

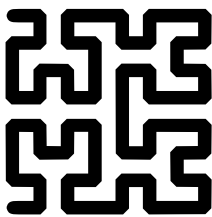


*Testy Różnych Interpretacji Kwantyfikatorów
Ogólnego i Egzystencjalnego
(RIKO i RIKE) –
raport z konstrukcji narzędzi*

Raport Badawczy

numer: 7(7)/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

Test Różnych Interpretacji Kwantyfikatora Ogólnego (RIKO)	4
1 Konstrukcja pozycji testowych	5
1.1 Dwie wersje testu	6
1.2 Konstrukcja pozycji właściwych	6
1.3 Konstrukcja pozycji wypełniających	7
2 Pilotaż	7
2.1 Wersja A	7
2.2 Wersja E	7
2.3 Różnice pomiędzy grupami	8
3 Własności psychometryczne	8
3.1 Wskaźnik trudności	8
3.1.1 Wersja A	8
3.1.2 Wersja E	8
3.2 Rzetelność	8
3.2.1 Wersja A	9
3.2.2 Wersja E	9
4 Analiza niepoprawnych odpowiedzi	10
4.1 Wersja A	10
4.2 Wersja E	10
5 Podsumowanie	11
Test Różnych Interpretacji Kwantyfikatora Egzystencjalnego (RIKE)	12
6 Konstrukcja pozycji testowych	12
6.1 Dwie wersje testu	13
6.2 Konstrukcja pozycji właściwych	13
6.3 Konstrukcja pozycji wypełniających	14
7 Pilotaż	14
7.1 Wersja S	14
7.2 Wersja C	15
7.3 Różnice pomiędzy grupami	15
8 Własności psychometryczne	15
8.1 Wskaźnik trudności	15
8.1.1 Wersja S	15
8.1.2 Wersja C	15

8.2	Rzetelność	16
8.2.1	Wersja S	16
8.2.2	Wersja C	17
9	Analiza odpowiedzi niezgodnych z kluczem odpowiedzi	17
10	Podsumowanie	17

Spis przykładów

1	Niepoprawne rozwiązania zadań w teście SRD	5
2	RIKO – pozycje właściwe	6
3	RIKE – pozycje właściwe	14

Spis tabel

1	RIKO – Schematy pozycji właściwych	6
2	RIKO – Schematy pozycji wypełniających	7
3	Wskaźnik trudności dla pozycji wypełniających – wersja A	8
4	Trudność pozycji wypełniających – wersja E	9
5	Rzetelność po usunięciu poszczególnych pozycji – wersja A	9
6	Rzetelność po usunięciu poszczególnych pozycji – wersja E	10
7	Zestawienie odpowiedzi niepoprawnych – wersja A	10
8	Zestawienie odpowiedzi niepoprawnych – wersja E	10
9	RIKE – Schematy pozycji właściwych	13
10	RIKO – Schematy pozycji wypełniających	14
11	Wskaźnik trudności dla pozycji wypełniających – wersja S	15
12	Wskaźnik trudności dla pozycji wypełniających – wersja C	16
13	Rzetelność po usunięciu poszczególnych pozycji – wersja S	16
14	Rzetelność po usunięciu poszczególnych pozycji – wersja C	17
15	Zestawienie odpowiedzi – wersja S	18

Test Różnych Interpretacji Kwantyfikatora Ogólnego (RIKO)

Test Różnych Interpretacji Kwantyfikatora Ogólnego (RIKO) powstał w celu sprawdzenia związku pomiędzy formą językową, w jakiej wyrażony jest kwantyfikator ogólny a poprawnością zadań zbudowanych w oparciu o proste sylogizmy. Analiza niepoprawnych rozwiązań w teście *Stylów Rozumowania Dedukcyjnego* (Paluszkiewicz, 2016) dokonywanych przez badanych pokazała, że jednym z najczęściej popełnianych błędów było interpretowanie zdań ogólnotwierdzących (Każde S jest P) jako wyrażających identyczność zakresów nazw S i P . Zdania ogólnotwierdzące są natomiast prawdziwe w przypadku inkluzji zbioru S i P , a więc w dwóch przypadkach: kiedy zakresy nazw S i P są identyczne oraz gdy S jest podzbiorem właściwym P . W przypadku rozwiązywania sylogizmów poprawnych uwzględnianie tych dwóch możliwości nie wpływa na poprawność rozwiązania. Natomiast w przypadku sylogizmów zbudowanych według schematów niepoprawnych pominięcie sytuacji, w której S jest podzbiorem P może prowadzić do rozwiązania błędnego.

W trakcie rozwiązywania zadań w teście SRD badani mogli wykonywać notatki. W przypadku błędnych rozwiązań zadań zawierających zdania ogólnotwierdzące, badani często wyrażali relację zachodzącą pomiędzy zakresami nazw poprzez użycie znaku równości (=). Takiego typu błędne rozwiązania zostały zaprezentowane na Przykładzie 1(a) i 1(c). Badani używali także różnych wersji diagramów Venna oraz kół Eulera, które (przynajmniej teoretycznie) mogłyby pomagać w uwzględnieniu obu interpretacji zdania ogólnotwierdzącego. Jednak także użycie graficznej formy reprezentacji relacji pomiędzy zakresami nazw także mogło prowadzić do wyboru rozwiązań niepoprawnych (por. Przykład 1(b)).

