	4,4	3,5	3,8	4,4	4,5	4,1	3,2	4,1	3,9	3,7	3,4	3,5	3,8	4,3	4,1	3,0	
d''	2,7	2,8	2,3	1,9	2,9	2,2	3,0	3,0	2,8	3,1	2,4	2,5	2,1	2,3	3,4	2,6	2,3	2,9	3,0	2,9	2,6	2,2	2,9	3,2	
-b''	-0,2	-0,2	-0,2	-0,1	-0,4	-0,2	-0,6	-0,6	-0,2	-0,6	-0,3	-0,1	-0,1	-0,1	-0,9	-0,2	-0,2	-0,4	-0,5	-0,5	-0,3	-0,1	-0,3	-0,9	
W	4,5	4,4	3,9	3,8	4,2	3,8	3,8	3,9	4,5	4,1	4,0	4,5	4,7	4,2	4,0	4,2	4,0	4,0	4,0	4,0	3,9	4,1	4,3	4,3	3,7
č	,71	,70	,56	,38	,59	,60	,51	,55	,70	,55	,59	,67	,61	,61	,52	,68	,61	,70	,59	,60	,64	,63	,70	,47	
	↑			↑			↑					↑	↑		↑									↑	
	a			b			c					d	e		f									g	

W tabeli VIII warto zwrócić uwagę na następujące cechy wskaźników:

- Wysoka *Konkluzyjność*, mimo że treść bodźca-stwierdzenia nr 1 jest nieco zbyt łatwo akceptowana (świadczy o tym wysoki poziom *Ważkości*  $W=4,5$ ). W obu pomiarach identyczne, optymalnie duże *Rozrzuty* i małe *Rozziewy*.
- Zbyt niska *Konkluzyjność*, wynikająca z dużego *Rozziewu* w pierwszym pomiarze (brak umiejętności komputerowych) i małego *Rozrzutu* w drugim pomiarze (nadrobienie braków dzięki zajęciom). Jednak ze względu na to, że bodziec-stwierdzenie nr 4 próbuje poziomy umiejętności, które z biegiem lat będą narastać – zdecydowałem o pozostawieniu tego wskaźnika, mimo że w czasie weryfikacji nie spełniał on kryterium doboru do skali.

- c. *Konkluzyjność* ledwo spełniająca kryterium (0,5) wynika z nadmiernego *Rozziewu* (porównaj ze wskaźnikiem 1), lecz przede wszystkim bierze się z „uproblemowienia” bodźca-stwierdzenia, stawiającego respondentów przed trudnym wyborem (świadczy o tym niski poziom *Ważkości*  $W=3,8$ ).
- d. Mały *Rozrzut* wypowiedzi oraz zerowy *Rozziew* charakteryzują wskaźniki z łatwym wyborem zgodnych opcji (por. z 13). *Konkluzyjność* jest jednak wysoka, gdyż zwiększył się *Rozrzut* wypowiedzi w drugim pomiarze.
- e. Zawyżony poziom *Ważkości* wypowiedzi świadczy tutaj o pewnej kurtuazji wobec prowadzącego zajęcia i o przecenianiu roli kursów komputerowych.
- f. Empirycznie wyznaczona *Konkluzyjność* jest tu bezzasadnie niska – nie da się bowiem wyjaśnić, skąd aż tak duży *Rozziew* wypowiedzi w oczywistej kwestii, jaką jest używanie komputera nie tylko na kursach Informatyki.
- g. Zbyt niska *Konkluzyjność*, lecz o pozostawieniu wskaźnika zdecydowała chęć badania świadomości zagrożeń wynikających z globalnej komputeryzacji.

Miarodajność wskaźników można szacować na pasmowym wykresie dyspersji i dysjunkcji, podzielonym na dwie strefy. Rozpiętość pasemek wyraża *Rozrzut* wypowiedzi, a długość pasemek czarnych ilustruje wielkość *Rozziewu*. Najlepiej jest, gdy całe pasmo jak najszerzej wypełnia strefę zgodności, a pole czarne ma jak najmniejszą powierzchnię. Oznacza to wysoką *konkluzyjność*.



Ryc. 75. Rozrzut wskaźników połączonych prób X i Y względem strefy zgodności; odchylenie ujemne [-d], rozziew [-b], odchylenie dodatnie [+d]

### Koherentność oznak – komparacja wskaźników

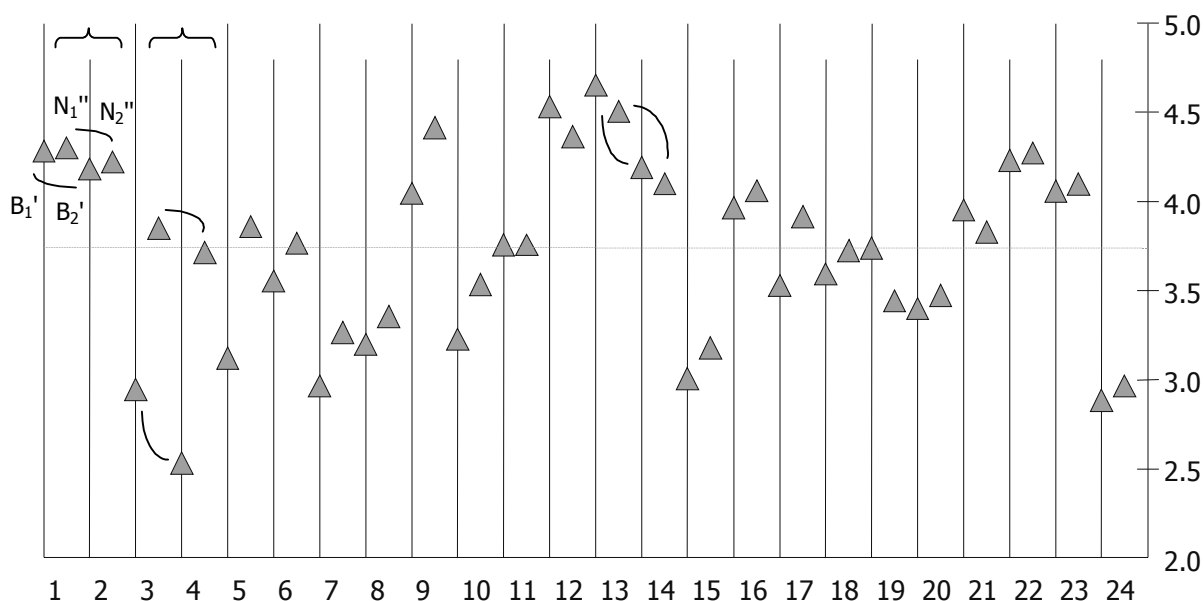
W procesie wartościowania przydatna jest ocena wewnętrznej spójności wypowiedzi. Polega ona na sprawdzaniu kolejnych przypadków w taki sposób, ażeby ustalić, na ile wypowiedzi tego samego respondenta różniły się wobec kontrolnych par wskaźników. Po uśrednieniu daje to parametr *Spójności* [c]. Trzeba jednak pamiętać, że pary wskaźników celowo próbują zróżnicowane aspekty czasoprzestrzenne, dlatego nie można oczekiwać spójności zupełnej. Parametr ten jest skalibrowany, stąd jego wartości 0,5; 0,67 i 0,83 wyrażają odpowiednio przeciętne zróżnicowanie 1,5 działki, 1 działkę i 0,5 działki skali. Podane w tabeli wartości empiryczne spójności wydają się niewysokie, jednak należy wziąć pod uwagę, że pojedyncze przypadki wyboru dwóch wypowiedzi mogą skrajnie różnić się aż o 6 działek, co znacząco obniża spójność średnią. Z tego powodu wartościowsza jest ocena *Zbieżności* [ $\pm c$ ], która wskazuje na proces wzmacniania, ewentualnie osłabiania zwartości wypowiedzi w czasie.

Spójność [c] par kontrolnych i Zbieżność [ $\pm c$ ] wypowiedzi z prób X, Y i łącznie Tab. IX

i:	Ocena	Aplauz	Ambicje	Intencje	Spokój	Odpór	Osąd	Wgląd	Obycie	Pewność	Zdolność	Ogłada
	13×14	1×2	11×12	21×22	9×10	23×24	15×16	17×18	3×4	5×6	7×8	19×20
$c_X'$	0,73	0,68	0,65	0,73	0,50	0,51	0,48	0,52	0,67	0,60	0,60	0,59
$c_X''$	0,75	0,71	0,69	0,76	0,55	0,55	0,51	0,50	0,80	0,72	0,61	0,51
$\pm c_X$	0,02	0,03	0,04	0,03	0,05	0,04	0,03	-0,02	0,14	0,11	0,01	-0,08
$c_Y'$	0,78	0,67	0,70	0,75	0,57	0,51	0,62	0,48	0,67	0,65	0,58	0,51
$c_Y''$	0,73	0,69	0,67	0,69	0,56	0,45	0,60	0,51	0,78	0,74	0,64	0,51
$\pm c_Y$	-0,06	0,02	-0,03	-0,06	-0,01	-0,06	-0,02	0,03	0,11	0,09	0,06	0,00
$c'$	0,75	0,68	0,67	0,74	0,53	0,51	0,54	0,50	0,67	0,62	0,59	0,56
$c''$	0,74	0,70	0,68	0,73	0,56	0,51	0,55	0,51	0,79	0,72	0,63	0,51
$\pm c$	-0,01	0,02	0,01	-0,01	0,02	0,00	0,01	0,01	0,12	0,10	0,03	-0,05
	↑ a		↑ b	↑ c		↑ d	↑ e	↑ f	↑ g	↑ g		↑ h

- Z tabeli IX odczytać można różne symptomy koherencji par wskaźników:
- Ocenianie afektywne zajęć jest wysoce spójne, lecz u licealistów rozbieżne.
  - Gotowość poznawcza jest mniej spójna niż gotowość do działania (por. c); obie deklaracje dyspozycji studentów są zbieżne, a u licealistów rozbieżne.
  - Bardzo niska spójność w obu próbach oznacza, że respondenci sami nie są wrażliwi na uzależnienie, lecz dostrzegają tę słabość u innych, przy czym także i tu wypowiedzi licealistów są w drugim pomiarze bardzo rozbieżne.
  - Osąd przydatności komputera w pierwszym pomiarze był o wiele bardziej spójny u licealistów niż u studentów, co budzi wątpliwość interpretacyjną, ponieważ trafność oceny zastosowań (por. f) jest w próbie Y niespójna.
  - Bardzo wysoka zbieżność obu składników testujących nabycie wprawy.
  - Niska spójność i zaskakująco silna w próbie X rozbieżność wypowiedzi wobec par kontrolnych, testujących roztropność zachowań przy komputerze.

Oprócz parametrycznej analizy koherencji, możliwa jest również analiza graficzna. O ile jednak statystyczne miary spójności składane są z elementarnych różnic pomiędzy dwiema kolejnymi wypowiedziami, o tyle na wykresach porównuje się jedynie skumulowane wartości średnie, a uśrednianie zawsze prowadzi do zacierania różnic indywidualnych. Komparacja średnich nie jest więc testem spójności wypowiedzi, lecz oceną podobieństwa par wskaźników. Służą głównie do badania właściwości bodźców-stwierdzeń tworzących skalę.



Ryc. 76. Poziomy kontrolnych par wskaźników dla połączonych prób X i Y  
– porównanie Bazy [B'] i Normy [N''] sąsiednich pozycji skali

Różnice pomiędzy średnimi poziomami wskaźników tworzących pary kontrolne stanowią odniesienie dla oceny logiki wypowiedzi. Jeśli poziomy norm nieparzystej i parzystej pozycji skali są podobne, to należy oczekiwać wyższej wartości empirycznej parametru *Spójności*. Największe podobieństwo zarówno Bazy ( $B_1' \approx B_2'$ ), jak i Normy ( $N_1'' \approx N_2''$ ) wykazuje para pozycji 1-2, lecz wyższą *Spójność* ma para 13-14, mimo że średnie różnią się bardziej. Znajduje to uzasadnienie, gdyż łatwiej jest spójnie ocenić lokalny kurs, w którym się uczestniczyło, niż globalny proces informatyzacji. Potwierdza się zasada, że średnia z różnic elementarnych jest trafniejszą statystyką od różnicy średnich ogólnych.

<i>Ślady</i>		<i>Wzorce</i>
sposoby objaśniania oznak	.....	strategia i taktyka interpretacji
wzorce parametrów	.....	normy standaryzujące skalę
względne poziomy średnich	.....	wartościowanie przez odniesienie
poziomy wysokie i niskie	.....	trójpunktowa analiza frakcji HML
zróznicowanie/zgodność opinii	.....	rozkładowa analiza konkluzyjności
koherentność wypowiedzi	.....	miary spójności a pary kontrolne

## Zmienność wypowiedzi – analiza różnicowa

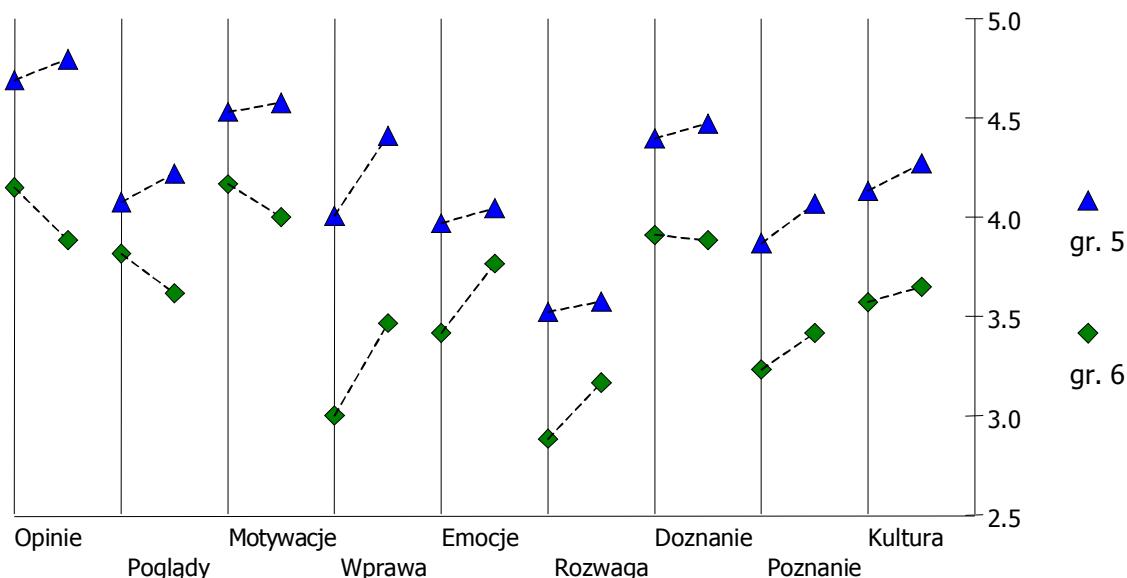
Wspominałem już wcześniej, że różnice pomiędzy poziomami różnych składników indykatum mają jedynie znaczenie wtórne, jako właściwość skali. Duże znaczenie ma natomiast zachodząca w czasie zmienność poziomu jednej i tej samej zmiennej. Umożliwia bowiem pośrednią ocenę oddziaływań, które powodują owe zmiany. Zgodnie z tą zasadą, nie ma potrzeby porównywania poziomów chwilowych, np. *Ambicji ze Zdolnościami*, gdyż sensowna jest jedynie ocena par składników komplementarnych, np. *Ambicji i Intencji*. Istotniejsza jest przede wszystkim analiza zmienności każdego komponentu indykatum. Najprostszą procedurą jest porównywanie wartości średnich M 'przed' i 'po', dające w rezultacie parametry różnicowe  $\pm M$ , które oznaczają wektory zmian. W analizie różnicowej ocenia się przede wszystkim intensywność i kierunek zmian, lecz zawsze w odniesieniu do tego, jaki był poziom na wejściu i wyjściu względem norm standaryzujących. Uwzględnić należy to, że wysokich przyrostów można oczekiwać tylko wtedy, gdy poziom początkowy był niski; jeśli natomiast był nadmiernie zawyżony, to mogą pojawić się uzasadnione spadki. Poszukiwanie przyczyn spadków jest już powinnością samego wykładowcy.

Zmiana poziomów średnich [ $\pm M$ ] i Korelat [ $\pm r$ ] składników w grupach E i F Tab. X

	Ocena	Aplauz	Ambicje	Intencje	Spokój	Odpór	Osąd	Wgląd	Obycie	Pewność	Zdolność	Ogłada
$N_X''$	4,15	4,29	4,11	4,11	3,87	3,60	3,74	3,83	3,79	3,81	3,40	3,49
g:	E											
$M_E'$	4,2	4,2	4,1	4,2	3,6	3,8	3,9	3,6	2,8	3,3	3,0	3,6
$M_E''$	4,4	4,3	4,0	4,1	3,7	3,6	4,1	3,7	3,6	3,5	3,3	3,5
$\pm M_E$	0,24	0,16	-0,10	-0,06	0,09	-0,25	0,19	0,07	0,84	0,26	0,31	-0,07
$\pm r_E$	0,15	0,35	0,20	0,40	0,38	0,48	0,35	0,29	0,58	0,28	0,02	0,55
g:	F											
$M_F'$	4,5	4,4	4,1	4,5	3,3	3,8	4,1	3,9	3,0	3,3	2,7	3,7
$M_F''$	4,0	4,1	3,8	4,2	3,6	3,4	3,9	3,9	3,2	3,4	3,1	3,7
$\pm M_F$	-0,53	-0,32	-0,37	-0,32	0,35	-0,40	-0,16	0,01	0,28	0,15	0,34	0,01
$\pm r_F$	0,38	0,29	0,53	0,15	0,01	0,46	0,24	0,13	0,54	0,52	0,28	0,38
	↑ a	↑ a	↑ b	↑ b	↑ c	↑ d	↑ e		↑ f	↑ f	↑ g	

Tabela X ilustruje analizę różnicową, z przykładem zmian spadkowych:

- W grupie F poziomy afektywne 'przed' były zawyżone, lecz spadki zbyt duże, co może dziwić, gdyż ten sam wykładowca z grupą E uzyskał przyrosty.
- Spadły poziomy motywacji, choć w grupie E nieznacznie i dopuszczalnie.
- Brak korelacji zmian ujawnił, że wystąpiły takie fluktuacje bezstresowości, kiedy to część studentów po roku zajęć oceniła stres przeciwnie niż *a priori*.
- W obu grupach duże spadki odporności przed możliwością uzależnienia się.
- Zmiany różne, lecz w obu grupach poziom końcowy i tak wyższy od normy.
- Silna korelacja oznacza zgodne opinie – nabycie wprawy było zbyt nikłe.
- Brak korelacji ujawnił, że wystąpiły fluktuacje przy samoocenie zdolności.



Ryc. 77. Zmiany poziomów średnich z porównaniem dwóch różnych grup

Wykres ilustruje inną sytuację, kiedy to dwie grupy uczniów w jednym liceum wybrały różne profile: informatyczny (gr. 5) i humanistyczny (gr. 6). Pomimo wspólnego środowiska i oczekiwanego podobieństwa grup, nie są one w ogóle homogeniczne pod względem *InfoKultury*. Ten sam nauczyciel, który spowodował przyrosty i tak już wysokich poziomów w grupie 5, nie potrafił utrzymać trzy spadkowe składniki w grupie 6 choćby na pierwotnym poziomie. Programy, treści i wymiary godzinowe zajęć były różne, natomiast zabrakło zróżnicowania kompetencji: grupę przygotowującą do matury z informatyki winien uczyć informatyk z przygotowaniem pedagogicznym, a grupę humanistyczną – raczej absolwent specjalności *pedagogika medialna i informatyka*<sup>47</sup>.

