

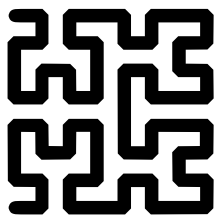


*Style Rozumowania  
Dedukcyjnego (SRD) –  
raport z konstrukcji narzędzia*

*Raport Badawczy*

number: 5(5)/2016; opublikowany: 2 grudnia 2016.

KATARZYNA PALUSZKIEWICZ



REASONING  
RESEARCH  
GROUP

Badanie jest częścią projektu *Modelowanie rozumowań abdukcyjnych* finansowanego ze środków Narodowego Centrum Nauki (nr 2013/10/E/HS1/00172).

## Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>3</b>
<b>2</b>	<b>Logiczna interpretacja <i>niektóre</i> w języku naturalnym</b>	<b>4</b>
2.1	Druga ankieta . . . . .	5
<b>3</b>	<b>Pseudosłowa</b>	<b>6</b>
<b>4</b>	<b>Konstrukcja pozycji testowych</b>	<b>6</b>
4.1	Forma wniosku . . . . .	8
4.2	Dwie wersje . . . . .	8
4.3	Ocena sędziów kompetentnych . . . . .	8
<b>5</b>	<b>Wybór i ostateczna forma pozycji testowych</b>	<b>9</b>
<b>6</b>	<b>Pilotaż</b>	<b>10</b>
6.1	Wersja A . . . . .	10
6.2	Wersja B . . . . .	10
6.3	Różnice pomiędzy grupami . . . . .	10
<b>7</b>	<b>Własności psychometryczne</b>	<b>11</b>
7.1	Rzetelność . . . . .	11
7.2	Wskaźniki trudności oraz wskaźniki dyskryminacji . . . . .	13
7.2.1	Wersja A . . . . .	13
7.2.2	Wersja B . . . . .	14
7.3	Różnica pomiędzy grupami . . . . .	14
<b>8</b>	<b>Analiza czynnikowa</b>	<b>14</b>
8.1	Wersja A . . . . .	14
8.2	Wersja B . . . . .	15
<b>9</b>	<b>Podsumowanie</b>	<b>16</b>

## Spis schematów

1	Przykładowa pozycja ankiety 1 . . . . .	4
2	Przykładowa pozycja ankiety 2 . . . . .	5
3	Schemat polisylogizmu . . . . .	7
4	Przykładowa pozycja testowa . . . . .	8
5	Wersja A: wykres osypiska . . . . .	12
6	Wersja B: wykres osypiska . . . . .	12

## Spis tabel

1	Przykładowa pozycja testowa – wynik ankiety 2 . . . . .	6
2	Sylaby użyte do konstrukcji pseudosłów . . . . .	7
3	Pozycje wybrane do wersji pilotażowej . . . . .	9
4	Wskaźniki trudności oraz dyskryminacji – wersja A . . . . .	14
5	Wskaźniki trudności oraz dyskryminacji – wersja B . . . . .	15
6	Pozycje testowe ładujące czynniki w wersji A . . . . .	15
7	Pozycje testowe ładujące czynniki w wersji B . . . . .	16

## 1 Wprowadzenie

*Style rozumowania dedukcyjnego* (SRD) jest testem, który został skonstruowany do pomiaru zdolności rozwiązywania zadań o charakterze dedukcyjnym. Konstrukcja pozycji testowych w teście SRD została oparta o polisylogizmy. Polisylogizmami nazywamy układ dwóch (lub więcej) sylogizmów, które prowadzą do pojedynczego wniosku. Układ ten jest zbudowany tak, że wniosek każdego poprzedniego sylogizmu staje się pierwszą przesłanką następnego (Bairan, 2005). Poprawność wniosku zależna jest od poprawności wszystkich sylogizmów składowych – jeżeli którykolwiek z nich jest zbudowany według niepoprawnego schematu, cały polisylogizm możemy uznać za niepoprawny.

Jednym z dwóch głównych celów konstrukcji testu było skonstruowanie narzędzia, które pozwoliłoby lepiej różnicować badanych o wysokim poziomie inteligencji płynnej (badanej Testem Matryc Ravena w wersji zaawansowanej). Przeprowadzone wcześniej badania (Urbański, Paluszkiwicz i Urbańska, 2014) z użyciem *Testu Umiejętności Rozumowania Dedukcyjnego* (TRD) pokazały, że rozwiązywanie prostych, dwuprzęsłankowych sylogizmów nie jest problematyczne dla badanych o wysokim poziomie inteligencji płynnej. Poprawność rozwiązania zadań w teście TRD sięgała w ich przypadku 90%.

Drugim celem konstrukcji narzędzia było sprawdzenie, czy poprawność odpowiedzi może być związana z formą językową zdań użytych w charakterze przesłanek i wniosków. Najbardziej problematyczna wydaje się interpretacja zdań szczegółowych zawierających wyrażenie *niektóre*. W klasycznej sylogistyce jest ono interpretowane jako *przynajmniej jeden, a może wszystkie*. Interpretacja ta nie jest zbieżna z interpretacją pragmatyczną wyrażenia *niektóre*, rozumianego jako *niektóre, ale nie wszystkie*. Trudno przewidzieć interpretację, jaką przyjmują badani w trakcie rozwiązywania zadań. Niepoprawne rozwiązania zadań, w których badani zinterpretowali wyrażenie *niektóre* pragmatycznie można mylnie zinterpretować jako niski poziom umiejętności rozumowania dedukcyjnego, podczas gdy „błąd” polegał na przyjęciu odmiennej od założonej przez badacza interpretacji językowej.

