PAWEŁ KLEKA, ROBERT SZYMAŃSKI, BARBARA KONAT

TRAFNOŚĆ PROCESU ODPOWIADANIA W POLSKIEJ ADAPTACJI SKALI POSTRZEGANEJ SIŁY ARGUMENTU (PAS-PL): BADANIE PILOTAŻOWE[1]

ABSTRACT. Paweł Kleka, Robert Szymański, Barbara Konat, *Trafność procesu odpowiadania w polskiej adaptacji Skali Postrzeganej Siły Argumentu (PAS-PL): badanie pilotażowe* [Response Process Validity in the Polish Adaptation of the Perceived Argument Strength Scale (PAS-PL): A Pilot Study] edited by Barbara Konat, Cristián Santibáñez, „Człowiek i Społeczeństwo" vol. LXI: *Emotion in Argumentation*, Poznań 2026, pp. 139–161, Adam Mickiewicz University. e-ISSN: 2956-5243, ISSN 0239-3271, https://doi.org/10.14746/cis.2026.61.10.

This article presents the translation and cross-cultural adaptation of the Polish version of the Perceived Argument Strength Scale (PAS-PL), along with a pilot study of response process validity in the context of arguments concerning COVID-19 vaccination. The adaptation followed international guidelines for the translation of psychometric instruments, using a bidirectional (forward–backward) translation procedure. Response processes were examined using the think-aloud method in a sample of ten students. Four categories of response problems were identified (confusion, disengagement, general response difficulties, and reactivity), and the reliability of protocol annotation was high (Krippendorff's $\alpha = 0.808$). In 83% of observations, no problems with interpretation or responding were detected. Exploratory analyses using a mixed-effects model indicated that problems occurred less frequently during the second administration of the questionnaire; a directionally consistent but statistically uncertain tendency was also observed for more frequent problems when selecting the midpoint response. The findings provide preliminary evidence for the response process validity of the PAS-PL and outline directions for further psychometric validation of the instrument.

Keywords: perceived argument strength, response process validity, think-aloud protocol, measurements validity, COVID-19 vaccination

[1] Finansowane częściowo z grantu nr 2020/39/D/HS1/00488 Polskiego Centrum Nauki.

Paweł Kleka, Uniwersytet im. Adama w Poznaniu, Wydział Psychologii i Kognitywistyki, Szamarzewskiego 89AB, 60-568 Poznań, e-mail: pawel.kleka@amu.edu.pl, https://orcid.org/0000-0003-0841-0015

Robert Szymański, Instytut Biologii Doświadczalnej im. Marcelego Nenckiego PAN, ul. Pasteura 3, 02-093 Warszawa, e-mail: r.szymanski@nencki.edu.pl, https://orcid.org/0000-0002-7556-3778

Barbara Konat, Uniwersytet im. Adama w Poznaniu, Wydział Psychologii i Kognitywistyki, Szamarzewskiego 89AB, 60-568 Poznań, e-mail: bkonat@amu.edu.pl, https://orcid.org/0000-0003-2370-4636

Wprowadzenie

Trafność pomiaru jest centralnym kryterium jakości narzędzi psychometrycznych. W ciągu ostatnich dekad konceptualizacja trafności uległa istotnym przeobrażeniom – od podejść opartych na sieciach nomologicznych (Cronbach i Meehl, 1955) po ujęcia ontologiczne, w których trafność jest własnością relacji między testem a mierzonym atrybutem (Borsboom i in., 2004). Zgodnie z tym drugim stanowiskiem test jest trafny wtedy, gdy mierzony atrybut rzeczywiście istnieje i wywiera przyczynowy wpływ na odpowiedzi respondentów (Borsboom, 2005). Badania procesu odpowiadania nie rozstrzygają samodzielnie tak rozumianej trafności, ale dostarczają ważnego, cząstkowego dowodu: pokazują, czy respondenci rozumieją pozycje i dochodzą do odpowiedzi w sposób zgodny z teoretycznymi założeniami narzędzia. Jest to etap wciąż rzadko realizowany w praktyce psychometrycznej, zwłaszcza w kontekście polskich adaptacji narzędzi (Kleka, 2021).

Skala Postrzeganej Siły Argumentu (ang. Perceived Argument Strength – PAS) autorstwa Xiaoquan Zhao i współpracowników (2011) mierzy subiektywnie odczuwaną perswazyjną moc argumentu. Skala składa się z dziewięciu pozycji testowych obejmujących ocenę wiarygodności, siły przekonywania i ważności argumentu, a także jego wpływu na postawy i intencje behawioralne respondenta. PAS została opracowana jako bardziej ekonomiczna i szerzej stosowalna alternatywa dla metody wypisywania myśli (ang. *thought-listing*), wymagającej ręcznego kodowania narracji respondentów (Zhao i in., 2011). Skala zyskała ugruntowaną pozycję w badaniach nad perswazją w obszarze zdrowia publicznego, w tym w kontekście kampanii promujących szczepienia (Bigsby i in., 2013; Dillard i in., 2007; Falk i in., 2015).

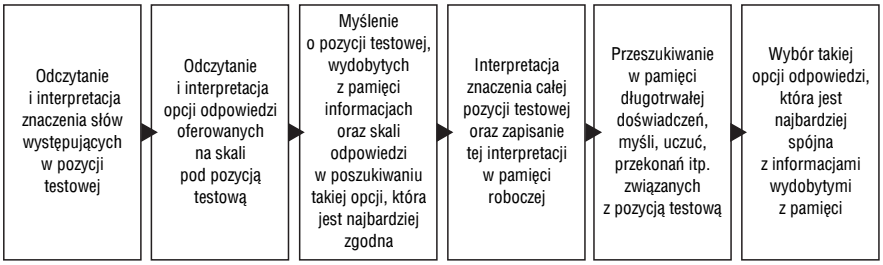
Znaczenie takiego pomiaru wykracza poza techniczną ocenę jakości narzędzia. W debatach o zdrowiu publicznym argumenty pełnią funkcję poznawczą, społeczną i retoryczną: wskazują, którym źródłom ufać, jakie ryzyka uznać za istotne oraz jakie konsekwencje działania lub zaniechania

są moralnie i praktycznie ważne. Ocena postrzeganej siły argumentu pozwala więc badać nie tylko skuteczność komunikatów perswazyjnych, lecz także sposób, w jaki odbiorcy rozpoznają wiarygodność, ważność i siłę przekonywania uzasadnień obecnych w sferze publicznej. W przypadku szczepień przeciw COVID-19 ma to szczególne znaczenie, ponieważ decyzje zdrowotne jednostek były powiązane z zaufaniem do instytucji, interpretacją wiedzy eksperckiej oraz napięciami między odpowiedzialnością indywidualną i zbiorową.

Pomimo rosnącego zapotrzebowania na psychometryczne narzędzia do badania perswazji zdrowotnej w populacji polskiej, dotychczas nie opublikowano polskiej adaptacji PAS. Luka ta jest tym bardziej dotkliwa, że przynajmniej jedna praca empiryczna z udziałem autorów niniejszego artykułu korzysta już z roboczego tłumaczenia skali (Obrębska i in., 2025), co wskazuje na potrzebę formalnego opracowania PAS-PL. Niniejszy artykuł odpowiada na tę potrzebę, opisując adaptację translatorską skali oraz pilotażowe badanie trafności procesu odpowiadania.