### Współbieżność oznak – komparacja składników

Zgodnie z modelem struktury indykatum, oba składniki tworzące pary komplementarne powinny wykazywać podobne właściwości statyczne i dynamiczne. Oznacza to podobieństwo zarówno norm, jak i tendencji. Zbiórczym parametrem dążności poziomów jest *Trend*  $[\pm t]$ , który daje wypadkową zmian zachodzących w trzech frakcjach  $[\pm H]$ ,  $[\pm M]$  i  $[\pm L]$ . Zmiany poziomów cech we frakcjach są nierównomierne, dlatego najkorzystniej jest analizować łącznie *Trend* i wzajemne przesunięcia  $\pm HML$ . W dalszej fazie interpretacji posłuży to do oceny trafności procesów oddziaływania, natomiast tutaj chodzi głównie o ocenę trafności modelu ekstrakcji i operacjonalizacji. Do tego celu scala się dane empiryczne wszystkich prób badawczych, a następnie porównuje pary składników pod kątem zgodnej z modelem współzmienności. Oprócz podobnych tendencji, zbliżoną wartość powinny mieć też współczynniki korelacji  $[\pm r]$ .



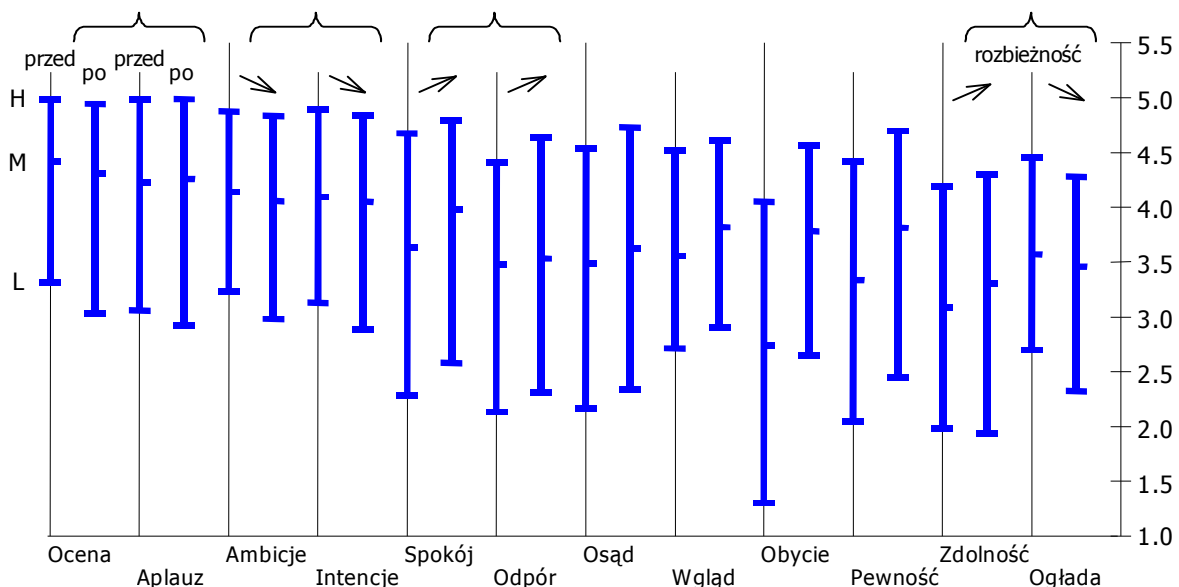
Zmiana [ $\pm$ HML], Trend [ $\pm$ t] i Bilans [ $\pm$ b] składników po scaleniu prób X i Y Tab. XI

	Ocena	Aplauz	Ambicje	Intencje	Spokój	Odpór	Osąd	Wgląd	Obycie	Pewność	Zdolność	Ogłada
N"	4,30	4,26	4,06	4,05	3,97	3,53	3,62	3,82	3,78	3,81	3,31	3,46
$\pm$ H	-0,05	0,00	-0,04	-0,06	0,12	0,23	0,19	0,09	0,52	0,28	0,11	-0,17
$\pm$ M	-0,12	0,03	-0,09	-0,04	0,34	0,06	0,13	0,26	1,04	0,47	0,23	-0,11
$\pm$ L	-0,28	-0,14	-0,25	-0,24	0,29	0,18	0,17	0,19	1,35	0,40	-0,04	-0,38
$\pm$ r	0,37	0,35	0,46	0,45	0,46	0,46	0,53	0,37	0,46	0,49	0,45	0,40
$\pm$ t	-0,10	-0,02	-0,08	-0,07	0,13	0,07	0,08	0,10	0,35	0,17	0,04	-0,10
$\pm$ b	-0,02	-0,01	-0,02	-0,03	0,06	0,00	0,01	0,03	0,26	0,07	0,03	-0,03
	↑ a		↑ b		↑ c		↑ d		↑ e		↑ f	

Tabela XI ujawnia podobieństwo komplementarnych par składników:

- Wprowadzie średnie nie są współbieżne, jednak tendencje globalne i bilans pary składników opiniujących są spadkowe, z podobną, niższą korelacją.
- Zmiany pary składników motywacyjnych są spadkowe we wszystkich trzech frakcjach HML, stąd zgodność tendencji i podobny współczynnik korelacji.
- Przyrostowe zmiany obu składników emocjonalnych, zwłaszcza wyzbycie się stresu; średni *Odpór* wzrósł niewiele, lecz wyraźnie przybrał we frakcji H.
- Podobne tendencje wzrostowe, chociaż tu średnia *Wglądu* [ $\pm$ M] zmieniła się bardziej niż we frakcji H, stąd zróżnicowanie współczynników korelacji.
- Silne tendencje wzrostowe, przede wszystkim poziomów *Obycia* z komputerem w dolnej frakcji L, co dało wyjątkowo wysoki, dodatni bilans zmian.
- Jedyna niewspółbieżna para składników, świadcząca o tym, że wzrastająca ocena zdolności powoduje obniżenie ogłady zachowań przy komputerze.

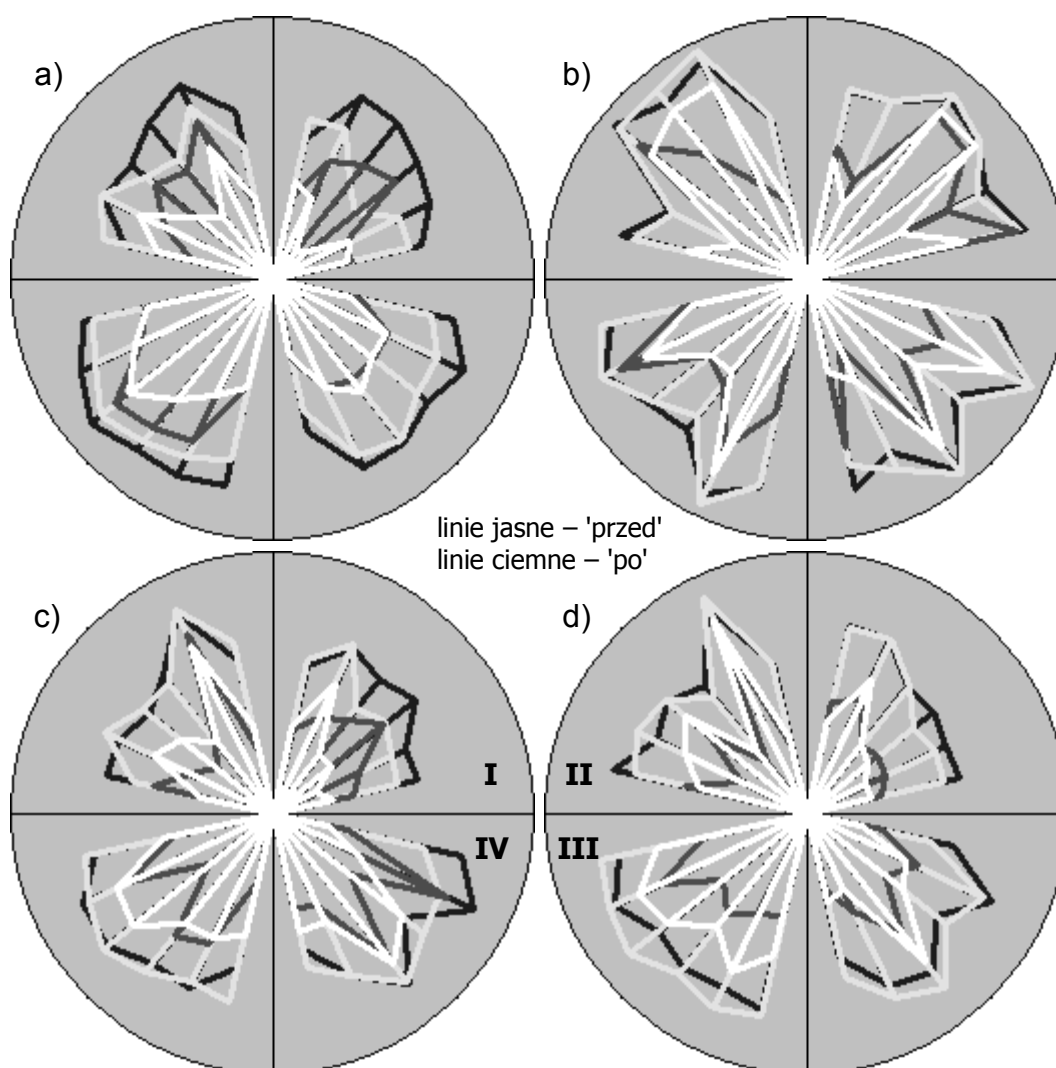
Analizę współbieżności można przeprowadzić na wykresach poziomów, obserwując zgodność kierunków wzajemnych przesunięć punktów H, M i L.



Ryc. 78. Współbieżność par składników – zmiany HML połączonych prób X i Y

### Harmonijność przemian – analiza gradientowa

Specyfiką ewaluacji splotowej jest próbkowanie takich komponentów in-dykatum, które charakteryzują się różnorodną zmiennością. I tak zmienność poziomów średnich nazwałem *zmianą*, wahliwość wypowiedzi elementarnych – *fluktuacją*, zgodność zmian składników – *współbieżnością*, niezgodność par czynników – *rozbieżnością*, a podobieństwo serii wyników – *powtarzalnością*. Cała ta różnorodność ma swe źródło w stochastycznej zmienności wskaźników. Ideałem teoretycznym byłby harmonijny rozwój wszystkich cech cząstkowych, dający na wykresie biegunowym rozrost i wygładzanie czterech „płatków ro-zety” (ryc. 79a). Tak korzystne przemiany są jednak możliwe tylko wtedy, gdy na początku poziomy cech były niskie. W praktyce „płatki” są postrzępione, co oznacza zróżnicowane natężenia wskaźników, a ponadto – przy wysokich poziomach 'przed' nie ma już przestrzeni na przyrosty (79b). Oczekiwany jest „rozrost płatków ciemnych”, głównie w percepcyjnych ćwiartkach I i IV (79c), a niepożądany jest zbyt duży spadek cech afektywnych w ćwiartce III (79d).



Ryc. 79. Wykresy biegunowe do analizy harmonii i kierunków zmian

### Dopasowanie procesów – bilansowanie przemian

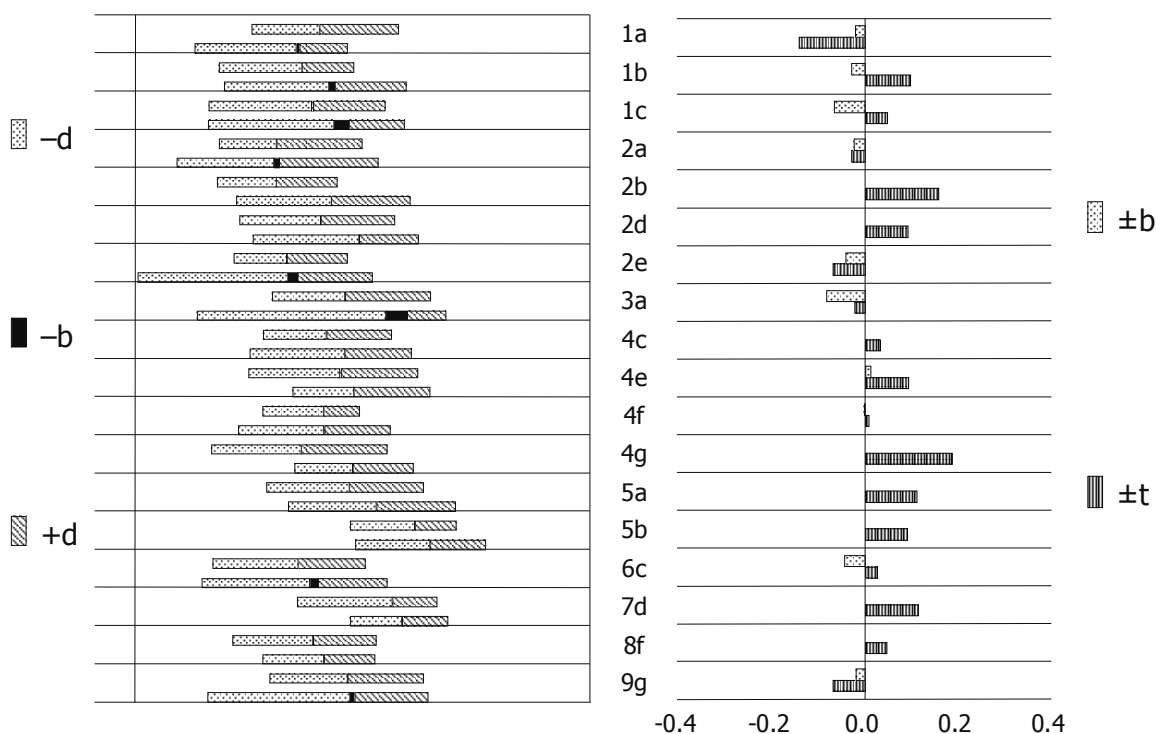
Jednym z podstawowych celów ewaluacji *InfoKultury* jest ocena jakości szkoleń informatycznych. Jednak ewaluacja nie polega na sprawdzaniu wiedzy i umiejętności, lecz na wartościowaniu procesu w całej złożoności. Obserwuje się zmiany cech mentalnych u słuchaczy i na tej bazie wnioskuje o jakości oddziaływań. Kluczowym problemem osoby prowadzącej zajęcia jest dopasowanie poziomu kursu do percepcyjnych możliwości uczących się. Ze względu na zróżnicowanie osób w grupie, znalezienie złotego środka nie jest łatwe. Wstępnie można oszacować poziom grupy na podstawie pomiaru 'przed', lecz skutki oddziaływań dają się ocenić dopiero 'po'. Jedną z metod jest analiza proporcji zmian, jakie zaszły w grupie rozwarstwionej na trzy frakcje: osób o wysokim, średnim i niskim poziomie indykatorum. Polega to na porównaniu *Zmian*  $\pm H$ ,  $\pm M$  i  $\pm L$ , przy czym oczekuje się proporcjonalnych przyrostów. Pamiętać jednak należy o tym, że na poziomie L jest większa przestrzeń do zmian, dlatego przyrosty  $\pm L$  powinny być nieco większe niż  $\pm H$ . Analizę ułatwia parametr *Zawilkości*  $[\pm i]$ , który przy idealnym dopasowaniu dąży do zera, powyżej  $+0,05$  świadczy o zajęciach trudnych, a poniżej zera – o zbyt łatwych.

Zmiany poziomów  $[\pm HML]$ , Zawilłość  $[\pm i]$  i Bilans  $[\pm b]$  wyników w klasach licealnych Tab. XII

k:	1a	1b	1c	2a	2b	2d	2e	3a	4c	4e	4f	4g	5a	5b	6c	7d	8f	9g
$\pm H$	-0,34	0,35	0,13	0,11	0,48	0,15	0,17	0,10	0,14	0,08	0,20	0,17	0,21	0,19	0,14	0,07	-0,01	0,03
$\pm M$	-0,16	0,18	0,15	-0,01	0,36	0,25	0,01	0,27	0,12	0,09	0,00	0,34	0,18	0,10	0,08	0,06	0,07	0,02
$\pm L$	-0,38	0,04	0,00	-0,28	0,13	0,09	-0,63	-0,50	-0,09	0,29	-0,16	0,55	0,14	0,04	-0,07	0,35	0,20	-0,41
$\pm i$	-0,02	0,10	0,05	0,09	0,12	0,05	0,17	0,19	0,08	-0,04	0,10	-0,04	0,06	0,09	0,06	-0,07	-0,04	0,11
$\pm b$	-0,02	-0,03	-0,07	-0,03	0,00	0,00	-0,04	-0,08	0,00	0,01	0,00	0,00	0,00	0,00	-0,05	0,00	0,00	-0,02
	↑ a		↑ b		↑ c		↑ d	↑ d				↑ e				↑ f	↑ f	

Z tabeli XII wysnuć można następujące wnioski o dopasowaniu zajęć:

- Bardzo duże spadki na wszystkich poziomach świadczą jednoznacznie źle o splocie okoliczności, które spowodowały aż tak niekorzystne przemiany.
- Proces pozornie dopasowany, lecz niekorzystny bilans świadczy o tym, że nastąpił w tej klasie rozryw opinii części uczniów wobec opinii większości.
- Przemiany pozytywne, mimo nierównomiernego przyrostu, najwyższego na poziomie H. Program zajęć zbyt ambitny, lecz szczęśliwie nie spowodowało to spadku na poziomie L uczniów słabszych (podobnie w klasach 1b i 5b).
- Charakterystyczne objawy niedopasowania procesu – zbyt ambitne zajęcia jak na możliwości uczniów słabszych wywołał silny spadek na poziomie L.
- Dobre efekty, lecz dopasowanie tylko fasadowe. Zbyt łatwy program zajęć umożliwił wprawdzie osiągnięcie przyrostów na wszystkich poziomach, jednak nie wykorzystano znacznie większego potencjału uczniów w tej klasie.
- Bardziej wyraziste objawy niedopasowania, świadczące o niezaawansowanym poziomie zajęć, bez zaspokojenia aspiracji uczniów najzdolniejszych.