Dostępne w literaturze badania dotyczące związków pomiędzy interpretacją wyrażenia *niektóre* a poprawnością rozumowań rzadko wykorzystują klasyczne sylogizmy (por. Feeney, 2004; Schmidt i Thompson, 2008). Zadania dedukcyjne o bardziej skomplikowanym charakterze (np. zbudowane w oparciu o polisylogizmy), zwłaszcza o abstrakcyjnej treści, są w zasadzie nieobecne.

### Niektóre wróble są ptakami.

- [ ] Co najmniej jeden wróbel jest ptakiem.
- [ ] Przynajmniej jeden wróbel jest ptakiem.
- [ ] Na pewno jakiś wróbel jest ptakiem.
- [ ] Na pewno co najmniej jeden wróbel jest ptakiem.

Schemat 1: Przykładowa pozycja ankiety 1

## 2 Logiczna interpretacja *niektóre* w języku naturalnym

Pierwszym krokiem konstrukcji pozycji testowych było odnalezienie wyrażenia najlepiej odpowiadającego logicznej interpretacji *niektóre* jako *przynajmniej jeden, a może wszystkie*. W tym celu przeprowadzone dwie ankiety. Badani udzielali odpowiedzi drogą internetową.

### Pierwsza ankieta

Troje sędziów kompetentnych (rodzimych użytkowników języka polskiego) w wyniku dyskusji zaproponowało cztery wyrażenia mogące oddawać zakładaną, logiczną interpretację *niektóre*:

- co najmniej jeden
- na pewno co najmniej jeden
- na pewno jakiś
- przynajmniej jeden

Badani zostali poproszeni o uszeregowanie czterech stwierdzeń precyzujących wyrażenie *niektóre* w skali od 1 (najlepiej odpowiadające) do 4 (najgorzej odpowiadające). Przykładowe zadanie zaprezentowane zostało na Schemacie 1.

Ankieta składała się z sześciu pozycji. Trzy z nich miały formę zdań szczegółowotwierdzących (Niektóre *A* są *B*), pozostałe trzy zdań szczegółowoprzeczących (Niektóre *A* nie są *B*). Uwzględniona została także prawdziwość zdań użytych w badaniu. Wykorzystano dwa zdania prawdziwe wyłącznie w interpretacji logicznej (np. Niektóre ptaki są wróblami.), dwa zdania prawdziwe w obu interpretacjach (np. Niektóre kobiety są brunetkami.), oraz dwa fałszywe w obu interpretacjach (np. Niektóre jabłka nie są owocami.). Zastosowanie takiej procedury miało na celu zwrócenie uwagi badanych na językową formę zdania, a nie jego znaczenie. Wszystkie pozycje testowe zaprezentowane zostały w Załączniku 1.

W badaniu wzięło udział 39 badanych (w tym 30 kobiet): 18 osób bez edukacji logicznej oraz 21 po podstawowych kursach logiki, jednak nie obejmujących klasycznej sylogistyki. Średni wiek osób badanych wyniósł 22,92 ( $SD = 5,62$ ). Badani byli studentami, najwięcej z nich ( $N = 20$ ) studiowało filologię polską.

### Na pewno jakiś kusel jest fafką.

1. Niektóre kusły są fafkami.
2. Na pewno żaden kusel nie jest fafką.
3. Istnieją takie kusły, które są fafkami i istnieją takie kusły, które nie są fafkami.
4. Istnieją kusły, które są fafkami, a być może wszystkie kusły są fafkami.
5. Na pewno jakiś kusel jest fafką, a być może wszystkie kusły są fafkami .
6. Na pewno każdy kusel jest fafką.

#### Schemat 2: Przykładowa pozycja ankiety 2

Wyniki ankiety nie pozwoliły jednoznacznie wybrać odpowiedniego wyrażenia. Wskazywały jednak, że najodpowiedniejsze dla oddania znaczenia logicznej interpretacji *niektóre jest co najmniej jeden*. Różnice pomiędzy liczbą wyborów *co najmniej jeden* i *na pewno co najmniej jeden* były jednak niewielkie. Zdecydowano się przeprowadzić kolejną ankietę.

### 2.1 Druga ankieta

Pozycje w drugiej ankiecie składały się ze zdania-bazy oraz sześciu innych zdań. Zdania-bazy zawierały jedną z czterech fraz dookreślających logiczną interpretację *niektóre*. Przygotowanych zostało osiem takich zdań. Cztery z nich miały formę twierdzącą (np. Na pewno jakieś  $A$  jest  $B$ .), a cztery przeczącą (np. Co najmniej jedno  $A$  nie jest  $B$ .). Dla każdego z nich dobrano sześć innych zdań, które miały (lub nie) takie samo znaczenie. Zadaniem badanych było wybranie zdań, które ich zdaniem miały dokładnie takie samo znaczenie jak zdanie-baza.

Do konstrukcji wszystkich zdań użyte zostały pseudosłowa. Przykładowe zadania prezentuje Schemat 2.

Użycie słów języka polskiego mogłoby spowodować otrzymanie niemiarodajnych wyników. Na przykład, badani mogliby wybierać zdania prawdziwe zamiast zdań o dokładnie takim samym znaczeniu co zdanie-baza. Pseudosłowa miały także na celu skupienie uwagi badanych na formie językowej zdań i znaczeniu zaproponowanych przez nas wyrażań. W niektórych badaniach zamiast pseudosłów wykorzystuje się literowe zmienne. Wydaje się jednak, że ich użycie może sugerować badanym konieczność podstawienie konkretnej nazwy w miejsce litery (potraktowanie takiej zmiennej jako „niewiadomej”, którą trzeba zastąpić). Zastosowanie takiej strategii może wpłynąć negatywnie na wyniki. Pseudosłowa wydają się najodpowiedniejszym rozwiązaniem – podkreślają strukturę językową zdań, jednocześnie zaznaczając, że mamy do czynienia ze zbiorami jakiś obiektów o pewnej nazwie, a także pozwalają uniknąć różnego rodzaju odchyłeń wyników.