Trafność konstruktu i procesy odpowiadania

Zgodnie z modelem trafności Denny'ego Borsbooma i współpracowników (2004; 2005) trafność testu wymaga relacji przyczynowej między mierzonym atrybutem a odpowiedziami respondentów. Paradygmat badań procesu odpowiadania (*response process research*), wywodzący się z pracy K. Andersa Ericssona i Herberta Simona (1980; 1993), a rozwijany w kontekście walidacji kwestionariuszy psychologicznych przez badaczy takich jak David French i współpracownicy (2007), Stuart Karabenick i współpracownicy (2007) oraz Catherine Darker i David French (2009), pozwala badać jeden element tej relacji: zgodność faktycznych procesów rozumienia i odpowiadania z procesami zakładanymi przez konstrukcję narzędzia. Fundamentalnym założeniem tego podejścia jest to, że odpowiedź na pozycję testową jest wynikiem wieloetapowego procesu poznawczego obejmującego: (1) odczytanie i zrozumienie pytania, (2) interpretację, (3) odzyskanie z pamięci lub wygenerowanie sądu, (4) mapowanie sądu na dostępną skalę odpowiedzi (por. Karabenick i in., 2007). Zakłócenia na którymkolwiek z tych etapów stanowią zagrożenie trafności pomiaru (rys. 1), ale ich brak nie jest jeszcze wystarczającym warunkiem pełnej trafności narzędzia.



Rysunek 1. Model przetwarzania informacji przy odpowiadaniu na pozycje testowe

Źródło: adaptacja na podstawie Karabenick i in., 2007.

Metoda myślenia na głos (*think-aloud*; Ericsson i Simon (1980)) pozwala na bezpośredni wgląd w te procesy: respondent werbalizuje swój tok myślenia podczas wypełniania kwestionariusza, a badacz analizuje protokoły pod kątem kategorii problemów świadczących o zakłóceniach w przetwarzaniu. French i współpracownicy (2007) oraz Darker i French (2009) z powodzeniem zastosowali tę metodę do walidacji kwestionariuszy mierzących zmienne z Teorii Planowanego Zachowania (Ajzen, 1991), wykazując, że niektóre pozycje generują systematyczne trudności interpretacyjne niewidoczne w standardowych analizach psychometrycznych. Paweł Kleka (2021) wykazał z kolei, że badania procesu odpowiadania mogą ujawnić problemy trafności narzędzi stosowanych w polskim kontekście kulturowym niedostrzegalne w analizie czynnikowej.

Uzasadnienie badania

Adaptacja PAS-PL jest uzasadniona z kilku wzajemnie powiązanych powodów. Po pierwsze, rosnące zapotrzebowanie na rzetelne narzędzia do analizy perswazji i argumentacji w obszarze zdrowia publicznego – widoczne szczególnie w kontekście kampanii szczepień – powoduje konieczność dostępności dobrze opisanych instrumentów w języku polskim. Po drugie, dotychczasowe polskie badania nad perswazją korzystały albo z narzędzi nieadaptowanych formalnie, albo z procedur zastępczych (np. pomiaru postaw), które nie uchwytują specyfiki siły argumentu jako konstruktu. Po trzecie, skala PAS, choć stosunkowo prosta w budowie, zawiera pozycje, których ekwiwalentność semantyczna w języku polskim nie jest oczywista – co czyni badanie procesu odpowiadania szczególnie wartościowym elementem walidacji.

Celem badania była ocena trafności procesu odpowiadania w polskiej adaptacji translatorskiej Skali Postrzeganej Siły Argumentu (PAS-PL) w kontekście argumentów dotyczących szczepień przeciw COVID-19. Badanie miało charakter pilotażowy i stanowiło pierwszy etap szerszego programu walidacji narzędzia. Odpowiadało na trzy pytania badawcze: (1) Czy respondenci napotykają trudności w interpretacji i udzielaniu odpowiedzi na pozycje PAS-PL? (2) Jaki jest charakter i natężenie tych trudności? (3) Czy wystąpienie trudności wiąże się z wyborem odpowiedzi środkowej lub z kolejnością wypełniania kwestionariusza?

Metoda

Plan badania

Badanie miało charakter obserwacyjny i przekrojowy, obejmując analizę procesu odpowiadania na pozycje PAS-PL z zastosowaniem metody myślenia na głos. Projekt wpisuje się w fazę testowania hipotez w metodologicznym podejściu Keitha Markusa i Denny'ego Borsbooma (2013), wypełniając lukę w badaniach walidacyjnych, które rzadko uwzględniają analizę procesów poznawczych zaangażowanych w odpowiadanie na pytania kwestionariuszowe. W tym ujęciu metoda myślenia na głos dostarcza dowodów dotyczących trafności procesu odpowiadania, lecz nie przesądza samodzielnie o istnieniu, jednorodności ani przyczynowej roli mierzonego atrybutu.

Przyjęto metodę mieszaną, łączącą dane jakościowe (protokoły myślenia na głos z anotacją kategorii problemów) z ilościowymi (liczebności problemów, wskaźniki zgodności anotacji, wyniki testów zależności). Badanie nie było prerejestrowane, a jego orientacja była eksploracyjna – co jest uzasadnione nowością polskiej adaptacji PAS i ogólnym brakiem badań procesu odpowiadania w kontekście narzędzi mierzących perswazję zdrowotną.

Uczestnicy

Rekrutacja odbywała się za pośrednictwem mediów społecznościowych. Jedynym kryterium włączenia był status studenta (studia pierwszego lub drugiego stopnia). W badaniu wzięło udział dziesięć osób (5 kobiet, 5 mężczyzn) w wieku 22–25 lat (M = 23,5, SD = 1,1). Sześć osób miało 23 lata, trzy – 25 lat, jedna – 22 lata. Próba obejmowała studentów nauk społeczno-ekonomicznych (50%), humanistycznych (30%) oraz ścisłych i medycznych

(20%), studiujących na publicznych uczelniach w Poznaniu. Wszyscy uczestnicy zadeklarowali zaszczepienie się przeciw COVID-19 (70% przyjęło trzy dawki, 30% – dwie dawki). Oznacza to, że badanie obejmowało osoby, które już podjęły zachowanie docelowe, a więc pozwala przede wszystkim ocenić proces odpowiadania w warunkach postdecyzyjnych. Jednorodność próby pod względem wieku, wykształcenia i statusu szczepienia stanowi istotne ograniczenie przy generalizacji wyników, zwłaszcza na osoby niezaszczepione lub niezdecydowane.

Polskie tłumaczenie Skali Postrzeganej Siły Argumentu

Skala PAS (Zhao i in., 2011) składa się z dziewięciu pozycji testowych ocenianych na pięciostopniowej skali Likerta. Narzędzie mierzy subiektywnie odczuwaną perswazyjną moc argumentu przez ocenę jego wiarygodności, siły przekonywania i ważności, a także wpływu na postawy i decyzje wobec zachowania docelowego. Wynik ogólny oblicza się jako średnią odpowiedzi na pozycje 1–5 i 8–9 oraz indeks przychylności myśli (będący przeliczeniem pozycji 6 i 7).

Tłumaczenie przeprowadzono zgodnie z rekomendacjami międzynarodowych standardów adaptacji narzędzi psychometrycznych (Hambleton i Zenisky, 2010; Iliescu, 2017) oraz z wytycznymi International Test Commission (2017). Zastosowano schemat tłumaczenia dwukierunkowego: dwie niezależne filolożki z wieloletnim doświadczeniem translatorskim dokonały przekładu z języka angielskiego na polski, po czym spotkały się z ekspertką z zakresu teorii argumentacji i językoznawstwa, aby wypracować wspólną wersję. Wersję polską przesłano następnie do dwóch kolejnych tłumaczek (badaczki argumentacji politycznej i zawodowe tłumaczki), które niezależnie opracowały tłumaczenie wsteczne (z polskiego na angielski) bez dostępu do oryginału. W ostatniej fazie zwołano panel z udziałem wszystkich tłumaczek i psychometryka, poświęcony porównaniu oryginału z tłumaczeniami wstecznymi.

