



UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE



UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE

Odpowiadanie nieuważne w badaniach samoopisowych - porównanie i walidacja metod identyfikacji.

Paweł Grygiel

Instytut Pedagogiki UJ; IBE

Grzegorz Humenny

Instytut Socjologii UJ; IBE

Marek Muszyński

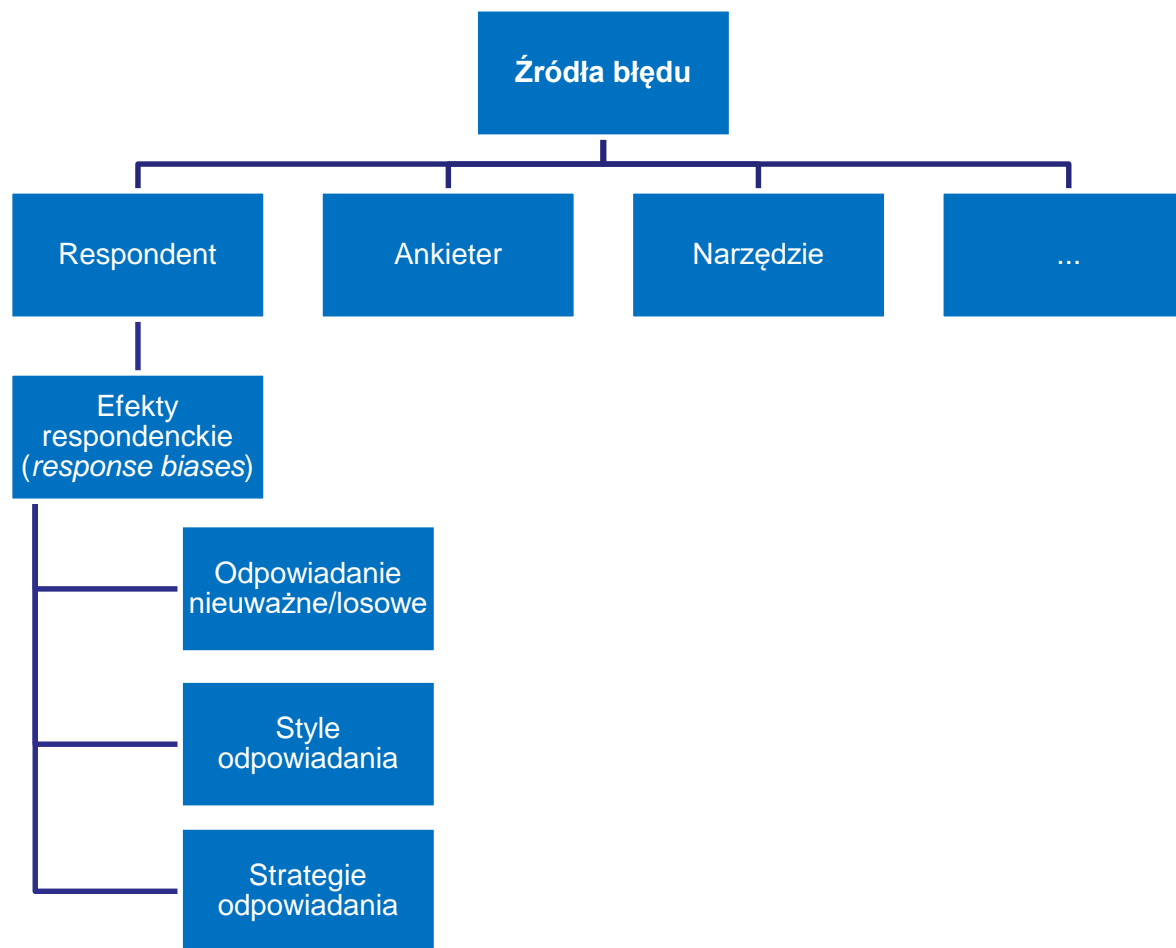
Instytut Socjologii UJ; Evidence Institute



1. Plan prezentacji.

- Zagrożenia pomiaru w metodach samoopisowych.
- Rodzaje efektów respondenckich.
- Odpowiadanie nieuważne i losowe- omówienie.
- Konsekwencje efektów respondenckich.
- Metody identyfikacji i korekty.
- Omówienie badań własnych.
- Dyskusja.

