

**SGH**

Szkoła Główna  
Handlowa  
w Warszawie

# **Efekt techniki, wrażliwość pytań a ocena poziomu własnych kompetencji – wstępne rezultaty eksperymentu metodologicznego**

dr Maja Rynko, SGH i dr hab. Tomasz Drabowicz, prof.  
UŁ

IV Kongres Statystyki Polskiej, Warszawa 2-4 lipca 2024

# Motywacja do badania

- Badania statystyczne powinny dostarczyć danych dobrej jakości...
- Specyficzne wyzwania dot. jakości danych w badaniach ankietowych
- Uwzględnienie kontekstu kulturowego / społecznego przy konstruowaniu narzędzi badawczych i ustalaniu sposobu zbierania danych
- W Polsce niewiele badań w zakresie eksperymentów metodologicznych
- Nasze badanie pokazuje istotne różnice w odpowiedziach na pytania dot. samooceny umiejętności przy metodach CAPI i PAPI (self-administered)

# Definicje - efekt techniki i pytanie wrażliwe

## Efekt techniki (*mode effect*)

- zastosowanie różnych sposobów zbierania materiałów (technik badawczych) może pociągać za sobą ryzyko uzyskania odmiennych, różnych danych (Sztabiński & Żmijewska-Jędrzejczyk, 2012: 38).

## Pytanie wrażliwe / drażliwe

- odnoszące się do regulowanych społecznie norm, których nie wypada nie respektować, bądź poruszające kwestie krępujące, w odczuciu respondenta nazbyt intymne (Boguszewski i Hipsz, 2012: 75).

# Mixed Mode Survey Design – Mixed Mode

- Procedura badania, w której dopuszcza się łączenie różnych technik wywiadu (np. PAPI i CAPI)
- 2 rodzaje Mixed Mode:
  - Sekwencyjne (*sequential*) - określamy jedną podstawową (preferowaną) technikę oraz dodatkowe, uzupełniające, które będą stosowane w dalszej kolejności
  - równoczesne (tzw. *concurrent*) stosowanie różnych technik zbierania informacji. Nie określa się wówczas techniki podstawowej, lecz traktuje wszystkie jak ekwiwalentne i wykorzystuje się je równocześnie, w tym samym czasie (Sztabiński & Żmijewska-Jędrzejczyk, 2012: 35).

# Efekt techniki – przegląd literatury

## Sztabiński & Żmijewska-Jędrzejczyk (2012)

- Eksperyment PAPI vs CATI w ESS PL R3 (+CAPI w innych krajach). Różnice dotyczą sposobu kontaktowania się i komunikowania się z respondentem oraz sformułowania pytań
- Istotne różnice we wskaźnikach realizacji badania (PAPI 70,1% vs 35,3% CATI)
- Przykładowe różnice w wynikach:
  - ponad 10 pp różnicy dot. pytania o udział w wyborach (na korzyść CATI)
  - blisko 20 pp różnicy w pytaniu o poczucie związku z jakąś partią polityczną
  - skracanie/spłaszczanie skali w CATI w pytaniu dot. poczucia zaufania do innych ludzi (w CATI odpowiedzi „00. Ostrożności nigdy za wiele” oraz punkt środkowy „05” wskazało 46% respondentów)
  - niewielkie różnice (i nieistotne statystycznie) rozkładów odpowiedzi w pytaniu o stosunek do osób o orientacji homoseksualnej

# Efekt techniki – przegląd literatury

## Boguszewski & Hipsz (2012)

- Analiza porównawcza technik PAPI i CAPI w regularnych badaniach CBOS: zmiana techniki realizacyjnej w CBOS w 2008r. nie przyniosła znaczących różnic ani w poziomie realizacji, ani w wynikach badań
- Nieistotne zmiany w rozkładach odpowiedzi na poniższe pytania:
  - deklaracji udziału w wyborach
  - samooceny wiary
  - określenia poglądów politycznych na 7-punktowej skali od lewicy do prawicy
  - oceny sytuacji materialnej własnego gospodarstwa domowego
- Jediną zauważalną zmianą był spadek udziału odpowiedzi „trudno powiedzieć” w badaniach CAPI

# Efekt techniki – przegląd literatury

## Heerwegh (2009)

- Ocena czy wywiady bezpośrednie dostarczają danych lepszej jakości niż CAWI
- Eksperyment metodologiczny przeprowadzony na 444 studentach Katholieke Universiteit Leuven w Belgii; pytania dotyczyły postaw wobec imigrantów i azylantów
- Grupa poddana oddziaływaniu odpowiadała na część pytań w wywiadzie bezpośrednim, a następnie kończyła ankietę metodą CAWI
- Próba zapewnienia jak największej porównywalności między metodą CAPI i CAWI (karty odpowiedzi w CAPI, wyświetlanie pojedynczych pytań w CAWI)
- W CAWI więcej „don't know” i więcej braków odpowiedzi; poziom „przytakiwania” (*acquiescence*) i braku różnicowania odpowiedzi (*non-differentiation*) podobny w obydwu grupach

