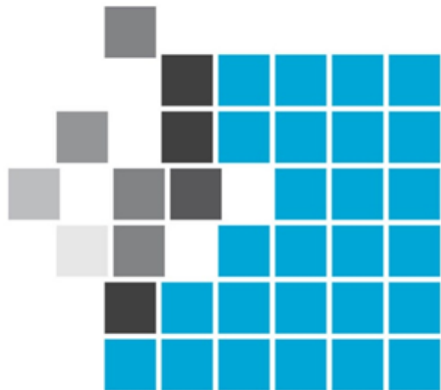




EDUKACYJNA WARTOŚĆ DODANA

Analiza własności psychometrycznych testu w heterogenicznych populacjach przy pomocy wielopoziomowej analizy czynnikowej

Tomasz Żółtak
Instytut Badań Edukacyjnych

Konferencja *Teraźniejszość i przyszłość psychometrii*,
Katowice 7-8.10.2015



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY





Założenie o homogeniczności populacji

- *Klasycznie* budując model zakładamy, że narzędzie pomiarowe (test) ma takie same własności psychometryczne dla wszystkich badanych (różnych grup badanych) – *mierzy to samo, w ten sam sposób*.
- W wielu sytuacjach jest to założenie problematyczne:
 - Międzynarodowe badania porównawcze: różnice pomiędzy adaptacjami narzędzia, różnice międzykulturowe.
 - Szeroko pojęte różnice *kulturowe* między różnymi grupami społecznymi (obeznanie z kontekstem, w którym osadzone są zadania, możliwość zrozumienia treści).
 - Występowanie *efektu oceniającego* w testach, w których rozwiązania zadań podlegają ocenie.
 - Analizy populacji *naturalnie* pogrupowanych (np. szkoły i oddziały w badaniach edukacyjnych), gdy są podstawy by uważać, że występuje segregacja ze względu na analizowany konstrukt pomiędzy grupami.



Założenie o homogeniczności populacji w analizie czynnikowej

W kontekście analizy czynnikowej założenie o homogeniczności populacji można rozpatrywać w odniesieniu do kilku elementów modelu:

- Wartości oczekiwanej konstruktu.
- Wariancji konstruktu.
- Trudności zadań.
- Dyskryminacji (ładunków czynnikowych) zadań.
- Ewentualnie innych parametrów opisujących zależność pomiędzy konstruktem a wynikami realizacji zadania (np. parametr pseudozgadywania w modelu IRT 3PL).



Różna waga łamania założenia o homogeniczności populacji

- Zróżnicowanie parametrów rozkładu konstruktu:
 - Stanowią wyzwanie dla poprawnej estymacji, ale same w sobie nie są niepożądaną cechą narzędzia pomiarowego.
- Zróżnicowanie parametrów opisujących związek między konstruktem a zmiennymi opisującymi wynik realizacji zadania:
 - Co do zasady stanowią niepożądaną cechę narzędzia pomiarowego.
 - Wyjątek: modele typu *multifaceted*, w modelach uwzględniających *efekt oceniającego*.
 - Analizy występowania takich zaburzeń noszą różne nazwy (i posługują się nieco innymi narzędziami statystycznymi) w różnych kontekstach teoretycznych:
 - IRT: *Differential Item Functioning*;
 - *klasyczna analiza czynnikowa*: *Measurement Invariance*.



Metody radzenia sobie z łamaniem założenia o homogeniczności populacji

Zróżnicowanie parametrów rozkładu konstruktów:

- Regresja latentna:
 - Pozwala opisać zróżnicowanie wartości oczekiwanej konstruktów (zarówno w funkcji cech mierzonych zmiennymi ciągłymi, jak porządkowymi/kategorialnymi).
 - Nie pozwala modelować zróżnicowania wariancji konstruktów pomiędzy grupami.
- Modele wielogrupowe z inwariancją pomiarową:
 - Pozwalają opisać zróżnicowanie zarówno wartości oczekiwanej jak i wariancji, ale tylko pomiędzy grupami.
- W odniesieniu do grup obie ww. metody mogą być **problematiczne przy dużej liczbie grup** (zwłaszcza przy niewielkiej liczbie obserwacji przypadających na grupę).



Metody radzenia sobie z łamaniem założenia o homogeniczności populacji

Zróżnicowanie parametrów trudności zadań wynikające z *efektu ocenającego*:

- Możliwe do zastosowania tylko w schematach badawczych, w których tą samą pracę ocenia kilku *egzaminatorów*.
- Wieloaspektowy (*multifaceted*) model Rascha, (*Generalised*) *Linear Logistic Test Model*.
- Bardziej złożone, specjalistyczne modele *efektu ocenającego*.