Vendler (1967) zwrócił uwagę na odmienne sposoby rozumienia angielskich słów *all* oraz *every*, które są używane to budowy zdań ogólnotwierdzących. Jego zdaniem *all* sugeruje kolektywną interpretację, *every* – dystrybutywną. Rozróżnienie to jest dobrze widoczne w następującym przykładzie (Brooks i Braine, 1996): zdanie „Wszystkie kwiaty są w wazonie” (ang. *All the flowers are in a vase*) sugeruje traktowanie wszystkich kwiatów jako zbioru w sensie kolektywnym i zinterpretowanie zdania jako „Wszystkie kwiaty są w jednym wazonie”. Z drugiej strony zdanie „Każdy kwiat jest w wazonie” (ang. *Every flower is in a vase*) skłania do potraktowania każdego kwiatu z osobna, i może lepiej odnosić się do sytuacji, gdy każdy z kwiatów znajduje się w osobnym wazonie. Podobny efekt może być zauważalny dla polskich odpowiedników tych wyrażen: *wszystkie* oraz *każdy*.

W polskiej literaturze dotyczącej sylogistyki zdania ogólnotwierdzące najczęściej są sformułowane przy użyciu wyrażenia *każdy*: *Każde S jest P* . Można je także sformułować jako: *Wszystkie S są P* . Celem konstrukcji testu RIKO było sprawdzenie, czy sformułowanie językowe kwantyfikatora w zdaniach ogólnotwierdzących może wpłynąć na poprawność rozwiązań i np. zwrócić uwagę badanych na konieczność rozpatrzenia dwóch sytuacji, w których zdanie *SaP* może być prawdziwe. Zgodnie z rozróżnieniem zaproponowanym powyżej, sformułowanie zdań w formie klasycznej (*Każde S jest P*) powinno zwracać uwagę badanych na pojedyncze obiekty należące do kategorii, a więc wspomagać inkluzywną

Każda glara jest kraliem.
Żaden kral nie jest cekrą.
Na pewno coś najmniej jedna fafka nie jest cekrą.
Żadna glara nie jest fafką.

GLARA → KRAL
NKRAL → CEKRA

GLARA = KRAL
KRAL ≠ CEKRA
GLARA ≠ FAFKA

prawda fałsz nie można tego stwierdzić

(a) Przykładowe rozwiązanie zadania 2, badany K19

Każda wsółka jest pafem.
Każda żanka jest wsółką.
Każde bero jest pafem.
Każda żanka jest berem.

V. P. z B
○

prawda fałsz nie można tego stwierdzić

(b) Przykładowe rozwiązanie zadania 7, badany K19

Każda wsółka jest pafem.
Każda żanka jest wsółką.
Każde bero jest pafem.
Każda żanka jest berem.

$z = w = p = b$

prawda fałsz nie można tego stwierdzić

(c) Przykładowe rozwiązanie zadania 7, badany IB18

Przykład 1: Niepoprawne rozwiązania zadań w teście SRD

interpretację zdań ogólnotwierdzących. Zdania z użyciem wyrażenia *wszystkie* natomiast sugerują traktowanie obiektów w kategorii zbiorów, co może wpływać na większą tendencję do pomijania inkluzywnego znaczenia zdań ogólnotwierdzących.

1 Konstrukcja pozycji testowych

Test składa się z 8 zadań. Wszystkie pozycje mają charakter dwuprzestankowych sylogizmów. W każdym występują trzy różne zmienne nazwowe (w schematach oznaczane jako S , P oraz M). Test składa się z dwóch rodzajów pozycji: właściwych oraz wypełniających. Poprawną odpowiedzią w przypadku pozycji właściwych jest zawsze *nie można tego stwierdzić*. Tak jednorodnie zbudowany test mógłby skłaniać badanych do zmian odpowiedzi w poszczególnych zadaniach tylko dlatego, by wszystkie odpowiedzi nie były identyczne. Zdecydowano się dodać pozycje wypełniające, których zadaniem było zróżnicowanie możliwych poprawnych odpowiedzi, przy jednoczesnym zachowaniu spójnej struktury testu.

Zadanie badanych zarówno w przypadku pozycji właściwych, jak i wypełniających było takie samo. Kierując się informacjami zawartymi w przesłankach mieli ocenić, czy wniosek jest prawdziwy, fałszywy czy nie można tego stwierdzić.

Pozycja 1	Pozycja 3	Pozycja 4	Pozycja 6	Pozycja 8
MaS	MaS	MaS	SeM	PaM
<u>MeP</u>	<u>MaP</u>	<u>PeM</u>	<u>MaP</u>	<u>MaS</u>
SeP	SaP	SeP	SeP	SaP

Tabela 1: RIKO – Schematy pozycji właściwych

Każdy kułok jest bafą.
Każdy kułok jest żorką.
 Każda bafa jest żorką.

(a) wersja E

Wszystkie kułoki są bafami.
Wszystkie kułoki są żorkami.
 Wszystkie bafy są żorkami.