Badani mogli zaznaczyć dowolną liczbę zdań, które ich zdaniem miało to samo znaczenie, co zdanie-baza. Wszystkie pozycje ankiety zaprezentowane zostały w Załączniku 2.

W badaniu wzięło udział 37 osób (w tym 23 kobiety): 27 osób bez edukacji logicznej

<b>Na pewno jakiś kuseł jest fafką.</b>	
Niektóre kusły są fafkami	24,36%
Na pewno żaden kuseł nie jest fafką	1,28%
Istnieją takie kusły, które są fafkami i istnieją takie kusły, które nie są fafkami	28,21%
Istnieją kusły, które są fafkami, a być może wszystkie kusły są fafkami	23,08%
Na pewno jakiś kuseł jest fafką, a być może wszystkie kusły są fafkami	23,08%
Na pewno każdy kuseł jest fafką	0%

Tabela 1: Wyniki zaznaczeń dla pozycji zaprezentowanej na Rysunku 2

oraz 10 po podstawowych lub zaawansowanych kursach logiki. Średni wiek osób badanych wyniósł 22,65 ( $SD = 2,28$ ). Badani byli studentami różnych kierunków; najwięcej z nich studiowało filologię niderlandzką ( $N = 3$ ).

Aby wybrać wyrażenie oddające logiczną interpretację *niektóre* porównane zostały zdania, wyrażające logiczną oraz pragmatyczną interpretację. W Przykładzie 2 logiczna interpretacja jest wyrażana na przykład w zdaniu *Istnieją kusły, które są fafkami, a być może wszystkie kusły są fafkami*. Pragmatyczna natomiast poprzez *Istnieją takie kusły, które są fafkami i istnieją takie kusły, które nie są fafkami*. Dla każdej pozycji ankiety zliczona została suma zaznaczeń zdań 1–6 dokonanych przez badanych. Następnie obliczono jaki procent z całości zaznaczeń stanowi zaznaczenie każdego ze zdań 1–6 z osobna. Przykładowa tabela wyników dla pozycji ze Schematu 2 została zaprezentowana w Tabeli 1.

W przypadku pozycji przedstawionej w tabeli, badani najczęściej wybierali zdanie *Istnieją takie kusły, które są fafkami i istnieją takie kusły, które nie są fafkami*. jako przekazujące tę samą informację, co zdanie *Na pewno jakiś kuseł jest fafką*. Najczęściej wybieranym sformułowaniem wyrażającym logiczną interpretację *niektóre* było wyrażenie *na pewno co najmniej*.

### 3 Pseudosłowa

Pseudosłowa to ciągi liter języka, którego struktura ortograficzna i fonologiczna przypomina słowo danego języka. W przeciwieństwie do słów, nie mają one jednak znaczenia. Konstruując pseudosłowa na potrzeby testu SRD najpierw przygotowano tabelę zawierającą wszystkie możliwe połączenia o postaci spółgłoska+samogłoska, zachowując reguły fonotaktyczne języka polskiego (por. Tabela 2). Następnie łączono maksymalnie trzy pary z tabeli, niekiedy uzupełniając je spółgłoską na końcu pseudosłowa lub pomiędzy sylabami.

### 4 Konstrukcja pozycji testowych

Wszystkie pozycje testowe mają formę polisylogizmów o trzech przesłankach i jednym wniosku. Punktem wyjścia do konstrukcji poszczególnych polisylogizmów było 35 par klasycznych, dwupresłankowych schematów sylogizmów. Wśród nich znalazło się 16 par, w których oba schematy były poprawne oraz 19 par, w których pierwszy schemat był

	<b>a</b>	<b>e</b>	<b>i</b>	<b>o</b>	<b>u</b>	<b>y</b>
<b>b</b>	ba	be	bi	bo	bu	by
<b>c</b>	ca	ce	ci	co	cu	cy
<b>d</b>	da	de	di	do	du	dy
<b>f</b>	fa	fe	fi	fo	fu	fy
<b>g</b>	ga	ge	gi	go	gu	–
<b>h</b>	ha	he	hi	ho	hu	–
<b>j</b>	ja	je	–	jo	ju	–
<b>k</b>	ka	ke	ki	ko	ku	ky
<b>l</b>	la	le	li	lo	lu	ly
<b>ł</b>	ła	łe	li	ło	łu	ły
<b>m</b>	ma	me	mi	mo	mu	my
<b>n</b>	na	ne	ni	no	nu	ny
<b>p</b>	pa	pe	pi	po	pu	py
<b>r</b>	ra	re	ri	ro	ru	ry
<b>s</b>	sa	se	si	so	su	sy
<b>t</b>	ta	te	ti	to	tu	ty
<b>w</b>	wa	we	wi	wo	wu	wy
<b>z</b>	za	ze	zi	zo	zu	zy

Tabela 2: Sylaby użyte do konstrukcji pseudosłów

poprawny, a drugi niepoprawny. Pary dobierano tak, aby wniosek pierwszego schematu sylogizmu stanowił jednocześnie pierwszą przesłankę drugiego. Na pojedynczy polisylogizm składają się dwie przesłanki pierwszego sylogizmu, druga przesłanka drugiego sylogizmu oraz wniosek drugiego sylogizmu. Konstrukcję przykładowego polisylogizmu ilustruje Schemat 3.