2. Zagrożenia pomiaru w metodach samoopisowych.





2. Zagrożenia pomiaru – efekty respondenckie

- (systematyczna) tendencja do udzielania odpowiedzi na pytania na podstawie innej niż ich treść (Paulhus, 1991)
- „trwała dyspozycja do reagowania w określony sposób w różnych obszarach treściowych” (Hornowska, 2007)
- związane lub niezwiązane z treścią itemów (Nichols, Greene i Schmolck, 1989)
- „nieuwaga” (*careless responding*; Meade i Craig, 2012)
- minimalizowanie wysiłku (*satisficing*; Krosnick, 1991)
- najczęściej związane z cechami respondenta lub **zależne od interakcji cech respondenta z kontekstem badania** (Baumgartner i Steenkamp, 2001; Mors, Kieruj i Vermunt, 2014; Schwarz, 1999; Weijters, 2006; Weijters i in., 2010; Ziegler & Buehner, 2009; Ziegler, 2015)

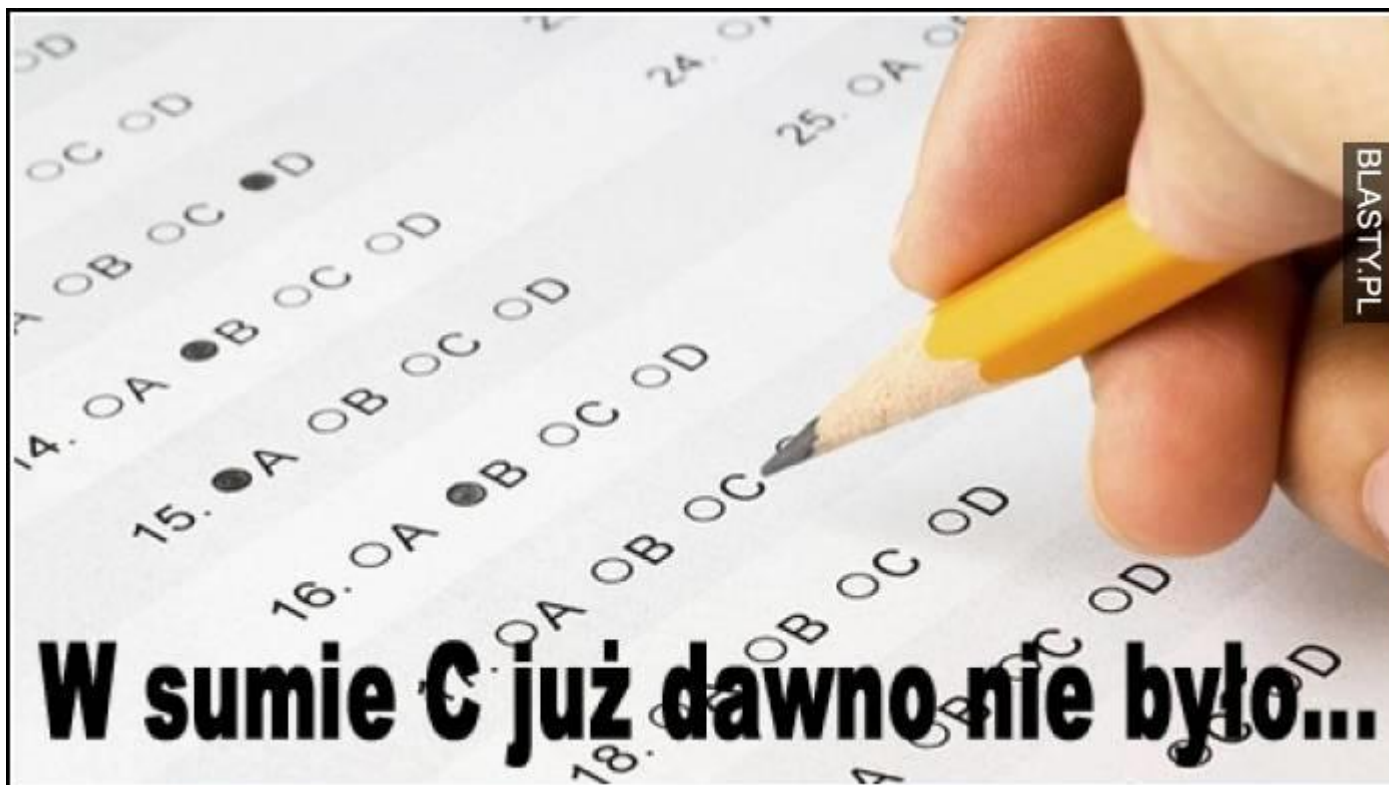


3. Przegląd efektów respondenckich

- odpowiadanie losowe, bądź nieuważne (*careless responding, inattentive responding*; Meade i Craig, 2012)
- style odpowiedzi (*response styles*; Weijters, 2006)
- strategie odpowiedzi (*response sets*; Paulhus, 2002)