(b) wersja A

Przykład 2: RIKO – pozycje właściwe

1.1 Dwie wersje testu

W celu sprawdzenia, czy istnieje związek pomiędzy językową formą kwantyfikatora ogólnego a poprawnością rozwiązań skonstruowano dwie wersje testu. W wersji E zdania ogólnotwierdzące miały postać klasyczną – *Każde S jest P*. W wersji A zastąpiono wyrażenie *każde* wyrażeniem *wszystkie*. Zdania o schemacie *SaP* przyjmowały formę *Wszystkie S są P*.

1.2 Konstrukcja pozycji właściwych

Test zawiera 5 pozycji właściwych. Schematy zastosowane w konstrukcji pozycji zostały zaprezentowane w Tabeli 1. Należy zauważyć, że wszystkie te schematy są niepoprawnymi schematami sylogizmów (tj. prawdziwość przesłanek nie gwarantuje prawdziwości wniosku). Zarówno przesłanki, jak i wnioski wszystkich sylogizmów są zdaniami ogólnymi. W każdym z sylogizmów występuje przynajmniej jedno zdanie ogólnotwierdzące, w dwóch z nich zdanie o schemacie *SaP* występuje jako wniosek. W przypadku pozostałych zdania-wnioski są zdaniami ogólnoprzeczącymi.

Następnym krokiem w konstrukcji pozycji było zastąpienie zmiennych nazwowych pseudosłowami (konstrukcja pseudosłów przebiegła w sposób analogiczny do opisanego w raporcie Paluszkiewicz (2016)) oraz liter (*a, i, e, o*) występujących w schematach odpowiednimi wyrażeniami. Format pozycji właściwych dla obu wersji testu prezentuje Przykład 2.

Zadania zostały skonstruowane tak, aby zależnie od wybranej przez badanego interpretacji zdań ogólnotwierdzących, wybierał on różne odpowiedzi. W przypadku prawidłowych rozwiązań zadań właściwych (pozwalających założyć, że badany dopuszczał inkluzywną interpretację zdań ogólnotwierdzących) badani mieli wybierać odpowiedź *nie można tego stwierdzić*. Natomiast w przypadku założenie tylko identycznościowej interpretacji zdań ogólnotwierdzących badani, niepoprawnie, powinni wybierać odpowiedź *prawda*. Taka konstrukcja pozycji pozwala pośrednio wnioskować o przyjmowanej przez badanych interpretacji kwantyfikatora ogólnego.

Pozycja 2	Pozycja 5	Pozycja 7
MeP	SaM	PaM
<u>SaM</u>	<u>PeM</u>	<u>SeM</u>
SaP	SiP	SeP

Tabela 2: RIKO – Schematy pozycji wypełniających

1.3 Konstrukcja pozycji wypełniających

W teście zawarte są 3 pozycje wypełniające. Pozycje te także mają postać dwuprzესłankowych sylogizmów. Dwa z nich (pozycje 2 oraz 5) to sylogizmy niepoprawne, w których prawidłową odpowiedzią jest *falsz*. Jedna pozycja (7) to sylogizm poprawny, o właściwej odpowiedzi *prawda*. Pozycje wypełniające dobrano tak, by interpretacja zdań ogólnotwierdzących (inkluzywna bądź równościowa) nie miała znaczenia dla poprawności rozwiązywania zadania. Tabela 2 prezentuje schematy pozycji wypełniających. Następnym krokiem było zastąpienie zmiennych nazwowych pseudosłowami i nadanie odpowiedniej formy językowej zdaniom ogólnotwierdzącym (w zależności od wersji testu).

2 Pilotaż

Grupę badaną stanowiło 30 studentów i absolwentów uczelni wyższych (19–35 lat). Średnia wieku wyniosła 23,28 ($SD = 3,43$). 15 osób wypełniało test RIKO w wersji A, pozostałe 15 osób – w wersji E. Rozwiązywanie testu RIKO było częścią badania dotyczącego Styli Epistemicznych. W trakcie jednego spotkania badani rozwiązywali test RIKO (wersję A lub E), w trakcie drugiego jedną z wersji testu RIKE. Wszystkie osoby były ochotnikami i otrzymały wynagrodzenie za udział w badaniu.

2.1 Wersja A

Średnia suma punktów uzyskiwanych przez badanych w całym teście wyniosła 4 ($SD = 1,93$) przy możliwej maksymalnej liczbie punktów 8 (pozycje właściwe oraz wypełniające). Rozkład wyników nie jest zgodny z rozkładem normalnym ($W = 0,846$; $p = 0,015$).

Średnia suma punktów uzyskiwanych tylko w pozycjach testowych (maksymalnie 5) wyniosła 1,67 ($SD = 1,72$). Rozkład tej zmiennej również nie jest zgodny z rozkładem normalnym ($W = 0,821$; $p = 0,007$).

2.2 Wersja E

Średnia suma punktów uzyskiwanych przez badanych w całym teście wyniosła 3,6 ($SD = 1,72$) przy możliwej maksymalnej liczbie punktów 8 (pozycje właściwe oraz wypełniające). Rozkład wyników jest zgodny z rozkładem normalnym ($W = 0,891$; $p = 0,069$).