Większość pozycji nie budziła wątpliwości – zgodność między tłumaczeniami była wysoka. Problematyczna okazała się pozycja pierwsza: angielskie *believable* nie ma jednoznacznego odpowiednika polskiego w kontekście oceny siły argumentu. Panel rozpatrywał cztery warianty tłumaczenia. Aby dokonać empirycznie uzasadnionego wyboru, przeprowadzono badanie rankingowe na próbie 27 studentów kognitywistyki, których poproszono o uszeregowanie wariantów od najbardziej do najmniej trafnego. Największe wsparcie – 63% głosów (tab. 1) – uzyskało tłumaczenie dosłowne

„Podany powód za […] jest wiarygodny", które włączono do pilotażowej wersji PAS-PL.

Tabela 1. Warianty tłumaczenia pozycji testowej nr 1

Wariant tłumaczenia	Miejsce 1 (%)	Miejsce 2 (%)	Miejsce 3 (%)	Miejsce 4 (%)
Podany powód za {…} jest wiarygodny.	63,0	18,5	14,8	3,7
Podany powód za {…} brzmi wiarygodnie.	22,2	33,3	18,5	25,9
Podany powód za {…} jest dla mnie wiarygodny.	11,1	29,6	40,7	18,5
Podany powód za {…} jest rzeczowy.	3,7	18,5	25,9	51,9

Uwaga. *N* = 27. Pierwowzór: The statement is a reason for {…} that is believable. Ostateczna pilotażowa wersja PAS-PL zestawiona z oryginałem jest przedstawiona w tabeli 2.

Źródło: badania własne.

Tabela 2. Skala PAS i jej polska adaptacja w wersji pilotażowej

PT	Perceived Argument Strength Scale (Zhao i in., 2011)	Skala Postrzeganej Siły Argumentu PAS-PL
1	The statement is a reason for {…} that is believable.	Podany powód za {…} jest wiarygodny.
2	The statement is a reason for {…} that is convincing.	Podany powód za {…} jest przekonujący.
3	The statement gives a reason for {…} that is important to me.	Podany powód za {…} jest dla mnie ważny.
4	The statement helped me feel confident about how best to {…}.	Stwierdzenie dało mi pewność, że wiem, jak najlepiej {…}.
5	The statement would help my friends {…}.	Stwierdzenie pomogłoby moim przyjaciołom w podjęciu decyzji o {…}.
6	The statement put thoughts in my mind about wanting to {…}.	Stwierdzenie wywołało u mnie myśl o tym, żeby {…}.
7	The statement put thoughts in my mind about not wanting to {…}.	Stwierdzenie wywołało u mnie myśl o tym, żeby NIE {…}.
8	Overall, how much do you agree or disagree with the statement?	W jakim stopniu ogólnie zgadzasz się lub nie zgadzasz się z tym stwierdzeniem?
9	Is the reason the statement gave for {…} a strong or weak reason?	Czy podany powód za tym, żeby {…}, jest słaby czy mocny?

Uwaga. PT = numer pozycji testowej.

Źródło: badania własne.

Korpus argumentów i procedura selekcji

Bazą wyjściową do wyboru materiału bodźcowego były treści promocyjne Narodowego Programu Szczepień adresowane do społeczeństwa polskiego pod hasłem #szczepimysię. Studenci kognitywistyki uczestniczący w kursie języka kognitywnego zebrali argumenty ze spotów promocyjnych, plakatów, materiałów informacyjnych i publicznych wypowiedzi ekspertów zdrowia publicznego. Uzupełnieniem były argumenty zaczerpnięte z oficjalnych komunikatów Ministerstwa Zdrowia, Narodowego Instytutu Zdrowia Publicznego oraz instytucji akademickich. Wszystkie argumenty ekstrahowano wyłącznie w formie tekstowej, bez towarzyszących elementów wizualnych i audio, aby wyeliminować wpływ medium na ocenę siły argumentu.

Wstępnie zebrano 92 zbiory przesłanek (sekwencje zdań logicznie powiązane z konkluzją „zaszczep się przeciw COVID-19"; zgodnie z rozróżnieniem Fransa van Eemerena i in. (2014) mogły mieć strukturę równoległą lub szeregową). Troje sędziów kompetentnych (dwie osoby na trzecim i jedna na piątym roku kognitywistyki) oceniało, czy każdy ze zbiorów stanowi argument za szczepieniem. Zgodność sędziów dla 92 obserwacji, obliczona współczynnikiem Krippendorffa (2019), wyniosła $\alpha = 0{,}77$ (SE = 0,085, 95% CI [0,583; 0,912]) – co uznano za wystarczające do podjęcia decyzji inkluzyjnych. Argument zachowywano, gdy co najmniej dwóch spośród trojga sędziów głosowało za jego utrzymaniem. Na tej podstawie odrzucono 25 argumentów, tworząc ostateczny korpus 67 pozycji; do badania właściwego wylosowano 20 zbiorów przesłanek (pełna lista w Załączniku).

Procedura badania

Badanie przeprowadzono zgodnie ze standardami etyki badań naukowych. Uczestnicy zostali poinformowani o celu, procedurze, rejestracji audio i prawie do rezygnacji; wszyscy wyrazili świadomą zgodę, a dane zanonimizowano. Ze względu na prosty plan badawczy, realizację w ramach struktury uczelni oraz skrajnie niskie prawdopodobieństwo wywołania negatywnych odczuć u osób badanych nie wystąpiono o odrębną opinię komisji etycznej.

Etap 1: instrukcja. Po udzieleniu zgody odczytywano instrukcję przygotowaną na podstawie szablonów stosowanych w badaniach myślenia na głos (Darker i French, 2009; French i in., 2007):

> Za chwilę rozpoczniemy badanie oceny argumentów zachęcających do szczepień przeciw COVID-19. Na potrzeby tego badania opracowaliśmy

> polską adaptację kwestionariusza Skali Postrzeganej Siły Argumentu. Chcemy sprawdzić, czy ludzie rozumieją pytania w ten sam sposób, w jaki rozumieją je twórcy kwestionariusza. Aby to zrobić, poproszę cię o myślenie na głos podczas wypełniania ankiety. Kwestionariusz składa się z 18 pytań, po dziewięć dla każdego z dwóch ocenianych za ich pomocą argumentów. Chcę, abyś powiedział(a) mi wszystko, co myślisz, czytając każde pytanie i decydując, jak na nie odpowiedzieć. Chciał(a)bym, abyś mówił(a) bez przerwy. Nie chcę, żebyś planował(a) to, co mówisz, ani nie próbował(a) mi tego wytłumaczyć. Po prostu zachwuj się tak, jakbyś był(a) sam(a) w pokoju. Jeśli będziesz milczeć przez dłuższy czas, poproszę cię, abyś mówił(a) dalej. Czy masz jakieś pytania dotyczące przebiegu badania?