**SGH** Oceniając wyniki badania należy mieć na uwadze specyfikę badanej grupy [www.sgh.waw.pl](http://www.sgh.waw.pl)

# Efekt techniki – przegląd literatury

## Newman et al. (2002)

- Porównanie odpowiedzi z wywiadów bezpośrednich z odpowiedziami z audi-CASI (*audio computer assisted self-interviewing*)
- Badaniu poddani uczestnicy programu „syringe exchange” (N = 1417); kwestionariusz podzielony na trzy części: “stigmatized behaviors”, “neutral behaviors”, “psychological distress”
- Więcej odpowiedzi z zakresu „stigmatized behaviors” w audio-CASI (social desirability bias?); “psychological distress” częściej raportowany w wywiadzie bezpośrednim (bezosobowy komputer nie zachęca od szczerości w odniesieniu do pytań o stan psychiczny?)
- Naturalnie ograniczona możliwość uogólniania wyników ze względu na specyficzność badanej próby



# Efekt techniki – przegląd literatury

## Guzy and Leitgöb (2015)

- Porównanie badania CATI i CAWI w rundzie 2010 badania *Crime and Victim Survey* (ICVS-2) w badanych krajach (DEN, GER, NED, SWE, UK)
- Ocenie poddano odpowiedzi dot. pięciu rodzajów przestępstw (*car theft, bicycle theft, personal theft, burglary, sexual offenses*)
- W metodzie CAWI istotnie wyższe wskaźniki dot. bycia ofiarą przestępstw, ale w dużym stopniu różnice te wynikają z błędów pokrycia i braków odpowiedzi; efekt techniki jest prawdopodobny jedynie dla odpowiedzi dot. kradzieży roweru i przestępstw na tle seksualnym

# Efekt techniki – przegląd literatury

## Lugtig et al. (2011)

- Porównanie metody CATI i CAWI na podstawie badania mixed-mode w Holandii w 2008 r.; tematyka badania dotyczyła tematyki zanieczyszczenia z przemysłu, rolnictwa i ruchu
- Próby CATI i CAWI losowane z rejestru adresów pocztowych, przy czym CATI ograniczało się do gospodarstw domowych z telefonem stacjonarnym (N=2685, RR=47%); osoby z próby CAWI otrzymywały list (listy) z linkiem i loginem do kwestionariusza (N=1347, RR=19%)
- Wykorzystana metodologia do oceny efektu techniki to estymacja przez dopasowanie (PSM)
- Wyniki wskazują na występowania efektu techniki (PSM pokazuje mniejsze różnice niż porównanie bezpośrednio wyników); respondenci CATI wybierali części odpowiedzi „ekstremalne” i ogólnie odpowiadali bardziej pozytywnie

# Efekt techniki – przegląd literatury

## Schork, Riillo, and Neumayr (2021)

- Porównanie metod CATI i CAWI w badaniu Luxembourgish Labour Force Survey (2015-2017); porównywano wyniki dotyczące zarówno miar obiektywnych (status zatrudnienia) jak i subiektywnych (zadowolenie z pracy, odpowiednie wynagrodzenie)
- Zastosowanie metodologii estymacji przez dopasowanie
- Wyniki wskazują odporność miar obiektywnych na technikę badania; dla miar subiektywnych odnotowano wyższe oceny zadowolenia w wywiadzie CATI
- Efekt techniki jest zależny od kontekstu ocenianej zmiennej

# Efekt techniki – przegląd literatury

## Haas, Eckman, and Bach (2021)

- Badanie dot. statystyki przedsiębiorstw i ocenę, czy zmiana metody zbierania danych zmniejszą obciążenie ankietowe przedsiębiorstw w Niemczech
- Porównanie metod PAPI i CAWI, przedsiębiorstwa miały też możliwość wyboru metody (4 porównania: *Paper-only versus Web-only*, *Choice-Paper versus Paper-only*, *Choice-Web versus Web-only*, *Choice-Paper versus Choice-Web*)
- Ankiety CAWI charakteryzowały się najkrótszym czasem wypełnienia ankiety niezależnie, czy przedsiębiorstwa miały wybór techniki badania czy nie

# Eksperyment metodologiczny w postPIAAC

- postPIAAC jest badaniem zrealizowanym w latach 2014-2015 przez IBE
- Krajowa kontynuacja międzynarodowego badania PIAAC
- PIAAC + postPIAAC stanowią tzw. krótki panel
- Próba postPIAAC to 5224 osoby w wieku 18-69 lat (nadreprezentacja osób młodych)
- Jednym z modułów tematycznych badania była samoocena umiejętności z zakresu rozumienia tekstu i rozumowania matematycznego
- Losowe przypisanie respondentów do techniki zbierania danych:
  - 2622 respondentów otrzymała arkusz papierowy samooceny
  - 2602 respondentów odpowiadała na pytania ankietera (CAPI)