Ryc. 80. Wykresy rozkładowej i różnicowej analizy wyników klas licealnych – Odchylenia [-d, +d], Rozziew [-b], Bilans [±b] i Trend [±t]

Każda z miar reprezentuje inne zjawisko empiryczne, warto więc je porównywać. Wykresy *Trendu* [±] najlepiej ilustrują przemiany, lecz nie ukazują poziomów. Z zawiłych wykresów rozkładu trudniej jest oszacować tendencje. Łączna analiza oznak daje więcej przesłanek do prawidłowej interpretacji: na przykład klasy 2d i 5b uzyskały podobny *Trend*, chociaż startowały z zupełnie różnych poziomów. Wszystkie klasy, w których przyrost jest znaczny i jednocześnie bilans zerowy, uzyskały wyniki dobre. Zbyt małe przyrosty po roku zajęć w klasach 4c i 8f oraz stagnacja w klasie 4f powinny skłonić do refleksji. Jednak najbardziej niepokoją silne tendencje spadkowe w klasach 1a, 2e i 9g. Z wykresu rozkładu wynika, że w 1a obniżyły się poziomy wszystkich frakcji HML, a w 2e i 9g tylko we frakcji dolnej L. Niekorzystne są również te wyniki, w których *Bilans* jest ujemny, choćby nawet *Trend* był dodatni (1b, 1c i 6c). Oznacza to, że przybyło uczniów kwestionujących standardy. Negatywne tendencje, wymagające naprawy oddziaływań, ujawniły się aż w ośmiu klasach.

Ślady	Wzorce
zmienność wypowiedzi	..... różnicowa analiza poziomów średnich
współbieżność zmian	..... trójpunktowa komparacja składników
harmonijność zmian	..... analiza gradientowa wskaźników
dopasowanie procesów	..... analiza przemian w frakcjach HML
bilansowanie przemian	..... komparacja różnic, rozkładów i trendów

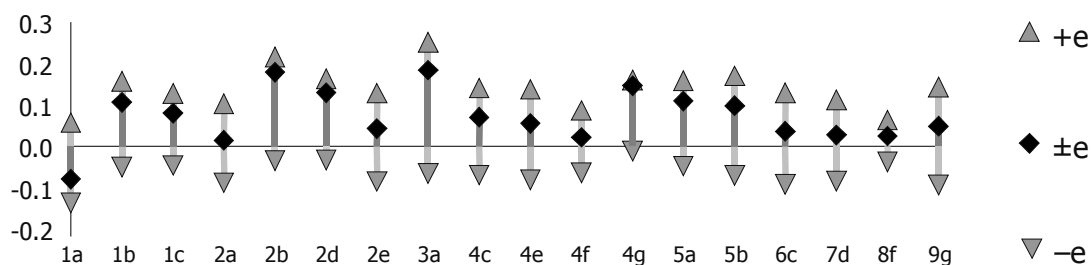
### Wahliwość wypowiedzi – analiza indeksowa

Ocena jakości zajęć może być mylna, jeśli w badanej klasie znajdzie się kilku uczniów o chwiejnych postawach i nieugruntowanej świadomości, bądź też kwestionariusze będą wypełniane nierzetelnie. Ich wpływ na uśrednione rezultaty pomiarów okazuje się czasem tak znaczący, że skutecznie zaciera efekty procesu edukacyjnego. Co gorsza – nawet duże wartości przyrostów są wówczas niewiarygodne (np. w klasie 3a). Najkorzystniejszą strategią ustalania parametrów służących do porównywania prób niehomogenicznych jest analiza indeksowa. Parametry fluktuacji są bowiem standaryzowane. Jeden z wektorów fluktuacji oznacza *Przyrost* [+e], a drugi *Spadek* [-e]. Oczekuje się, że przyrost powinien być większy, zatem różnica wyznacza *Efekt* [ $\pm e$ ]. Łącznie suma ich długości określa *Zasięg* [e] ilustrujący intensywność oddziaływań. Analizę indeksową przeprowadza się razem z analizą różnicową (ryc. 80 i 81).

Zasięg [e], Przyrost [+e], Spadek [-e] i Efekt [ $\pm e$ ] wyników w klasach licealnych Tab. XIII

k:	1a	1b	1c	2a	2b	2d	2e	3a	4c	4e	4f	4g	5a	5b	6c	7d	8f	9g
$\pm i$	-0,02	0,10	0,05	0,09	0,12	0,05	0,17	0,19	0,08	-0,04	0,10	-0,04	0,06	0,09	0,06	-0,07	-0,04	0,11
e	0,20	0,21	0,18	0,19	0,25	0,19	0,21	0,32	0,21	0,22	0,15	0,17	0,21	0,24	0,22	0,19	0,10	0,23
+e	0,06	0,16	0,13	0,10	0,21	0,16	0,13	0,25	0,14	0,14	0,09	0,16	0,16	0,17	0,13	0,11	0,06	0,14
-e	-0,14	-0,05	-0,05	-0,09	-0,04	-0,03	-0,08	-0,07	-0,07	-0,08	-0,06	-0,01	-0,05	-0,07	-0,09	-0,08	-0,04	-0,09
$\pm e$	-0,08	0,11	0,08	0,01	0,18	0,13	0,04	0,18	0,07	0,06	0,02	0,15	0,11	0,10	0,04	0,03	0,03	0,05
	↑ a			↑ b	↑ c			↑ d			↑ e	↑ f				↑ g	↑ h	

- W tabeli XIII należy zwrócić uwagę na zróżnicowane parametry efektów:
- Jedyna klasa licealna, w której wektor spadku okazał się znacznie dłuższy od wektora przyrostu, w wyniku czego rezultat wypadkowy jest ujemny. Tak bardzo złe wyniki uzyskują potwierdzenie również w miarach trendu.
  - Podobna długość wektorów fluktuacji powoduje, że efekt jest bliski zeru. Dla porównania analiza różnicowa wykazuje niewielkie tendencje ujemne.
  - Najlepszy rezultat wypadkowy efektu, potwierdzony także wysokim dodatnim trendem, ponadto uzyskany na zajęciach o podwyższonej trudności [ $\pm i$ ].
  - Pozornie bardzo dobry efekt, lecz uzyskany kosztem wyobcowania frakcji L i ujawnienia tendencji spadkowych (tab. XII), co podważa trafność wyniku. W klasie tej ujawniło się niedopasowanie i nadzwyczaj silne oddziaływanie.
  - Efekt (0,02) niezgodny z różnicą (0,09 – 0,06) nie jest błędem, lecz skutkiem zaokrąglenia po obliczaniu z większą precyzją (dotyczy to także innych tabel).
  - Dobry efekt, potwierdzony również w analizie różnicowej, jednak uzyskany na zajęciach zbyt łatwych, o czym świadczy ujemny parametr zawilności.
  - Poziom jeszcze bardziej niedopasowany do wyższych możliwości słuchaczy, stąd efekt niższy od spodziewanego po różnicowej analizie tendencji.
  - Skutek nikłych efektów zarówno zbyt słabego oddziaływania nauczyciela [e], jak też niedopasowania i niespełnienia oczekiwań frakcji H (por. z tab. XII).



Ryc. 81. Wykres indeksowej analizy wyników klas licealnych wybranych z próby Y – Przyrost [+e], Spadek [-e] i Efekt [ $\pm e$ ]

### Rozbieżność symptomów – komparacja czynników

Charakterystycznym zjawiskiem w ewaluacji splotowej jest to, że pewne komponenty indykatoru mogą wykazywać tendencje do zmian niezbieżnych. Rozbieżność może ujawniać się w tym, że w danej próbie badawczej poziomy jednych czynników przyrastają, a innych maleją, choć nie w każdej populacji. Uzasadnione jest porównywanie nie tylko par czynników komplementarnych, lecz także wzajemnie wszystkich sześciu czynników pod kątem ich zmienności. Podczas interpretacji odróżnić należy ewentualne spadki spowodowane wyjaśnioną teoretycznie właściwością konkretnego czynnika ( $\rightarrow$  *Motywacja*), od spadków, które są symptomami niekorzystnych procesów i oddziaływań. Również silne przyrosty mogą być immanentną cechą czynnika ( $\rightarrow$  *Wprawa*).

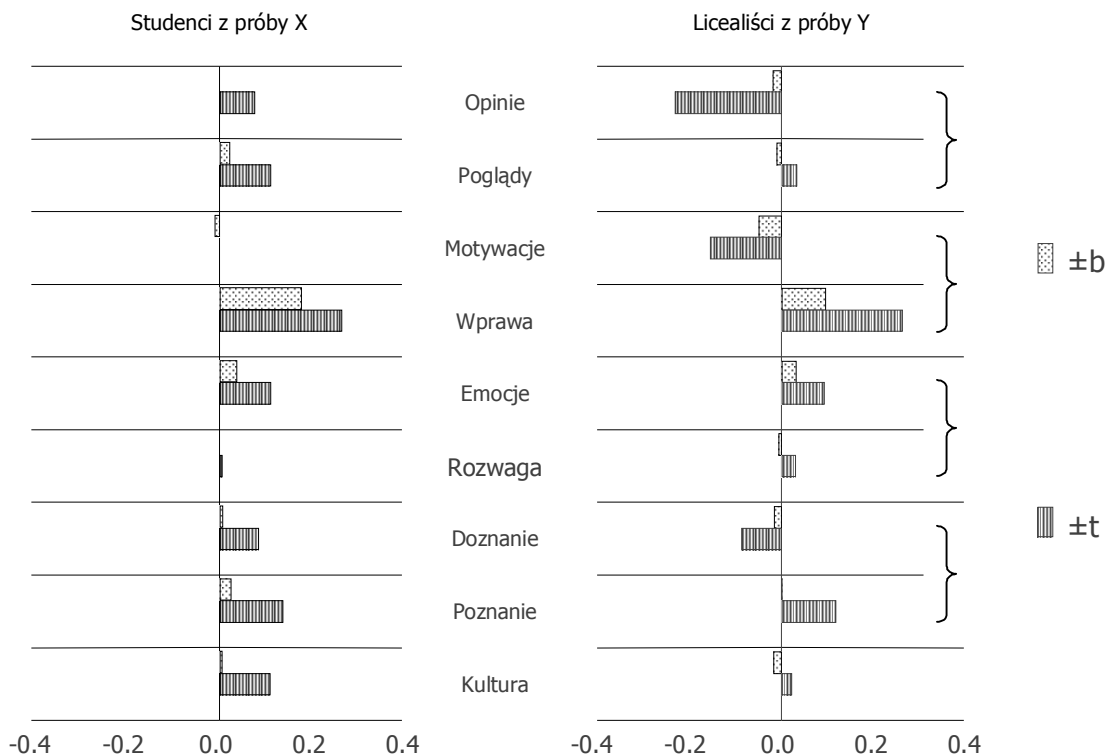
Różne kierunki zmian poziomów [ $\pm HML$ ], Bilans [ $\pm b$ ] i Trend [ $\pm t$ ] dla prób X i Y Tab. XIV

	Czynniki						Czynniki globalne		Wyniki Kultura
	Opinie	Poglądy	Motywacje	Wprawa	Emocje	Rozwaga	Doznanie	Poznanie	
$\pm H_x$	0,13	0,12	-0,03	0,38	0,16	0,01	0,11	0,18	0,13
$\pm M_x$	0,08	0,26	0,06	0,82	0,29	0,02	0,14	0,37	0,25
$\pm L_x$	0,10	0,31	-0,03	0,81	0,24	-0,01	0,16	0,36	0,24
$\pm b_x$	0,00	0,02	-0,01	0,18	0,04	0,00	0,01	0,02	0,00
$\pm t_x$	0,07	0,11	0,00	0,26	0,11	0,00	0,08	0,14	0,11
$\pm H_y$	-0,14	0,12	-0,06	0,45	0,15	0,14	-0,05	0,25	0,09
$\pm M_y$	-0,21	0,12	-0,21	0,68	0,09	0,10	-0,11	0,30	0,10
$\pm L_y$	-0,54	-0,06	-0,46	0,68	0,28	-0,05	-0,23	0,18	-0,08
$\pm b_y$	-0,02	-0,01	-0,05	0,09	0,03	-0,01	-0,02	0,00	-0,02
$\pm t_y$	-0,23	0,03	-0,15	0,26	0,09	0,03	-0,09	0,12	0,02
	↑ a	↑ b	↑ c	↑ d	↑ e	↑ f	↑ g	↑ h	↑ i

W tabeli XIV dostrzec można symptomy globalne dla dwóch populacji:  
 a. Pogorszenie *Opinii* o informatyzacji u licealistów, a poprawa u studentów. Spadki w frakcji  $H_y$  można wyjaśnić korektą nadmiernej fascynacji 'przed', lecz w frakcji  $L_y$  są nazbyt duże, co doprowadziło do ujemnego bilansu.

- b. Poprawiła się trafność *Poglądów*, z wyjątkiem dolnej frakcji  $L_Y$ . Niewielki ujemny bilans świadczy, że paru uczniów zmieniło poglądy na przeciwne.
- c. Nieliniowa tendencja *Motywacji* u studentów – przyrost w frakcji średnio pobudzonych  $M_X$ , a spadki w frakcjach silnie  $H_X$  i słabo pobudzonych  $L_X$ . U licealistów wysoce ujemny bilans świadczy o wyraźnym zniechęceniu.
- d. Oczekiwane, znaczące i podobne nabycie *Wprawy* w obu próbach. Wyższy bilans oznacza, że studenci przed kursem byli mniej obcy z komputerem.
- e. Pozytywny symptom harmonizowania *Emocji* dzięki nabyciu doświadczeń.
- f. *Rozwagę*, ujmowaną tu jako komponent kultury informatycznej, udaje się jedynie i należy kształtować znacznie wcześniej, a nie dopiero na studiach.
- g. U studentów harmonijny we wszystkich frakcjach przyrost *Doznania*, a u licealistów rozbieżny spadek, powodujący zwiększenie rozzewu odczuć.
- h. Z racji nabytych na kursie doświadczeń, tendencje *Poznania* zawsze muszą być dodatnie. Niższy przyrost w dolnej frakcji  $L_Y$  niżli w górnej  $H_Y$  oznacza, że dla części mniej zdolnych licealistów zajęcia były nieco za trudne.
- i. Wynik ogólny studentów, w odniesieniu do normy standaryzującej *Trend* (str. 220, tab. V), zalicza się do wysokich, a wynik licealistów – do niskich. Na ten ostatni silny wpływ miała dywergencja w dolnej frakcji  $L_Y$ .

Rozbieżność symptomów w liceach świadczy o potrzebie poprawy jakości tamtejszych zajęć informatycznych. Częściowym usprawiedliwieniem jest to, że miary tendencji są czułe na pojedyncze, najbardziej niekorzystne przypadki.



Ryc. 82. Tendencje zmian par czynników, porównawczo dla prób X i Y – Bilans [±b] i Trend [±t]

## Efektywność procesów – dynamika przemian

W ewaluacji procesów edukacyjnych musi być dostępna taka strategia, która umożliwi nauczycielowi rzetelną ocenę jakości własnych oddziaływań, bez względu na niekorzystny wpływ wyników osiąganych przez uczniów naj-słabszych. Konieczność optymalnego dopasowania poziomu zajęć wcale nie oznacza potrzeby „wyrównywania w dół”. Znana jest jednak swoista słabość ewaluacji demokratycznej, kiedy to uczniowie wyżej oceniają tego nauczyciela, który mniej wymaga. Mimo to autentycznie dobrzy dydaktycy i wychowawcy są doceniani przez podopiecznych i jednocześnie osiągają dobre rezultaty. Na potrzeby rozwikłania powyższego problemu bardziej przydatne są statystyki indeksowe, w których największy udział mają opinie uczniów z frakcji górnej. W ocenie efektywności bierze się pod uwagę głównie wektory fluktuacji, choć dla porównania odnosi się je do miar tendencji, czulszych na frakcję dolną.

Trend [ $\pm t$ ] i Efekt [ $\pm e$ ] do porównań wyników grup z różnych populacji

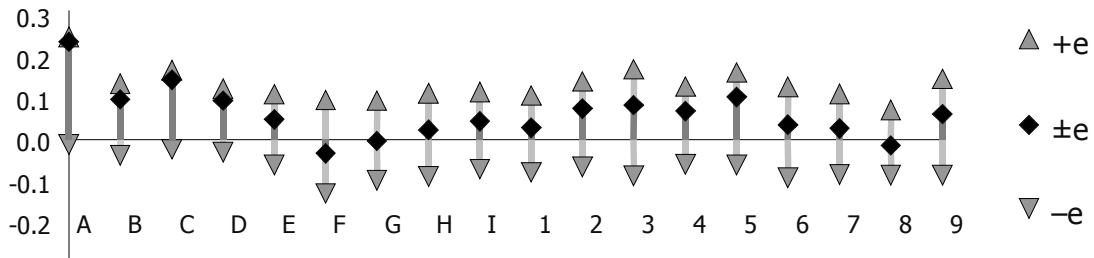
Tab. XV

g:	Grupy studentów z próby X									Grupy licealistów z próby Y									$\Sigma_{XY}$ ogół
	A	B	C	D	E	F	G	H	I	1	2	3	4	5	6	7	8	9	
$\pm i$	0,05	0,06	0,02	-0,03	-0,07	0,06	-0,08	-0,02	-0,03	0,04	0,10	0,10	0,02	0,06	0,06	-0,07	-0,01	0,08	0,03
$\pm t$	0,24	0,07	0,13	0,11	0,07	-0,02	0,07	0,03	0,09	-0,01	0,01	-0,03	0,07	0,08	0,02	0,11	0,00	-0,02	0,07
+e	0,25	0,14	0,17	0,12	0,11	0,10	0,09	0,11	0,12	0,11	0,14	0,17	0,13	0,16	0,13	0,11	0,07	0,15	0,15
-e	-0,01	-0,04	-0,02	-0,03	-0,06	-0,13	-0,10	-0,09	-0,07	-0,08	-0,07	-0,09	-0,06	-0,06	-0,09	-0,08	-0,09	-0,08	-0,06
$\pm e$	0,24	0,10	0,14	0,09	0,05	-0,03	0,00	0,02	0,04	0,03	0,08	0,08	0,07	0,10	0,04	0,03	-0,01	0,06	0,09
	↑				↑	↑	↑				↑	↑				↑	↑	↑	↑
	a				b	c	d				e	e				b	f	e	g

W tabeli XV warto zwrócić uwagę na strategię szacowania efektywności:

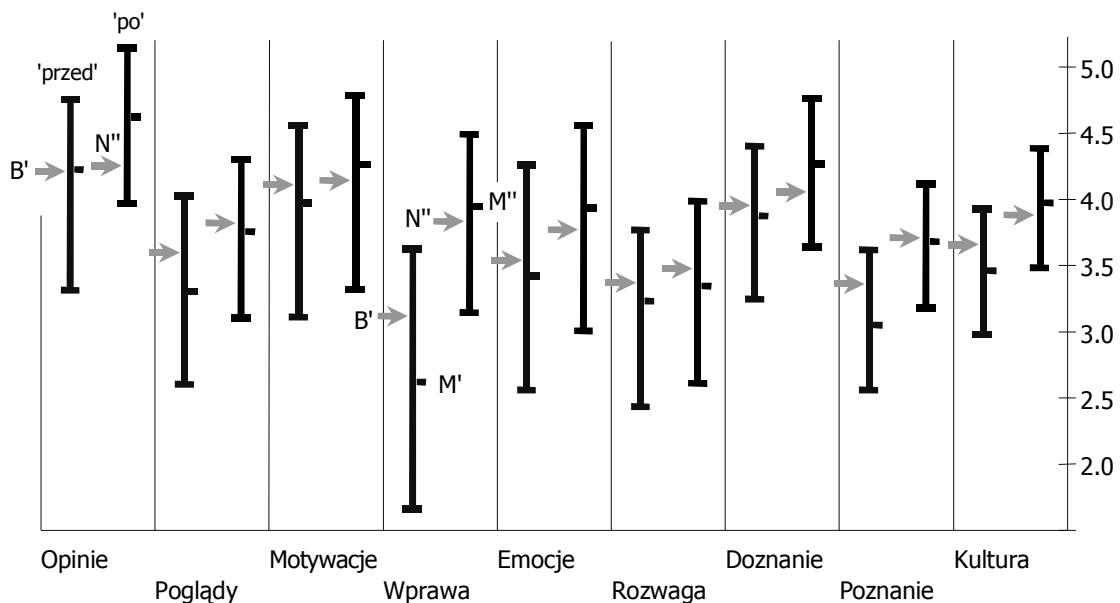
- Bardzo wysokie i całkowicie zgodne parametry *Efektu* i *Trendu* w grupie A świadczą jednoznacznie o doskonałej jakości oddziaływań wykładowcy.
- Wprawdzie *Efekt* i *Trend* są dodatnie, lecz ujemny parametr *Zawilości* [ $\pm i$ ] wskazuje, że poziom zajęć był zbyt niski, co deprecjonuje efektywność.
- Indeksy i tendencje ujemne oznaczają jednoznacznie złą jakość zajęć.
- Zerowy *Efekt* mimo korzystnego *Trendu*, to typowy przykład zbyt niskiego poziomu zajęć, odpowiadającego wyłącznie słuchaczom z dolnej frakcji.
- Pozornie dobre efekty ogólne, lecz uzyskane zbyt dużym kosztem wyalienowania frakcji dolnej i przymusem realizacji zajęć nadmiernie ambitnych. Tego typu „efektywność” bez harmonizowania grupy jest tylko fasadowa.
- Znamienny przykład stagnacji i bylejakości zajęć. Po całym roku szkolnym tendencje są zerowe, a indeks wypadkowy fluktuacji [ $\pm e$ ] bliski zeru.
- Z połączenia 1150 przypadków populacji X i Y wyłania się poziom norm standaryzujących, do których odnieść można efektywność lokalnych zajęć. Wynika z nich, że estymator *Trendu* jest globalnie nieco niższy od *Efektu*, poziom zajęć korzystnie sytuowany jest nieco powyżej średnich możliwości, a z indeksów oddziaływań zdecydowanie przeważają te pozytywne [ $+e$ ].