Średnia suma punktów uzyskiwanych tylko w pozycjach testowych (maksymalnie 5) wyniosła 1 ($SD = 1,46$). Rozkład tej zmiennej nie jest zgodny z rozkładem normalnym

Pozycja	Wskaźnik trudności
1	40%
3	40%
4	33%
6	27%
8	27%

Tabela 3: Wskaźnik trudności dla pozycji wypełniających – wersja A

($W = 0,739; p = 0,001$).

2.3 Różnice pomiędzy grupami

Porównano poziom wykonania pozycji właściwych pomiędzy dwoma wersjami testu. Wynik testu U Manna-Whitneya ($U = 82,5; p = 0,217$) pozwala stwierdzić brak istotnych statystycznie różnic w poziomie poprawności rozwiązań w obu wersjach testu.

Aby sprawdzić, osoby badane nie różnią się poziomem wykonania zadań sylogistycznych, sprawdzono także poziom wykonania pozycji wypełniających. Nie stwierdzono istotnych statystycznie różnic pomiędzy grupami ($U = 143; p = 0,217$).

3 Własności psychometryczne

3.1 Wskaźnik trudności

3.1.1 Wersja A

Test RIKO w wersji A okazał się testem trudnym. Średnia trudność pozycji właściwych wyniosła 33%. Wskaźniki trudności dla poszczególnych pozycji prezentuje Tabela 3. Analiza trudności pozycji wypełniających pozwala sądzić, że szczególnie trudne dla badanych okazały się pozycje wymagające odpowiedniej interpretacji zdań ogólnotwierdzących, a nie zadania polegające na rozwiązywaniu sylogizmów w ogóle. Trudności pozycji wypełniających wyniosły kolejno: pozycja 2 – 87%; pozycja 5 – 60%; pozycja 7 – 87%.

3.1.2 Wersja E

Test RIKO w wersji E okazał się testem bardzo trudnym. Średnia trudność pozycji testowych wyniosła 19%. Szczegółowe wartości wskaźnika trudności prezentuje Tabela 4. Podobnie jak w przypadku wersji A, w wersji E pozycje wypełniające okazały się dla badanych zdecydowanie mniej trudne od pozycji właściwych (pozycja 2: 80%; pozycja 5: 87%; pozycja 7: 93%).

3.2 Rzetelność

Do oceny rzetelności obu wersji testu wybrano współczynnik λ_2 Guttmana, który uwzględnia trudność pozycji wchodzących w skład narzędzia. Ponieważ obie wersje testu mają

Pozycja	Wskaźnik trudności
1	20%
3	27%
4	13%
6	20%
8	13%

Tabela 4: Trudność pozycji wypełniających – wersja E

Pozycja	λ_2 Guttmana po usunięciu
1	0,786
3	0,712
4	0,658
6	0,702
8	0,737

Tabela 5: Rzetelność po usunięciu poszczególnych pozycji – wersja A

niejednorodną strukturę (obecność dwóch rodzajów pozycji), obliczona została rzetelność narzędzi tylko dla pozycji wypełniających.

3.2.1 Wersja A

Rzetelność skali składającej się z pozycji właściwych wyniosła 0,765. Można uznać to za satysfakcjonujący poziom rzetelności. Rzetelność całego narzędzia (8 pozycji) wyniosła 0,703. Taki spadek rzetelności może świadczyć o różnym charakterze pozycji wypełniających oraz właściwych, co jest własnością pożądaną.

Tabela 5 prezentuje rzetelność narzędzia przy wykluczeniu poszczególnych pozycji. Wykluczenie pozycji 1 pozwala na nieznaczne zwiększenie rzetelności narzędzia.

3.2.2 Wersja E

Rzetelność skali składającej się z pozycji właściwych wyniosła 0,786. Można uznać to za satysfakcjonujący poziom rzetelności. Rzetelność całego narzędzia (8 pozycji) wyniosła 0,738. Taki spadek rzetelności może świadczyć o różnym charakterze pozycji wypełniających oraz właściwych, co jest własnością pożądaną.

Usunięcie pozycji 1 pozwala na znaczne zwiększenie rzetelności narzędzia. Jednak ze względu na niewielką ilość pozycji oraz konieczność zachowania spójności pomiędzy wersjami zdecydowano się na nie usuwanie pozycji.

Pozycja	λ_2 Guttmana po usunięciu pozycji
1	0,835
3	0,649
4	0,754
6	0,761
8	0,674

Tabela 6: Rzetelność po usunięciu poszczególnych pozycji – wersja E

Pozycja	l. odpowiedzi niepoprawnych	l. odpowiedzi <i>prawda</i>	%
1	9	7	77,78
3	9	9	100
5	10	8	80
6	11	9	81,82
8	11	6	54,55

Tabela 7: Zestawienie odpowiedzi niepoprawnych – wersja A

4 Analiza niepoprawnych odpowiedzi

4.1 Wersja A

Jak zostało wspomniane powyżej, pozycje właściwe zostały tak skonstruowane, by przyjęcie różnych interpretacji kwantyfikatora ogólnego skutkowało różnymi wyborami odpowiedzi. Przeanalizowano częstości wyborów odpowiedzi niepoprawnych *falsz* oraz *prawda*. Wyniki zostały zaprezentowane w Tabeli 7. Większość badanych wybierających odpowiedzi niepoprawne decydowała się na wybór odpowiedzi *prawda*, a więc takiej, która sugeruje nie uwzględnienie inkluzywnej interpretacji kwantyfikatora ogólnego.