# Samooceena umiejętności w postPIAAC

- 1) Wykonywanie prostych rachunków takich jak dodawanie i odejmowanie
- 2) Wylizywanie reszty przy płatności gotówką w sklepie
- 3) Szacowanie wymiarów bez użycia miarki (np. ten stół ma ok. 1 metr szerokości)
- 4) Obliczanie powierzchni o regularnych kształtach (np. powierzchni ścian lub podłogi przy zakupie płytek)
- 5) Obliczanie procentu z danej kwoty przy użyciu kalkulatora (np. 23% podatku)
- 6) Szacowanie w pamięci procentu z danej kwoty bez wykonywania dokładnych obliczeń (np. 20% rabatu na zakup obuwia)
- 7) Liczenie średnich (np. średnia ocen albo dochód na członka rodziny)
- 8) Wykonywanie obliczeń wymagających większej liczby operacji matematycznych (np. wybór najtańszego planu taryfowego spośród ofert sieci komórkowych)
- 9) Analiza i interpretacja danych statystycznych (np. wykresy kursu walut, ceny mieszkań w różnych miastach/dzielnicach)

- A Nie umiem, nie potrafię
- B Potrafię, ale sprawia mi to duże trudności
- C Potrafię, ale sprawia mi to nieznaczne trudności
- D Potrafię, przychodzi mi to z łatwością

# Samooceena umiejętności w postPIAAC

- 1) Znajdowanie informacji w krótkim tekście (np. o liczbie szklanek mąki w przepisie na ciasto lub o zdobywcy bramki w relacji z meczu)
- 2) Czytanie ze zrozumieniem instrukcji obsługi dołączanych do urządzeń (np. telewizora lub pralki)
- 3) Odczytywanie informacji z wykresu (np. o poziomie bezrobocia w województwie)
- 4) Łączenie kilku źródeł informacji (np. zbieranie informacji o celu wycieczki z gazet, przewodników, ulotek, internetu)
- 5) Wychwytywanie w tekście argumentów za i przeciw (np. w artykule prasowym)
- 6) Czytanie ze zrozumieniem dokumentów (np. umowa najmu mieszkania, umowa ubezpieczenia czy regulamin przewozów)
- 7) Streszczanie dużej ilości tekstu
- 8) Logiczne myślenie, analiza faktów

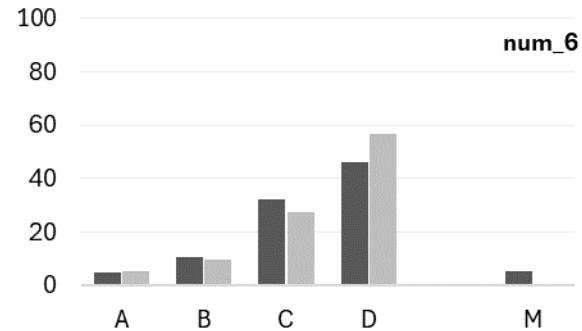
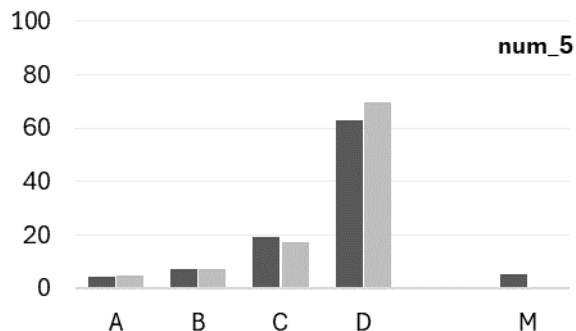
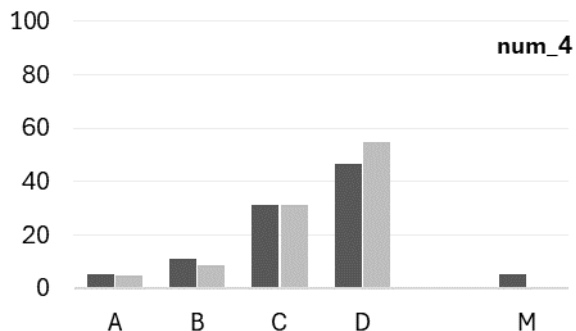
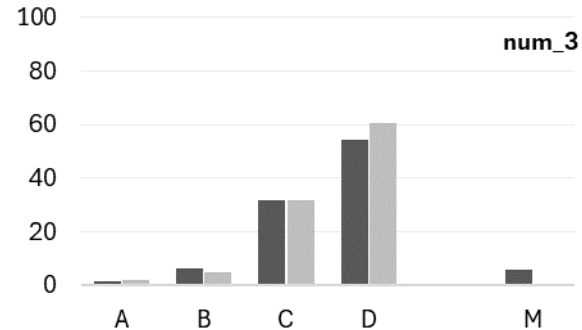
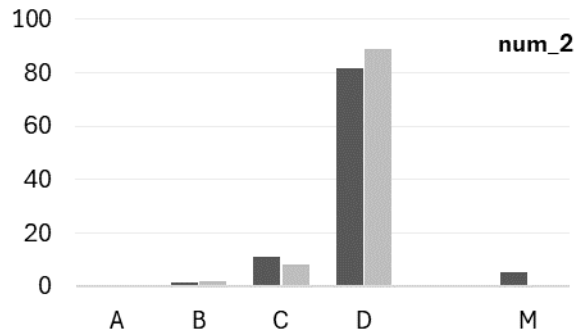
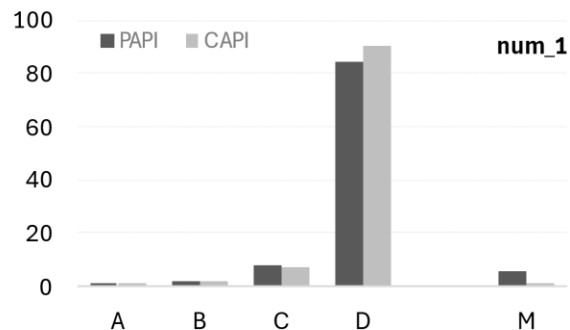
A Nie umiem, nie potrafię