Ryc. 83. Efektywność procesów porównawczych dla wszystkich grup z prób X i Y  
– Przyrost [+e], Spadek [-e] i Efekt [ $\pm e$ ]

Podstawową zaletą miar indeksowych jest to, że umożliwiają one porównywanie efektów uzyskiwanych w grupach niehomogenicznych, a także ocenę różnic pomiędzy odmiennymi populacjami. Długości indeksów wypadkowych na tle wektorów składowych fluktuacji najlepiej jest porównywać na wykresie ilustrującym dynamikę przemian (ryc. 83). Najefektywniejsze są te wyniki, gdzie długość wektorów skierowanych w górę jest największa, a jednocześnie wektory skierowane w dół są najkrótsze. Zadaniem prowadzącego zajęcia jest dynamizowanie przemian tak, aby *Efekt* [ $\pm e$ ] osiągał wartość co najmniej 0,05. W ocenie dynamiki zmian pojedynczej grupy, jej rezultaty odnosi się do norm 'przed' i 'po' (ryc. 84). Podatność zmian jest bowiem silnie zależna od stanu początkowego. W grupie A możliwa była aż tak duża zmiana, ponieważ niemal wszystkie średnie wartości poziomów 'przed' [M'] były poniżej *Bazy* [B']. Dobre rezultaty uzyskano dzięki temu, że w pomiarze 'po' większość poziomów M'' przesunęła się powyżej *Normy* [N'']. Z porównania wynika, które z czynników wymagają jeszcze dalszego kształtowania – są to *Poglądy* i *Rozwaga*.



Ryc. 84. Czynniki, wyniki połówkowe oraz wynik ogólny grupy A  
– poziomy w odniesieniu do Bazy i Normy populacji X

### Powtarzalność wyników – formowanie przemian

W ewaluacji procesów społecznych duże znaczenie mają badania wieloletnie, zwłaszcza wtedy, gdy w zachodzących procesach wyodrębnić można naturalne cykle zmienności. Taka sytuacja ma miejsce w instytucjonalnych formach edukacji, kiedy to konkretny kurs lub rodzaj zajęć powtarzany jest dla kolejnych grup bądź roczników. Możliwość powtarzania pomiarów stanowi ważną strategię weryfikacji wyników, a ponadto ułatwia formowanie przemian poprzez optymalizację oddziaływań. W strategii tej porównuje się rezultaty uzyskane przez grupy homogeniczne, koniecznie w tym samym środowisku, chociaż pomiar może obejmować kolejne lata podobnie realizowanego procesu edukacyjnego. Najlepiej, jeśli są to słuchacze tego samego wykładowcy. Oczekuje się podobieństwa statystyk potwierdzających zbliżone oddziaływania.

Zmiana średnich [ $\pm M$ ] a Selektowność [ $\check{s}$ ] i powtarzalność wyników

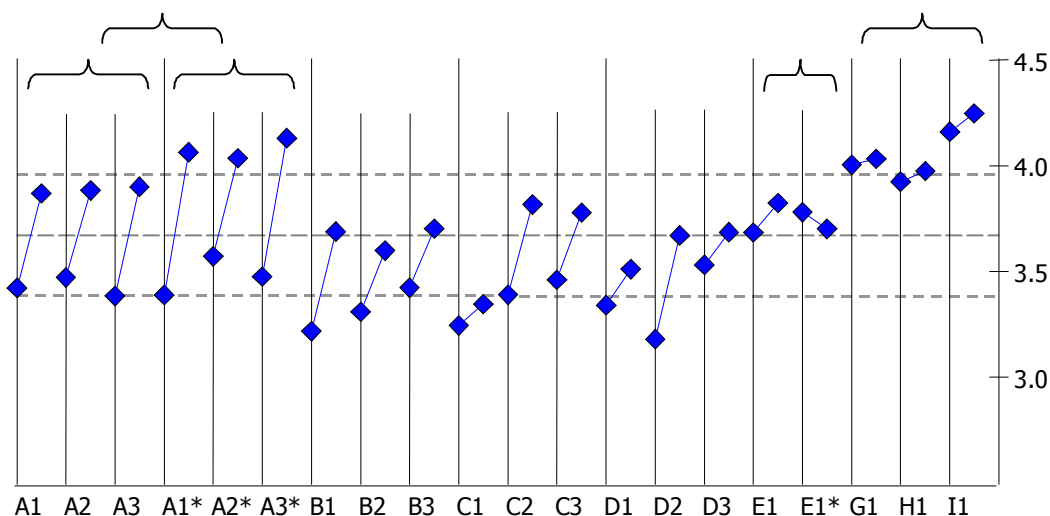
Tab. XVI

	klasy wykładowcy A						grupa	klasy wykładowcy D				grupa	ogół
	Aa	Ab	Ac	Aq	Ar	As	A	Da	Db	Dc	Dd	D	X
$\pm M$	0,45	0,41	0,52	0,68	0,47	0,65	0,51	0,16	0,28	0,32	0,15	0,23	0,25
$\check{s}$	0,46	0,27	0,35	0,32	0,38	0,39	0,15	0,45	0,33	0,39	0,44	0,20	0,09
$\pm t$	0,18	0,19	0,26	0,31	0,23	0,31	0,24	0,05	0,16	0,16	0,06	0,11	0,11
+e	0,23	0,21	0,22	0,31	0,25	0,32	0,25	0,1	0,14	0,15	0,11	0,12	0,16
-e	-0,02	0	-0,02	-0,01	-0,02	0	-0,01	-0,03	-0,04	-0,01	-0,04	-0,03	-0,04
$\pm e$	0,21	0,20	0,20	0,31	0,23	0,32	0,24	0,08	0,10	0,14	0,06	0,09	0,11
	↑ a	↑ b		↑ c		↑ c	↑ d			↑ e	↑ e	↑ f	↑ g

Tabela XVI zawiera następujące oznaki związane z powtarzalnością:

- Przez trzy kolejne lata realizacji zajęć z trzema klasami Aa, Ab i Ac o tej samej specjalności wykładowca A uzyskał identyczne, wysokie efekty [ $\pm e$ ].
- W klasach Ab÷As zmiany średnich [ $\pm M$ ] są dużo większe od *Selektowności* [ $\check{s}$ ], co oznacza różnice istotne już w obrębie nielicznych prób elementarnych.
- Ten sam wykładowca A z inną specjalnością (położnictwo) uzyskał rekordowe efekty, możliwe tylko dzięki wyjątkowej osobowości słuchaczek<sup>48</sup>.
- O ile na poziomie klas *Zmiany* bardzo rzadko są większe od *Selektowności*, o tyle na poziomie grup powinny być normą istotności statystycznej.
- Wykładowca D przez trzy kolejne lata uzyskiwał przyrost efektywności, po czym w czwartym roku z klasą Dd nastąpiło zachwianie skuteczności.
- Zmiana poziomu dla całej grupy D jest istotna statystycznie: [ $\pm M$ ] > [ $\check{s}$ ]. Jeden gorszy rok lub klasa i *Efekt* [ $\pm e$ ] spadł poniżej normy dla populacji X.
- Charakterystyczne zjawisko bardzo wąskiego przedziału *Selektowności* dla licznej próby powstałej z połączenia wszystkich grup studentów.

Analizę porównywalności najkorzystniej jest przeprowadzać na wykresie punktowym, ilustrującym przesunięcia w czasie poziomów średnich każdej z prób elementarnych (ryc. 85), porównując rezultaty klas w obrębie grupy.



Ryc. 85. Zmiany poziomów *InfoKultury* formowane przez różnych wykładowców

Łatwo tu dostrzec podobieństwo skuteczności oddziaływania wykładowcy A. Klamry spinają trzy kolejne roczniki A1, A2 i A3 oraz A1\*, A2\* i A3\* jako przejaw powtarzalności wieloletniej, a górna klamra wskazuje podobieństwo dwóch grup homogenicznych (pielęgniarstwo i położnictwo\*). Grupy B, C i D są też homogeniczne, lecz każdą z grup prowadził inny wykładowca. W klasach B1, B2 i B3 w każdym roku poziom 'przed' był coraz wyższy, lecz poziom 'po' nie nadążał za tą tendencją. U wykładowcy C w jednej klasie C1 rezultaty są nikłe, a u wykładowcy D przyrosty słabe aż w dwóch klasach D1 i D3. Najbardziej niespójne są oddziaływania wykładowcy E. W tym samym roku akademickim z klasą E1 uzyskał minimalny przyrost, a już w klasie E1\* odnotował spadek. Trzeba jednak dodać, że były to różne specjalności (ekonomia a marketing\*). Rezultaty trzech ostatnich grup (G1, H1 i I1) zamieściłem tu dla porównania poziomów osiąganych przez słuchaczy specjalności informatycznych. W tych próbach charakterystyczne są już jedynie niewielkie przyrosty.

### Wiarygodność wniosków – komparacja wyników

W badaniach podstawowych większą wagę przykładają się do symptomów i statystyk umożliwiających uogólnianie wniosków na populację. W praktyce wyższą wartość ma ewaluacja zjawisk lokalnych, przy których również musi być dostępny mechanizm weryfikacji wyników. W strategii skal dwuważonych wmontowanych jest szereg metod komparacyjnych, z których ta ostateczna polega na porównaniu wyników uzyskanych dwiema technikami – różnicową i indeksową. Pierwsza daje rezultaty nadmiernie wrażliwe na niesprzyjające wypowiedzi osób przeciwstawiających się standardom, a druga jest bardziej czuła na pozytywne zmiany u osób o wyższym poziomie badanych cech. Jeśli rezultaty są podobne, to wzrasta pewność co do prawdziwości wnioskowania.

Trafność [â] wyników lokalnych jako podstawa wiarygodności

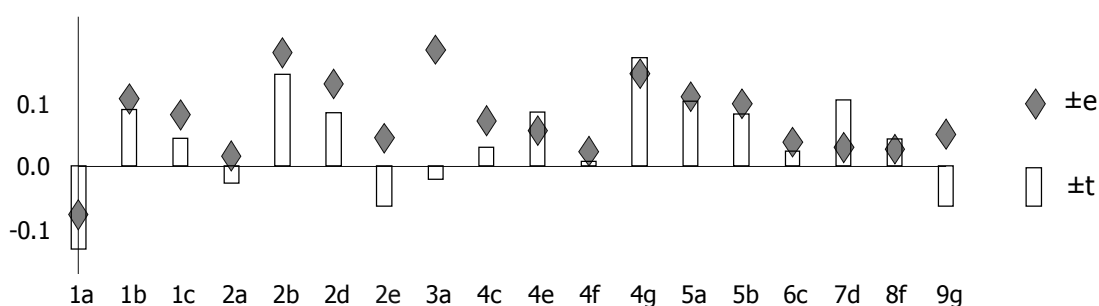
Tab. XVII

k:	wysoka pewność, że wyniki są dobre				wysoka pewność, że wyniki są złe				wyniki są niepewne			ogół
	Aq	Ar	Cc	Cd	Ff	Hh	Ih	3b	2e	3a	9g	X+Y
$\pm M$	0,68	0,47	0,43	0,49	-0,08	-0,14	0,00	-0,01	0,01	0,27	0,02	0,18
$\pm t$	0,31	0,23	0,18	0,22	-0,02	-0,08	-0,02	-0,03	-0,07	-0,02	-0,07	0,07
$\pm e$	0,31	0,23	0,17	0,21	-0,03	-0,08	-0,01	-0,01	0,04	0,18	0,05	0,09
$\hat{a}$	1,00	1,00	1,00	0,99	0,99	0,99	0,99	0,99	0,89	0,79	0,88	0,98
	↑ a				↑ b		↑ c		↑ d	↑ e		

Tablica XVII ukazuje trzy rodzaje sytuacji związanych z oceną trafności:

- Wysoka zgodność *Trendu* [ $\pm t$ ] i *Efektu* [ $\pm e$ ], przy jednoczesnych znacznych tendencjach dodatnich, świadczy o dużej wiarygodności dobrych wyników.
- Analogicznie wysoka zgodność, lecz przy tendencjach ujemnych, upewnia w przekonaniu, że wyniki są spadkowe, zatem należy poszukać przyczyn.
- Przyczyna może być zasadna, np. bardzo wysoki poziom 'przed' w klasie Ih.
- Znaczna niezgodność [ $\pm t$ ]  $\neq$  [ $\pm e$ ], zwłaszcza gdy jeden parametr jest ujemny, a drugi dodatni oznacza, że wnioski mogą być nietrafne. Klasy 2e i 9g są zdominowane przez silny wpływ opozycyjnych uczniów z frakcji dolnej.
- Rekordowa niezgodność w klasie 3a wzięła się stąd, że zdominowała ją silna frakcja górna, podczas gdy uczniowie słabsi zostali wyalienowani.

Na rycinie 86 wyraźnie widoczne są różnice *Trendu* i *Efektu*, a zwłaszcza taka właściwość, że indeksy [ $\pm e$ ] są z reguły korzystniejsze niż miary tendencji.

Ryc. 86. Porównanie Trendu [ $\pm t$ ] i Efektu [ $\pm e$ ] jako test trafności wyników

<i>Ślady</i>	<i>Wzorce</i>
wahania wypowiedzi	indeksowa analiza fluktuacji
rozbieżność symptomów	komparacja tendencji czynników
efekty badań globalnych	porównywanie prób badawczych
rezultaty badań lokalnych	dynamika przemian na tle norm
powtarzalność wyników	wieloletnie formowanie zmian
wiarygodność wniosków	komparacja trendów i efektów

## RÓŻNE IMPLEMENTACJE MODELU-WZORCA

Model metodologiczny ewaluacji nabiera znaczenia wyłącznie wtedy, gdy na jego założeniach da się przeprowadzić badania praktyczne o odpowiedniej jakości. Weryfikacja modelu odbywa się poprzez odwzorowanie strategicznych koncepcji oceniania wartościującego na płaszczyznę realnie przeprowadzanych procesów operacjonalizacyjnych, eksploracyjnych i interpretacyjnych. Czynność przechodzenia od idei do realizacji nazwijmy *zaimplementowaniem*, a wytwory i rezultaty owych działań *implementacjami*. Zatem, implementacja może odnosić się zarówno do samego procesu odwzorowania modelu, jak i do produktu, będącego efektem procesu i wzorcem praktycznej realizacji. Zakres i zupełność odwzorowania zależą zawsze od konkretnych celów ewaluacji, od sytuacji badawczej i od rodzaju indykatum. Przedstawię zatem przykłady różnych wariantów implementacji, dopasowanych do ściśle określonych potrzeb:

- wykorzystanie gotowego, standaryzowanego wzorca ewaluacji *InfoKultury*;
- adaptacja i modyfikacja rozszerzająca indykatum na obszar *MedioKultury*;
- całkowita zmiana indykatum na sferę postaw wobec *Paleo-* i *NeoTelewizji*;
- przystosowanie modelu zawężonego na potrzeby ewaluacji *Nowej Matury*;
- odmiana modelu, opracowana na potrzeby walidacji *Kryteriów Oceniania*.

Każda z implementacji wymagała użycia innych narzędzi i różnych technik, przypomnę więc, na jakich filarach opiera się pełna strategia postępowania. Podczas tworzenia nowego kwestionariusza, konstruowanego na metodologii splotowej i skali dwuważonych ocen, należy wykonać następujące czynności:

1. Opracować teoretyczny model ekstrakcyjnej struktury indykatum.
2. Dokonać rozwarstwienia zmiennej ogólnej na zmienne komplementarne.
3. Ze zmiennych globalnych wyodrębnić tworzące je czynniki i składniki.
4. Zdefiniować każdy z komponentów indykatum i relacje między parami.
5. Wyznaczyć wątki próbkowania poprzez dobór treści tez elementarnych.
6. Wybrać bodźce-stwierdzenia, które pobudzają do reakcji-wypowiedzi.
7. Zweryfikować empirycznie konkluzyjność wskaźników i udoskonalić je.
8. Znormalizować skalę na statystycznie reprezentatywnej i licznej próbie.

Powyższy zbiór zawiera wprawdzie procedury obligatoryjne, jednak w konkretnych implementacjach pewne elementy są zapożyczane lub redukowane.

Najprostszą sytuacją jest wykorzystanie skali opracowanej i zweryfikowanej przez innego badacza, o ile udostępnił on schemat interpretacji. Jeśli gotowe narzędzie mierzy indykatum podobne do tego, jakie my zamierzamy badać, to można próbować dokonać adaptacji skali. Zwykle jednak okazuje się, że przydatne są jedynie pomysły na konstrukcje bodźców-stwierdzeń. Znacznie częściej zapożyczają się uznawane modele teoretyczne struktur bądź taksonomie służące do obierania i definiowania zmiennych (ad 1). Alternatywą rozwarstwiania indykatum na dwie zmienne komplementarne (ad 2) jest mechaniczny podział skali (parzyste/nieparzyste) lub dużo lepsze mierzenie porównawcze dwóch komplementarnych obiektów indykatum. Redukcja modelu do mniejszej liczby warstw komponentów indykatum jest konieczna, jeśli nie daje się uzyskać aż czterech poziomów ekstrakcji (ad 3). Każdy komponent musi być zdefiniowany, lecz nie zawsze możliwa jest predykcja relacji (ad 4). Tezy elementarne w formie bodźców-stwierdzeń mogą być zastępowane przez pytania (ad 5 i 6), lecz ze względu do funkcję pobudzania nie jest to korzystne. W wyjątkowych sytuacjach można pominąć kryterium konkluzyjności (ad 7), jeśli trzeba przebadać zjawisko o nieustalonych standardach ewaluatywnych. Ustalenie norm standaryzujących skalę (ad 8) nie jest potrzebne, jeżeli służy ona wyłącznie do lokalnego porównania: wytworów intelektualnych, zjawisk komplementarnych albo cech grup homogenicznych. Rezygnacja z pewnych strategii metodologicznych musi być jednak w pełni uzasadniona, przy czym może to spowodować konieczność odejścia od nazwy – *miar skalowany*.