4.2 Wersja E

Analiza odpowiedzi dla wersji E również pokazała zdecydowaną przewagę odpowiedzi *prawda* w zakresie odpowiedzi niepoprawnych. Efekt ten jest bardziej widoczny w wersji E; w przypadku trzech pozycji 100% odpowiedzi niepoprawnych wskazywało na niewystarczającą, tylko identycznościową interpretację zdania ogólnotwierdzącego.

Pozycja	l. odpowiedzi niepoprawnych	l. odpowiedzi <i>prawda</i>	%
1	12	12	100
3	11	11	100
5	13	13	100
6	12	10	83,33
8	13	12	92,31

Tabela 8: Zestawienie odpowiedzi niepoprawnych – wersja E

5 Podsumowanie

Obie wersje testu okazały się narzędziami o zadowalającej rzetelności, pozwalającej wykorzystywać je w badaniach naukowych. Zarówno wersja A, jak i wersja E są narzędziami dość trudnymi. Pozwala to na traktowanie ich w kategoriach narzędzi selekcyjnych, pozwalających wyłonić osoby charakteryzujące się wysokimi wynikami w zakresie badanej cechy. Pomimo braku różnic w poziomie rozwiązania obu wersji, analizując zestawienia odpowiedzi niepoprawnych (por. Tabele 7 i 8) można zauważyć pewien trend błędnych odpowiedzi. W przypadku wersji E badani znacznie częściej wybierali błędne odpowiedzi związane z identycznościową interpretacją zdań ogólnotwierdzących. Należy wykonać badania na większej grupie osób i sprawdzić, czy takie zróżnicowanie ma charakter przypadkowy, czy może wskazuje na faktyczne różnice pomiędzy wersjami testu.

Test Różnych Interpretacji Kwantyfikatora Egzystencjalnego (RIKE)

W klasycznej sylogistyce kwantyfikator egzystencjalny, wyrażany poprzez wyrażenie *niektóre* jest rozumiany jako *przynajmniej jeden, a może wszystkie*. Zgodnie z tą interpretacją, zdanie „Niektóre wróble są ptakami” uznajemy za zdanie prawdziwe. Jednak jeżeli zinterpretujemy wyrażenie *niektóre* pragmatycznie jako *tylko niektóre* to samo zdanie musimy uznać za fałszywe. Trudności w interpretacji zdań szczegółowych bywają czasem wskazywane jako jeden z powodów niepoprawnych rozwiązań zadań dedukcyjnych opartych na sylogizmach.

Celem konstrukcji *Testu Różnych Interpretacji Kwantyfikatora Egzystencjalnego* (RIKE) było ustalenie czy i w jakich kontekstach pragmatyczna interpretacja zdań szczegółowych może stanowić wyjaśnienie błędów w rozwiązaniach zadań opartych na sylogizmach. Analiza zadań w teście *Stylów Rozumowania Dedukcyjnego* (Paluszkiewicz, 2016) pokazała, że w większości przypadków trudno wskazać powód, dla którego przyjęcie interpretacji pragmatycznej miało pogarszać rozwiązywanie zadań. Analiza wyników grup rozwiązującymi testy ze zdaniem szczegółowym wyrażonymi tradycyjnie (*Niektóre S są P*) oraz takich, w których wyeliminowano możliwość interpretacji pragmatycznej (*Na pewno co najmniej jedno S jest P*) nie wykazała różnic w poziomie poprawności rozwiązywania zadań. Może to sugerować, że interpretacja pragmatyczna istotnie wpływa na poziom rozwiązywania zadań, w których wykorzystywane są klasyczne zdania kategoriowe (por. Feeney, 2004; Schmidt i Thompson, 2008), jednak w zadaniach polegających na ocenie poprawności sylogizmów nie ma znaczenia.

Kolejnym rzadko uwzględnianym aspektem jest treść zdań używanych w badaniach; najczęściej są one przeprowadzane na nazwach konkretnych. W przypadku zdania „Niektórzy pszczelarze są pisarzami” przyjmowanie interpretacji pragmatycznej jest związane nie tylko z interpretacją wyrażenia *niektóre*, ale także odniesieniem się do wiedzy o świecie. Wydaje się więc, że wyeliminowanie możliwości odnoszenia się do wiedzy badanych i użycie abstrakcyjnej treści zdań może niwelować ewentualny efekt wybierania pragmatycznej interpretacji. Zarówno w przypadku testu SRD, jak i testu RIKE zdecydowano się na wyeliminowanie nazw mających konkretne znaczenie.