B Potrafię, ale sprawia mi to duże trudności

C Potrafię, ale sprawia mi to nieznaczne trudności

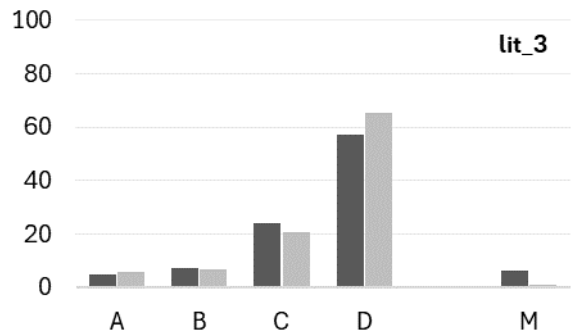
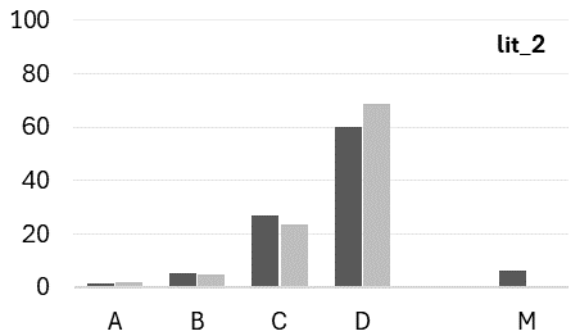
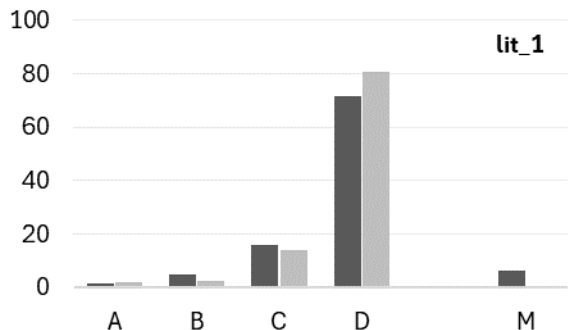
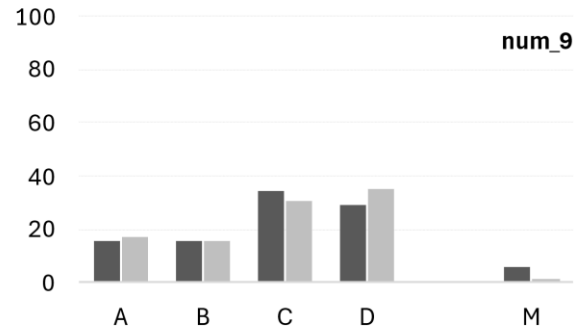
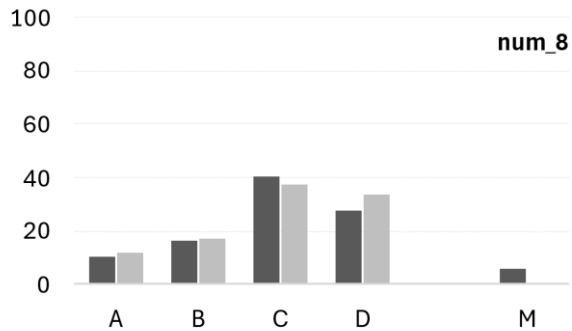
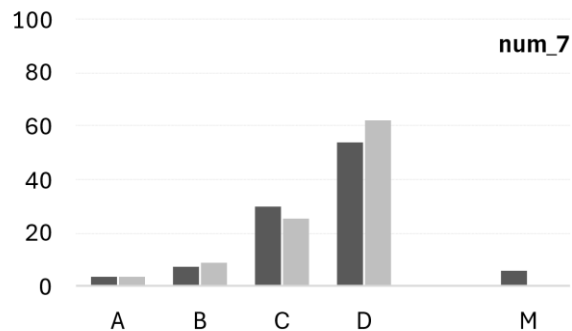
D Potrafię, przychodzi mi to z łatwością