### **Wykorzystanie wzorca – ewaluacja *InfoKultury***

Kompletne odwzorowanie modelu stanowi implementacja zrealizowana na potrzeby ewaluacji *Kultury informatycznej*. Stanowi ją zbiór metod, technik i narzędzi udostępnionych łącznie z kwestionariuszem i oprogramowaniem. Narzędzie to okazało się przydatne nie tylko w badaniach podstawowych, lecz także w praktyce szkolnej do oceny jakości zajęć informatycznych. Z jednej strony kwestionariusz umożliwia pomiar nasilenia postaw, tj. predyspozycji do emocjonalnego reagowania i oceniania zjawiska komputeryzacji. Z drugiej strony umożliwia wyznaczenie poziomu uświadomienia w ważnej dziedzinie wiedzy, jaką jest informatyka. *Kultura informatyczna* uzewnętrznia wartości uznawane dziś za istotne w rozwiniętych społeczeństwach, a tym samym wytycza standardy ewaluatywne współczesnych populacji kognitariusz<sup>49</sup>.

*Kwestionariusz Kultury Informatycznej* próbkuje samoocenę oraz ocenę zjawisk ogólnych, jednak nie mierzy wiedzy o komputerze ani zakresu umiejętności jego obsługi. To ostatnie przecież nauczyciel może oceniać wprost na podstawie obserwacji i sprawdzianów. W przeciwieństwie do kompetencji, osobistych nastawień i uświadomień nie daje się tak łatwo odkryć na lekcji.

Niezbędny jest pomiar wypowiedzi jako wskaźników pośrednich, z użyciem starannie dopracowanego narzędzia. Taką właśnie skalę i metodę pomiaru, wyprowadzoną z teorii psychologii poznawczej, polecam nauczycielom informatyki. Proponuję też pewne konwencje i ujednolicone formaty, aby ułatwić przetwarzanie danych i współpracę osobom zainteresowanym.

Demokratyczna ewaluacja procesu edukacyjnego opiera się na badaniu wypowiedzi podmiotów kształcenia: uczniów, studentów, słuchaczy kursów. Wypowiedzi te wyrażają opinie silnie uzależnione od subiektywnych odczuć. Choćby owe opinie z punktu widzenia nauczyciela wydawały się niesłuszne, to faktu tego lekceważyć mu nie wolno. Za wzrost *kompetencji komputerowych* nauczyciel wystawia uczniom stopnie szkolne, natomiast za wywołane u nich *przemiany mentalne*, dotyczące kultury informatycznej, refleksyjny edukator powinien oceniać sam siebie. Ewaluacja efektów oddziaływań edukacyjnych pomaga nauczycielom harmonizować rozwój uczniów poprzez równoważenie przyswajania wiedzy i umiejętności z formowaniem wartościowych postaw.

W definicjach akcentowana jest względna stabilność postaw, tymczasem z empirii wyraziście wyłania się fluktuacja nastawień wobec zajęć w szkole. Dochodzi do paradoksu, kiedy to raczej osiągnięcia uczniów pozostają na stałym poziomie, a nie postawy. Problem balansowania między doznawaniem a poznawaniem komputera przez uczniów jest wyzwaniem dla współczesnego nauczyciela. Zamiast martwić się nikłymi efektami realizacji złych podstaw programowych, lepiej ażeby nauczyciel skupił się na strukturach i treściach ponadczasowych, próbując własnych metod harmonizowania emocji uczniów, subtelnego stymulowania motywacji, kształtowania roztropnej świadomości, a zwłaszcza wyrobienia zdolności do generowania autowiedzy.

Każdy edukator w swej praktyce stosuje dyrektywy metodyki nauczania danego przedmiotu oraz adekwatne technologie kształcenia. Jeśli jest pewien swoich kompetencji, to właśnie z ewaluacji dowie się, czy rzeczywiście taki sposób realizacji zajęć przynosi pozytywne skutki. Wykryte pomiarem trendy sprzeczne z oczekiwaniami mają swoje przyczyny, których podłoże najlepiej rozpozna właśnie osoba ucząca daną grupę. Otrzymuje wszak wprost informacje o tendencjach konkretnych czynników i składników, może więc starać się je w następnym roku zrównoważyć. Oczywiście z uwzględnieniem faktu, że nie ma grup i sytuacji dydaktycznych wiernie powtarzalnych.

### **Modyfikacja indykatum – ewaluacja *MedioKultury***

Kolejna implementacja stanowi przykład adaptacji istniejącego narzędzia na potrzeby ewaluacji innego indykatum, przy czym obiektami próbkowania są też analogiczne cechy mentalne: postawy i uświadomienie, lecz odnoszone do obszaru mediów. Przyjrzyjmy się *Kwestionariuszowi Kultury Medialnej*:

*Prosimy o rzetelne wyrażanie własnych odczuć wobec zagadnień!*

1. Odczuwam pewną niechęć do tak ekspansywnej obecnie technicyzacji.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
2. Tylko uzdolnieni technicznie skorzystają z nowoczesnych technologii.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
3. Mam pozytywne doświadczenia w posługiwaniu się różnymi mediami.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
4. Potrafił(a)bym zaprogramować kanały tunera satelitarnego.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
5. Dał(a)bym radę wystąpić z pokazem medialnym przed dużym audytorium.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
6. Jestem w stanie nauczyć się realizacji wideo techniką jednokamerową.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
7. Umiejętność montażu telewizyjnego można zdobyć samemu.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
8. Nowo zakupiony sprzęt AV można uruchamiać bez czytania instrukcji.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
9. Jest mi źle w środowisku, gdy inni lepiej znają się na aparaturze.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
10. Obawiał(a)bym się zaprezentowania przed kamerą lub mikrofonem.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
11. Wyrażam gotowość uczenia się sztuki tworzenia mediów i programowania.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
12. Chcę rzetelnie studiować zawile psychologiczne aspekty mediów.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
13. Mam świadomość konieczności powszechnej edukacji medialnej.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
14. Wątpię w przydatność algorytmizacji procesów dydaktycznych.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
15. Uważam, że na razie dostępne multimedia są mało użyteczne w edukacji.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
16. W szkole media powinny być używane na wszystkich przedmiotach.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
17. Już dziś wiem, że do wideoprojektora można przyłączyć diaskop.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
18. Komputer (hardware) może radzić, jak rozwiązywać życiowe problemy.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
19. Ekspozycja ujęcia aparatem fotograficznym wymaga koncentracji uwagi.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
20. Długotrwałe oglądanie obrazu nawet na dużym monitorze jest szkodliwe.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
21. Dysponuję czasem na głębszą analizę przekazów medialnych.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
22. Mass media lub media użytkowe pochłaniają zbyt dużo mojego czasu.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak
23. Od mediów można się uzależnić jak od narkotyków.  
absolutnie nie    nie    chyba nie    brak zdania    chyba tak    tak    absolutnie tak
24. Odczuwam na sobie rodzaj zniewolenia przez media lub mass media.  
absolutnie nie    nie    raczej nie    brak zdania    raczej tak    tak    absolutnie tak

*Dziękujemy za udział w badaniach poziomu Kultury medialnej!*



W narzędziu dostrzec można duże podobieństwo do treści sformułowań zawartych w Kwestionariuszu na str. 123. Genezą powstania tej drugiej skali była potrzeba pomiaru predyspozycji studentów wybierających specjalność *Pedagogika medialna* (a po zmianie nazwy: *Edukacja medialna i informatyka*). Możliwość przeróbki narzędzia z próbkowania *InfoKultury* na *MedioKulturę* wydawała się oczywista, lecz zastosowanie tej drugiej skali jest ograniczone tylko do wąskiego grona studentów danej specjalności, podczas gdy przez obowiązkowe zajęcia informatyczne przechodzi każdy uczeń i student. Treści bodźców-stwierdzeń dotyczą tu zagadnień specjalistycznych, realizowanych w toku studiów. Przykład ten jest więc ilustracją możliwości przystosowania narzędzia uniwersalnego na specyficzne potrzeby lokalne.

Pomimo dużego podobieństwa, ta druga skala ma inne schematy polaryzacji i interpretacji, których nie mogę tu ujawnić, gdyż studenci czytający tę książkę będą respondentami. Ponadto nie ma potrzeby standaryzacji tej skali, gdyż w tym wypadku istotne są porównania wyłącznie prób homogenicznych. Pragnę jednak zwrócić uwagę na kilka istotnych szczegółów Kwestionariusza. Po pierwsze – sformułowania niektórych bodźców-stwierdzeń zostały jeszcze udoskonalone, czego nie warto było już robić w funkcjonującym pierwowzorze. Dotyczy to zwłaszcza tezy nr 24, pierwotnie zbyt mało konkluzyjnej. Tezę 18. uzupełniono o słowo *hardware* dla większej jednoznaczności, przy czym może ona pełnić funkcję bufora pomiędzy obiema skalami. Po drugie – próbkowane są uświadomienia i postawy wobec szerokiego spektrum obiektów medialnych, podczas gdy pierwowzór dotyczył tylko jednego medium – tj. komputera. Po trzecie – bez ujawniania zasady, zwracam uwagę na sposób sformułowania bodźca-stwierdzenia nr 17, wprowadzający ciekawy mechanizm autokontroli.

Narzędzie to wykorzystałem do pomiaru zależności pomiędzy poziomami predyspozycji mentalnych do studiowania specjalności a punktacją wynikającą z deklarowanej (przy okazji rekrutacji) aktywności w obranej dziedzinie i samooceny zdatności funkcjonowania w tym profilu. Współczynniki korelacji w dwóch rocznikach osiągały wysokie wartości 0,46 i 0,49, co może świadczyć o rzetelności wypowiedzi deklaracyjnych, które trudno jest weryfikować. Jeśli uzyskam zgodę studentów, gdy już kończyć będą studia, to sprawdzę jeszcze, czy ich średnie ocen ze studiów korelują równie silnie z predyspozycjami.

### **Zmiana indykatum – ewaluacja *Paleo-* i *NeoTelewizji***

Ważne zastosowanie ewaluacji splotowej ukazał dr Marian Michałowski<sup>50</sup>. W badaniach nad edukacyjną przydatnością nowych technologii telewizyjnych wzbogacił główny eksperyment, wprowadzając ewaluację w dwóch funkcjach:

- jako metodę służącą zrównoważeniu grupy eksperymentalnej z kontrolną;
- jako test zmian postaw i świadomości pod wpływem czynnika medialnego.

Ewaluacja służyła jako istotne uzupełnienie głównego celu badań, jakim było ustalenie różnic w edukacyjnych oddziaływaniach tradycyjnej *PaleoTelewizji* i przebijającej się *NeoTelewizji*. Podstawowymi wątkami pomiaru były różnice poznawcze, związane z uczeniem się i zapamiętywaniem, oraz afektywne, dotyczące reakcji na elementy nowego języka telewizji. Pominę obszerne wątki poznawcze, zwracając uwagę jedynie na istotny aspekt metodologiczny. Otóż zastosowano tam pomiary zarówno przed, jak i po zadziałaniu czynnika eksperymentalnego, dodatkowo z testem odroczonym. Ważnym metodologicznie zabiegiem było to, że respondenci zostali wstępnie zbadani pod kątem postaw wobec obiektów badań, dzięki czemu możliwe było pierwotne zrównoważenie obu grup. W odróżnieniu od tych quasi-eksperymentów, gdzie osoby do grup przydziela się wyłącznie losowo (metodą „część chybł, niż trafił”), tutaj obie próby przed oddziaływaniem były homogeniczne pod względem indykatum. Do zrównoważenia posłużył pomiar za pomocą następujących próbników:

1. Odczuwam raczej niechęć do przekazów telewizyjnych.
2. W zasadzie telewizja publiczna jest mi obojętna.
3. Ekscytuje mnie rozmach i dynamika przekazów telewizyjnych.
4. Telewizja długo jeszcze będzie ludzi fascynować.
5. Urzeka mnie bogactwo form wizualnych.
6. Urzeka mnie grafika komputerowa wykorzystywana w telewizji.
7. Lubię patrzeć na nowatorskie formy realizacji telewizyjnej.
8. Warstwa wizyjna wideoklipu często przeszkadza mi w odbiorze muzyki.
9. Telewizja pomaga w rozwiązywaniu codziennych spraw ludzi.
10. Telewizja dość często powoduje uzależnienia.
11. Telewizja tylko w niewielkim stopniu zaspokaja moje aspiracje poznawcze.
12. Dostrzegam spory ładunek edukacyjny w programach telewizyjnych.
13. Generalnie narracja przekazów telewizyjnych jest zbyt szybka.
14. Zawsze rozumiem treść przekazów telewizyjnych.
15. Kolorowe, błyskające, dynamiczne ozdobniki niewiele wnoszą do przekazów.
16. Przekazy telewizyjne mogłyby się obyć bez komputerowych efektów trikowych.
17. Chciał(a)bym więcej czasu poświęcić na oglądanie telewizji.
18. Telewizja jest najczęściej złodziejem czasu.
19. Wyobrażam sobie własny dom zupełnie bez telewizji.
20. Korzyści wynikające z oglądania telewizji są minimalne.
21. Pozytywną cechą telewizji jest duża ilość i różnorodność programów.
22. Lubię często zmieniać kanały za pomocą pilota.
23. Do oglądania wybieram pozycje z wcześniejszych zapowiedzi programowych.
24. W większości przypadków programy telewizyjne mnie nudzą.

Powyższe bodźce-stwierdzenia wybrano ze zbioru 42 sformułowań poddanych weryfikacji w badaniu pilotażowym. Dotyczą one zarówno ogólnie telewizji, jak też elementów znamienych dla *Paleo-* i *NeoTelewizji*. W takiej implementacji potrzebne było rozwarstwienie indykatum jedynie na trzy kategorie globalne.

### Przystosowanie modelu – ewaluacja *Nowej Matury*

Jedna z zastosowanych przeze mnie implementacji miała na celu wykorzystanie skali dwuważonej do ewaluacji zupełnie innych cech jakościowych, z modyfikacją także parametrów. Dotyczyła ona przeprowadzanego przez trzy kolejne lata eksperymentu porównywania tzw. *próbnej matury nowatorskiej* z wcześniejszą tradycyjną<sup>51</sup>. Głównym czynnikiem interweniującym była forma egzaminu z matematyki – pierwszego dnia testowa, a drugiego problemowa. Obie formy oceniali uczniowie i nauczyciele. W takim pomiarze nie chodziło o zmiany w czasie, lecz o różnice wypowiedzi dokładnie tych samych osób wobec obu odmiennych form. Z tego względu, przy zachowaniu metodologii ewaluacji splotowej, uległy modyfikacji niektóre nazwy miar skali dwuważonej. I tak *Zmianę* zastąpiły *Różnice*, w miejsce *Trendu* weszła *Preferencja* danej formy matury, a *Wzrost/Spadek* zostały zastąpione przez opcje *Za/Przeciw testowi*. Respondenci oceniali dwie grupy czynników symptomatycznych dla jakości matury: merytoryczno-przedmiotowych i formalno-organizacyjnych. Interesujące jest to, że pierwsze wypadły dobrze, a drugie już znacznie gorzej. Te kategorie, które nie uzyskały aprobaty większości, oznaczyłem gwiazdką:

Uczniowie	Nauczyciele
<i>Zgodność zadań i nastawień</i>	<i>Zgodność zadań z programem</i>
<i>Atrakcyjność zestawu zadań</i>	<i>Jakość doboru zestawu zadań</i>
<i>Percepcja treści zadań</i>	<i>Przekrojowość zagadnień</i>
<i>Jasność sformułowań</i>	<i>Jasność sformułowań</i>
<i>Trafność stopnia trudności zadań *</i>	<i>Trafność stopnia trudności zadań</i>
<i>Wystarczalność czasu *</i>	<i>Wystarczalność czasu</i>
<i>Spójność ocen i oczekiwań *</i>	<i>Spójność ocen i oczekiwań *</i>
<i>Przystępność formuły egzaminu</i>	<i>Zbieżność ocen z testu ze szkolnymi *</i>
<i>Osąd formuły problemowej *</i>	<i>Osąd formuły problemowej *</i>
<i>Osąd formuły testowej *</i>	<i>Osąd formuły testowej</i>
<i>Akceptacja dwu etapów matury *</i>	<i>Akceptacja dwu etapów matury</i>
<i>Samodzielność uczniów</i>	

Dzięki trzykrotnemu eksperymentowi z ewaluacją formatywną okazało się, co należałoby jeszcze udoskonalić przed wprowadzeniem formuły testowej jako obowiązującej, ażeby przybliżyć do siebie wartości uznawane przez cztery różne gremia: decydentów od spraw formalnych, metodyków od treści matematycznych, uczniów zdających egzamin i nauczycieli poprawiających prace. Niestety – jedyny jak dotąd zrealizowany postulat to wydłużony czas pisania testu, który nadal zawiera zadania zbyt „uproblemowane” jak na test.

Dla ewaluatora ważne jest to, że – wskutek braku społecznej akceptacji dla zbyt trudnej matury z matematyki – dotąd nie ukształtowały się standardy ewaluatywne i normy wartości wskaźników, wobec czego nie można w pełni zastosować metod kwantyfikacji i estymacji wzorcowych poziomów odniesienia.