W puli 35 polisylogizmów znalazło się po 8 pozycji o wniosku *SiP*, *SoP*, 9 o wniosku *SaP* oraz 10 o wniosku *SeP*. W tak skonstruowanych schematach polisylogizmów występowały cztery zmienne nazwowe (*A*, *B*, *C*, *D*). W ostatnim kroku konstrukcji zmienne nazwowe zastępowano pseudosłowami. Wszystkie polisylogizmy zostały zaprezentowane w Załączniku 3.

### Schematy sylogizmów    Schemat polisylogizmu

<i>AaB</i>	
<u><i>CaA</i></u>	<i>AaB</i>
<i>CaB</i>	<i>CaA</i>
	<u><i>DiC</i></u>
<i>CaB</i>	<i>DiB</i>
<u><i>DiC</i></u>	
<i>DiB</i>	

Schemat 3: Schemat poprawnego polisylogizmu oparty na schematach dwóch poprawnych sylogizmów

Każdy matuk jest bama  
Każda tida jest matukiem  
Każdy rutek jest tida  
Czy każdy rutek jest bama?

Na pewno tak    Na pewno nie    Nie można tego stwierdzić

Schemat 4: Przykładowa pozycja testowa

#### 4.1 Forma wniosku

Wnioski polisylogizmów zostały sformułowane w formie pytań rozstrzygnięcia (np. *Czy każdy topek jest glarą?*). Zadaniem badanych było wybranie jednej z trzech odpowiedzi: *na pewno tak*, *na pewno nie*, *nie można tego stwierdzić*. Przykładowe zadanie prezentuje Schemat 4.

#### 4.2 Dwie wersje

Jak wspomniano wcześniej, test SRD miał pozwolić na sprawdzenie związku pomiędzy językową formą wyrażenia *niektóre* a poprawnością rozwiązań zadań o charakterze dedukcyjnym. W tym celu sformułowano każdy z 35 sylogizmów w dwóch wersjach. W wersji A zdania szczegółowoprzeczące i szczegółowotwierdzące miały formę klasyczną: *Niektóre A są B* oraz *Niektóre A nie są B*. W wersji B użyto wybranego wcześniej wyrażenia, oddającego logiczną interpretację – *na pewno co najmniej jeden*. W drugiej wersji zdania o schemacie *SiP* miały formę *Na pewno co najmniej jedno A jest B*, a zdania o schemacie *SoP* — *Na pewno co najmniej jedno A nie jest B*.

Schematy polisylogizmów użyte do konstrukcji pozycji w obu wersjach były identyczne, różniła je jedynie forma językowa zdań szczegółowych.

#### 4.3 Ocena sędziów kompetentnych

Wszystkie 35 sylogizmów zostało rozwiązanych przez dwie grupy sędziów: 10 bez treningu logicznego oraz 2 po intensywnym treningu w zakresie logiki formalnej. Dodatkowo 2 sędziów po intensywnym treningu w zakresie logiki formalnej dokonało oceny pozycji pod względem formalnym i językowym. Połowa sędziów otrzymała do rozwiązania wersję A, połowa — wersję B. Kolejność sylogizmów w arkuszu dla każdego sędziego była inna, aby zapobiec np. gorszemu rozwiązywaniu końcowych pozycji ze względu na zmęczenie badanych.

Po przeanalizowaniu uwag dotyczących formy pozycji testowych wprowadzono dwie zmiany. Po pierwsze poprawiono pozycje tak, by pseudosłowa w obrębie jednego polisylogizmu rozpoczynały się różnymi literami. Po drugie, zmieniono formę wniosku z pytań na zdania twierdzące. Sędziowie zwracali uwagę, że odpowiedź na pytania mające formę *Czy na pewno co najmniej jedno A jest B?* jest problematyczna ze względu na obecność wyrażenia *na pewno* w możliwych odpowiedziach.



Lp.	(35)
1.	14
2.	25
3.	11
4.	22
5.	35
6.	33
7.	19
8.	5
9.	27
10.	1

Tabela 3: Pozycje wybrane do wersji pilotażowej i ich numery z pełnej puli 35 sylogizmów

## 5 Wybór i ostateczna forma pozycji testowych

Wybierając 10 pozycji testowych z 35 zaproponowanych wzięta została pod uwagę poprawność rozwiązań poszczególnych zadań przez 10 sędziów bez treningu logicznego. W obu wersjach testu łatwość pozycji wahała się od 0% do 100%. Oznacza to, że część pozycji była rozwiązywana przez wszystkich sędziów, a część nie została poprawnie rozwiązana przez żadnego.

Na podstawie analizy poprawności rozwiązań wybrano 10 pozycji testowych o następujących cechach:

1. dwie pozycje łatwe – poprawność rozwiązania w obu wersjach ok. 90%;
2. dwie pozycje o tej samej poprawności rozwiązania w obu wersjach ;
3. dwie pozycje, w których poprawność była większa w wersji A (*niektóre*);
4. dwie pozycje, w których poprawność była większa w wersji B (*na pewno co najmniej jeden*);
5. dwie dodatkowe pozycje zbudowane na podstawie niepoprawnego polisylogizmu.

Ze względu na sposób konstrukcji pozycji testowych, wszystkie wybrane pozycje testowe miały tylko dwie możliwe formy odpowiedzi: w przypadku poprawnych polisylogizmów poprawną odpowiedzią było *na pewno tak*, natomiast w przypadku niepoprawnych – *nie można tego stwierdzić*. Aby zróżnicować możliwe odpowiedzi w dwóch poprawnych polisylogizmach zmieniono zdanie-wniosek na zdanie sprzeczne (zgodnie z relacjami zachodzącymi pomiędzy zdaniami kategorycznymi według kwadratu logicznego). Dzięki temu w przypadku tych pozycji poprawną odpowiedzią było *na pewno nie*. Wszystkie pozycje testowe włączone do pilotażowej wersji testu znajdują się w Załączniku 4. Tabela 3 zawiera informacje o numerach, jakie pozycje wybrane do wersji pilotażowej posiadały w pełnej puli 35 sylogizmów (pula dostępna w Załączniku 3).