Etap 2: rozgrzewka. Przed badaniem właściwym przeprowadzano ćwiczenie myślenia na głos z prostym zadaniem arytmetycznym (obliczenie ceny frytek na podstawie ceny drożdżówki i zapiekanki), służące zaznajomieniu uczestnika z metodą. W jednym przypadku uczestnik miał trudność z werbalizacją; wymagało to powtórzenia instrukcji i zademonstrowania przykładu.

Etap 3: badanie właściwe. Każdy uczestnik wypełniał kwestionariusz PAS-PL dwukrotnie, oceniając dwa różne argumenty, jednocześnie werbalizując tok myślenia. Odpowiedzi zaznaczano na tablecie, nagrania audio rejestrował dyktafon. Badacz przebywał kilka metrów od uczestnika, aby ograniczyć pokusę dialogu, reagował wyłącznie na pytania techniczne. Średni czas sesji wyniósł 12 min 50 s (*SD* = 7 min 8 s, min. = 5 min, max = 24 min 19 s).

Anotacja protokołów i analiza danych

Z plików audio usunięto fragmenty niezwiązane z zadaniem (rozgrzewka, przerwy techniczne). Transkrypcje wykonano przy wsparciu programu Descript (www.descript.com). Każdy protokół podzielono na obserwacje obejmujące proces odpowiadania jednej osoby na jedno pytanie (*N* = 180 obserwacji łącznie). Anotacji dokonywało dwoje niezależnych anotatorów: osoba z doświadczeniem w badaniach argumentacji (studentka trzeciego roku kognitywistyki) oraz zawodowy dziennikarz (student czwartego roku socjologii) bez wcześniejszego doświadczenia anotatorskiego.

Schemat kodowania oparto na kategoriach zaproponowanych przez Darker i Frencha (2009), wyznaczając go na podstawie 54 losowo wybranych obserwacji (służących jako materiał treningowy). Wyróżniono dwie główne grupy problemów:

- problemy z interpretacją pozycji testowych: (a) zdezorientowanie – brak zrozumienia lub znaczna niepewność semantyczna; (b) oderwanie – spontaniczne rozumowanie wykraczające poza treść pytania;

- problemy z odpowiadaniem: (c) ogólny problem z odpowiedzią – wahanie, niemożność podjęcia decyzji; (d) reaktywność – zmiana odpowiedzi pod wpływem pytań kolejnych lub sytuacji badawczej.

Po ustaleniu schematu pozostałe obserwacje anotowało niezależnie dwoje anotatorów, kończąc proces panelem uzgodnieniowym. Rzetelność anotacji obliczono współczynnikiem α Krippendorffa (2019) dla trzech par sędziów (dwoje anotatorów oraz główny badacz, który anotował wszystkie obserwacje niezależnie). Analizę statystyczną przeprowadzono w R (wersja 4.2.0) z użyciem pakietów icr (zgodność anotacji), lme4 (modele mieszane) i tidyverse (manipulacja danymi). Do badania związków między wystąpieniem problemu a zmiennymi towarzyszącymi zastosowano logistyczny model mieszany, w którym zmienną zależną było wystąpienie co najmniej jednego problemu w danej obserwacji, predyktorami stałymi były kolejność wypełnienia kwestionariusza i wybór odpowiedzi środkowej, a uczestnik był uwzględniony jako efekt losowy. Model ten odzwierciedla zagnieżdżenie 180 obserwacji w 10 osobach badanych. Wyniki i skrypty udostępniono w repozytorium Open Science Framework (https://doi.org/10.17605/OSF.IO/X3CNA).

Wyniki

Rzetelność anotacji protokołów

Rzetelność schematu anotacji była wysoka. Współczynnik α Krippendorffa obliczony dla wszystkich trzech sędziów wyniósł 0,808 (SE = 0,058, 95% CI [0,681; 0,910]). Analiza par ujawniła nieco niższą zgodność między anotatorami A i B (α = 0,719), przy porównywalnie wysokiej zgodności dla par A–C i B–C (odpowiednio α = 0,853 i α = 0,856). Wyniki te przedstawia tabela 3.

Niższa zgodność pary A–B wymaga ostrożnej interpretacji. Spośród 126 obserwacji kodowanych niezależnie anotatorzy A i B byli zgodni w 116 przypadkach, a 10 przypadków było rozbieżnych. Większość tych rozbieżności (9 z 10) dotyczyła granicy między brakiem problemu a wystąpieniem problemu, a nie wyboru między dwiema różnymi kategoriami problemów. Najczęściej sporne były przypadki klasyfikowane przez jednego anotatora jako brak problemu, a przez drugiego jako ogólny problem z odpowiedzią (5 przypadków). W 10 rozbieżnych przypadkach trzeci sędzia zgadzał się równie często z anotatorem A, jak z anotatorem B (po 5 przypadków),

co nie wskazuje na jednostronne rozstrzyganie niezgodności przez głównego badacza.

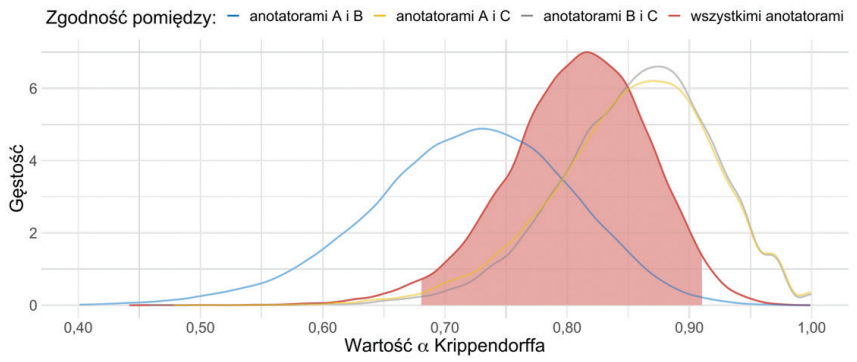
Tabela 3. Wyniki analizy zgodności anotacji protokołów myślenia na głos

Anotatorzy	α Krippendorffa	SE	95% CI (dolna)	95% CI (górna)
Wszyscy (A, B i C)	0,808	0,058	0,681	0,910
Anotatorzy A i B	0,719	0,083	0,541	0,867
Anotatorzy A i C	0,853	0,066	0,706	0,969
Anotatorzy B i C	0,856	0,062	0,722	0,967

Uwaga. *N* = 126 obserwacji (dla 7 spośród 10 protokołów; pozostałe protokoły kodowano wspólnie podczas ustalania schematu). Błędy standardowe i przedziały ufności uzyskano na podstawie 20 000 symulacji metodą bootstrap.

Źródło: badania własne.

Rysunek 2 przedstawia rozkład zgodności w podziale na anotatorów; zbieżność ocen potwierdza, że schemat anotacji był interpretowalny przez osoby z różnym doświadczeniem.



Rys. 2. Rozkład zgodności w podziale na anotatorów

Źródło: badania własne.

Kategorie i rozkład problemów

W 83% obserwacji (149 z 180) nie zidentyfikowano żadnych problemów z interpretacją lub odpowiadaniem. Najczęstszym problemem był ogólny problem z odpowiadaniem (8% obserwacji, n = 14), manifestujący się wahaniem respondentów przy wyborze odpowiedzi. Na zbliżonym poziomie wystąpiło zdezorientowanie (7%, n = 13), dotyczące przede wszystkim trudności w interpretacji słów kluczowych, takich jak „wiarygodny" czy „przekonujący". Kategorie „oderwanie" (n = 1, ~1%) i „reaktywność" (n = 3, ~2%) występowały sporadycznie. Rozkład problemów dla poszczególnych pozycji testowych przedstawia tabela 4.