6 Konstrukcja pozycji testowych

W skład testu wchodzi 8 pozycji mających formę dwuprzęsłankowych sylogizmów. Każdy z nich zawiera trzy zmienne nazwowe (w schematach oznaczane jako *S*, *P* oraz *M*). Pozycje mają dwójaki charakter: pięć z ośmiu pozycji to tzw. pozycje właściwe, a trzy – pozycje wypełniające. Motywacją do takiego sposobu konstrukcji testu, podobnie jak w przypadku RIKE, była potrzeba zróżnicowania możliwych poprawnych odpowiedzi w teście.

W przypadku obu typów pozycji zadanie badanych było takie samo. Kierując się informacjami zawartymi w przesłankach mieli ocenić, czy wniosek jest prawdziwy, fałszywy czy nie można tego stwierdzić.

Pozycja 2	Pozycja 4	Pozycja 5	Pozycja 7	Pozycja 8
MaP	MeP	SeM	PaM	PeM
<u>SaM</u>	<u>SaM</u>	<u>MaP</u>	<u>SeM</u>	<u>SaM</u>
SiP	SoP	PoS	SoP	SoP

Tabela 9: RIKE – Schematy pozycji właściwych

6.1 Dwie wersje testu

W celu sprawdzenia, czy istnieje związek pomiędzy językową formą kwantyfikatora szczegółowego a poprawnością rozwiązań skonstruowano dwie wersje testu. W wersji S zdania szczegółowe miały postać klasyczną – *Niektóre S są P* oraz *Niektóre S nie są P*. W wersji C zastąpiono wyrażenie *niektóre* wyrażeniem *na pewno co najmniej jeden*. Zdania o schemacie *SiP* przyjmowały formę *Na pewno co najmniej jedno S jest P*, a zdanie o schemacie *SoP* – *Na pewno co najmniej jedno S nie jest P*.

6.2 Konstrukcja pozycji właściwych

Pozycje właściwe zostały dobrane w taki sposób, aby w zależności od wybranej interpretacji zdań szczegółowych (pragmatycznej lub logicznej) badani decydowali się na wybór różnych odpowiedzi. Założenie to okazało się bardzo mocne – udawało się znaleźć schematy sylogizmów, w których przyjęcie mogło (ale nie musiało) wpływać na wybieraną odpowiedź. Nie pozwalały one jednak na jednoznaczne stwierdzenie, czy wybór odpowiedzi zależy od interpretacji zdań szczegółowych czy też od innych czynników.

Ocena prawdziwości zdań szczegółowych (*SiP*, *SoP*) w interpretacji pragmatycznej (*tylko niektóre*) i logicznej (*niektóre lub wszystkie*) różni się w przypadku, gdy:

- zakresy nazw użytych do konstrukcji sylogizmów są identyczne lub zachodzi między nimi stosunek podrzędności (*SiP*);
- zakresy nazw się wykluczają (*SoP*).

W obu przypadkach przyjęcie interpretacji pragmatycznej wyrażenia *niektóre* skutkuje stwierdzeniem fałszywości zdania, natomiast logicznej – jego prawdziwości.

Przykładami sylogizmów, których poprawność opiera się głównie na założeniu o logicznej interpretacji wyrażenia *niektóre* są tzw. tryby podrzędne. Tryby podrzędne powstają z trybów głównych, w wyniku zamiany wniosku ogólnego na szczegółowy. Tabela 9 prezentuje pięć schematów trybów pobocznych wykorzystanych do konstrukcji pozycji właściwych.

Poprawne odpowiedzi dla wszystkich schematów sylogizmów były takie same. Zgodnie z klasyczną interpretacją wyrażenia *niektóre*, poprawną odpowiedzią była *prawda*.

Następnie schematy wypełniono pseudosłowami (zastępując zmienne zdaniowe) oraz odpowiednimi wyrażeniami (zastępującymi litery *a*, *i*, *e*, *o*). Zdania szczegółowe sformułowano na dwa różne sposoby, zależnie od wersji testu. Format pozycji właściwych prezentuje Przykład 3.

Każdy pakus jest wozakiem.
Każdy czumas jest pakusem.
 Na pewno co najmniej jeden czumas jest wozakiem.

(a) Wersja C

Każdy pakus jest wozakiem.
Każdy czumas jest pakusem.
 Niektóre czumasy są wozakami.

(b) Wersja S

Przykład 3: RIKE – pozycje właściwe

Pozycja 1	Pozycja 3	Pozycja 6
MeP	MeS	SaM
<u>SaM</u>	<u>PeM</u>	<u>PeM</u>
SeP	SeP	SiP

Tabela 10: RIKE – Schematy pozycji wypełniających

6.3 Konstrukcja pozycji wypełniających

Aby zróżnicować rodzaj poprawnych odpowiedzi, pozycje wypełniające dobrano w sposób następujący: w dwóch z nich (1 oraz 6) poprawną odpowiedzią był *falsz*, w jednej (3) – nie można tego stwierdzić. Ponadto pozycje dobrano tak, aby interpretacja wyrażenia *niektóre* nie wpływała na poprawność wyboru odpowiedzi. Schematy pozycji wypełniających prezentuje Tabela 10.