W pozycjach numer 8 oraz 9 dokonano korekty w zakresie użytych pseudosłów tak, aby każde z nich rozpoczynało się na inną literę. W pozycjach 8 oraz 10 zamieniono wnioski na sprzeczne z poprawnymi. W pozycji 8 nastąpiła zmiana wniosku ze zdania o schemacie *SeP* na zdanie o schemacie *SiP*, natomiast w 10 – ze zdania o schemacie *SiP* na zdanie o schemacie *SeP*.

## 6 Pilotaż

### 6.1 Wersja A

W badaniu pilotażowym wzięło udział 93 osoby, w tym 31 kobiet. Osoby badane były studentami, najwięcej osób studiowało inżynierię bezpieczeństwa ( $N = 18$ ). Badani nie otrzymali edukacji logicznej w ogóle lub w zakresie nie większym niż jeden semestr. Średni wiek osób badanych wyniósł 21,06 ( $SD = 3,68$ ).

Średnia suma punktów uzyskiwanych przez badanych to 5,16 ( $SD = 2,45$ ). Test Kołmogorowa-Smirnowa pozwala uznać, że rozkład wyników ogólnych uzyskanych w teście nie jest zgodny z rozkładem normalnym ( $Z = 0,139$ ;  $p < 0,000$ ).

### 6.2 Wersja B

W badaniu pilotażowym wzięło udział 36 osób, w tym 24 kobiety. Badani byli studentami: najwięcej osób studiowało inżynierię bezpieczeństwa ( $N = 16$ ) oraz kognitywistykę ( $N = 14$ ). Jedna osoba nie zadeklarowała kierunku studiów. Studenci kognitywistyki przeszli intensywny kurs w zakresie logiki formalnej, pozostali uczestnicy badania nie otrzymali edukacji logicznej w ogóle lub w podstawowym zakresie. Średni wiek osób badanych wyniósł 20,78 ( $SD = 1,73$ ).

Średnia suma punktów uzyskiwanych przez badanych to 5,89 ( $SD = 2,42$ ). Wynik testu Shapiro-Wilka pozwala uznać, że rozkład wyników ogólnych uzyskanych w teście jest zgodny z rozkładem normalnym ( $W = 0,960$ ;  $p = 0,209$ ).

### 6.3 Różnice pomiędzy grupami

Jak wspomniano we wstępie, jednym z celów konstrukcji narzędzia było sprawdzenie, czy językowa forma zdań szczegółowych może mieć związek z poprawnością rozwiązań w teście SRD. Aby przetestować tę hipotezę, sprawdzono czy występują różnice pomiędzy grupami. Zastosowano test U Manna-Whitneya, ponieważ porównywane grupy nie były równoliczne, a rozkład zmiennej w jednej z grup niezgodny z rozkładem normalnym.

Analiza nie wykazała istotnych różnic pomiędzy grupami ( $U = 1381,0$ ;  $p = 0,121$ ). Pozwala to sądzić, że forma językowa zdań szczegółowych nie ma związku z poziomem poprawności rozwiązywania zadań w teście SRD.

## 7 Własności psychometryczne

Zgodnie z założeniem, wszystkie pozycje testowe miały mierzyć ten sam konstrukt – zdolność do rozwiązywania zadań dedukcyjnych. Wyniki badania pilotażowego pokazały, że trudność zadań jest zróżnicowana (por. Tabela 5a). W trakcie konstrukcji narzędzia nie zakładano jego wielowymiarowości, jednak pozycje testowe w obu wersjach mają różną moc dyskryminacyjną (por. Tabela 5b). Pozwala to zakładać, że inne wskaźniki rzetelności (tj. alfa Cronbacha czy lambda 2 Guttmana) nie są odpowiednimi miarami rzetelności testu; ich zastosowanie mogłoby prowadzić do niedoszacowania rzetelności (Kamata, Turhan i Darandari). Odpowiednią miarą rzetelności dla narzędzi, w których nie został spełniony tzw. *tau-equivalent condition* (pozycje nie mają podobnej mocy dyskryminacyjnej) jest omega McDonalda (*McDonald's omega estimates of general and total factor saturation*). Pomimo, że obie wersje testu są jednowymiarowe, to do wykonania obliczeń współczynnika rzetelności omega należało wyróżnić odpowiednią ilość czynników.

Do określenia liczby czynników oraz sprawdzenia założeń została wykonana analiza głównych składowych (PCA; *principal component analysis*). Miara Kaiser-Meyer-Olkin adekwatności doboru próby dla obu wersji testu (wersja A:  $KMO = 0,695$ , wersja B:  $KMO = 0,659$ ) uzyskała wartości pozwalające na wyróżnianie czynników w analizowanych danych. Kaiser (1974) zaleca uznawanie wartości większych od 0,5 za akceptowalne. Uzyskane wyniki można zinterpretować jako średnie. Test sferyczności Bartletta dla obu wersji był istotny (wersja A:  $p < 0,000$ , wersja B:  $p = 0,002$ ) – istnieją więc istotne korelacje pomiędzy zmiennymi i warto przeprowadzić analizę czynnikową. W przypadku obu wersji testu, wykres osypiska jest najbardziej stromy pomiędzy pierwszym a drugim czynnikiem. Kryterium to sugeruje rozwiązanie jednoczynnikowe. Natomiast zastosowanie kryterium Kaisera wskazuje, że najodpowiedniejszą liczbą czynników dla analizowanych danych są trzy lub cztery czynniki (por. Rysunek 5 i 6).