Tabela 4. Liczebności kategorii problemów dla każdej pozycji testowej

PT	Brak problemu	Zdezorientowanie	Oderwanie	Problem z odpowiedzią	Reaktywność
1	16	3	0	1	0
2	15	5	0	0	0
3	18	1	0	0	1
4	16	1	0	3	0
5	16	0	0	4	0
6	17	1	1	0	1
7	18	2	0	0	0
8	16	0	0	3	1
9	17	0	0	3	0
Suma	149 (83%)	13 (7%)	1 (1%)	14 (8%)	3 (2%)

Uwaga. PT = numer pozycji testowej.

Źródło: badania własne.

Analiza jakościowa protokołów pozwoliła zidentyfikować specyficzne trudności dla poszczególnych pozycji:

- Pozycja 1 („Podany powód za {…} jest wiarygodny"): troje uczestników miało trudności z interpretacją słowa „wiarygodny" w kontekście oceny argumentu, sygnalizując, że termin ten kojarzy im się raczej z autentycznością źródła niż z jakością rozumowania. Problem dotyczy więc nie tylko perspektywy oceny, ale także przedmiotu oceny: respondenci mogą oceniać wiarygodność źródła, prawdziwość przesłanki albo jakość uzasadnienia. Dodanie formuły „dla mnie" mogłoby

wyraźniej zlokalizować ocenę w perspektywie respondenta, ale nie usuwa tej wieloznaczności. Wynik ten sugeruje potrzebę dalszego testowania alternatywnych sformułowań odwołujących się bardziej bezpośrednio do prawdziwości, prawdopodobieństwa lub zaufania do podanego powodu;

- Pozycja 2 („Podany powód za {…} jest przekonujący"): pięcioro uczestników wyrażało dezorientację, zastanawiając się, czy pytanie różni się od poprzedniego. W trzech przypadkach respondenci sygnalizowali zamiar zmiany odpowiedzi na pozycję 1, co wskazuje na reaktywność wynikającą z semantycznej bliskości pozycji 1 i 2;
- Pozycja 5 („Stwierdzenie pomogłoby moim przyjaciołom {…}"): czworo uczestników miało trudności z perspektywą trzeciej osoby – oszacowaniem reakcji przyjaciół na argument. Sugeruje to konieczność albo uściślenia grupy referencyjnej (np. „niezaszczepionych przyjaciół"), albo rozważenia usunięcia tej pozycji z puli przeznaczonej do obliczania wyniku ogólnego w wersji dla populacji w większości zaszczepionej;
- Pozycje 6 i 7 (indeks przychylności myśli): w przypadku pozycji 6 jedna osoba angażowała się w rozważania o cudzych reakcjach (oderwanie). Pozycja 7 – zawierająca przeczenie – była dwukrotnie niezrozumiana przez tę samą uczestniczkę, co potwierdza znane z literatury ryzyko błędnej interpretacji pozycji z zaprzeczeniem (Wetzel i in., 2016). Interpretacja tych pozycji jest dodatkowo ograniczona przez status szczepienia uczestników. Ponieważ wszyscy badani byli już zaszczepieni, pytania o myśli dotyczące (nie)szczepienia się odnosiły się do decyzji już podjętej, a nie do obecnie rozważanego zachowania. Wyniki sugerują więc, że pozycje 6 i 7 mogą tracić trafność procesową w badaniach osób po decyzji szczepiennej, ale nie pozwalają rozstrzygnąć, jak funkcjonują u osób niezaszczepionych lub niezdecydowanych.

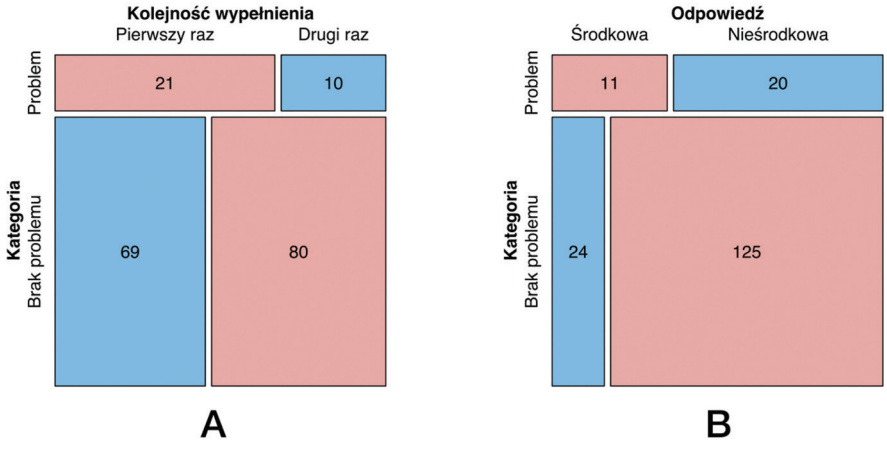
Korelaty wystąpienia problemów

Łączna liczba problemów na osobę wahała się od 0 do 11 (Md = 2,5, Q1 = 1, Q3 = 4), wskazując na znaczne różnice indywidualne w doświadczaniu i werbalizowaniu trudności.

Ze względu na zagnieżdżenie obserwacji w osobach związki między wystąpieniem problemu a zmiennymi towarzyszącymi oszacowano za pomocą logistycznego modelu mieszanego z losowym interceptem dla uczestnika (rys. 3).

1. Kolejność wypełnienia: problemy występowały częściej przy pierwszym niż przy drugim wypełnieniu kwestionariusza (21/90 vs 10/90 obserwacji). Po uwzględnieniu zróżnicowania między osobami drugie wypełnienie wiązało się z niższymi szansami wystąpienia problemu (OR = 0,34, 95% CI [0,14; 0,85], p = 0,020; równoważnie: pierwsze vs drugie wypełnienie OR = 2,90, 95% CI [1,18; 7,13]). Efekt ten interpretujemy jako możliwy przejaw uczenia się i habituacji – przy drugiej ekspozycji respondenci dysponowali już roboczą reprezentacją pozycji testowych.

2. Wybór odpowiedzi środkowej: problemy występowały częściej przy odpowiedziach środkowych niż przy odpowiedziach nieśrodkowych (11/35 vs 20/145 obserwacji). W modelu mieszanym efekt ten był zgodny kierunkowo z wynikami Darker i Frencha (2009), ale nie osiągnął konwencjonalnego progu istotności statystycznej (OR = 2,47, 95% CI [0,93; 6,57], p = 0,070). Wynik ten należy więc traktować jako eksploracyjną przesłankę, że odpowiedź środkowa może w części przypadków sygnalizować trudność w interpretacji lub mapowaniu sądu na skalę odpowiedzi.



Rysunek 3. Kolejność wypełnienia kwestionariusza a wystąpienie problemu – zestawienie proporcji

Źródło: badania własne.

Dyskusja

Trafność procesu odpowiadania w PAS-PL

Wyniki badania dostarczają empirycznych przesłanek na rzecz trafności procesu odpowiadania w polskiej adaptacji translatorskiej Skali Postrzeganej Siły Argumentu. Odsetek odpowiedzi wolnych od zidentyfikowanych problemów wynosił 83%, co jest porównywalne z wynikami uzyskiwanymi w analogicznych badaniach z użyciem protokołów myślenia na głos dla kwestionariuszy Teorii Planowanego Zachowania (Darker i French, 2009; French i in., 2007). Uzyskany wynik sugeruje, że PAS-PL jest w dużej mierze zrozumiała dla osób badanych w pilotażu oraz uruchamia procesy odpowiadania zgodne z założeniami teoretycznymi skali.

Jednocześnie analiza jakościowa ujawniła problemy o charakterze systematycznym, skoncentrowane wokół pozycji 1, 2, 5, 6 i 7. Wskazuje to, że obserwowane trudności nie są losowe, lecz wynikają z właściwości konkretnych sformułowań. Najbardziej wyraźnym przykładem jest pozycja 2 („przekonujący"), która przez część respondentów była oceniana nie niezależnie, lecz w odniesieniu do odpowiedzi udzielonej wcześniej na pozycję 1. Takie sekwencyjne porównywanie pozycji narusza założenie o niezależności pomiarów w bloku testowym i może prowadzić do zaniżenia rzeczywistej zmienności odpowiedzi. Podobne efekty reaktywności semantycznej obserwowali French i współpracownicy w badaniach nad kwestionariuszami intencji behawioralnych (French i in., 2007).

Uzyskane wyniki należy interpretować jako jeden element szerszego argumentu walidacyjnego. W perspektywie Borsbooma i współpracowników (2004; 2005) zgodność procesów odpowiadania z założeniami narzędzia jest ważna, ponieważ ogranicza ryzyko, że odpowiedzi są generowane przez czynniki uboczne, takie jak niezrozumienie pozycji, reaktywność lub trudność mapowania sądu na skalę. Nie jest to jednak wystarczający test trafności w sensie ontologicznym, ponieważ nie rozstrzyga o strukturze atrybutu ani o jego przyczynowym wpływie na odpowiedzi. W tym kontekście zebrane dane stanowią wstępny dowód trafności procesu odpowiadania w PAS-PL, wskazując jednocześnie na potrzebę dopracowania wybranych pozycji przed szerokim zastosowaniem badawczym.

Interpretując uzyskane wyniki, należy uwzględnić ograniczenia metody myślenia na głos. Literatura wskazuje, że protokoły współbieżne, choć bardziej wiarygodne niż retrospektywne, mogą pomijać część automatycznych

lub słabo uświadamianych procesów poznawczych, a także wpływać na przebieg wykonywania zadań (Hu i Gao, 2017; Russo i in., 1989). Oznacza to, że rzeczywisty odsetek problemów procesowych może być nieco wyższy niż zidentyfikowany w niniejszym badaniu, jednak zbieżność uzyskanych wyników z danymi z wcześniejszych badań sugeruje, że skala tych zniekształceń jest ograniczona i nie podważa głównych wniosków.

Odpowiedź środkowa jako wskaźnik trudności

Kierunek zależności między problemami z odpowiadaniem a wyborem opcji środkowej jest zgodny z modelem odpowiedzi środkowej jako strategii redukcji trudności poznawczej (Wetzel i in., 2016), choć w modelu mieszanym efekt ten nie osiągnął konwencjonalnego progu istotności statystycznej. Wynik ten ma więc charakter eksploracyjny, ale pozostaje ważny interpretacyjnie: odsetek odpowiedzi środkowych może być w pewnych przypadkach wskaźnikiem jakości pozycji testowej, a nie wyrazem faktycznie neutralnej postawy. W przyszłych badaniach walidacyjnych PAS-PL warto rozważyć analizę wzorców odpowiedzi środkowych jako diagnostyczne kryterium uzupełniające.

Habituacja i kolejność wypełnienia

Mniejsza liczba problemów przy drugim wypełnieniu kwestionariusza jest zgodna z teorią uczenia się proceduralnego: respondenci, którzy zetknęli się wcześniej z formatem i językiem skali, interpretują kolejne wypełnienie sprawniej (Ericsson i Simon, 1980). Efekt ten należy uwzględnić w projektach badawczych, w których PAS-PL jest stosowana w schemacie powtarzanych pomiarów – wyniki z pierwszej i kolejnych sesji mogą nie być w pełni porównywalne pod względem procesu odpowiadania.

Ograniczenia badania

Niniejsze badanie ma kilka istotnych ograniczeń. Po pierwsze, próba była mała i jednorodna: wszyscy uczestnicy byli studentami, w podobnym wieku i – co szczególnie ważne – wszyscy zaszczepieni. Ogranicza to możliwość generalizacji na grupy bardziej zróżnicowane demograficznie, kulturowo lub pod względem postawy wobec szczepień. Konsekwencja tego ograniczenia jest szczególnie istotna dla pozycji 6 i 7, które odnoszą się do myśli o szczepieniu lub nieszczepieniu się. W niniejszym badaniu pozycje te były oceniane

przez osoby po decyzji szczepiennej, a więc w warunkach postdecyzyjnych. Nie można na tej podstawie wnioskować, że te same procesy odpowiadania wystąpią u osób, które dopiero rozważają szczepienie, są niezdecydowane albo są mu przeciwne. Po drugie, metoda myślenia na głos sama w sobie może modyfikować procesy poznawcze, które ma mierzyć (Durning i in., 2013), choć efekt ten jest trudny do wyeliminowania i jest uznawany za inherentne ograniczenie paradygmatu. Po trzecie, zastosowany schemat anotacji był jednoetykietowy (każda obserwacja otrzymywała co najwyżej jedną kategorię problemów), co mogło zaniżyć liczebności w przypadkach, gdy dwa typy trudności współwystępowały w tej samej wypowiedzi. Wieloetykietowa anotacja stanowiłaby wartościowe rozszerzenie.

Implikacje i kierunki przyszłych badań

Zebrane wyniki wskazują na kilka istotnych kierunków dalszego rozwoju polskiej adaptacji Skali Postrzeganej Siły Argumentu. W pierwszej kolejności zasadne jest dopracowanie treści wybranych pozycji, które generowały problemy procesowe. Dotyczy to zwłaszcza pozycji 1, w której samo dodanie sformułowania „dla mnie" może nie wystarczyć do usunięcia wieloznaczności słowa „wiarygodny". W kolejnych badaniach warto porównać warianty odwołujące się wprost do tego, czy podany powód wydaje się respondentowi prawdziwy, prawdopodobny lub godny zaufania, oraz sprawdzić, czy zmiana ta nie oddala pozycji od znaczenia oryginalnego *believable*. Dalszego dopracowania wymaga także pozycja 5, w której warto uściślić grupę referencyjną, oraz pozycja 7, której interpretacja przez część respondentów sugeruje zasadność warunkowego stosowania w zależności od statusu szczepienia. Zmiany te mogą ograniczyć niejednoznaczności semantyczne i zmniejszyć ryzyko systematycznych zniekształceń w procesie odpowiadania.

Równolegle konieczne jest przeprowadzenie dalszych badań walidacyjnych na większej i bardziej zróżnicowanej próbie, obejmującej m.in. osoby niezaszczepione, niezdecydowane oraz starsze. Włączenie osób o różnym statusie szczepienia jest warunkiem oceny, czy pozycje dotyczące intencji i przychylności myśli zachowują trafność procesową także wtedy, gdy decyzja zdrowotna nie została jeszcze podjęta. Pozwoli to nie tylko na zwiększenie mocy wnioskowania, lecz także na ocenę niezmienności pomiarowej PAS-PL między grupami, co jest warunkiem jej rzetelnego stosowania w badaniach porównawczych. Kolejnym krokiem powinno być uwzględnienie standardowych analiz psychometrycznych, w tym potwierdzającej analizy czynnikowej weryfikującej jednoczynnikową strukturę

zaproponowaną przez autora oryginalnej skali (Zhao i in., 2011), a także analizy rzetelności wewnętrznej i stabilności wyników w schemacie test–retest. Uzupełnieniem tych działań byłaby weryfikacja trafności kryterialnej poprzez zbadanie związków PAS-PL z innymi wskaźnikami siły argumentu, takimi jak rezultaty metod opartych na generowaniu myśli.