3. Przegląd efektów respondenckich





3.1 Przegląd efektów respondenckich- odpowiadanie nieuważne

- definiowane najczęściej jako odpowiadanie bez pełnego zaangażowania (*careless responding*), odpowiadanie nieuważne (*inattentive responding*) lub odpowiadanie losowe (*random responding*) (Curran, 2015; Fronczyk, 2014; Goldsmith i Clark, 2005; Meade i Craig, 2012; Meyvis, Oppenheimer i Davidenko, 2009; Osborne i Blanchard, 2010)
- optymalizowanie/minimalizowanie wysiłku, „zadowalanie” (*satisficing*; Krosnick, 1991)
- do pewnego czasu bagatelizowane w badaniach, jako obejmujące znikomą część respondentów (Johnson, 2005)
- mała wiedza na temat rodzajów tego efektu (Meade & Craig, 2012)



3.2 Przegląd efektów respondenckich- odpowiadanie nieuważne- przyczyny/źródła

- niezmotywowani badani (szczególnie w badaniach niskiej stawki)
- niedopasowane narzędzie i/lub procedura badawcza (np. zbyt długi kwestionariusz)
- dystraktory (np. głośna sala)
- kontekst społeczny

(Meade & Craig, 2012)

- Niektóre z tych przyczyn są szczególnie istotne przy ankietach internetowych.



3.3 Przegląd efektów respondenckich- odpowiadanie nieuważne- powszechność występowania

- kilka % (Johnson, 2005)
- do 50% (Meyvis i in., 2009)
- około 10% (Maniaci & Rogge, 2014; Meade & Craig, 2012)
- problem nabiera szczególnej wagi w ankietach internetowych (Fleischer i in., 2015; Johnson, 2005):
 - większa anonimowość
 - mniejsza rozliczalność (*accountability*)
 - łatwość odpowiadania
 - traktowanie badania bardziej zabawowo, mniej serio
 - MTurk, boty



4. Konsekwencje efektów respondenckich

- zagrożenie trafności wnioskowania (Holden, 2007; w tym trafności kryterialnej)
- systematyczny błąd pomiaru (CIV; *spurious variance*; Schmidt, Le & Ilies, 2003; Ziegler i in., 2011)
- więcej losowego błędu pomiaru (Meade & Craig, 2012)
- zaniżone lub zawyżone średnie dla poszczególnych pytań (*theta shift*)
- skośne rozkłady zmiennych
- zaburzona struktura czynnikowa/struktura latentna narzędzia (Davies i in., 2015; *general factor of personality*- GFP; Burrus i in., 2011; Khorramdel i von Davier, 2014)- np. skorelowanie czynników ortogonalnych, ładowanie przez pozycje zwykłe i odwrócone dwóch różnych czynników (Huang, 2015; Woods, 2006)
- zaburzone korelacje z innymi zmiennymi (Pokropek, 2014)
- przekłamane miary spójności wewnętrznej (zaniżone lub zawyżone) (Fleischer i in., 2015)
- niekompletne odpowiedzi, wysoki odsetek braków danych (*nonresponse error*)



4. Konsekwencje efektów respondenckich

- dane powstałe w „inny” sposób (*protocol invalidity*)
- obniżenie jakości danych
- ryzyko błędów wnioskowania, zarówno I, jak i II rodzaju
- nieinterpretowalne dane

(Johnson, 2005; Kurtz, 2001; Maniaci & Rogge, 2014)



5. Metody radzenia sobie ze skutkami efektów respondenckich

- proaktywne (prewencja, zapobieganie)
 - reaktywne (remedium)
 - a priori
 - post hoc
-
- dewiacja, odstępstwo
 - (pod)grupa
 - źródło wariacji



5.1 Metody radzenia sobie ze skutkami efektów respondenckich- losowanie nieuważne i losowe

- proaktywne (prewencja, zapobieganie)
- odpowiednia konstrukcja narzędzi pomiarowych
- język narzędzi
- dostosowana procedura badawcza
- kontekst (warunki) wypełniania narzędzi
- motywacja badanych
- manipulacja instrukcją