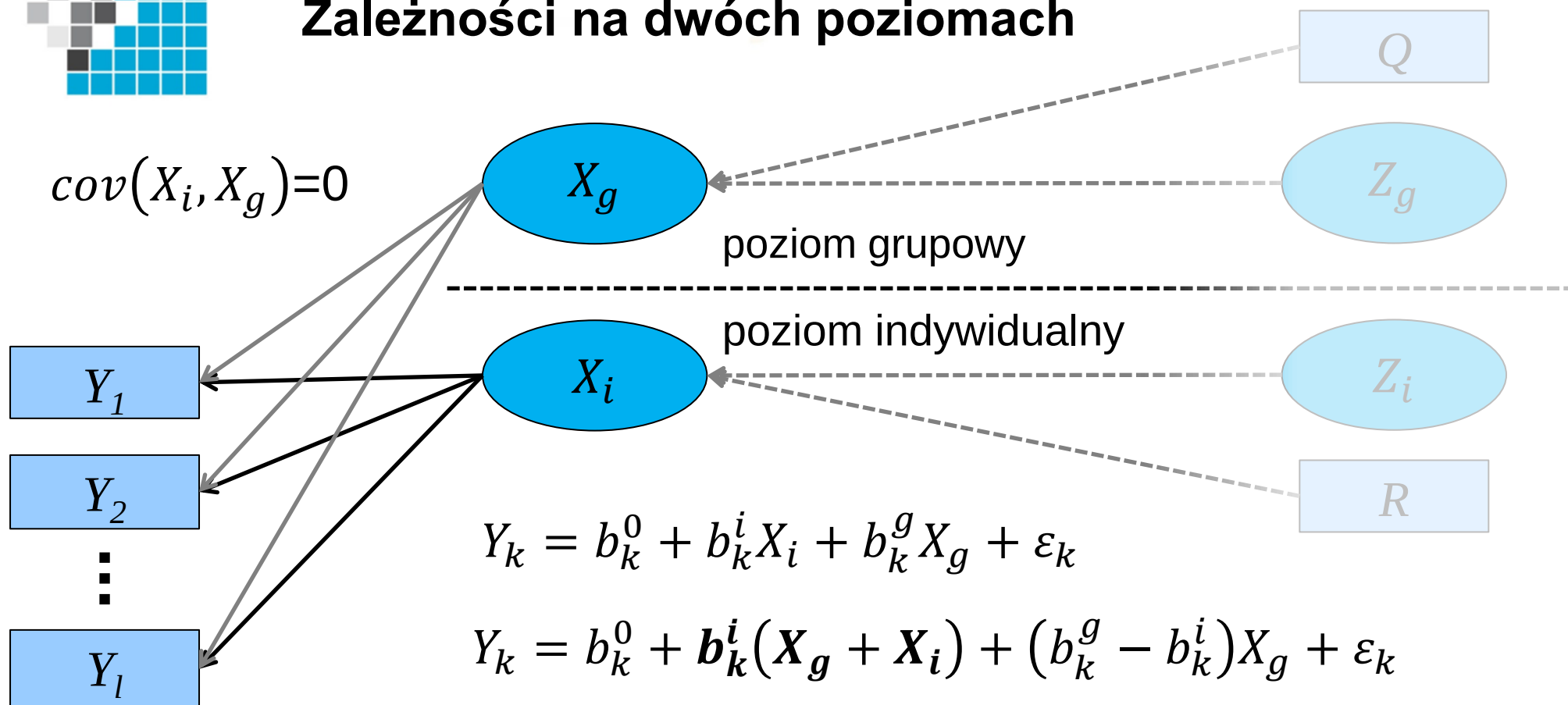
Czego chcieć więcej?

Analiza oddziaływania grup na tworzące je jednostki:

- Możliwość łatwej i efektywnej obliczeniowo oceny stopnia segregacji ze względu na konstrukt pomiędzy analizowanymi grupami.
 - Przy założeniu identycznego przebiegu zależności między konstruktem a zmiennymi obserwowalnymi na poziomie indywidualnym i grupowym możliwa interpretacja w kategoriach oddziaływania grupy na natężenie konstruktów jej członków.
- Możliwość uwzględnienia zróżnicowanego oddziaływania grup (specyfiki grup) na wyniki poszczególnych zadań.
 - W zasadzie narusza założenie o inwariancji *skalarnej*.
 - Uwalnianie stałych dla poszczególnych zadań domyślnie mniej elastyczne niż w modelach wielogrupowych.
 - Przy pewnej dozie *dobrej woli* można założyć, że jest to forma korekty na specyfikę grup. Np. dla testów wiedzy uwzględnia skupienie się w różnych grupach na różnych partiach materiału.