### **Odmiana modelu – walidacja *Kryteriów Oceniania***

Ostatni z prezentowanych przykładów ilustruje, w jaki sposób ewaluację zastosować można w procesach decyzyjnych. W konkretnym przypadku chodziło o ustalenie obiektywnych i standaryzowanych kryteriów oceniania wypracowań maturalnych z języka polskiego. Z jednej strony jednolite *Kryteria* są absolutnie niezbędne dla egzaminatorów oceniających prace, lecz z drugiej strony – ich opracowanie jest niezwykle trudne. Zespoły ekspertów przedstawiały kolejne projekty, a nauczyciele poloniści odnosili się do nich krytycznie. Z sondażu opinii wyłaniał się obraz kluczowych kwestii, wymagających rozstrzygnięcia: A lub B. Zaproponowałem przeprowadzenie strategii ewaluacji polegającej na *dwuważonej komparacji par wskaźników dysjunkcyjnych*<sup>52</sup>:

A. Matura winna różnicować, stąd ocena musi uwzględniać porównanie z poziomem innych prac.  
 B. Matura winna sprawdzać, stąd ocena pracy musi wynikać wyłącznie ze spełnienia kryteriów.  
 A. Kryteria powinny być szczegółowsze, ażeby ujmowały z osobna każdą z pożądaných cech.  
 B. Kryteria powinny być bardziej ogólne, z podobnie dużym zasięgiem cech jak standardy.  
 A. Kryterium wyjątkowo niepasujące do formy omawianego utworu może zostać pominięte.  
 B. W ustalaniu oceny musi być uwzględnione odniesienie do wszystkich bez wyjątku kryteriów.  
 A. Trzeba wskazać te kryteria, których niespełnienie z góry wyklucza możliwość zaliczenia.  
 B. Trzeba wskazać kryteria, które mogą być niedostateczne, a mimo to pracę wolno zaliczyć.  
 A. Warto kryteriom przypisać wagi punktowe, aby ustalić ważkość danej cechy wypracowania.  
 B. Zamiast zróżnicowanej punktacji, lepiej wszystkie kryteria oceniać w skali stopni szkolnych.  
 A. Kryteria winny wyznaczać tylko dwa skrajne progi, tj. brak zaliczenia oraz ocenę celującą.  
 B. Warto rozpiąć kryteria na każdy ze stopni szkolnych, od niedostatecznego do celującego.  
 A. Zamiast przeciwstawiać sobie odmienne cechy, lepiej oceniać natężenie jakości każdej z nich.  
 B. Dosłowne i metaforyczne odczytywanie to prawidłowy układ skrajnych jakości na skali ocen.  
 To zaledwie fragment narzędzia, jakie należałoby użyć, ażeby ustalić wstępne założenia *Kryteriów*. Respondenci występujący w roli sędziów kompetentnych wybieraliby z czterech opcji: *nie*, *raczej nie*, *raczej tak* względnie *tak*. Ta teza elementarna, która uzyskałaby wyższą aprobatę, byłaby obowiązującą normą.

Charakterystyczną właściwością par dysjunkcyjnych jest to, że odrębnie stanowią stwierdzenia w zasadzie słuszne, jednak obu postulatów nie można wdrożyć jednocześnie, gdyż wzajemnie się wykluczają. Przed instytucjonalną walidacją *Kryteriów* konieczne było w owych kwestiach ustalenie polaryzacji opinii nauczycieli, którzy na maturze pełnią rolę egzaminatorów. Trudno jest bowiem zmusić do stosowania wytycznych, które nie są akceptowane. Niestety – prace na ewaluacją i walidacją *Kryteriów* przerwano, a *Nową Maturę* i tak wzdrożono, bez dopracowania zasad, jako urzędniczy eksperyment na żywym organizmie abiturientów. W rezultacie – matura z języka polskiego jest ze wszystkich egzaminów najmniej obiektywna, co łatwo wykazać, np. poprzez odniesienie zadań do tzw. *wzorca odpowiedzi*. Nie sprawdza się kompetencji językowych, lecz to, czy uczeń odgadnie, o co chodziło autorowi zadania<sup>53</sup>.

## Skorowidz z translacją słów kluczowych

- abstrahowanie, 133
- addytywność wskaźników, 66
- afektywność, 150
- agregacja wskaźników, 129
- alfa Cronbacha, 196
- Ambicje* 'Ambition – zmienna, 146
- ambicje a intencje, 147
- amplitudy fluktuacji, 83, 188
- analiza
  - gradientowa, 179, 234
  - ilościowa, 82, 134
  - indeksowa, 237
  - jakościowa, 84, 85, 138
  - rozkładowa, 224, 236
  - różnicowa, 231, 236
- Aplauz* 'Applause – zmienna, 146
- aplauz a ocena, 147
- argumenty zmiennej losowej, 144
- aspiracje, 149
- asymetria rozkładu, 171
- automatyzacja analizy, 205
- autonormalizacja skali, 223
- autorytaryzm walidacji, 23
- badania, 14
  - przekrojowe, 20
  - quasi-eksperymentalne, 20
- badanie a ewaluacja, 14
- Baza* 'Basis – parametr [B], 168, 222
- Bilans* 'Balance – parametr [ $\pm b$ ], 182, 235
- bilans zmian wypowiedzi, 58, 235
- błędy jakości, 53
- błędy kwantyzacji, 41
- błędy sondaży jednorazowych, 42
- bodźce-stwierdzenia, 31, 59, 67
- budowa kwestionariusza, 111
- cecha populacji, 17
- cechy mentalne, 16
- cel deskryptywny i formatywny, 17
- cele ewaluacji, 97
  - praktyczne, 98
  - teoretyczne, 20
- charakterystyka ewaluacji, 27
- Częstość realna* 'Real Frequency
  - rozkład [ $f(x)$ ], 172
- częstość próbkowania, 42
- czułość skali, 26
- Czynniki* 'Factors – zmienne, 148
  - tabele, 212, 238
  - wykresy, 224, 226, 232, 239, 241
- definicje pojęć kluczowych
  - cechy mentalne, 16
  - Ewaluacja splotowa*, 14, 15
  - ewaluacja demokratyczna, 15
  - InfoKultura*, 16
  - postawa, 17, 48
  - skala dwuważona*, 52
  - Skala dwuważonych ocen*, 54
  - standardy ewaluacyjne, 15
  - uświadomienie, 49
  - uświadomienie informatyczne, 17
- definicje pojęć metodologicznych
  - badanie, 14
  - bodźce-stwierdzenia, 67
  - kwalifikacja a kwalizacja, 84, 86
  - kwantyfikacja a kwantyzacja, 82
  - kwantyfikatory, 18, 25
  - metodologia ewaluacji, 13
  - ocenie wartościujące, 14
  - oznaka, 153
  - pomiar, 18
  - reakcje-wypowiedzi, 68
  - symptomy, 24
  - techniki ewaluacji, 13
  - wskaźniki, 17, 60
  - zmienna, 143
  - zmienna ogólna, 17
- definicje pojęć statystycznych, 82, 83
  - cecha populacji, 17
  - miara, 154
  - momenty, 159, 160
  - nazwy parametrów, 154
  - operatory działań, 155
  - parametr, 154
  - populacja, 15
  - próby badawcze, 15
  - przekroje analityczne, 77
  - statystyki ewaluacji, 13
  - symbole parametrów, 154, 155
  - zbiory danych empirycznych, 156
- definiowanie i ekstrakcja indykatorum, 111
- demokratyzacja ewaluacji, 23
- depolaryzacja wskaźników, 128
- deskrypcja, 86
- dewiacja oznak, 83, 162
- dewiacja rozkładu, 224
- dobór wskaźników, 59, 114
- dokładność miar zagregowanych, 189
- dołączanie danych, 203
- dominanta, 172, 176
- dopasowanie procesu, 235

- dostrajanie skali, 35  
*Doznanie 'Affection* – zmienna, 50, 150  
 doznanie a poznanie, 18, 48, 151  
 drukowanie, 207  
   tabel, 209  
   wykresów, 214  
 dwuwymiarowość wypowiedzi, 56  
 dynamika fluktuacji, 84, 190  
 dynamika przemian, 240  
 dysjunkcja oznak, 83, 161, 227  
 dyspersja oznak, 83, 164, 224  
 dystrybuanta, 172, 177  
*Efekt 'Effect* – parametr  $[\pm e]$ , 190, 237  
 efektywność procesów, 240  
 ekspansja czasoprzestrzeni, 30  
 eksperyment, 20, 44  
   *Próbnej Matury*, 251  
 eksplikacja, 86  
 eksploracja zasobów, 199  
 ekstrakcja zmiennych, 29, 50, 111  
 ekstraspекcja a introspekcja, 30, 65  
*Emocje 'Emotion* – zmienna, 148  
 emocje a rozważa, 149  
 estymacja  
   pozycji trójpunktowa, 170  
   przedziałowa, 168  
   rozkładu, 172  
 ewaluacja, 10, 14, 15  
   demokratyczna, 23  
   naukowa, 21  
   wieloletnia, 126  
*Ewaluacja splotowa 'Splice Evaluation*  
   *InfoKultury*, 19, 246  
   *Kryteriów Oceniania*, 252  
   *MedioKultury*, 247  
   *Nowej Matury*, 251  
   *Paleo- i NeoTelewizji*, 249  
 fazy analizy, 205  
*Fluktuacja 'Fluctuation*, 83  
   – wektory  $[\vec{F}]$ , 187  
 fluktuacja a zmiana, 187  
 folder zasobów, 202  
 formowanie przemian, 242  
 formułowanie wniosków, 141  
 funkcje  
   analizy statystycznej, 209  
   operowania zbiorami, 202  
   wizualizacji graficznej, 214  
 futurospekcja a retrospekcja, 30, 65  
 generalizacja, 86  
 globalność czy lokalność, 30  
 gotowość potencjalna, 48  
 gradient, 179  
 grafika wektorowa, 207  
*Grupa*, próba badawcza, 15, 110  
 grupowanie danych, 127  
 harmonijność przemian, 234  
 idealizacja, 86  
 implementacje, 245  
   modyfikacja indykatum, 247  
   odmiana modelu, 252  
   przystosowanie modelu, 251  
   wykorzystanie wzorca, 246  
   zmiana indykatum, 249  
 impuls, 149  
 indeksacja wskaźników, 54  
 indeksy iteracyjne, 156  
 indeksy selekcyjne, 156  
 indeksy względne, 89, 178, 188  
*Indykatum 'Indicatum*, 16, 17  
*InfoKultura 'InfoCulture*, 16, 151  
 informacja o autorze, 207  
 instalacja pakietu, 200  
*Intencje 'Intention* – zmienna, 146  
 intencje a ambicje, 147  
*Intensywność 'Intensity* – parametr  $[I]$ , 160  
 intensywność wypowiedzi, 56, 83  
 intensywność zmian, 185  
 interferencja strategii oceniania, 24  
 interpretacja  
   parametrów, 134  
   wykresów, 136  
   zjawisk, 139, 219  
 introspekcja a ekstraspекcja, 30, 65  
 intuicja, 148  
 istotność różnicowania, 194  
*Jakość 'Quality* – parametr  $[Q]$ , 160  
 jakość a intensywność, 38, 53  
 jakość eksploracyjna, 91  
 jakość interpretacyjna, 93  
 jakość wypowiedzi, 56  
 jury oceniające, 45  
 justowanie estymatorów, 36  
 kalibracja skali, 55  
 kierunki zmian, 179  
*Klasa* – próba elementarna, 110  
 kodowanie wypowiedzi, 126  
 kody literowe prób, 110, 210, 213  
 kognitywność, 150  
 koherentność oznak, 196, 229  
 komparacja  
   czynników, 40, 238  
   składników, 232  
   wartości i ważkości, 56  
   wskaźników, 229  
   wyników, 243  
 komplementarne pary składników, 39  
 komplementarność wskaźników, 65

- kompletność wskaźników, 64
- komponenty
- postaw, 48
  - uświadczenia, 49
  - zmiennych, 146
- kondensowanie, 133
- koniunkcja oznak, 83, 166, 229
- Konkluzyjność* 'Conclusiveness, 38, 69, 84
- parametr [č], 193, 210, 227
- konkluzyjność a dyskryminacyjność, 10
- konstruowanie
- skali pomiarowej, 59
  - tabel, 131
- kontrolne pary wskaźników, 39, 229
- konwencje oznaczania, 155
- konwersja skali, 88, 167
- kopiowanie do schowka, 206
- tabel, 209
  - wykresów, 214
- korekta danych, 203
- korelacja wypowiedzi, 83
- Korelat* 'Correlate – parametr [±r], 184
- kres górny i dolny skali, 87, 89
- kryteria empiryczne, 52
- kryteria subiektywne, 22
- Kultura* 'Culture – zmienna, 151
- Kultura informatyczna, 16, 151
- kwalifikacja a kwalizacja, 84, 86
- kwantyfikacja a kwantyzacja, 82
- Kwantyfikatory* 'Quantifiers, 18, 25
- kwantyfikatory dobroci skali, 26
- buforowy, 192, 196
  - kumulacyjny, 192, 195
  - pomocniczy, 166
  - selekcyjne, 39, 55, 192, 193, 194
- kwantyfikatory kalibrujące, 55
- kwantyfikatory próbkujące, 55
- kwantyzacja, 41, 55
- Kwestionariusz InfoKultury*
- 'InfoCulture Questionnaire, 18, 123
- Kwestionariusz MedioKultury* 'MediaCulture Questionnaire, 248
- legenda na ilustracjach, 214
- licencja na używanie pakietu, 13, 207
- liczebność prób, 110, 210, 213
- ładowanie zasobów, 202
- macierz danych, 77, 196
- mediana, 172
- menu, analiza statystyczna, 209
- Czynniki* 'Factors, 212
  - Składniki* 'Elements, 211
  - Wskaźniki* 'Indicators, 210
  - Wyniki* 'Findings, 213
- menu, oferta usług, 207, 209, 214
- English/Polish*, 208
- Info...*, 207
- Opcje* 'Options, 214
- Przekroje* 'Profiles, 209
- menu, polecenia, 202
- Dołącz* 'Append, 203
  - Drukuj* 'Print, 207
  - Kopiuj* 'Copy, 206
  - Pokaż* 'View, 206
  - Twórz* 'Create, 205
  - Usuń* 'Delete, 204
  - Wczytaj* 'Load, 202
  - Zamknij* 'Close, 208
  - Zapisz* 'Save, 204
- menu, wizualizacja, 214
- Dynamika* 'Dynamics, 218
  - Dyspersje* 'Dispersion, 217
  - Kierunki* 'Directions, 217
  - Poziomy* 'Levels, 216
  - Średnie* 'Means, 216
  - Tendencje* 'Tendencies, 218
- metodologia ewaluacji, 13
- metody analizy i syntezy, 106
- metody ewaluacji, 27
- ekspansja czasoprzestrzeni, 30
  - ekstrakcja symptomów, 28
  - koncypowanie dwubieżne, 46
- miarodajność wskaźników, 227
- miary statystyczne, 154
- chwilowe, 158
  - dystansowe, 178
  - ilościowe, 82
  - indeksowe, 178
  - normujące, 43
  - różnicowe, 178
- minimaks, 34
- moc dyskryminacyjna, 23
- moc pomiarowa, 227
- model-wzorzec ewaluacji, 19
- momenty zmiennej losowej, 159, 160
- Motywacje* 'Motive – zmienna, 148
- motywacje a wprawa, 149
- napisy na pasku tytułowym, 201
- narzędzia pomiarowe, 104
- nazwy parametrów, 154
- nazwy plików na pasku stanu, 201
- nierównomierność zmian, 181
- niespójność immanentna, 183
- niwelacja niespójności, 183
- Norma* 'Norm – parametr [N"], 168, 211, 222
- normalizacja skali, 211, 223
- normy standaryzujące, 220
- obiekt badań, 16

- obszary ufności, 53  
 obwiednia fluktuacji, 190  
*Obycie* 'Contact – zmienna, 146  
 obycie a pewność, 147  
*Ocena* 'Approval – zmienna, 146  
 ocena a aplauz, 147  
 ocena wytworów intelektualnych, 45  
 ocenianie  
     różnicujące, 22  
     sprawdzające, 22  
     wartościujące, 14, 23  
     zatwierdzające, 23  
*Odchylenia realne* 'Real Deviations  
     – parametry [-d] [+d], 162, 224  
 odczytywanie wykresów, 136  
*Odpór* 'Hardiness – zmienna, 146  
 odpór a spokój, 147  
 odtwarzanie ze schowka, 206  
*Ogłada* 'Sagacity – zmienna, 146  
 ogłada a zdolność, 147  
*Ogól*, próba badawcza, 110  
 okres pomiarów, 42  
 opcje  
     domyślne, 205  
     grafiki, 206, 207  
     właściwości tabel, 209  
     właściwości wykresów, 215  
 opcje wypowiedzi, 87, 89, 114  
 operatory działań, 155  
*Opinie* 'Opinion – zmienna, 148  
 opinie a poglądy, 148  
 oprogramowanie *shareware*, 13  
 optymalizacja skali, 68  
*Osąd* 'Verdict – zmienna, 146  
 osąd a wgląd, 147  
*Oznaki* 'Signs, 153  
     uświadomienia, 49  
     wartości, 25  
 oznaki a symptomy, 25  
 oznakowanie zasobów, 124  
 pakiet oprogramowania, 13  
 pamięć podręczna, 206  
 parametry statystyczne, 154  
     bazowe, 156  
     fluktuacji, 189  
     niemianowane, 191  
     zmiany, 181, 185  
 pary komplementarne  
     czynników, 238  
     składników, 212, 232  
 pary kontrolne, 51  
     wskaźników, 65, 211  
 pasek stanu, 201  
 pasek tytułowy, 201  
 permutacje, 45  
*Pewność* 'Aplomb – zmienna, 146  
 pewność a obycie, 147  
 plany badawcze, 19, 20  
     grup skrzyżowanych, 44  
     oceny z pełną permutacją, 45  
 pliki pakietu, 13  
 podmioty ewaluacji, 102  
 podmioty kształcenia, 24  
 podprzedziały skali, 53  
 podstawy uogólnień, 91  
*Poglądy* 'Notion – zmienna, 148  
 poglądy a opinie, 148  
 polaryzacja bodźców-stwierdzeń, 32  
 polaryzacja opinii, 25  
 pomiar, 18, 124  
 populacja, 15  
 porównywalność efektów, 88  
 porównywanie  
     cech mentalnych, 43  
     grup i klas, 43  
     obiektów podobnych, 45  
*Postawa* 'Attitude, 17, 48  
 powtarzalność wyników, 242  
*Poziomy* 'Levels, 82  
     – parametry [HML], 170  
 poziomy odniesienia, 221  
*Poznanie* 'Cognition – zmienna, 50, 150  
 poznanie a doznanie, 18, 48, 151  
 prawa autorskie, 12  
 predykcja, 86  
 predylekcyjna definicja postaw, 48  
 problemy badawcze, 9, 99  
     naukowe, 20  
     praktyczne, 100  
 proces tworzenia skali, 28  
 progi wymagań, 22  
 projekt badawczy, 96  
 próba badawcza, 107  
 próbkowanie, 41  
 przedmiot ewaluacji, 100  
 przedziały ocen, 87  
 przedziały ufności, 53, 168  
 przekroje analityczne, 77  
 przekroje statystyczne, 109  
 przemiany mentalne, 21  
 przesłanki walidacyjne, 141  
 przesunięcie rozkładu, 180  
 przetwarzanie danych, 128  
 przyciski ikoniczne, 209  
 przykłady obliczeń  
     *Agregacja wskaźników*, 158  
     *Bilans*, 183  
     *Częstość realna*, 173