5.2 Metody radzenia sobie ze skutkami efektów respondenckich- losowanie nieuważne i losowe

- reaktywne (remedium)
- a priori:
 - **itemy „pułapki” (*bogus items*)** (Oppenheimer i in., 2009)
Jeśli czytasz pytania uważnie, pomini ten rząd.
Zaznacz w tym pytaniu pierwszą odpowiedź z lewej.
Jeśli to czytasz to...
 - **skale samoopisowe** (Maniaci & Rogge, 2014; Meade & Craig, 2012; Schinka i in., 1997)
- post hoc:
 - **czasy reakcji**
 - indywidualne (*within-subject*) **miary spójności odpowiedzi** (np. indeks Cattella- rzetelność połówkowa; psychometryczne synonimy i antonimy; wariancja, etc.)
 - **analiza przypadków odstających** (np. dystans Mahalanobisa)
 - **miary zróżnicowania odpowiedzi** (*straight-liners*)



5.3 Metody radzenia sobie ze skutkami efektów respondenckich- losowanie nieuważne i losowe- zastosowane metody

- „item pułapka”
- pakiet *careless* R (Yentes & Wilhelm, 2018)
 - irv (Intra-individual Response Variability)
 - dystans Mahalanobisa
 - long (wartość najdłuższego ciągu identycznych odpowiedzi)
 - indeks Cattella (*even-odd*, !)
 - psychometryczne synonimy/antonimy (!)
- pakiet *PerFit* (Tendeiro, 2018)
 - lz(poly)
 - u3(poly)
 - G(poly) (Guttman error)
- miara dr^* (Mansolf i Reise, 2018; Reise i in., 2016)



5.4 Metody radzenia sobie ze skutkami efektów respondenckich- losowanie nieuważne i losowe- walidacja rozwiązań

- **korelacje zbieżne i rozbieżne z innymi zmiennymi**
- struktura latentna narzędzi
- własności psychometryczne narzędzi
- analiza klas/profilu latentnych



6. Zebrane dane

BADANIE KOMPETENCJE 2018



- około 25 000 uczniów (VI, VII SP, II GIM) ze szkół z całej Polski
- testy umiejętności (mat, pol, ang)
- skale kontekstowe (lęk przed matematyką, postawy wobec czytania, angielski poza szkołą)



6.1 Zebrane dane- skala samoopisowa

- radość z czytania (*Joy of reading*)
- PISA 2009
- 11 itemów (5 odwróconych)

- Czytam wtedy, kiedy muszę
- Czytanie to ulubione zajęcie
- Rozmowy o książkach
- Trudno doczytać książkę do końca
- Książka na prezent
- Czytanie to strata czasu
- Chodzenie do księgarni
- Czytam tylko gdy potrzebuję
- Nie mogę czytać dłużej niż kilka minut
- Lubię wyrażać opinie o książkach
- Wymiana książkami z przyjaciółmi