7 Pilotaż

Grupę badaną stanowiło 30 studentów i absolwentów uczelni wyższych (19–35 lat). Średnia wieku wyniosła 23,28 ($SD = 3,43$). 15 osób wypełniało test RIKE w wersji S, pozostałe 15 osób – w wersji C. Wszystkie osoby były ochotnikami i otrzymały wynagrodzenie za udział w badaniu. Badani rozwiązywali także test RIKE, sesje badawcze były rozdzielone przynajmniej tygodniowym odstępem. Pozwala to sądzić, że rozwiązywanie RIKE nie wpływało na znajomość zadań lub wykształcenie określonej strategii rozwiązywania zadań o postaci sylogizmów.

7.1 Wersja S

Średnia suma punktów uzyskiwanych przez badanych w całym teście wyniosła 3,53 ($SD = 1,85$) przy możliwej maksymalnej liczbie punktów 8 (pozycje właściwe oraz wypełniające). Rozkład wyników jest zgodny z rozkładem normalnym ($W = 0,938$; $p = 0,354$).

Średnia suma punktów uzyskiwanych tylko w pozycjach testowych (maksymalnie 5) wyniosła 1,27 ($SD = 1,58$). Rozkład tej zmiennej nie jest zgodny z rozkładem normalnym ($W = 0,798$; $p = 0,002$).

Pozycja	Wskaźnik trudności
2	47%
4	13%
5	13%
7	33%
8	20%

Tabela 11: Wskaźnik trudności dla pozycji wypełniających – wersja S

7.2 Wersja C

Średnia suma punktów uzyskiwanych przez badanych w całym teście wyniosła 5,27 ($SD = 1,67$) przy możliwej maksymalnej liczbie punktów 8 (pozycje właściwe oraz wypełniające). Rozkład wyników jest zgodny z rozkładem normalnym ($W = 0,921; p = 0,198$).

Średnia suma punktów uzyskiwanych tylko w pozycjach testowych (maksymalnie 5) wyniosła 3,33 ($SD = 1,59$). Rozkład tej zmiennej nie jest zgodny z rozkładem normalnym ($W = 0,828; p = 0,009$).

7.3 Różnice pomiędzy grupami

Przeanalizowano wyniki pod kątem różnicy w poziomie wykonania pozycji właściwych w obu wersjach. Stwierdzono, że istnieje statystycznie istotna różnica w poziomie poprawności rozwiązań w pomiędzy wersjami testu ($U = 40,00; p = 0,002$). Wersja C jest rozwiązywana lepiej niż wersja S.

Aby sprawdzić, czy różnica w poziomie wykonania nie ma związku z charakterystyką osób badanych w danych grupach porównano także poziom wykonania pozycji wypełniających. Nie stwierdzono istotnych statystycznie różnic pomiędzy grupami ($U = 133,5; p = 0,389$). Pozwala to sądzić, że na poprawność rozwiązań pozycji właściwych wpływała jedynie forma językowa wyrażenia *niektóre*.

8 Własności psychometryczne

8.1 Wskaźnik trudności

8.1.1 Wersja S

Test RIKE w wersji S okazał się testem trudnym. Średnia trudność pozycji wyniosła 25,2%. Wskaźnik trudności dla poszczególnych pozycji prezentuje Tabela 11. Przeanalizowano także poziom trudności wypełniaczy. Okazały się one zdecydowanie łatwiejsze od pozycji właściwych. Wskaźnik trudności dla pozycji 1 wyniósł 93,33%, pozycji 3 oraz 6 – 66,67%.

8.1.2 Wersja C

Test RIKE w wersji C jest testem niezbyt trudnym. Średnia trudność pozycji wyniosła 66,6%. Pozwala to sądzić, że językowa forma zdań szczegółowych, indukująca interpretację logiczną

Pozycja	Wskaźnik trudności
2	93%
4	73%
5	47%
7	60%
8	60%

Tabela 12: Wskaźnik trudności dla pozycji wypełniających – wersja C

Pozycja	λ_2 Guttmana po usunięciu pozycji
1	0,806
3	0,770
4	0,770
6	0,663
8	0,776

Tabela 13: Rzetelność po usunięciu poszczególnych pozycji – wersja S

zmniejsza trudność zadań zbudowanych na podstawie schematów trybów podrzędnych. Sprawdzono także wskaźnik trudności dla pozycji wypełniających. Pozycje 1 i 3 okazały się nieco łatwiejsze od większości pozycji właściwych; ich trudność wyniosła 80%. Pozycja 6 natomiast okazała się najtrudniejszą w całym teście; jej wskaźnik trudności wyniósł 33,33%. W wersji S ta sama pozycja była rozwiązywana dwukrotnie lepiej. W przypadku wersji C najczęściej wybieraną niepoprawną odpowiedzią było *nie można tego stwierdzić*.

Wnioskiem poprawnie wyprowadzonym z przesłanek w przypadku pozycji 6 było zdanie o schemacie *SeP*, natomiast wnioskiem zastosowanym w przypadku tej pozycji było zdanie o schemacie *SiP* (sprzeczne z *SeP*). Możliwe, że językowa forma zdania szczegółowotwierdzącego (*na pewno co najmniej jeden*) będącego wnioskiem w przypadku tej pozycji utrudniła rozwiązanie zadania.