- Efekt*, 190  
*Fluktuacja*, 187  
*Jakość*, 160  
*Konkluzyjność*, 193  
*Korelat*, 184  
*Odchylenia realne*, 162  
*Odchylenie standardowe*, 164  
*Oznaka nierzetelności*, 198  
*Poziomy HML*, 170  
*Przyrost*, 189  
*Rozrzut*, 165  
*Rozziew*, 161  
*Rzetelność*, 197  
*Selektywność*, 194  
*Skośność*, 171  
*Spadek*, 189  
*Spójność*, 166  
*Standaryzacja rozkładu*, 176  
*Średnie kwadratowe*, 157  
*Trafność*, 195  
*Trend*, 186  
*Ufność realna*, 169  
*Zasięg*, 190  
*Zawiłość*, 182  
*Zmiana*, 180  
*Przyrost 'Increase* – parametr [+e], 188, 237  
 pytania badawcze, 20  
 reakcje-wypowiedzi, 68  
 refleksja, 149  
 retrospekcja a futurospekcja, 30, 65  
 rezultaty zmiennych globalnych, 150  
 rodzaje wykresów, 214  
 role i zadania w ewaluacji, 15  
 rozbieżność symptomów, 238  
 rozdzielczość ekranu, 200  
 rozdzielczość skali, 173, 196  
 rozkład, 83  
 rozkład liczebności, 164, 170  
 rozproszenie oznak, 224  
*Rozrzut 'Dispersion* – parametr [d], 164, 210, 226  
*Rozwaga 'Thought* – zmienna, 148  
 rozwaga a emocje, 149  
*Rozziew 'Bias* – parametr [-b], 161, 210, 227  
 rozziew wypowiedzi, 38, 58  
 różnice a wektory, 191  
 różnicowanie respondentów, 23  
*r-Pearsona*, 184  
*Rzetelność 'Reliability*, 84  
   – parametr [r], 196, 213  
 rzetelność wskaźnikowania, 73  
 scalanie danych, 203  
 schemat
- agregacji, 129, 211, 212  
 ekstrakcji, 29, 51, 79  
 interpretacji, 144, 146, 148, 150  
 polaryzacji, 128  
 schowek, 206  
*Selektywność 'Selectivity*, 84  
   – parametr [š], 194  
 selektywność skali, 26, 53  
 sfery wartościowania, 80  
*shareware*, oprogramowanie, 13  
 siła wnioskowania, 193, 227  
 skala  
   dwuważona, 52  
   ilorazowa, interwałowa, porządkowa, 53  
 skala a kwestionariusz, 59  
*Skala dwuważonych ocen*, 10, 52  
   '*Two-weighted ratings scale*, 54  
*Skala sumowanych ocen Likerta*, 52  
 skale ocen szkolnych, 87  
 skalowalność oznak, 88  
 składanie wektorów, 191  
*Składniki 'Elements* – zmiennie, 146  
   tabele, 211, 229, 231, 233  
   wykresy, 223, 233  
*Skośność 'Skewness*, 83  
   – parametr [s], 171, 225  
 skrypt ewaluacyjny ocen, 87  
 słuszność wnioskowania, 195  
*Spadek 'Decrease* – parametr [-e], 188, 237  
 splot cech mentalnych, 29, 49  
*Spokój 'Calmness* – zmienna, 146  
 spokój a odpór, 147  
*Spójność 'Consistency* – parametr [c], 166, 229  
 standardy, 22  
   ewaluatywne, 15, 56, 91, 211, 222  
 standaryzacja parametru, 185  
 standaryzacja rozkładu, 176  
 standaryzacja skali, 220  
 statystyki ewaluacji, 13  
 strategia pomiaru, 105  
 strategie ewaluacji, 27, 50  
   *dwuważenie*, 57  
   *minimaksowa*, 73  
   *splotowa*, 10  
 strategie oceniania, 22  
 strefa zgodności, 227  
 stres, 147  
 struktura książki, 10  
 struktura postaw, 48  
 strukturyzacja danych, 126  
 strukturyzacja oznak, 131  
 style wyjaśniania, 30

- style wypowiedzania się, 21  
 subiektywizm kryteriów, 22  
 sumowanie warunkowe, 156  
 symbole parametrów, 154, 155  
*Symptomy* 'Symptoms, 24  
 symptomy a oznaki, 25  
 synteza  
   jakości, 86  
   wartości, 140  
 ścieżki koncygowania, 46  
 ścieżki przejść przez opis, 11  
*Ślady... Wzorce*, 18, 27, 47, 58, 75, 86, 94,  
   110, 127, 135, 142, 152, 157, 165, 177,  
   186, 191, 198, 213, 218, 230, 236, 244  
 średnie, 156  
   empiryczne, 160  
   oczekiwane, 168, 222  
   przeskalowane, 167  
*Tabele* 'Tables, 209  
 technika buforowania, 90  
 technika minimaksowa, 34  
 technika pomiaru, 124  
 techniki ewaluacji, 13, 27  
   analiza dwuwymiarowa, 38  
   justowanie estymatorów, 35  
   komparacja oznak, 39  
   porównywanie obiektów, 43  
   próbkiowanie stanów, 41  
   ustalanie znaków walencyjnych, 31  
   ważenie zdań, 33  
 tendencje przemian, 83  
 tolerancja średnich, 168  
*Trafność* 'Aptness, 84  
   – parametr [â], 195, 244  
 trafność wskaźników, 32, 61  
*Trend* 'Trend – parametr [±t], 185  
 trudność oceniania, 21  
 tworzenie narzędzia, 111, 245  
*Ufność realna* 'Real Confidence, 82  
   – parametry [-c] [+c], 168  
 uogólnienie  
   ewaluacji, 245  
   indykatum, 151  
   wyników, 91  
 uprzyśtępnianie wyników, 87  
 ustalanie znaków walencyjnych, 31  
 usuwanie plików, 204  
*Uświadomienie* 'Consciousness, 49  
 uświadomienie informatyczne, 17  
 uzależnienie, 147  
 wady sondaży społecznych, 35  
 wady strategii korelacyjnych, 26  
 wagi, 53, 158  
 wahliwość przemian, 187  
 wahliwość wypowiedzi, 237  
 walencja, ambiwalencja, 32  
*Walidacja* 'Validation, 23  
 wartości, 23, 25, 140  
 wartości oczekiwane, 220  
 wartościowanie poziomów, 222  
 wartościowość zmian, 182  
*Wartość* 'Value – parametr [V], 167  
 wartość i ważkość wypowiedzi, 57  
 ważenie zdań, 35  
*Ważkość* 'Weight – parametr [W], 167  
 ważkość i wartość wypowiedzi, 57  
 wczytanie danych lub wykresu, 202  
 wektory a różnice, 191  
 wektory fluktuacji  
   składowe, 188  
   wynikowe, 191, 240  
 wektory składowe trendu, 186  
 wersje językowe pakietu, 208  
 weryfikacja empiryczna skali, 68  
 weryfikacja wskaźników, 34, 70, 114  
*Wgląd* 'Insight – zmienna, 146  
 wgląd a osąd, 147  
 wiarygodność wniosków, 243  
 wiedza, 148  
 wielkości prób, 210  
 wizualizacja oznak, 136  
 właściwości wskaźników, 61  
 właściwości wykresów, 215  
*Wprawa* 'Practice – zmienna, 148  
 wprawa a motywacje, 149  
 wskaźnik buforowy, 90  
*Wskaźniki* 'Indicators – zmienne, 17,  
   60, 144  
   tabele, 210, 222, 225, 227  
   wykresy, 228, 230, 234  
 wskaźniki uświadomienia, 49  
 współbieżność oznak, 232  
 współczynnik korelacji, 184  
 współczynnik rzetelności, 213  
 współzależność zmian, 184  
 wybór narzędzia, 52  
 wybór symptomów, 24  
 wykładnia indykatum, 16  
 wykładnia metody, 14  
 wykresów opcje, 215  
 wykresów rodzaje, 137, 214, 218  
   biegunowy kierunków, 179, 217  
   histogram a rozkład, 177  
   odcinkowy poziomów, 216  
   pasmowy rozrzutu, 217  
   punktowy średnich, 216  
   słupkowy tendencji, 218  
   wektorowy dynamiki, 218

- wykresy  
 amplitudy, 191, 218, 238, 241  
*duuuważenie*, 57  
 dynamika, 191, 218, 238, 241  
 efekty, 238, 241, 244  
 estymacja, 177  
 fluktuacja, 188, 189  
 gradient, 217, 234  
 histogramy, 32, 38, 56, 70, 161, 163  
 kierunki, 179, 217, 234  
 komparacja, 39, 236, 241, 244  
 korelacja, 185  
 kwantyzacja, 41  
 momenty, 159  
 normy, 230, 241  
 odchylenia, 164  
 poziomy, 39, 216, 226, 233  
 przedziały ufności, 169
- wykresy (cd.)  
 rozkłady, 71, 159, 164, 171, 172  
 rozrzuty, 35, 37, 217, 228, 236  
 selektywność, 195  
 skale wag i ocen, 167, 170  
 skośność, 171  
 średnie, 43, 44, 216, 223, 232, 243  
 tendencje, 40, 218, 236, 239, 244  
 zmiany HML, 181, 186, 241
- wymiary wypowiedzi, 56
- Wyniki 'Findings* – zmienne, 85, 150  
 tabele, 213, 235, 237, 240, 242, 244  
 wykresy, 236, 238, 241, 243, 244
- wyznaczniki jakości skali, 192
- wyznaczniki wartości, 18
- względność poziomów, 88
- względność wypowiedzi, 221
- wzory  
*Baza* [B'], 168  
*Bilans* [ $\pm b$ ], 182  
*Częstość realna* [ $f(x)$ ], 172  
*Efekt* [ $\pm e$ ], 190  
*Fluktuacja* [ $\bar{F}$ ], 187  
*Intensywność* [I], 160  
*Jakość* [Q], 160  
*Konkluzyjność* [ $\check{c}$ ], 193  
*Korelat* [ $\pm r$ ], 184  
*Norma* [N''], 168  
*Odchylenia realne* [-d] [+d], 162  
*Odchylenie standardowe* [ $\sigma$ ], 163  
*Poziomy* [HML], 170  
*Przyrost* [+e], 188  
*Rozrzut* [d], 165  
*Rozziew* [-b], 161  
*Rzetelność* [ $\check{r}$ ], 196
- Selektywność* [ $\check{s}$ ], 194  
*Skośność* [s], 171  
*Spadek* [-e], 188  
 średnie, 156  
*Trafność* [ $\check{a}$ ], 195  
*Trend* [ $\pm t$ ], 185  
*Ufność realna* [-c] [+c], 169  
 wagi [w], 158  
*Wartość* [V], 167  
*Ważkość* [W], 167  
*Zasięg* [e], 190  
*Zawiłość* [ $\pm i$ ], 181  
*Zbieżność* [ $\pm c$ ], 183  
*Zmiana* [A ], 180
- zadania badawcze, 20
- zadania ewaluatorów, 14  
 i role w ewaluacji, 15
- zakres czasowy ewaluacji, 103
- założenia  
 metodologiczne, 10  
 pomiarowe, 10  
 redakcyjne, 9
- zamykanie programu, 208
- zapamiętanie opcji, 215
- zapisanie do pliku, 204, 214
- zasady używania pakietu, 13
- Zasięg 'Extent* – parametr [e], 190, 240
- zasięg przestrzenny ewaluacji, 103
- zasoby, 202
- zaspokojenie, 149
- Zawiłość 'Intracacy*, 83  
 – parametr [ $\pm i$ ], 181, 213, 235
- zaznaczanie komórek tabeli, 206, 207
- Zbieżność 'Convergence* – parametr [ $\pm c$ ], 183
- zbieżność wypowiedzi, 229
- zbiory danych empirycznych, 156
- Zdolność 'Ability* – zmienna, 146
- zdolność a ogłada, 147
- zgodność respondentów, 25
- Zmiana 'Alteration*, 83  
 – wektory [A ], 180
- zmiana a fluktuacja, 187
- zmienna ogólna, 17
- zmiennie  
 cząstkowe, 146  
 globalne, 150  
 splotowe, 148
- zmiennność wypowiedzi, 231
- znak sumy, 156
- znaki i oznaczenia, 155
- róźnicowanie wewnątrzspójne, 196
- róźnicowanie wypowiedzi, 25

## Wykaz rycin

1. Trzy wersje ekstrakcji symptomów splotowych dla skali 24-pozycyjnej .....	29
2. Ustalanie polaryzacji bodźców-stwierżeń na podstawie histogramów .....	32
3. Ważenie zdań na podstawie empirycznego rozrzutu wskaźników .....	35
4. Justowanie estymatorów w celu unormowania wskaźników .....	37
5. Histogram ilustrujący zjawisko rozziwu wypowiedzi .....	38
6. Komparacja: a) poziomu wskaźników, b) współbieżności składników .....	39
7. Komparacja tendencji par czynników i wyników połówkowych .....	40
8. Próbkowanie dwukrotne z kwantyzacją progów natężenia cechy .....	41
9. Porównanie skuteczności oddziaływania tego samego nauczyciela .....	43
10. Eksperyment z grupami skrzyżowanymi i trzema pomiarami .....	44
11. Schematy koncipowania dwubieżnego .....	46
12. Schemat ekstrakcji czynników, składników i wskaźników indykatum .....	51
13. Graficzna reprezentacja wzorca skali dwuważonej .....	53
14. Odmienność histogramów jakości i intensywności wypowiedzi .....	56
15. Wartość wypowiedzi a standardy ewaluatywne .....	57
16. Ważkość wypowiedzi .....	57
17. Rozziew wypowiedzi .....	58
18. Rozkłady obligujące do odrzucenia niekonkluzyjnych wskaźników .....	70
19. Aproksymowane rozkłady wskaźników <i>InfoKultury</i> .....	71
20. Zwarta struktura danych surowych w formacie .dxt .....	76
21. Organizacja zmiennych $X'$ i $X''$ w macierzach dwuwymiarowych .....	77
22. Przekrój przez przypadki z wyodrębnieniem grupy [A] i dwóch klas [a, b] .....	78
23. Przekroje potrzebne do wyodrębnienia wskaźników oraz składników .....	79
24. Graficzna interpretacja momentów zwykłych, centralnych i absolutnych .....	159
25. Graficzna interpretacja miary rozziwu określającej niezgodność w próbie .....	161
26. Odchylenia realne wierniej odwzorowują dewiacje rozkładu .....	163
27. Realne miary dewiacji i dyspersji rozkładu empirycznego .....	164
28. Konwersja ze skali wag na skalę ocen .....	167
29. Przedziały ufności realnej zależne od liczebności, rozrzutu i skośności .....	169
30. Wyższa wiarygodność liczebności w przedziale HL niż ufność do poziomu M .....	170
31. Skośność – przykłady asymetrii a) lewostronnej, b) prawostronnej .....	171
32. Gęstość prawdopodobieństwa i dystrybuanta rozkładu realnego .....	172
33. Histogram rozkładu empirycznego i krzywa rozkładu estymowanego .....	177
34. Wykres biegunowy – obraz zmian wskaźników .....	179
35. Przesunięcia momentów i wektory zmian poziomów HML .....	180
36. Przykłady interpretacji zawilości oddziaływań edukacyjnych .....	181
37. Silna korelacja między wynikami 'przed' i 'po' .....	185
38. Wektory składowe trendu dla trzech różnych sytuacji empirycznych .....	186
39. Wektory składowe fluktuacji wypowiedzi kolejnych $n$ - respondentów .....	188
40. Usytuowanie parametrów fluktuacji .....	189
41. Tworzenie wektorów wynikowych fluktuacji dla $k$ - grup .....	191
42. Selektowność – oszacowanie istotności różnic pomiędzy średnimi .....	195

43. Rozdzielczość i koherentność macierzy wskaźników .....	196
44. Program instalacyjny pakietu .....	200
45. Okno startowe Analizatora .....	201
46. Menu zasobów .....	202
47. Wczytywanie .....	202
48. Korekta błędu w zbiorze danych .....	203
49. Scalanie danych .....	203
50. Zapisywanie i usuwanie zasobów .....	204
51. Schemat sekwencji przy analizie automatycznej .....	205
52. Zaznaczanie aktywnego obszaru tabeli .....	206
53. Rozkład empiryczny odtworzony ze schowka .....	206
54. Dwa rodzaje wykresów – półtonowy i kreskowy .....	207
55. Okienko informacyjne .....	207
56. Opisy tabel i wykresów w języku angielskim .....	208
57. Menu tabel .....	209
58. Przekroje – próby i miary .....	210
59. Selekcyjne kwantyfikatory wskaźników .....	210
60. Porównywanie wskaźników dwóch grup z poziomami normy .....	211
61. Porównywanie składników z poziomami normy .....	211
62. Czynniki, wyniki połówkowe i wynik ogólny dwóch grup .....	212
63. Tabela porównawcza wyników dziewięciu grup oraz ogółu .....	213
64. Menu wykresów .....	214
65. Opcje wykresów .....	215
66. Średnie trzech grup .....	216
67. Poziomy N" i H"M"L" .....	216
68. Rozrzut i rozziw wskaźników .....	217
69. Kierunki zmian .....	217
70. Tendencje .....	218
71. Dynamika przemian .....	218
72. Autonormalizacja poziomu składników w trakcie procesu edukacji .....	223
73. Autonormalizacja poziomu czynników i wyników ogólnych .....	224
74. Poziomy HML czynników i wyników porównawczo prób X i Y .....	226
75. Rozrzut wskaźników połączonych prób X i Y względem strefy zgodności .....	228
76. Poziomy kontrolnych par wskaźników dla połączonych prób X i Y .....	230
77. Zmiany poziomów średnich z porównaniem dwóch różnych grup .....	232
78. Współbieżność par składników – zmiany HML połączonych prób X i Y .....	233
79. Wykresy biegunowe do analizy harmonii i kierunków zmian .....	234
80. Wykresy rozkładowej i różnicowej analizy wyników klas licealnych .....	236
81. Wykres indeksowej analizy wyników klas licealnych wybranych z próby Y .....	238
82. Tendencje zmian par czynników, porównawczo dla prób X i Y .....	239
83. Efektywność procesów porównawczo dla wszystkich grup z prób X i Y .....	241
84. Czynniki, wyniki połówkowe oraz wynik ogólny grupy A .....	241
85. Zmiany poziomów <i>InfoKultury</i> formowane przez różnych wykładowców .....	243
86. Porównanie Trendu i Efektu jako test trafności wyników .....	244

## Wykaz tabel

I. Schemat interpretacji wskaźników Kwestionariusza Kultury Informatycznej .....	144
II. Schemat interpretacji składników Kwestionariusza Kultury Informatycznej .....	146
III. Schemat interpretacji czynników Kwestionariusza Kultury Informatycznej .....	148
IV. Schemat agregowania rezultatów połówkowych <i>InfoKultury</i> .....	150
V. Normy standaryzujące Kwestionariusz Kultury Informatycznej .....	220
VI. Baza i Norma wskaźników u studentów X, licealistów Y i gimnazjalistów Z .....	222
VII. Odchylenia realne i Skośność rozkładu wskaźników po scaleniu prób X i Y .....	225
VIII. Rozrzut, Rozziew i Konkluzyjność wskaźników po scaleniu prób X i Y .....	227
IX. Spójność par kontrolnych i Zbieżność wypowiedzi z prób X, Y i łącznie .....	229
X. Zmiana poziomów średnich i Korelat składników w grupach E i F .....	231
XI. Zmiana, Trend i Bilans składników po scaleniu prób X i Y .....	233
XII. Zmiany poziomów, Zawilość i Bilans wyników w klasach licealnych .....	235
XIII. Zasięg, Przyrost, Spadek i Efekt wyników w klasach licealnych .....	237
XIV. Różne kierunki zmian poziomów, Bilans i Trend dla prób X i Y .....	238
XV. Trend i Efekt do porównań wyników grup z różnych populacji .....	240
XVI. Zmiana średnich a Selektowność i powtarzalność wyników .....	242
XVII. Trafność wyników lokalnych jako podstawa wiarygodności .....	244