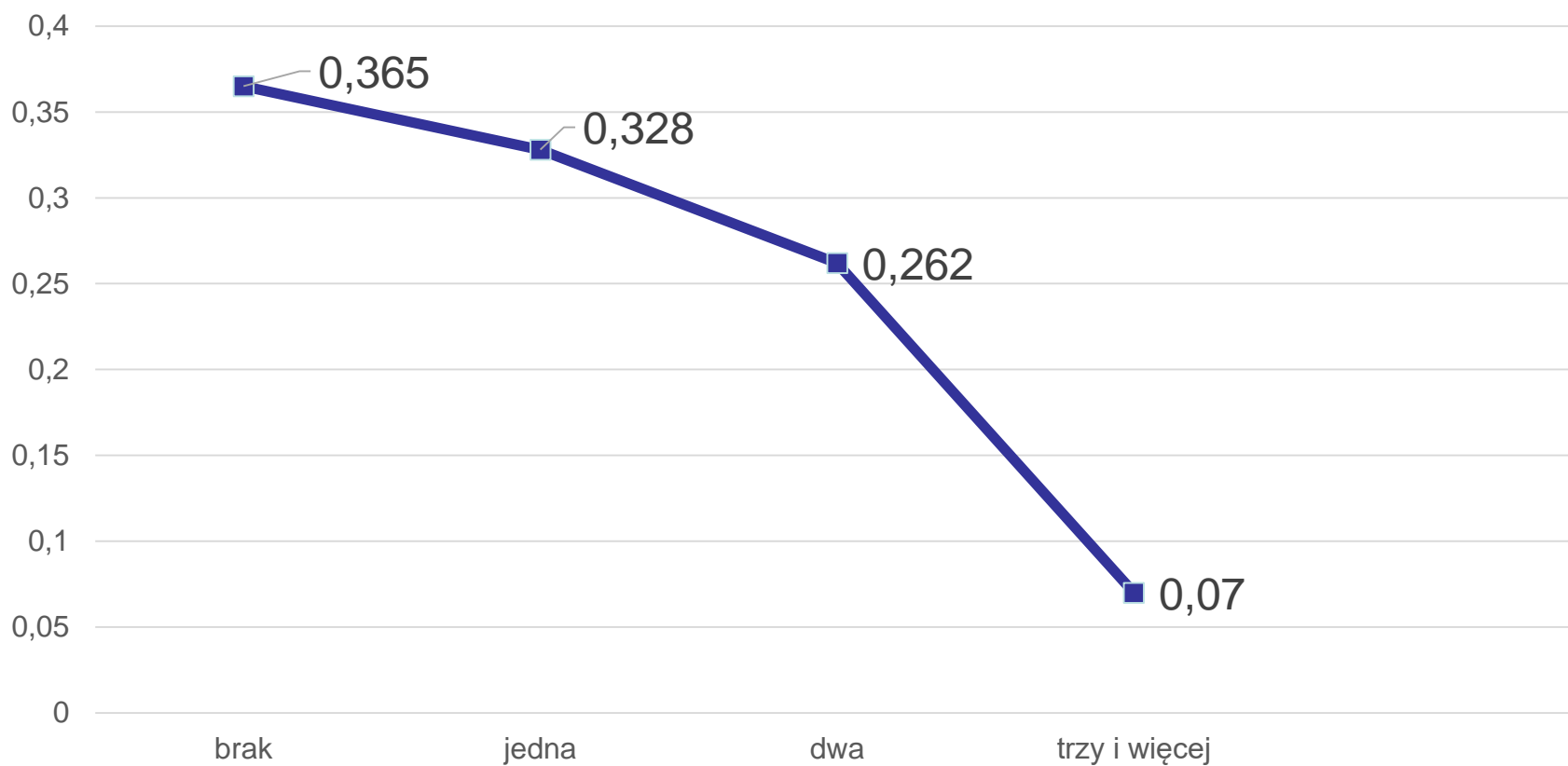


7. Wyniki na danych empirycznych

		Korelacja poziomu umiejętności polonistycznych z radością z czytania wzorzec odpowiedzi				
	kryterium	%	nietypowy	typowy	łącznie	
Wpływ usunięcia respondentów o nietypowych odpowiedziach na relacje między zmiennymi	pułapka	13,9%	,190**	,352**	,348**	
	dr (1,8%)	1,8%	0,039	,354**		
	dr (6,1%)	6,1%	,152**	,359**		
	irv	2,6%	0,034	,353**		
	mah (2,5%)	2,5%	,140**	,354**		
	mah (3,7%)	3,7%	,133**	,356**		
	long >6	2,0%	0,056	,351**		
	long >4	5,2%	,121**	,351**		
	Lz	7,9%	-0,006	,374**		
	U3	10,6%	,305**	,354**		
	G	9,3%	,133**	,370**		
	dr bif	1,8%	,186**	,350**		,338**
	dr bif	5,0%	,201**	,353**		



Korelacja osiągnięć z polskiego i radości z czytania a liczba kryteriów zaliczających do grupy nietypowych





Średni poziom umiejętności polonistycznych
(skala 500,100)

a

wzorzec odpowiedzi

Różnica między
respondentami o
nietypowych
wektorach odpowiedzi
w średnim poziomie
umiejętności

kryterium	nietypowy	typowy	d-Cohena	wielkość efektu
pułapka	413,01	480,67	0,66	medium
dr (1,8%)	407,93	472,41	0,67	medium
dr (6,1%)	419,88	474,56	0,53	medium
irv	418,15	472,64	0,49	small
mah (2,5%)	409,89	472,85	0,64	medium
mah (3,7%)	414,60	473,43	0,58	medium
long >6	392,99	472,86	0,74	medium
long >4	416,54	474,23	0,55	medium
lz	430,39	474,76	0,42	small
u3	437,40	475,25	0,35	small
G	429,12	475,54	0,46	small
dr bif	414,05	472,29	0,57	medium
dr bif	420,59	473,90	0,52	medium