# Samoocena umiejętności – rozkłady odpowiedzi

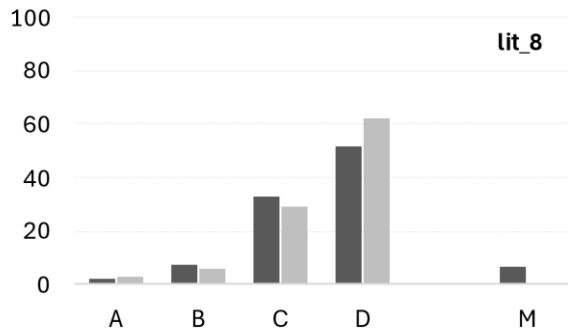
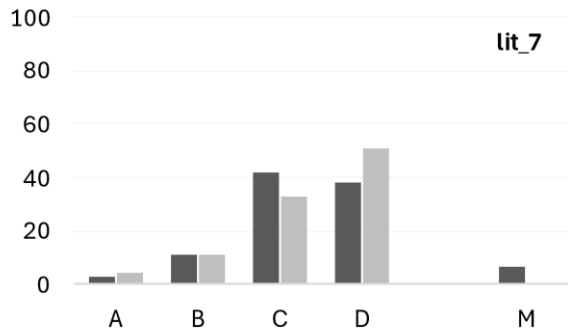
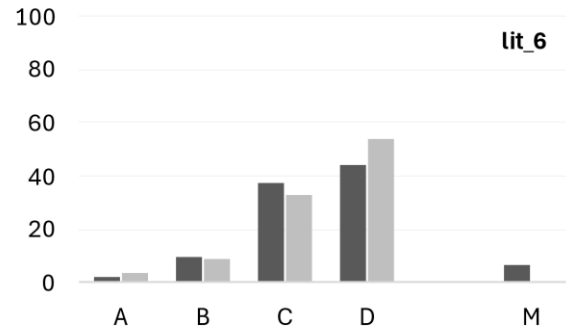
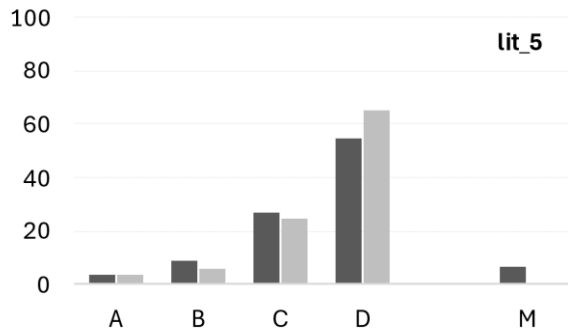
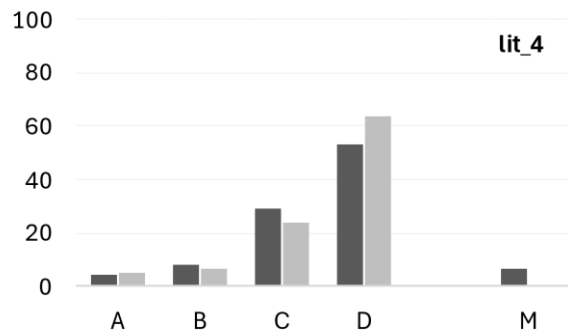




# Samocena umiejętności – rozkłady odpowiedzi

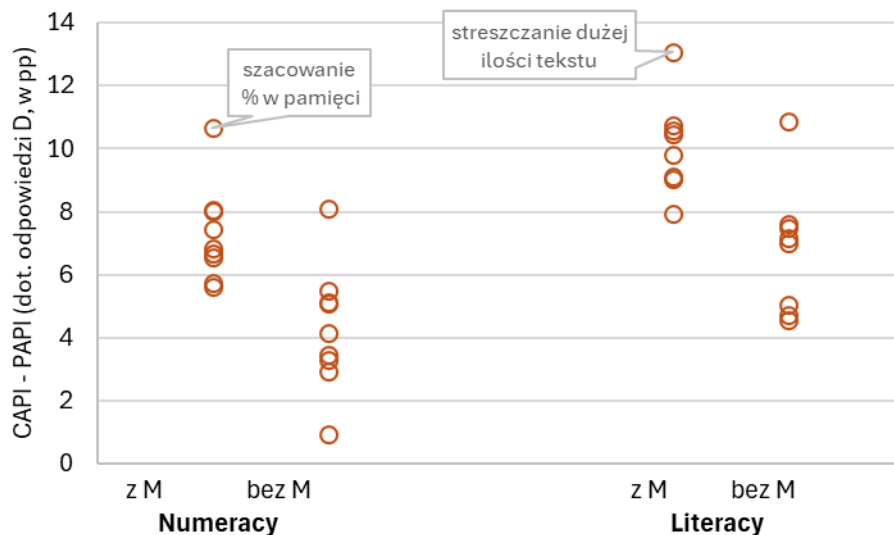


# Samocena umiejętności – rozkłady odpowiedzi



# Analiza rozkładów odpowiedzi

- W CAPI obserwujemy częściej odpowiedź D „Potrafię przychodzi mi to z łatwością”
  - różnice między literacy i numeracy (w literacy większe rozbieżności)
  - różnice między pytaniami (mniejsze rozbieżności dla łatwiejszych pytań)
- Uwzględnianie lub nie braków odpowiedzi w analizie rozkładów odpowiedzi zmienia wyniki, ale różnice pozostają nadal nieignorowalne
- Wytłumaczenie:
  - *social desirability bias*?
  - skracanie skali odpowiedzi?