8.2 Rzetelność

Analizę rzetelności obu wersji testu przeprowadzono z wykorzystaniem współczynnika λ_2 Guttmana. Uwzględnia trudność pozycji wchodzących w skład narzędzia, więc jest odpowiedniejszą miarą rzetelności niż α Cronbacha. Ponieważ obie wersje testu mają niejednorodną strukturę (obecność dwóch rodzajów pozycji), obliczona została rzetelność narzędzi tylko dla pozycji wypełniających.

8.2.1 Wersja S

Rzetelność skali, w której skład wchodzi pozycje wypełniające wyniosła 0,8. Możemy uznać ją za dość wysoką. Rzetelność całego narzędzia (8 pozycji) jest niższa i wynosi 0,705. Jest to efekt pożądany, ponieważ świadczy o różnym charakterze pozycji wypełniających oraz właściwych.

Pozycja	λ_2 Guttmana po usunięciu pozycji
2	0,836
4	0,822
5	0,803
7	0,627
8	0,627

Tabela 14: Rzetelność po usunięciu poszczególnych pozycji – wersja C

Tabela 13 prezentuje rzetelność narzędzia przy wykluczeniu poszczególnych pozycji. Wykluczenie pozycji 1 pozwala na uzyskanie bardzo niewielkiego wzrostu rzetelności.

8.2.2 Wersja C

Rzetelność skali, w której skład wchodzi pozycje wypełniające wyniosła 0,804. Możemy uznać ją za dość wysoką. Rzetelność całego narzędzia (8 pozycji) jest niższa i wynosi 0,651. Jest to efekt pożądany, ponieważ świadczy o różnym charakterze pozycji wypełniających oraz właściwych.

Tabela 14 prezentuje rzetelność narzędzia przy wykluczeniu poszczególnych pozycji. Wykluczenie pozycji 1 pozwala na uzyskanie bardzo małego wzrostu rzetelności. Jednak ze względu na niewielką liczbę pozycji wchodzących w skład narzędzia oraz konieczność zachowania spójności obu wersji zdecydowano się nie usuwać żadnej z pozycji.

9 Analiza odpowiedzi niezgodnych z kluczem odpowiedzi

Kryterium poprawności dla wszystkich pozycji właściwych stanowiły warunki poprawności sylogizmów. Zakładają one m.in. logiczną interpretację wyrażenia *niektóre* jako „niektóre lub wszystkie”. Interpretacja ta nie jest zgodna z pragmatyczną interpretacją – „tylko niektóre”. Analiza odpowiedzi niezgodnych z kluczem przeprowadzona dla wersji S pokazuje, że w przypadku zadań zbudowanych według schematów trybów podrzędnych badani częściej wybierali odpowiedzi zgodne z pragmatyczną interpretacją wyrażenia *niektóre*. Trudno jednak te odpowiedzi nazywać błędnymi – rozbieżność pomiędzy odpowiedziami przewidzianymi w kluczu odpowiedzi, a odpowiedziami badanych opierała się na różnej interpretacji wyrażen językowych, a nie deficycie w zdolności rozumowania dedukcyjnego. Zgodnie z przyjętą interpretacją kwantyfikatora szczegółowego badani rozwiązywali te zadania prawidłowo.

10 Podsumowanie

Obie wersje testu RIKE okazały się narzędziami o satysfakcjonującym poziomie rzetelności, pozwalające wykorzystywać je w badaniach naukowych. Analiza wyników w obu wersjach testu pokazuje, że w przypadku bardzo konkretnych rodzajów sylogizmów językowa

Pozycja	l. odpowiedzi niezgodnych	l. odpowiedzi <i>falsz</i>	%
2	8	8	100
4	13	12	92,3
5	13	10	76,92
7	10	10	100
8	12	12	100

Tabela 15: Zestawienie odpowiedzi – wersja S

forma może wpływać na poprawność odpowiedzi. Należy jednak zauważyć, że pozycje wykorzystane w konstrukcji narzędzi miały bardzo jednolite schematy, których formalna poprawność opierała się przede wszystkim na założeniu logicznej interpretacji wyrażenia *niektóre*. W przypadku pozostałych schematów sylogizmów efekt przyjmowania pragmatycznej interpretacji nie jest tak widoczny, zwłaszcza gdy zmienne nazwowe występujące w zadaniach są pseudosłowami. Wyklucza to wpływ wiedzy badanych na rozwiązywanie zadań.

Bibliografia

- Brooks, P. J., Braine, M. D. (1996). What do children know about the universal quantifiers all and each?. *Cognition*, 60(3), 235–268.
- Feeney, A., Scafton, 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.
- Paluszkiewicz, K. (2016). *Style Rozumowania Dedukcyjnego – raport z konstrukcji narzędzia*. Raport badawczy, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu.
- Schmidt, J. R., Thompson, V. A. (2008). "At least one" problem with "some" formal reasoning paradigms. *Memory & cognition*, 36(1), 217–229.
- Vendler, Z. (1967). Each and every, Any and All. W: Z. Vendler (red.), *Linguistic in Philosophy* (ss. 70–96). Ithaca: Cornell University Press.