Różnica między
respondentami o
nietypowych
wektorach odpowiedzi
w średnim poziomie
umiejętności

	Średni poziom radości z czytania (skala 0,1) a wzorzec odpowiedzi			
	kryterium nietypowy	typowy	d-Cohena	wielkość efektu
pułapka	-0,277	0,048	0,39	small
dr (1,8%)	0,112	0,001	-0,19	negligible
dr (6,1%)	-0,051	0,006	0,08	negligible
irv	-0,127	0,006	0,41	small
mah (2,5%)	0,033	0,002	-0,05	negligible
mah (3,7%)	0,030	0,002	0,03	negligible
long >6	-0,197	0,007	0,48	small
long >4	-0,287	0,019	0,54	medium
lz	0,075	-0,003	-0,15	negligible
u3	-0,082	0,013	0,09	negligible
G	0,036	-0,001	0,04	negligible
dr bif	-0,057	0,010	-0,08	negligible
dr bif	-0,087	0,014	-0,11	negligible



8. Dane symulacyjne

- wygenerowano dane:
- 11 itemów
- jeden czynnik
- ładunki czynnikowe w granicach 0,5-0,8
- łącznie 1400 obserwacji „typowych”

oraz:

- 200 losowy rozkład jednostajny
- 200 losowy rozkład normalny
- 100 losowy rozkład skrajny + (4,5)
- 100 losowy rozkład skrajny – (1,2)
- łącznie 600 obserwacji „nietypowych”
- łącznie: 2000 obserwacji (70/30 typowe/nietypowe)



8.1 Dane symulacyjne-

Potencjał poszczególnych miar do odróżniania wzorców zgodnych z modelem jednoczynnikowym od wzorców zaburzonych

Średnia wartość miary niedopasowania a wzorzec odpowiedzi				
kryterium	zaburzony	zgodny z modelem	d-Cohena	wielkość efektu
dr	3,559	2,800	-1,05	large
irv	0,971	0,898	-0,27	small
mah	15,786	8,941	-1,17	large
long	2,873	2,739	-0,12	negligible
lz	-0,773	0,693	1,16	large
u3	0,227	0,107	-1,24	large
G	56,975	32,630	-0,85	large



8.1 Dane symulacyjne-

Potencjał poszczególnych miar do odróżniania wzorców zgodnych z modelem podwójnego czynnika od wzorców zaburzonych

Średnia wartość miary niedopasowania a wzorzec odpowiedzi				
kryterium	zaburzony	zgodny z modelem	d-Cohena	wielkość efektu
dr	3,695	2,741	-1,17	large
irv	0,977	0,904	-0,27	small
mah	15,836	8,919	-1,21	large
long	2,877	2,727	-0,14	negligible
lz	-0,768	0,690	1,17	large
u3	0,228	0,106	-1,31	large
G	57,300	33,066	-0,85	large



9. Podsumowanie

Potrzeba:

- metody:
 - taniej,
 - łatwej do zastosowania,
 - o dużej trafności wnioskowania,
 - możliwej do zastosowania (także) *a posteriori (post hoc)*,
 - do różnych skal i kontekstów badania
- metod proponuje się wiele, ale...
- mało badań proponujących **ich porównania, walidację, a także syntezę**
- **mało badań w kontekście „niskiej stawki”**
- **mechanizmy zjawiska wciąż niedostatecznie zbadane**

(Meade & Craig, 2012; Johnson, 2005)



9. Podsumowanie

- im więcej wskaźników tym lepiej
- wzorce odpowiedzi „nietypowych” (*protocol invalidity*) są różne
- różne wskaźniki zdają się wskazywać na różnych badanych
- w sumie niewielki (?) wpływ na badaną zależność (dlaczego?)
- inne metody walidacji (spójność wewnętrzna, struktura latentna, multigroup SEM, ?)



10. Dyskusja





UNIWERSYTET JAGIELLOŃSKI
W KRAKOWIE

Dziękujemy za uwagę!

Paweł Grygiel

Instytut Pedagogiki UJ; IBE

Grzegorz Humenny

Instytut Socjologii UJ; IBE

Marek Muszyński

Instytut Socjologii UJ; Evidence Institute