# Modelowanie cechy nieobserwowalnej

- Przyjmujemy założenie o istnieniu dwóch nieobserwowalnych cech – literacy i numeracy (analogicznie jak w PIAAC)
- Wykorzystujemy metodologię IRT do oszacowania rozkładu nieobserwowalnych cech dla respondentów badania
- W szczególności stosujemy model grm (*graded response model*), w którym prawdopodobieństwo odpowiedzi k lub wyższej na pytanie i przez osobę j zadane jest wzorem:

$$\Pr(Y_{ij} \geq k | \theta_j) = \frac{\exp\{a_i(\theta_j - b_{ik})\}}{1 + \exp\{a_i(\theta_j - b_{ik})\}} \quad \theta_j \sim N(0, 1)$$

(a<sub>i</sub> discrimination parameter of item i; b<sub>ik</sub> difficulty parameter; theta<sub>j</sub> latent trait of person j)

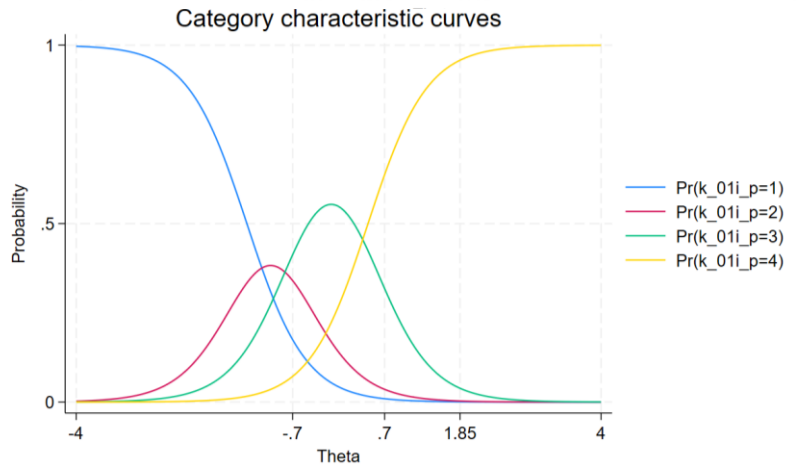
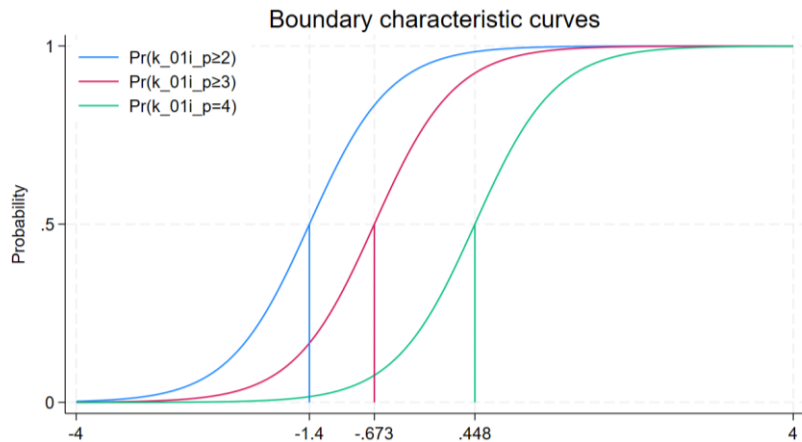
# Modelowanie cechy nieobserwowalnej

- Przyjmujemy założenie o istnieniu dwóch nieobserwowalnych cech – literacy i numeracy (analogicznie jak w PIAAC)
- Wykorzystujemy metodologię IRT do oszacowania rozkładu nieobserwowalnych cech dla respondentów badania
- W szczególności stosujemy model grm (*graded response model*), w którym prawdopodobieństwo odpowiedzi k lub wyższej na pytanie i przez osobę j zadane jest wzorem:

$$\Pr(Y_{ij} \geq k | \theta_j) = \frac{\exp\{a_i(\theta_j - b_{ik})\}}{1 + \exp\{a_i(\theta_j - b_{ik})\}} \quad \theta_j \sim N(0, 1)$$

(a<sub>i</sub> discrimination parameter of item i; b<sub>ik</sub> difficulty parameter; theta<sub>j</sub> latent trait of person j)