## Przypisy z literaturą źródłową i rozszerzającą

- <sup>1</sup> Koncepcja modelu-wzorca ewaluacji – Ubermanowicz S.: *Specyfika Skali ważonych ocen*, [w] „Neodidagmata” XXIII 1996-1997, s. 63-81, Wydawnictwo Naukowe UAM, Poznań 1997.  
Prezentacja metody i wyników – Ubermanowicz S.: *Ewaluacja zajęć informatycznych*, [w] XIII Konferencja „Informatyka w szkole”, s. 101-109, wydanie UWr-UMCS, Lublin 1997.
  - <sup>2</sup> Pierwzór narzędzia – Vidacovic D., Paprzycki M.: *Using Computers in Calculus Teaching*, [w] „The Journal of Computing in Small Colleges” 1993, 8, s. 34-45.
  - <sup>3</sup> Prezentacja wyników adaptacji kulturowej – Paprzycki M., Vidacovic D., Ubermanowicz S.: *Comparing Attitudes Toward Computers of Polish and American Prospective Teachers*, [w] SITE-95 San Antonio, Publ. AACE, Charlottesville 1995.
  - <sup>4</sup> Koncepcja włączenia do skali wskaźników uświadomienia – Ubermanowicz S., Paprzycki M.: *Między stylistyką a statystyką w teście kultury informatycznej*, [w] „Neodidagmata” XXII 1994-1995, s. 87-106, Wydawnictwo Naukowe UAM, Poznań 1996.  
Prezentacja koncepcji zmian kwestionariusza – Paprzycki M., Ubermanowicz S.: *Studying Computer Awareness*, [w] SITE-96 Phoenix, s. 841-846, Publ. AACE, Charlottesville 1996.
- - -
- <sup>5</sup> Próby zdefiniowania pojęcia ewaluacji – Malenda A.: *Ewaluacja*, [w] *Encyklopedia pedagogiczna XXI wieku*, tom 1, s. 1091-1094, Warszawa 2003, a także Radnor H.: *Ewaluacja*, Wyd. MEN, Warszawa 1996.
  - <sup>6</sup> Zbiory różnych tekstów z wykładnią ewaluacji – Korporowicz L. (red.): *Ewaluacja w edukacji*, Program TERM, Oficyna Naukowa, Warszawa 1997 oraz Mizerek H., Hildebrandt A. (red.): *Ewaluacja w szkole: wybór tekstów*, Program TERM, Olsztyn 1997.
  - <sup>7</sup> Krytyczna recenzja pseudonaukowego podejścia do procesu ewaluacji – Ubermanowicz S.: *B. Niemierko, Próbną maturą 1997 – Ewaluacja metodologiczna*, [w] „Neodidagmata” XXIV 1998-1999, s. 186-192, Wydawnictwo Naukowe UAM, Poznań 1999.
  - <sup>8</sup> Opracowania dotyczące ewaluacji są w Polsce najczęściej materiałami szkoleniowymi lub dokumentami projektów badawczych, np.: *Raport końcowy ewaluacji zewnętrznej szkoleń na rzecz zarządzania oświatą (TERM) 1995-1997*, PHARE 1997 oraz *Ewaluacja wdrażania reformy systemu edukacji: materiały seminaryjne*, PHARE, Warszawa 1999.  
Publikacje zbiorcze odnoszą się przede wszystkim do oceny jakości systemu edukacyjnego: Cierzniewska R. (red.): *Ewaluacja jakości pracy nauczyciela*, Wyd. Akad. Bydgoskiej, 2003. Kropiwnicki J. (red.): *Mierzenie jakości w szkole*, Wyd. Nauczycielskie, Jelenia Góra 2002. Wenta K. (red.): *Diagnoza i ewaluacja w reformie edukacyjnej*, Uniwersytet Szczeciński, 2002. Wenta K., Zeidler W. (red.): *Diagnoza pedagogiczno-psychologiczna wobec zagrożeń transformacyjnych*, Uniwersytet Szczeciński, 2003.  
Także artykuły dotyczą głównie edukacji: Denek K.: *Ewaluacja rezultatów edukacji szkolnej*, cz. 1-3 [w] „Wychowanie na Co Dzień” 1998, nr 10/11, nr 12 oraz 1999, nr 1/2, ponadto *Metodologiczne aspekty diagnozy i ewaluacji edukacyjnej*, [w] „Nowe w Szkole” 2000/2001, nr 5 (i inne tegoż autora). Szymela K.: *Teoria i praktyka ewaluacji wewnętrznej*, [w] „Pedagogika Pracy” 1998, t. 33. Szypryt-Nowicka H.: *Nie od razu Kraków zbudowano, czyli po co ewaluacja*, „Język Polski w Liceum” 2001/2002, nr 3. Furgoł S., Szafraniec M.: *Zastosowanie ewaluacji w badaniu jakości pracy szkoły*, [w] „Dyrektor Szkoły” 2001, nr 7/8.

- Korporowicz L.: *Ewaluacja – zaproszenie do rozwoju*, [w] „Edukacja i Dialog” 2000, nr 6.
- Kubiak W.: *Działania edukacyjne przedmiotem ewaluacji pedagogicznej*, [w] „Edukacja Humanistyczna” 2002, nr 1/2.
- Wadowski D.: *Proces ewaluacji w praktyce szkolnej*, [w] „Język Polski w Liceum” 2002/2003, nr 2.
- Lisiecka Z.: *O potrzebie ewaluacji programów dydaktycznych i wychowawczych*, [w] „Język Polski w Liceum” 2003/2004, nr 4, ponadto *W stronę ewaluacji: szkoła w obliczu zmiany edukacyjnej*, [w] „Język Polski w Liceum” 2001/2002, nr 1.
- Majewski M.: *Ewaluacja – jej funkcje i znaczenie*, „Asocjacje” 1999, nr 9; *Ewaluacja – jak to łatwo powiedzieć*, „Personel” 2000, nr 7 oraz *Ewaluacja – narzędzie wprowadzania i oceny standardów jakości*, [w] „Edukacja i Dialog” 2000, nr 6.
- <sup>9</sup> Monografie naukowego podejścia do badań – Frankfort-Nachmias Ch., Nachmias D.: *Metody badawcze w naukach społecznych*, wyd. I, Zysk i S-ka Wydawnictwo s.c., Poznań 2001.
- Brzeziński J.: *Metodologia badań psychologicznych*, wyd. IV, WN PWN, Warszawa 2003.
- <sup>10</sup> Charakterystyka różnych modeli oceniania – Ornstein A.C., Hunkins F.P.: *Program szkolny – założenia, zasady, problematyka*, s. 313-345, WSiP, Warszawa 1998.
- <sup>11</sup> Wykładnia wartości procesów edukacyjnych oraz kierunki jej kształtowania – Denek K.: *Aksjologiczne aspekty edukacji szkolnej*, wyd. III, Wyd. Adam Marszałek, Toruń 2002.
- <sup>12</sup> Rola stylu wyjaśniania w bipolarnych wymiarach: czasokres (chwilowość | stałość), zasięg (uniwersalność | ograniczoność), personalizacja (zewnętrzna | wewnętrzna) – Seligman M.E.: *Optymizmu można się nauczyć*, Wydawnictwo Media Rodzina of Poznań, 1993.
- <sup>13</sup> Inferencyjne pary strategii badawczych: abstrahowanie | kondensowanie, analiza | synteza, stany | procesy i inne – Bazewicz M., Collen A.: *Podstawy metodologiczne systemów ludzkiej aktywności i informatyki*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1995.
- <sup>14</sup> Wydaje się zasadne wprowadzenie nazw *kwalifikacja* i *kwalizacja* dla procesów przeciwbieżnych *kwantyfikacji* i *kwantyzacji* – zob. słowniki wyrazów obcych np. W. Kopalińskiego.
- <sup>15</sup> Trójskładnikowa definicja postawy – Mika S.: *Psychologia społeczna*, PWN, Warszawa 1982.
- <sup>16</sup> Różne ujęcia cech kognitywnych, afektywnych i behawioralnych – Maruszewski T. (red.): *Poznanie, afekt, zachowanie*, Wydawnictwo Naukowe PWN, Warszawa 1993.
- <sup>17</sup> Aktualnie dominujące ujęcie postawy – Wojciszke B.: *Psychologia postaw i ocen*, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2001.
- <sup>18</sup> Zbiór różnorodnych wykładni świadomości – Nunn Ch.: *Awareness, What it is, What it does*, Routledge, London and New York 1996.
- <sup>19</sup> Braki uświadomienia dotyczą nie tylko komputeryzacji, lecz ogólnie mediów i mass mediów – Strykowski W., Skrzydlewski W. (red.): *Media i edukacja w dobie integracji*, Wyd. eMPI<sup>2</sup>, Poznań 2002 oraz *Kompetencje medialne społeczeństwa wiedzy*, Wyd. eMPI<sup>2</sup>, Poznań 2004.
- Strykowski W. (red.): *Media a edukacja*, t. 1–3, Wyd. eMPI<sup>2</sup>, Poznań 1997, 1998, 2000.
- <sup>20</sup> Znaczenie procesów kształtowania świadomości ekologicznej – Domka L.: *Kryzys środowiska a edukacja dla ekorozwoju*, Wyd. Naukowe UAM, Poznań 1996.
- <sup>21</sup> Relatywność zachowań badanych – Brzeziński J., Siuta J. (red.): *Społeczny kontekst badań psychologicznych i pedagogicznych. Wybór tekstów*, Wyd. Naukowe UAM, Poznań 1991.
- <sup>22</sup> Dostępne są zaawansowane techniki analizy połówkowej skal – zobacz np. opis modułu *Analiza rzetelności/pozycji* w pakiecie STATISTICA firmy StatSoft: [www.statsoft.com](http://www.statsoft.com)



- <sup>23</sup> Wzorce schematów badawczych – Żechowska B.: *Wybrane metodologiczne wzory badań empirycznych w pedagogice*, Skrypt Uniwersytetu Śląskiego, nr 363, Katowice 1985.
- <sup>24</sup> Ekstrakcja komponentów indykatorum – Zakrzewska M.: *Analiza czynnikowa w budowaniu i sprawdzaniu modeli psychologicznych*, Wyd. Naukowe UAM, Poznań 1994 oraz Górniak J.: *Analiza czynnikowa i analiza głównych składowych*, [w] „ASK” 1998, nr 7, s. 83-102.
- <sup>25</sup> Skala Likerta oraz inne strategie badawcze – Robson C.: *Real World Research. A Resource for Social Scientists and Practitioner-Researchers*, Blackwell, Oxford–Cambridge 1993.
- <sup>26</sup> Wykładnia standardów ewaluacyjnych – Gołąb A., Reykowski J.: *Studia nad rozwojem standardów ewaluacyjnych*, Wyd. PAN, Ossolineum, Wrocław (i in.) 1985.
- <sup>27</sup> Wskaźnikowanie zmiennych – Sztumski J.: *Wstęp do metod i technik badań społecznych*, Wyd. Uniwersytetu Śląskiego, Katowice 1997.
- <sup>28</sup> Symptomy postaw i ich wskaźnikowanie – Marody M.: *Sens teoretyczny a sens empiryczny pojęcia postawy. Analiza metodologiczna zasad doboru wskaźników w badaniach nad postawami*, PWN, Warszawa 1976.
- <sup>29</sup> Operacjonalizacja pojęć informatycznych – Ubermanowicz S., Bielawska H.: *Czynnościowe kształtowanie pojęć*, [w] „Neodidagmata” 2003, nr 25/26, Wyd. Naukowe UAM, Poznań 2003.
- <sup>30</sup> Kwestionariusze, kwantyfikatory – Sztabiński P.B.: *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych*, Wydawnictwo IFiS PAN, Warszawa 1997.
- <sup>31</sup> Zagadnienia jakości narzędzia – Brzeziński J. (red.): *Problemy teorii, rzetelności, konstrukcji i analizy wyników testów psychologicznych*, [w] „Biblioteka Psychologa Praktyka”, t. II, Wyd. Polskiego Towarzystwa Psychologicznego, Warszawa 1988.
- <sup>32</sup> Różnice pomiędzy ujęciem ilościowym a jakościowym – Pilch T., Bauman T.: *Zasady badań pedagogicznych. Strategie ilościowe i jakościowe*, Wydawnictwo „Żak”, Warszawa 2001.
- <sup>33</sup> Metody interpretacji jakościowych zasobów empirycznych – Milles M.B., Huberman A.M.: *Analiza danych jakościowych*, Wydawnictwo Trans Humana, Białystok 2000.
- <sup>34</sup> Istota i sposoby wyznaczania indeksów statystycznych – Ignatczyk W., Chromińska M.: *Statystyka. Teoria i zastosowanie*, wydanie Wyższej Szkoły Bankowej, Poznań 1999.
- <sup>35</sup> Kompendium wiedzy zogniskowanej na czynnościowym podejściu do badań – Babbie E.: *Badania społeczne w praktyce*, Wydawnictwo Naukowe PWN, Warszawa 2003.
- <sup>36</sup> Cele oraz pułapki ewaluacji w praktyce edukacyjnej – Brzezińska A., Brzeziński J. (red.): *Ewaluacja procesu kształcenia w szkole wyższej*, Wyd. Fundacji Humaniora, Poznań 2000 oraz Brzezińska A., Brzeziński J., Elias A. (red.): *Ewaluacja a jakość kształcenia w szkole wyższej*, Wydawnictwo SWPS Academica, Warszawa 2004.
- <sup>37</sup> Technika doboru prób badawczych – Steczkowski J.: *Metoda reprezentacyjna w badaniu zjawisk ekonomiczno-społecznych*, Wydawnictwo Naukowe PWN, Warszawa–Kraków 1995.
- <sup>38</sup> Kodeks postępowania etycznego w prowadzeniu badań naukowych – *Dobre obyczaje w nauce. Zbiór zasad*, wyd. II, Komitet Etyki w Nauce przy Prezydium PAN, Warszawa 1996.
- <sup>39</sup> Zasady definiowania pojęć – Hornowska E.: *Operacjonalizacja wielkości psychologicznych. Założenia – struktura – konsekwencje*, Wydawnictwo Ossolineum, Wrocław 1989.
- <sup>40</sup> Tworzy się systemy eksperckie, lecz personifikacja komputera to przesada – Johnson-Laird Ph.: *Komputer i umysł. Wstęp do nauk poznawczych*, Wydawnictwo Protex, Poznań 1999.

- <sup>41</sup> Porównaj inną definicję *opinii*, wyzbytą z aspektów emocjonalnych i niesłusznie utożsamianą z *poglądami* – Aronson E.: *Człowiek istota społeczna*, Wyd. Naukowe PWN, Warszawa 2002.
- <sup>42</sup> Ocenianie impulsywne, afektywność przed kognytywnością – Zajonc R.: *Uczucia a myślenie: nie trzeba się domyślać, by wiedzieć, co się woli*, „Przegląd Psychologiczny” 1985, nr 1.
- <sup>43</sup> Obawy przed statystyką są powszechne nie tylko w Polsce, zob. *Test lęku przed podręcznikami do statystyki – Statistics Textbook Anxiety Rating Test* – Schacht S.P.: *Statistics Textbooks: Pedagogical Tools or Impediments to Learning?* [w] „Teaching Sociology” 1990, nr 18.
- <sup>44</sup> Kompendium zagadnień i przykładów z dziedziny statystyki – Ferguson G.A., Takane Y.: *Analiza statystyczna w psychologii i pedagogice*, Wyd. Naukowe PWN, Warszawa 1997.
- <sup>45</sup> Wartości statystyki  $U(\alpha)$  dla dowolnego poziomu istotności  $[\alpha]$  można obliczyć w arkuszu kalkulacyjnym np. MS Excel na podstawie formuły =ROZKŁAD.NORMALNY.S.ODW(1- $\alpha$ /2) – zob.: *Tworzenie formuł: Funkcje statystyczne*, [w] *Microsoft Excel – Pomoc* (format cyfrowy).
- <sup>46</sup> Serwer grafiki nie daje możliwości wyłączenia wyświetlania co drugiej linii łączącej średnie, lecz do publikacji warto wykresy udoskonalić, przetwarzając je w zaawansowanym edytorze – zob. np. temat: *Modyfikuj rysunek clip art*, [w] *Microsoft Word – Pomoc* (format cyfrowy).
- <sup>47</sup> Toczy się dyskusja o treściach *Informatyki, Technologii informacyjnej i Edukacji medialnej* w szkole oraz o kompetencjach nauczycieli – Ubermanowicz S.: *Edukacja informatyczna: na rozdrożu, czy na bezdrożach?* [w] „Gazeta-IT” 2005, nr 9 (39), [http://www.gazeta-it.pl/edukacja/git35/edu\\_inf.html](http://www.gazeta-it.pl/edukacja/git35/edu_inf.html); Sysło M.M.: *Standardy przygotowania nauczycieli informatyki i technologii informacyjnej*, projekt Rady ds. Edukacji Informatycznej i Medialnej, 2004; Syllabus: *Edukacja medialna i informatyka* na Wydziale Studiów Edukacyjnych UAM, 2005.
- <sup>48</sup> Owe słuchaczki Studium Medycznego ujawniły znakomite cechy osobowości, sprzyjające realizacji trudnego programu edukacyjnego – Ubermanowicz S.: *Uczyć informatyki tak, by dotknąć jej sedna*, [w] materiały XI Konferencji „Informatyka w szkole”, Kielce 1995.
- <sup>49</sup> Pojęcie *kognitariusz* i inne kategorie: wartości, motywacje, emocje – Kozielski J.: *Koncepcje psychologiczne człowieka*, Wydawnictwo Akademickie „Żak”, Warszawa 2000.
- <sup>50</sup> Ewaluacja splotowa w funkcji wspierającej eksperyment – Michałowski M.: *Nowe technologicznie realizacje telewizyjne i ich przydatność w programach edukacyjnych*, praca doktorska dostępna w Zakładzie Technologii Kształcenia UAM, Poznań 2004.
- <sup>51</sup> Raport z ewaluacji – Ubermanowicz S.: *Wyniki pomiaru próbnej matury z matematyki*, [w] Hryhorowicz Z. (red.): *Nowa Matura – matematyka*, Kuratorium Oświaty, Poznań 1997.
- <sup>52</sup> Jakościowa ewaluacja kryteriów oceniania wypracowań z języka polskiego (s. 73-83) oraz ewaluacja splotowa próbnej matury z matematyki (s. 83-91) – Ubermanowicz S.: *Ewaluacja Nowej Matury*, [w] „Neodidagmata” XXIV 1998-1999, Wyd. Naukowe UAM, Poznań 1999.
- <sup>53</sup> Typowy przykład zaczerpnięty z materiałów OKE: Należało zastąpić słowo „skompilowanych” wyrazem rdzennie polskim. Jako wzorzec poprawnej odpowiedzi podano (!?) „utworzonych”. Uczeń celujący wie, że kompilacja to zapożyczenie i scalenie, a więc jedynie przetworzenie. Słowo zalecane jako wzorcowe w ogóle nie oddaje sensu użytego pierwotnie i nie pasuje do kontekstu: „... z list 'utworzonych' od Palestyńczyków”. Tekst rzekomo źródłowy pochodzi aż z trzeciej ręki i nie wiadomo, czy np. nie dotyczył komputerowej kompilacji z posortowaniem.