Zależności na dwóch poziomach



$$Y_k = b_k^0 + b_k^i X_i + b_k^g X_g + \varepsilon_k$$

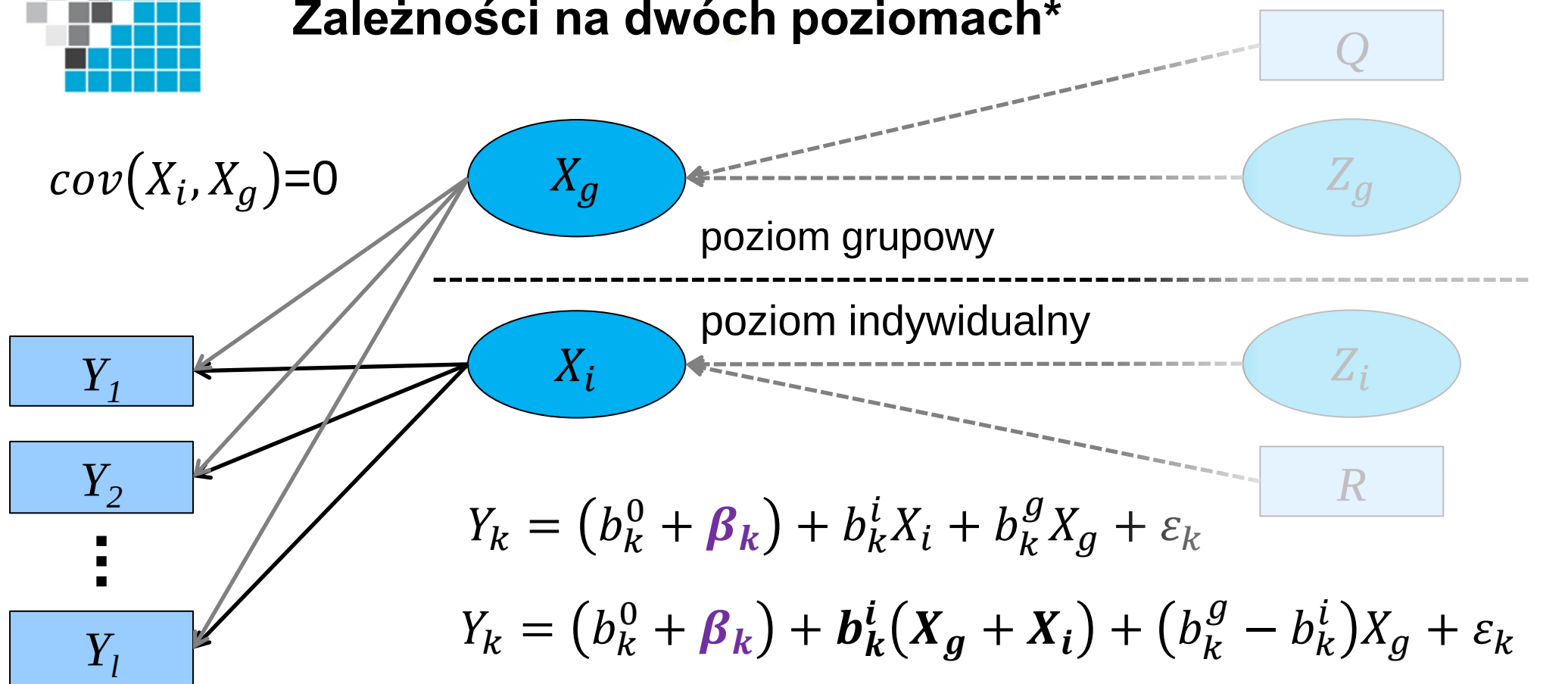
$$Y_k = b_k^0 + \mathbf{b}_k^i (X_g + X_i) + (b_k^g - b_k^i) X_g + \varepsilon_k$$

$$Y_k = [b_k^0 + (\mathbf{b}_k^g - \mathbf{b}_k^i) X_g] + \mathbf{b}_k^i (X_g + X_i) + \varepsilon_k$$

$$b_k^g = b_k^i \Leftrightarrow Y_k = b_k^0 + \mathbf{b}_k^i (X_g + X_i) + \varepsilon_k$$