# Wyniki oszacowania modelu IRT (numeracy)



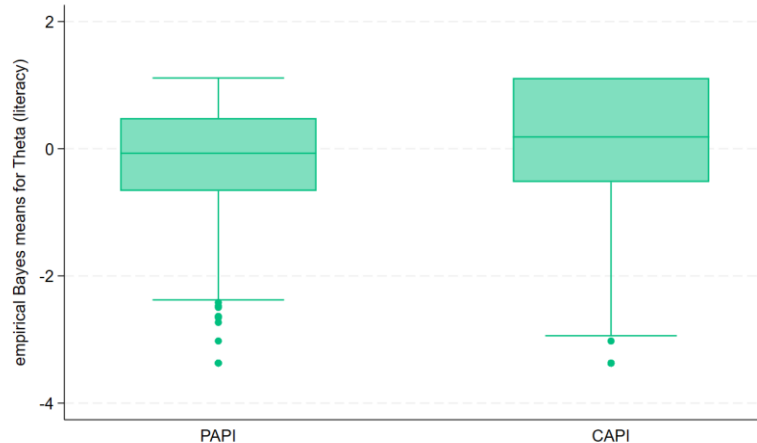
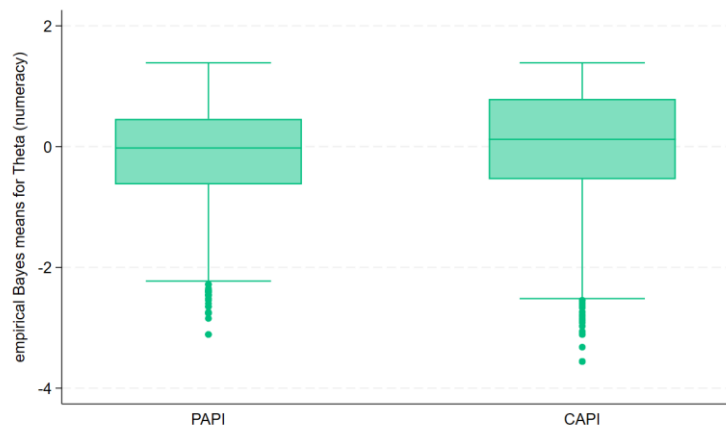
Graded response model  
Log likelihood = -35771.218

Number of obs = 5,096

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
<b>k_01a_p</b>						
Discrim	1.683355	.0768941	21.89	0.000	1.532645	1.834065
Diff						
>=2	-3.962368	.1685633			-4.292746	-3.63199
>=3	-2.982924	.1044582			-3.187659	-2.77819
=4	-1.852566	.0577291			-1.965713	-1.739419
<b>k_01b_p</b>						
Discrim	1.535502	.070246	21.86	0.000	1.397822	1.673182
Diff						
>=2	-4.456685	.2134651			-4.875068	-4.038301
>=3	-3.22868	.1217897			-3.467384	-2.989977
=4	-1.784595	.0590372			-1.900306	-1.668884
<b>k_01c_p</b>						
Discrim	1.162087	.0426383	27.25	0.000	1.078518	1.245657
Diff						
>=2	-3.828736	.1411897			-4.105463	-3.552009
>=3	-2.39087	.0813737			-2.550359	-2.23138
=4	-.2473504	.0308312			-.3077785	-.1869224
<b>k_01d_p</b>						
Discrim	1.721477	.0517386	33.27	0.000	1.620071	1.822883
Diff						

Źródło: opracowanie własne na podstawie danych postPIAAC

# Porównanie oszacowań poziomów umiejętności



Źródło: opracowanie własne na podstawie danych postPIAAC

# Studia czy CAPI? ;-)

	Model 1 (numeracy)		Model 2 (literacy)	
	Coeff.	SE	Coeff.	SE
<b>papi_capi</b>	0.149	0.023	0.207	0.023
<b>gender</b>	-0.300	0.024	0.026	0.023
<b>age</b>	0.001	0.001	-0.008	0.001
<b>edu3</b>				
<b>primary</b>	-0.582	0.030	-0.439	0.030
<b>tertiary</b>	0.253	0.027	0.249	0.027
<b>_cons</b>	0.261	0.059	-0.051	0.058

Źródło: opracowanie własne na podstawie danych postPIAAC



# Model poznawczy kształtowania odpowiedzi

1. Kodowanie – proces formowania wspomnień z doświadczeń (może mieć miejsce na długo przed badaniem)
2. Zrozumienie – proces interpretacji pytania, identyfikacji jego znaczenia
3. Wyszukiwanie – proces przypominania informacji niezbędnych do udzielenia odpowiedzi na pytanie
4. Opiniowanie – proces łączenia lub uzupełniania informacji na etapie wyszukiwania
5. Sprawozdanie – proces wyboru i komunikacji odpowiedzi.

# Social desirability, social approval

- a tendency to avoid criticism and to portray oneself in a favourable social light., which has been defined as an individual's perception or expectation of what most others approve or disapprove of. **Social desirability effects are** likely to exist when sensitive questions are asked, and are **more likely to be observed in interviewer-administered than in self-administered surveys** (Aquilino 1994; Aquilino and Lo Sciuto 1990; Cialdini 2007, Felderer et al. 2019; Gnamb and Kaspar 2017; Kreuter et al. 2008; Tourangeau 2018).
- the understanding of what constitutes a sensitive question can differ across individuals and modes of data collection (Kreuter et al. 2008).

# Wnioski

- Pokazujemy istotny efekt techniki
- Wyjaśnieniem jest najprawdopodobniej efekt *social desirability*
- Efekt techniki jest większy w dziedzinie *literacy* – większa wrażliwość pytań (mniejszy problem w przyznaniu się do deficytów umiejętności w dziedzinie *numeracy...*)
- Dalsze prace metodologiczne:
  - Modelowanie odpowiedzi na poszczególne pytania (modele logitowe)
  - poprawa specyfikacji modelu regresji liniowej dla  $\theta$
  - modelowanie DIF (*differential item functioning*)

# Bibliografia

- AAPOR (2008) Standard Definitions: Final Dispositions of Case Codes and Outcome Reports for Surveys. (5th.edn) Lenexa, Kansas: American Association for Public Opinion Research.
- Aquilino, W.S.: Interview mode effects in surveys of drug and alcohol use: a field experiment. *Public Opin. Q.* 58, 210–240 (1994). <https://doi.org/10.1086/269419>
- Aquilino, W.S., Lo Sciuto, L.A.: Effects of interview mode on self-reported drug use. *Public Opin. Q.* 54, 362–393 (1990). <https://doi.org/10.1086/269212>
- Boguszewski, R., Hipsz, N. Od kwestionariusza papierowego do wspomaganego komputerowo. *Analiza porównawcza technik PAPI i CAPI. Przegląd Socjologiczny tom LXI/1 (2012), str. 65-82.*
- Felderer, B., Kirchner, A., Kreuter, F.: The effect of survey mode on data quality: disentangling nonresponse and measurement error bias. *J. Off. Stat.* 35, 93–115 (2019). <https://doi.org/10.2478/jos-2019-0005>
- Gnamb, T., Kaspar, K.: Socially desirable responding in web-based questionnaires: a meta-analytic review of the Candor hypothesis. *Assessment* 24, 746–762 (2017). <https://doi.org/10.1177/1073191115624547>
- Guzy, Nathalie, and Heinz Leitgöb. 2015. "Assessing Mode Effects in Online and Telephone Victimization Surveys." *International Review of Victimology* 21(1):101–31. doi: 10.1177/0269758014547995.
- Haas, Georg-Christoph, Stephanie Eckman, and Ruben Bach. 2021. "Comparing the Response Burden between Paper and Web Modes in Establishment Surveys." *Journal of Official Statistics* 37(4):907–30. doi: 10.2478/jos-2021-0039.
- Heerwegh, D. 2009. "Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects." *International Journal of Public Opinion Research* 21(1):111–21. doi: 10.1093/ijpor/edn054.
- Jones, J., G. Snijders, and G. Haraldsen. 2013. "Surveys and Business Surveys." In *Designing and Conducting Business Surveys*, edited by G. Snijders, G. Haraldsen, J. Jones, and D.K. Willimack: 1–33. Hoboken, NJ: John Wiley & Sons.
- Kreuter, F., Presser, S., Tourangeau, R.: Social desirability bias in CATI, IVR, and Web surveys: the effects of mode and question sensitivity. *Public Opin. Q.* 72, 847–865 (2008). <https://doi.org/10.1093/poq/nfn063>
- Lugtig, Peter, Gerty J. L. M. Lensvelt-Mulders, Remco Frerichs, and Assyn Greven. 2011. "Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey." *International Journal of Market Research* 53(5):669–86. doi: 10.2501/JMR-53-5-669-686.
- Newman, Jessica Clark, Don C. Des Jarlais, Charles F. Turner, Jay Gribble, Phillip Cooley, and Denise Paone. 2002. "The Differential Effects of Face-to-Face and Computer Interview Modes." *American Journal of Public Health* 92(2):294–97.
- Schork, Joachim, Cesare A. F. Riillo, and Johann Neumayr. 2021. "Survey Mode Effects on Objective and Subjective Questions: Evidence from the Labour Force Survey." *Journal of Official Statistics* 37(1):213–37. doi: 10.2478/jos-2021-0009.
- Sztabiński, F., & Żmijewska-Jędrzejczyk, T. Mixed Mode Survey Design: Problem efektu techniki. *Przegląd Socjologiczny, tom LXI/1 (2012), str. 31-63.*
- Tourangeau R, Rips L and Rasinski KA (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tourangeau, R.: Confidentiality, privacy and anonymity. In: Vannette, D.L., Krosnick, J.A. (eds.) *The Palgrave Handbook of Survey Research*, pp. 501–507. Palgrave Macmillan, Cham (2018)
- Willimack, D., and E. Nichols. 2010. "A Hybrid Response Process Model for Business Surveys." *Journal of Official Statistics* 26: 3–24. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/a-hybrid-response-process-modelfor-business-surveys.pdf>