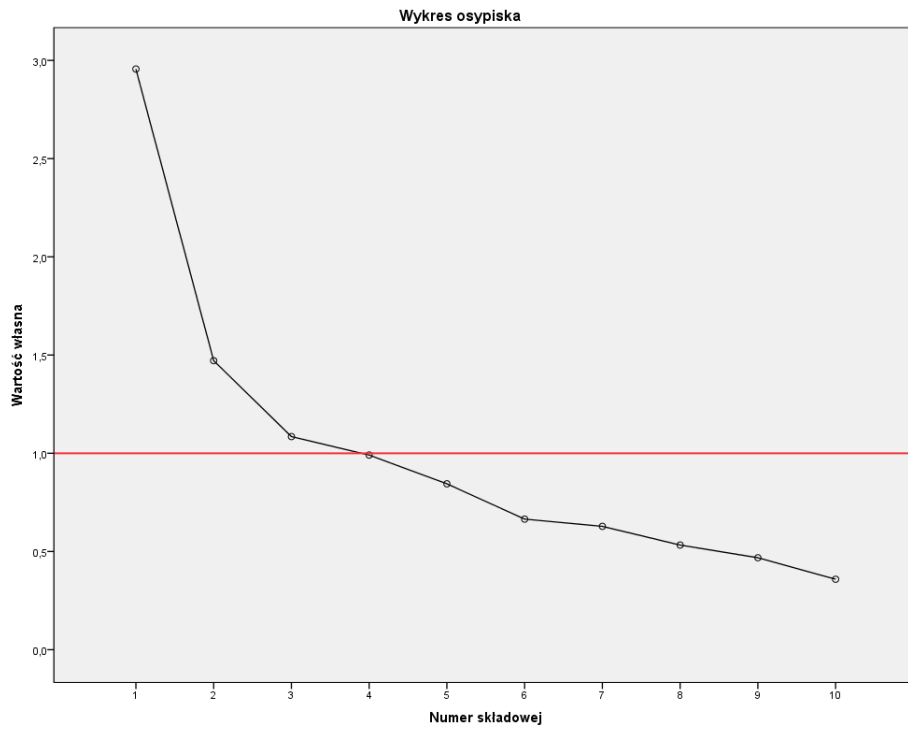
Interpretacja miary  $KMO$  wskazuje, że wyróżnianie czynników nie jest niezbędne w przypadku braku podstawy teoretycznej do ich wyróżniania. Brak jednoznacznego rozstrzygnięcia w odniesieniu do odpowiedniej liczby czynników wyróżnionych danych również sugeruje, że możemy uznać rozwiązanie jednoczynnikowe za odpowiednie. Jednak w przypadku wybranej miary rzetelności niezbędne jest wyróżnienie czynników; w takiej sytuacji, stosując kryterium Kaisera, wyróżnić można trzy czynniki i taka wartość została wykorzystana do obliczenia omegi McDonalda.

### 7.1 Rzetelność

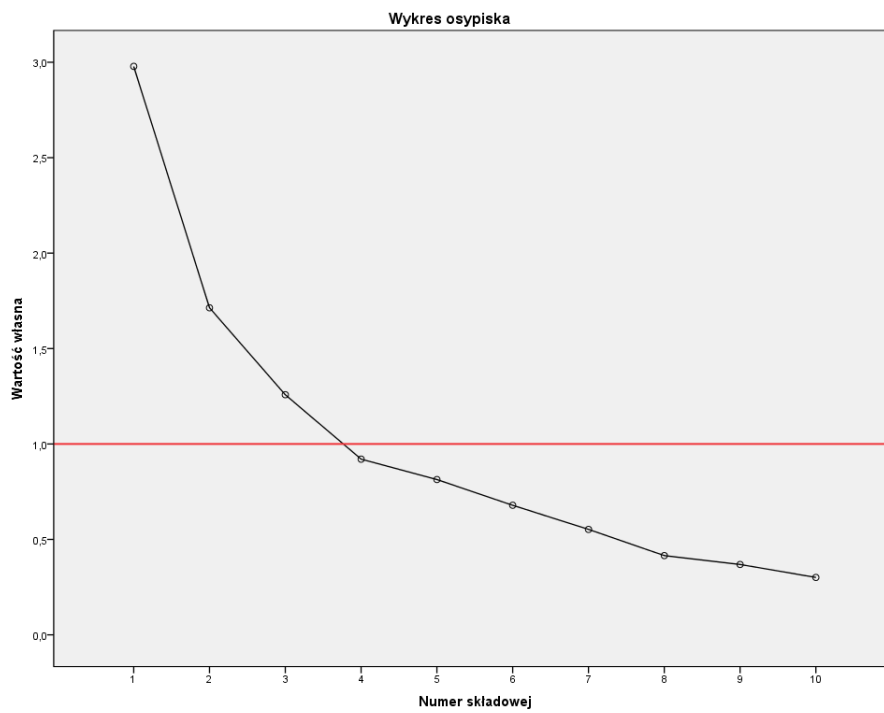
Do wykonania obliczeń współczynnika rzetelności wykorzystano opcję *omega* z pakietu *psych*, dostępnego w środowisku R. Obliczenia dla obu wersji testu wykonano dla 3 czynników, analizą czynnikową metodą reszt minimalnych (*minimum residual (OLS) solution*) oraz rotacją oblimin<sup>1</sup>.

---

<sup>1</sup>wybrane opcje zapisano przy pomocy następującej komendy: `omega(nfactors = 3, fm = "minres", n.iter = 1, p = 0.05, rotate = "oblimin")`



Schemat 5: Wersja A: wykres osypiska – czerwoną linią zaznaczono kryterium Kaisera



Schemat 6: Wersja B: wykres osypiska – czerwoną linią zaznaczono kryterium Kaisera

Współczynnik omega dla wersji A wyniósł 0,78.