Kolejny etap prac powinien również objąć przygotowanie raportu technicznego dokumentującego adaptację PAS-PL w odniesieniu do wytycznych International Test Commission. Taki raport powinien porządkować warunki formalne adaptacji, zakres dotychczas wykonanych prac translatorskich, status empirycznej weryfikacji równoważności konstruktu, procedurę docelowego stosowania skali, zasady skalowania i interpretacji wyników oraz dokumentację kolejnych analiz walidacyjnych. W niniejszym artykule elementy te omawiamy tylko w zakresie niezbędnym do przedstawienia pilotażowego badania trafności procesu odpowiadania.

Interpretacja uzyskanych wyników wymaga jednocześnie uwzględnienia specyfiki zastosowanej metody badawczej. Jak pokazują wcześniejsze badania, odsetek błędów w protokołach myślenia na głos zależy m.in. od złożoności zadania, doświadczenia uczestników, formy instrukcji oraz kontekstu społecznego badania (Barkaoui, 2010; Brinkman, 1993; Li, 2004; Russo i in., 1989). W literaturze podkreśla się również, że protokoły współbieżne, choć mniej podatne na zapominanie i fabrykację niż retrospektywne, mogą pomijać część procesów automatycznych lub wpływać na przebieg wykonywania zadania (Hu i Gao, 2017). Z tego względu uzyskany w niniejszym badaniu odsetek odpowiedzi wolnych od problemów należy traktować jako ostrożny, lecz wiarygodny wskaźnik jakości procesowej skali.

Z perspektywy badań nad komunikacją publiczną PAS-PL może być użyteczna nie tylko jako kwestionariuszowy wskaźnik jakości komunikatu, lecz także jako narzędzie analizy tego, jak odbiorcy przekształcają argumenty społeczne i eksperckie w subiektywne oceny wiarygodności, ważności i siły przekonywania. Ma to znaczenie dla badań nad perswazją zdrowotną, ale również dla szerszych analiz sporów publicznych, w których skuteczność argumentu zależy od relacji między treścią uzasadnienia, zaufaniem do źródeł wiedzy oraz sytuacją społeczną odbiorcy.

Uwzględnienie tych uwarunkowań metodologicznych oraz podjęcie wskazanych działań walidacyjnych może przyczynić się do przekształcenia PAS-PL z narzędzia wstępnie zweryfikowanego w lepiej udokumentowany instrument badawczy, użyteczny zarówno w badaniach nad perswazją, jak i w szerszym kontekście analiz postaw i procesów argumentacyjnych.

Literatura

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Barkaoui, K. (2010). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, *28*(1), 51–75. https://doi.org/10.1177/0265532210376379

Bigsby, E., Cappella, J.N., Seitz, H.H. (2013). Efficiently and effectively evaluating public service announcements: Additional evidence for the utility of perceived effectiveness. *Communication Monographs*, *80*(1), 1–23. https://doi.org/10.1080/03637751.2012.739706

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511490026

Borsboom, D., Mellenbergh, G.J., Van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brinkman, J.A. (1993). Verbal protocol accuracy in fault diagnosis. *Ergonomics*, *36*(11), 1381–1397. https://doi.org/10.1080/00140139308968007

Cronbach, L.J., Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281. https://doi.org/10.1037/h0040957

Darker, C.D., French, D.P. (2009). What sense do people make of a theory of planned behaviour questionnaire? A think-aloud study. *Journal of Health Psychology*, *14*(7), 861–871. https://doi.org/10.1177/1359105309340983

Dillard, J.P., Weber, K.M., Vail, R.G. (2007). The relationship between the perceived and actual effectiveness of persuasive messages: A meta-analysis with implications for formative campaign research. *Journal of Communication*, *57*(4), 613–631. https://doi.org/10.1111/j.1460-2466.2007.00360.x

Durning, S.J., Artino Jr, A.R., Beckman, T.J., Graner, J., Van Der Vleuten, C., Holmboe, E., Schuwirth, L. (2013). Does the think-aloud protocol reflect thinking? Exploring functional neuroimaging differences with thinking (answering multiple choice questions) versus thinking aloud. *Medical Teacher*, *35*(9), 720–726. https://doi.org/10.3109/0142159x.2013.801938

Eemeren, F.H., van, Garssen, B., Krabbe, E.C.W., Snoeck Henkemans, A.F., Verheij, B., Wagemans, J.H.M. (2014). *Handbook of Argumentation Theory*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9473-5

Ericsson, K.A., Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson, K.A., Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Falk, E.B., O'Donnell, M.B., Tompson, S., Gonzalez, R., Dal Cin, S., Strecher, V., Cummings, K.M., An, L. (2015). Functional brain imaging predicts public health campaign success. *Social Cognitive and Affective Neuroscience*, *11*(2), 204–214. https://doi.org/10.1093/scan/nsv108

French, D.P., Cooke, R., Mclean, N., Williams, M., Sutton, S. (2007). What do people think about when they answer Theory of Planned Behaviour questionnaires? A think aloud study. *Journal of Health Psychology*, *12*(4), 672–687. https://doi.org/10.1177/1359105307078174

Hambleton, R.K., Zenisky, A.L. (2010). Translating and adapting tests for cross-cultural assessments. W: D. Matsumoto, F.J.R.E. van de Vijver (red.), *Cross-cultural Research Methods In Psychology* (s. 46–70). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511779381.004

Hu, J., Gao, X.A. (2017). Using think-aloud protocol in self-regulated reading research. *Educational Research Review*, *22*, 181–193. http://doi.org/10.1016/j.edurev.2017.09.004

Iliescu, D. (2017). Translation designs. W: *Adapting Tests in Linguistic and Cultural Situations* (s. 355–414). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316273203.009

Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R., De Groot, E., Gilbert, M.C., Musu, L., Kempler, T.M., Kelly, K.L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151.

Kleka, P. (2021). *Trafność pomiaru w praktyce psychometrycznej. Badanie empiryczne*. Poznań: Wydawnictwo Nauk Społecznych i Humanistycznych UAM w Poznaniu.

Krippendorff, K. (2019). Content Analysis: An Introduction to Its Methodology (wyd. 4). Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781071878781

Li, D. (2004). Trustworthiness of think-aloud protocols in the study of translation processes. *International Journal of Applied Linguistics*, *14*(3), 301–313. https://doi.org/10.1111/j.1473-4192.2004.00067.x

Markus, K.A., Borsboom, D. (2013). Frontiers of test validity theory: Measurement, causation, and meaning. New York: Routledge. https://doi.org/10.4324/9780203501207

Obrębska, M., Konat, B., Kleka, P., Kiljan, K., Dembska, N. (2025). *Personality traits and argumentation style as factors in perceived argument strength in natural conditions of competitive debating: The role of psychological matching of speaker and listener*. SSRN. https://doi.org/10.2139/ssrn.5416998

Russo, J.E., Johnson, E.J., Stephens, D.L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*(6), 759–769. https://doi.org/10.3758/BF03202637

Wetzel, E., Böhnke, J.R., Brown, A. (2016). Response biases. W: F.T.L. Leong, D. Iliescu (red.), *The ITC International Handbook of Testing and Assessment* (s. 349–363). Oxford: Oxford University Press.

Zhao, X., Strasser, A., Cappella, J.N., Lerman, C., Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. *Communication Methods and Measures*, *5*(1), 48–75. https://doi.org/10.1080/19312458.2010.547822

Załącznik
LISTA ARGUMENTÓW UŻYTYCH W BADANIU

Tabela A1
Argumenty wybrane do badania

ID	Argument
1	Szczepionka jest ogromną szansą na uodpornienie społeczeństwa na zakażenie oraz zdobycie kontroli nad transmisją wirusa SARS-CoV-2. Wdrożenie masowych szczepień, przy wysokim procencie osób zaszczepionych, pozwoli na stałe powrócić do pełnej funkcjonalności szkół, sklepów, obiektów sportowych i kulturalnych.
4	Nie bagatelizujmy śmiertelnego wirusa, jakim jest COVID-19. Codziennie na świecie z powodu zakażenia umierają tysiące ludzi. Tylko szczepienia chronią nas przed ciężkim przebiegiem choroby i zgonami.
7	Choroba COVID-19 może być przyczyną groźnych powikłań. Nie testuj swojego zdrowia.
10	Wciąż za mało z nas się zaszczepiło, a ponad 99% zgonów na koronawirusa to osoby niezaszczepione.
11	Szczepienie przeciwko COVID-19 redukuje ryzyko ciężkiego przebiegu choroby oraz powikłań. Skutkami COVID-19 mogą być uszkodzenia nerek, uszkodzenia układu nerwowego, zakrzepy naczyń krwionośnych etc.
15	Pokonanie pandemii COVID-19 i powrót do normalności są możliwe, gdy przerwiemy łańcuch zakażeń. Aby efekty były jak najlepsze i jak najszybsze, wszyscy musimy być solidarni. Każdy z nas może zatrzymać pandemię. Przyjęcie szczepionki to nie tylko ochrona nas samych – to także ochrona naszych rodziców, dziadków, dzieci i przyjaciół.
26	Po zaszczepieniu ryzyko transmisji koronawirusa jest zmarginalizowane. W obecnej sytuacji pandemicznej kluczowa jest m.in. wysoka skuteczność ochrony przed ciężkimi objawami koronawirusa.
27	Z bardzo dużym prawdopodobieństwem – sięgającym nawet 95% – szczepienie uchroni cię przed zakażeniem COVID-19. Im więcej będzie zaszczepionych osób, tym szybciej osiągniemy odporność populacyjną. Szczepiąc się, zyskasz wewnętrzny spokój wynikający z bezpieczeństwa własnego i najbliższych. Szczepionki są bezpieczne. Szczepiąc się, przyczyniasz się do szybszego znoszenia ograniczeń i powrotu nas wszystkich do normalnego życia.
28	Do dzisiaj nie zostało wynalezione lekarstwo na COVID-19, a jedynym wyjściem z tej sytuacji jest osiągnięcie odporności populacyjnej oraz szybkie przerwanie transmisji wirusa. Możemy tego dokonać jedynie poprzez szczepienie jak największej liczby osób. Jak zapewniają specjaliści, korzyści ze szczepień wielokrotnie przewyższają ryzyko. Teraz możemy sprawić, że świat stanie się wolny od zgonów powodowanych koronawirusem SARS-CoV-2, a rzeczywistość powróci do normalności.
31	Szczepienia są bezpieczne i zmniejszają ryzyko hospitalizacji i zgonu z powodu COVID-19.
36	Szczepionki przeszły wszystkie wymagane etapy badań, podobnie jak pozostałe szczepionki dostępne na rynku. Kryteria oceny są bardziej szczegółowe niż te, którym podlegają leki. Szczepionki bez żadnych wyjątków podlegają restrykcyjnym procesom oceny wyników badań nieklinicznych (na zwierzętach) i wymaganym 3 etapom badań klinicznych (na ludziach). Jeżeli szczepionka została dopuszczona do obrotu, to znaczy, że spełniła wszystkie wymagania oceny jej jakości, bezpieczeństwa i skuteczności.

ID	Argument
39	W historii rozwoju medycyny szczepienia ochronne są jednym z największych osiągnięć. Szczepionki to odkrycie, które uratowało życie milionom ludzi. Historia pokazuje, że szczepionki przeciw COVID-19 mogą wywrzeć podobny wpływ na nasze życie.
42	To jest szczepionka tzw. zabita, która nie powoduje żadnych powikłań czy problemów u osób, które mają choroby przewlekłe współistniejące. To jedna z najbezpieczniejszych szczepionek.
46	Szczepienia przeciw COVID-19 uratują życie bardzo wielu ludzi. Szczepienie przeciw COVID-19 złagodzi przebieg ewentualnej infekcji. Szczepionki są skuteczną bronią przeciw chorobom zakaźnym.
49	Na przestrzeni lat szczepienia uratowały życie milionom ludzi na całym świecie. Pandemia COVID-19 przypomniała nam, jak ważna jest ochrona przed groźnymi chorobami zakaźnymi.
52	Szczepionki są ściśle monitorowane. Jeżeli cokolwiek by się pojawiło, że ten stosunek korzyści do ryzyka byłby niewłaściwy, że większe ryzyko niesie ze sobą szczepionka, Europejska Agencja Leków natychmiast wycofa taką szczepionkę.
54	Każdego roku na całym świecie szczepienia ratują 2–3 mln istnień ludzkich. Strategia szczepień w cyklu całego życia uznaje rolę szczepień jako strategię zapobiegania chorobom i maksymalizacji zdrowia przez całe życie. Zaszczepienie się jest obowiązkiem nas wszystkich i naszą odpowiedzialnością. Szczepienia są niezbędnym narzędziem zdrowia.
58	Aby osiągnąć odporność populacyjną, zaszczepić powinno się jak najwięcej osób. Szczepionki są najbardziej skutecznym rozwiązaniem ograniczającym ryzyko transmisji kolejnych wariantów wirusa SARS-CoV-2.
61	Żeby osiągnąć odporność populacyjną. Żeby wrócić do normalności. Żeby zacząć chodzić na imprezy. Żeby zacząć uprawiać sport.
66	Każdy z nas może powstrzymać pandemię. Przyjęcie szczepionki to nie tylko ochrona nas samych. To także ochrona naszych rodziców, dziadków, dzieci i przyjaciół.

Uwaga. Argumenty zostały losowo wybrane z ostatecznego korpusu 67 argumentów. ID odpowiadają numerom z oryginalnego korpusu.

Pozycje PAS-PL

Skala zawiera 9 pozycji testowych ocenianych na pięciopunktowej skali Likerta (od 1 = zdecydowanie się nie zgadzam, do 5 = zdecydowanie się zgadzam; dla pozycji 9: od 1 – bardzo słaby, do 5 – bardzo mocny):

1. Podany powód za {…} jest wiarygodny.
2. Podany powód za {…} jest przekonujący.
3. Podany powód za {…} jest dla mnie ważny.
4. Stwierdzenie dało mi pewność, że wiem, jak najlepiej {…}.
5. Stwierdzenie pomogłoby moim przyjaciołom w podjęciu decyzji o {…}.
6. Stwierdzenie wywołało u mnie myśl o tym, żeby {…}.
7. Stwierdzenie wywołało u mnie myśl o tym, żeby NIE {…}.
8. W jakim stopniu ogólnie zgadzasz się lub nie zgadzasz się z tym stwierdzeniem?
9. Czy podany powód za tym, żeby {…}, jest słaby czy mocny?