Zależności na dwóch poziomach*



$$Y_k = (b_k^0 + \beta_k) + b_k^i X_i + b_k^g X_g + \varepsilon_k$$

$$Y_k = (b_k^0 + \beta_k) + b_k^i (X_g + X_i) + (b_k^g - b_k^i) X_g + \varepsilon_k$$

$$Y_k = [(b_k^0 + \beta_k) + (b_k^g - b_k^i) X_g] + b_k^i (X_g + X_i) + \varepsilon_k$$

$$b_k^g = b_k^i \Leftrightarrow Y_k = (b_k^0 + \beta_k) + b_k^i (X_g + X_i) + \varepsilon_k$$



Wielopoziomowa analiza czynnikowa - ograniczenia

- Trudności z interpretacją dla modeli zakładających zróżnicowany wpływ czynnika grupowego na trudność poszczególnych zadań:
 - Czy można rozsądnie argumentować na rzecz prowadzenia pomiaru przy niespełnionym założeniu o inwariancji *skalarnej*?
 - Domyślna specyfikacja modeli wielopoziomowej analizy czynnikowej (lub SEM) w programach statystycznych zakłada właśnie taki zróżnicowany wpływ.
 - Bardzo niewiele refleksji teoretycznej nad interpretacją modeli w tej formie w literaturze.
- Mimo wszystko obciążająca obliczeniowo (zwłaszcza przy kategoryalnych wskaźników konstruktu).
 - Szczególnie w przypadku, gdy chcemy opisać zróżnicowany wpływ grup na poszczególne zadania.



Wielopoziomowa analiza czynnikowa - przykład

- Analiza wyników 3 ostatnich serie testu matryc Ravena (wersja standard) na ogólnopolskiej reprezentatywnej próbie uczniów klas I gimnazjów.
- 5560 uczniów ze 150 szkół.
- Estymacja WLSMV w programie Mplus 7.3.
- Nie można porównywać indeksów dopasowania pomiędzy modelem jednopoziomowym a wielopoziomowymi!

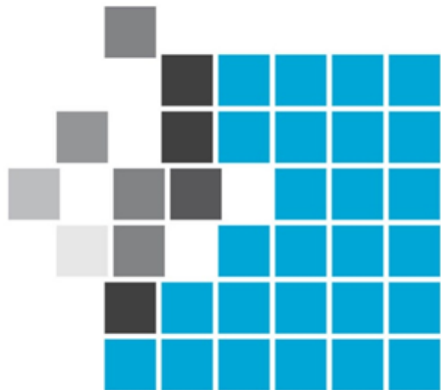
model	l. param.	RMSEA	CFI	TLI	WRMR
jednopoziomowy	48	0,040	0,925	0,917	3,094
jednopoziomowy (w specyfikacji dwupoz.)	48	0,022	0,883	0,883	3,827
dwupoziomowy z jednorodnym wpływem	49	0,023	0,878	0,878	2,365
dwupoziomowy ze zróżn. wpływem	73	0,023	0,875	0,869	2,341



Wielopoziomowa analiza czynnikowa: wariacje na temat

Można sobie wyobrazić (i wyestymować) model wielopoziomowej analizy czynnikowej analogiczny do modelu wielogrupowego z całkowicie zniesionym założeniem o inwariancji pomiarowej, tj. wprowadzający efekt losowy na parametry dyskryminacji, ale:

- W takim wypadku nie ma sensu mówienie, że badany konstrukt mierzy (znaczy) to samo w poszczególnych grupach.
 - Problem dobrze znany i opisany na gruncie badania inwariancji pomiarowej.
- Model jest bardzo obciążający obliczeniowo i niestabilny numerycznie, chyba że dysponujemy dużą liczbą obserwacji na grupę.
 - W związku z tym jeśli w ogóle estymowany, to raczej metodami bayesowskimi



EDUKACYJNA WARTOŚĆ DODANA

Dziękuję za uwagę!

t.zoltak@ibe.edu.pl

Instytut Badań Edukacyjnych

ul. Górczewska 8, 01-180 Warszawa

tel.: (22) 241 71 00, e-mail: ewd@ewd.edu.pl



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



EDUKACYJNA
WARTOŚĆ
DODANA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