Współczynnik omega dla wersji B wyniósł 0,81.

## 7.2 Wskaźniki trudności oraz wskaźniki dyskryminacji

Wskaźnik trudności (nazywany niekiedy wskaźnikiem łatwości) to wyrażona w procentach proporcja osób, które odpowiedziały poprawnie na daną pozycję testową (Hornowska, 2007). Był obliczany zgodnie z następującym równaniem:

$$T = \frac{n_i}{N} = p_i \times 100\%$$

gdzie:  $T$  oznacza wskaźnik trudności;  $n_i$  liczbę osób, które odpowiedziały poprawnie na daną pozycję;  $N$  ogólną liczbę osób, które odpowiedziały na daną pozycję;  $p_i$  natomiast proporcję odpowiedzi poprawnych do wszystkich odpowiedzi. Im wyższy wskaźnik trudności, tym więcej osób odpowiedziało prawidłową na daną pozycję. Ponieważ pozycje w teście SRD są ze sobą skorelowane zaleca się, aby średnia trudność całego testu oscylowała w wokół 50%.

Do oceny mocy dyskryminacyjnej pozycji wybrano wskaźnik dyskryminacji, który opiera się porównaniu dolnej i górnej grupy badanych. Grupy zostały wyróżnione poprzez wybranie 27% badanych z najwyższymi oraz najniższymi wynikami, gdyż taka procedura jest uważana za najoptymalniejszą (Kelly, 1939; za: Hornowska, 2007, s. 177). Wskaźnik dyskryminacji obliczano zgodnie ze wzorem:

$$D = p_u - p_l$$

gdzie:  $D$  oznacza wskaźnik dyskryminacji;  $p_u$  proporcję odpowiedzi poprawnych w górnej grupie badanych;  $p_l$  natomiast proporcję odpowiedzi poprawnych w dolnej grupie badanych.

### 7.2.1 Wersja A

Wskaźnik średniej trudności dla wersji A wynosi 51,61%. Oznacza to, że test dość dobrze różnicuje badanych na całym kontinuum zmienności cechy (zdolności do rozwiązywania zadań dedukcyjnych). Trudność poszczególnych pozycji waha się od 22,58% do 83,87%, więc test charakteryzuje się dużym zróżnicowaniem w zakresie trudności pozycji. Wskaźniki dla poszczególnych pozycji zostały zaprezentowane w Tabeli 4a.

Do obliczenia wskaźnika dyskryminacji wybrano 25 osób z najwyższymi wynikami oraz 25 osób z najniższymi wynikami, następnie obliczono odpowiednie proporcje. Wszystkie pozycje w tej wersji testu charakteryzują się wskaźnikiem dyskryminacji równym lub większym od 0,4 (pozwalającym włączyć daną pozycję do testu bez poddawania jej modyfikacjom) (Hornowska, 2007). Wskaźniki dyskryminacji dla wszystkich pozycji zostały zaprezentowane w Tabeli 4b.

Lp.	Wskaźnik trudności	Lp.	Wskaźnik dyskryminacji
1.	83,87%	1.	0,44
2.	47,31%	2.	0,8
3.	36,56%	3.	0,4
4.	22,58%	4.	0,56
5.	76,34%	5.	0,64
6.	35,48%	6.	0,76
7.	30,11%	7.	0,56
8.	66,67%	8.	0,6
9.	43,01%	9.	0,76
10.	74,19%	10.	0,4
średnia	51,61%		

(a) Wskaźnik trudności

(b) Wskaźnik dyskryminacji

Tabela 4: Wskaźnik trudności oraz wskaźnik dyskryminacji dla poszczególnych pozycji w wersji A

### 7.2.2 Wersja B

Wskaźnik średniej trudności dla tej wersji testu wynosi 58,89%. Oznacza to, że test dość dobrze różnicuje badanych na całym kontinuum zmienności cechy (zdolności do rozwiązywania zadań dedukcyjnych). Trudność poszczególnych pozycji waha się od 27,78% do 86,11%, więc test charakteryzuje się dużym zróżnicowaniem w zakresie trudności pozycji. Wskaźniki dla poszczególnych pozycji zostały zaprezentowane w Tabeli 5a.

Do obliczenia wskaźnika dyskryminacji wybrano 10 osób z najwyższymi wynikami oraz 10 osób z najniższymi wynikami, następnie obliczono odpowiednie proporcje. Hornowska (2007) wskazuje, że pozycje testowe można włączyć do testu jeżeli wskaźnik mocy dyskryminacyjnej jest większy bądź równy 0,4. Trzy pozycje w teście osiągnęły moc dyskryminacyjną na poziomie 0,3, co wskazuje, że należy poddać je niewielkim modyfikacjom. Pomimo tego postanowiono nie wprowadzać zmian w tych pozycjach, aby zachować spójność formy zadań oraz ich jednorodny charakter. Wskaźniki dla wszystkich pozycji zostały zaprezentowane w Tabeli 5b.

### 7.3 Różnica pomiędzy grupami

## 8 Analiza czynnikowa

Przeprowadzono eksploracyjną analizę czynnikową metodą głównych składowych (PCA; *principal component analysis*) z rotacją Varimax.

### 8.1 Wersja A

Czynnik 1 jest ładowany przez cztery z pięciu pozycji, w których poprawną odpowiedzią jest *nie można tego stwierdzić*. W czynniku 2 znajduje się jeden sylogizm poprawny (1)

Lp.	Wskaźnik trudności	Lp.	Wskaźnik dyskryminacji
1.	80,56%	1.	0,3
2.	52,78%	2.	1
3.	55,56%	3.	0,7
4.	27,78%	4.	0,3
5.	77,78%	5.	0,6
6.	33,33%	6.	0,6
7.	41,67%	7.	0,7
8.	75,00%	8.	0,3
9.	58,33%	9.	0,9
10.	86,11%	10.	0,4
średnia	58,89%		

(a) Wskaźnik trudności

(b) Wskaźnik dyskryminacji

Tabela 5: Wskaźnik trudności oraz wskaźnik dyskryminacji dla poszczególnych pozycji w wersji B

Czynnik 1		Czynnik 2		Czynnik 3	
4	(0,591)	1	(0,791)	2	(0,625)
6	(0,647)	5	(0,807)	3	(0,687)
7	(0,781)	10	(0,472)	4	(0,439)
9	(0,609)			10	(-0,480)

Tabela 6: Pozycje testowe ładujące poszczególne czynniki w wersji A. W nawiasach podano ładunki czynnikowe.

oraz dwa sylogizmy niepoprawne (5,10), w których prawidłową odpowiedzią jest *na pewno nie*. W przypadku czynnika 3 trudno zidentyfikować wspólną cechę dla ładujących go pozycji – pozycja 2 to sylogizm niepoprawny, natomiast 3 – poprawny. Wysoko ten czynnik ładuje także pozycja 4. Warto zauważyć, że pozycja numer 8 nie ładuje żadnego czynnika powyżej satysfakcjonującego poziomu (0,4).

## 8.2 Wersja B

Czynnik pierwszy jest ładowany przez pozycje, w których poprawną odpowiedzią jest *nie można tego stwierdzić*, przy czym należy zauważyć, że ładunek czynnikowy pozycji 6 dla czynnika pierwszego był niższy niż zalecane 0,4 (wyniósł 0,393). Ze względów merytorycznych został jednak włączony jako ładujący ten czynnik. Czynnik 2 jest ładowany przez dwa poprawne polisylogizmy (1,3) oraz dwa niepoprawne (5, 10). W przypadku dwóch niepoprawnych sylogizmów ładujących czynnik 2 poprawną odpowiedzią jest *na pewno nie*. Są to pozycje, które powstały w wyniku zamienienia zdania-wniosku na sprzeczne (por. Wybór i ostateczna forma pozycji testowych), w poprawnym sylogizmie. Możemy więc uznać, że są one bardziej podobne do sylogizmów poprawnych niż do sylogizmów niepoprawnych, których prawidłowa odpowiedź to *nie można tego stwierdzić*. Czynnik 3

Czynnik 1		Czynnik 2		Czynnik 3	
2	(0,706)	1	(0,773)	8	(0,768)
4	(0,677)	3	(0,700)	6	(-0,781)
6	(0,393)	5	(0,807)		
7	(0,629)	10	(0,472)		
9	(0,707)	4	(-0,440)		

Tabela 7: Pozycje testowe ładujące poszczególne czynniki w wersji B. W nawiasach podano ładunki czynnikowe.

jest ładowany tylko przez pozycję 8 oraz ujemnie przez pozycję 8. Trudno zinterpretować ten czynnik merytorycznie, jedyną cechą odróżniającą pozycję 8 od pozostałych jest fakt, że wszystkie przesłanki są zdaniami ogólnymi, a wniosek zdaniem szczegółowym.

Brak jasnych kryteriów merytorycznych oraz fakt, że jeden z czynników jest ładowany tylko przez jedną pozycję sugeruje, że założenie o jednoczynnikowej strukturze testu jest prawidłowe.

## 9 Podsumowanie

Obie wersje testu okazały się rzetelnymi narzędziami, o satysfakcjonujących wskaźnikach trudności. Pomimo, że niektóre pozycje w wersji B wykazywały nieco zbyt niską moc dyskryminacyjną zdecydowano nie wносить poprawek, aby zachować spójność formy i treści pomiędzy obiema wersjami. Uzyskany współczynnik trudności wydaje się odpowiedni dla lepszego różnicowania badanych o wyższym poziomie inteligencji.

Brak różnic w poziomie wykonania obu wersji może świadczyć o braku związku pomiędzy formą językową kwantyfikatora szczegółowego a poprawnością rozwiązań. Możliwe, że pomimo interpretowania kwantyfikatora szczegółowego na różne sposoby, także inne czynniki wpływały na poziom rozwiązań, niwelując efekt wpływu formy językowej. Aby przetestować założenie o związku formy językowej kwantyfikatora a poziomem poprawności rozwiązań należy skonstruować test, który w bardziej widoczny sposób będzie można wyróżnić wpływ wyboru interpretacji (pragmatycznej lub logicznej) na poprawność rozwiązywania zadań. Na przykład konstruując takie pozycje testowe, w których wybierana interpretacja będzie znajdowała odzwierciedlenie w wybieranej odpowiedzi.



## Bibliografia

- Bairan, B. P. (2005). *An Introduction to Syllogistic Logic*. Goodwill Trading.
- Feeney, A., Scrafton, S., Duckworth, A., Handley, S. J. (2004). The story of some: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(2), 121.
- Hornowska, E. (2007). *Testy psychologiczne: teoria i praktyka..* Warszawa: Wydawnictwo naukowe „Scholar”.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- Kamata, A., Turhan, A., Darandari, E. (2003). Estimating reliability for multidimensional composite scale scores. In *Annual meeting of American Educational Research Association*, Chicago, IL.
- Schmidt, J. R., Thompson, V. A. (2008). "At least one" problem with "some" formal reasoning paradigms. *Memory & cognition*, 36(1), 217–229.
- Urbański, M., Paluszkiewicz, K., Urbańska. (2014). *Deductive Reasoning and Learning: a Cross-Curricular Study*. Raport badawczy, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu.