

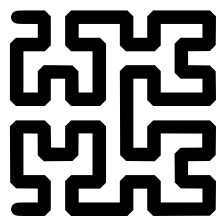


Test Giętkości Dedukcyjnej
(TGD) – raport z konstrukcji
narzędzia

Raport Badawczy

numer: 3(3)/2016; opublikowany: 12 czerwca 2016.

NATALIA ŻYLUK



REASONING
RESEARCH
GROUP

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

SPIS TREŚCI

1	Wprowadzenie	3
2	Proces konstrukcji narzędzia	3
2.1	Badana zmienna	3
2.2	Format pozycji testowych	3
2.2.1	Klasyczne zdania kategoryczne, ich rodzaje oraz związki między nimi	3
2.2.2	Konstrukcja pojedynczej pozycji testowej	4
2.3	Wstępny dobór pozycji testowych do pilotażowej wersji narzędzia	6
2.4	Ocena sędziów kompetentnych – ostateczny wybór pozycji do wersji pilotażowej narzędzia	7
2.5	Opracowanie treści instrukcji	8
2.6	Sposób oceny poprawności rozwiązania zadania	8
3	Badanie pilotażowe	9
3.1	Statystyki opisowe	9
3.2	Własności psychometryczne narzędzia	10
3.3	Popelniane błędy	11
4	Finalna wersja testu	11
5	Bibliografia	13

1 WPROWADZENIE

Test Giętkości Dedukcyjnej powstał jako narzędzie mierzące umiejętność poprawnego wskazywania zestawów przesłanek, z których logicznie wynikałby zadany wniosek. Przez wskazanie zestawów przesłanek rozumie się wybór z określonej grupy zdań takich ich kombinacji, z których wniosek, również będący zdaniem, mógłby zostać dedukcyjnie wyprowadzony. Co istotne, zestawy te muszą być zestawami „minimalnymi”, co oznacza, że nie mogą zawierać zdań, po usunięciu których wniosek nadal wynikałby z nich logicznie.

Do pilotażowej wersji narzędzia włączono 10 pozycji testowych, które wybrano z puli 93, skonstruowanych w oparciu o schematy wnioskowań bezpośrednich oraz pośrednich z wykorzystaniem klasycznych zdań kategorycznych. Tak opracowany kwestionariusz wykorzystany został w badaniu pilotażowym. Na bazie wyników badania pilotażowego podjęto decyzję o kształcie narzędzia finalnego (8 pozycji). *Test Giętkości Dedukcyjnej* może być stosowany w toku badań nad różnymi aspektami poznawczego funkcjonowania jednostki, w szczególności nad różnymi typami inteligencji.

2 PROCES KONSTRUKCJI NARZĘDZIA

2.1 BADANA ZMIENNA

W psychologii przez giętkość poznawczą (*cognitive flexibility*) najogólniej rozumie się zdolność do przełączania pomiędzy myśleniem o odmiennych pojęciach oraz do myślenia o dwóch pojęciach jednocześnie [Scott, 1962]. Nazwę badanej zmiennej ukuło bazując na definicji powyższego konstrukt – uznano, że umiejętność do znajdowania różnych zestawów przesłanek, z których logicznie wynikałby zadany wniosek jest rodzajem specyficznie rozumianej „giętkości” (wymaga bowiem przełączania między różnymi potencjalnymi kandydatami na zestaw przesłanek oraz jednoczesnego myślenia o nich). Zważywszy na charakter związku pomiędzy zdaniami, z których wyprowadzany miałby być określony w zadaniu wniosek, przyjęto, że owa giętkość nazwana zostać może dedukcyjną. Jakkolwiek giętkość dedukcyjna byłaby zdolnością o dość jasnej charakterystyce logicznej (opisać ją bowiem można poprzez odwołanie się do relacji logicznego wynikania), jej operacjonalizacja w języku psychologicznym, nawiązującym choćby do funkcjonujących w literaturze teorii inteligencji, jest zadaniem nietrywialnym i wymagającym dalszych badań.

2.2 FORMAT POZYCJI TESTOWYCH

2.2.1 KLASYCZNE ZDANIA KATEGORYCZNE, ICH RODZAJE ORAZ ZWIĄZKI MIĘDZY NIMI

Jak nadmieniono we wstępie, w procesie konstrukcji narzędzia wykorzystano *klasyczne zdania kategoryczne*, konkretnie zaś były to subsumpcyjne zdania kategoryczne – zdania podmiotowo-orzecznikowe o stosunkach pomiędzy zakresami podmiotu i orzecznika. Gałęzią logiki zajmującą się związkami pomiędzy zdaniami tego rodzaju *sylogistyka*. Klasyczne zdania kategoryczne podzielić można wedle *jakości* na zdania twierdzące oraz przeczące oraz wedle *ilości* na zdania ogólne i szczegółowe. W konsekwencji powyższego można mówić o czterech typach zdań rozważanego rodzaju (S i P są zmiennymi nazwowym):

- ogólnotwierdzące (Każde S jest P , zapisywane jako SaP);
- ogólnoprzeczące (Żadne S nie jest P , zapisywane jako SeP);
- szczegółowotwierdzące (Niektóre S są P , zapisywane jako SiP);
- szczegółowoprzeczące (Niektóre S nie są P , zapisywane jako SoP).

Rozważane zdania ogólnotwierdzące oraz ogólnoprzeczące rozumiane są w duchu ich mocnej interpretacji, tj. zakłada się, że desygnaty podmiotów owych zdań faktycznie istnieją (wedle tej interpretacji zdanie „Każdy jednorożec jest zwierzęciem kopytnym” jest fałszywe). Co również warte podkreślenia, w zdaniach szczegółowych wyrażenie „niektóre” interpretowane jest jako „co najmniej niektóre” (a nie „tylko niektóre”, jak zakładałaby interpretacja pragmatyczna).

Pomiędzy przytoczonymi typami klasycznych zdań kategorycznych zachodzą związki prawdziwościowe opisywane przez dwa typy wnioskowań: *bezpośrednie* (o jednej przesłance) oraz *pośrednie* (o dwóch przesłankach). Do wnioskowań pierwszego rodzaju zalicza się związki kwadratu logicznego oraz prawa konwersji, obwersji, kontrapozycji i inwersji, do wnioskowań drugiego rodzaju zalicza się *sylogizmy*.

Bazując na schematach wszystkich przytoczonych rodzajów wnioskowań generowano pozycje testowe wchodzące w skład opisywanego narzędzia. Jako opis że wspomnianych wnioskowań (w szczególności ich schematów) nie jest przedmiotem niniejszego raportu, opisując sposób budowania TGD zakładać będziemy, że czytelnik zaznajomiony jest z ich charakterystyką (zainteresowany tym tematem sięgnąć może np. po pracę Żarneckiej-Białej [1995]).

2.2.2 KONSTRUKCJA POJEDYNCZEJ POZYCJI TESTOWEJ

Każda pojedyncza pozycja testowa wchodząca w skład wstępnej puli pytań składała się z sześciu klasycznych zdań kategorycznych zdań: jednego zdania pod kreską (wniosku) oraz pięciu zdań nad kreską (potencjalnych przesłanek). W znaczącej większości przypadków zdania nad kreską dobierano w taki sposób, aby przy wykorzystaniu ich oraz zadanego z góry wniosku można było zbudować poprawne schematy wnioskowań bezpośrednich lub pośrednich. Opracowano również kilka zadań z zerową liczbą poprawnych zestawów przesłanek. Wygenerowano pozycje testowe, w przypadku których możliwe było znalezienie od 0 do 6 zestawów zdań nad kreską. Co istotne, dane zdanie ze zbioru nad kreską mogło być „użyte” w różnych zestawach przesłanek w jednym zadaniu. Jak wspomniano we wstępie, jako jeden zestaw przesłanek uznawano „minimalny” zbiór zdań nad kreski – taki, którego przyjęcie wystarczało do uznania wniosku pod kreską.

Wygenerowano od 20 do 29 pozycji testowych dla każdej grupy zadań wyróżnionych ze względu na rodzaj zdania będącego wnioskiem: 29 pozycji z wnioskiem ogólnotwierdzącym, 24 pozycji z wnioskiem ogólnoprzeczącym oraz po 20 pozycji z wnioskami szczegółowotwierdzącymi i szczegółowoprzeczącymi. Pozycje testowe opracowywano wykorzystując symboliczny zapis zmiennych nazwowych (S , M , P ; w każdej pozycji testowej wykorzystywano tylko 3 zmienne). W przesłankach stosowano negację zdaniową (\neg) oraz nazwową ($'$),

w pojedynczych przypadkach wniosków używano negacji zdaniowe.

Przykładowe pozycje dla każdego rodzaju wniosku:

Wniosek ogólnotwierdzący:

- (1) $P'aS'$
- (2) MaP
- (3) SaM
- (4) $\neg SoP$
- (5) $\frac{\neg SeP}{SaP}$

Rozwiązaniem są kombinacje: 1 (wniosek jest równoważny z efektem swojej kontrapozycji); 2 i 3 (obie przesłanki wraz z wnioskiem tworzą tryb sylogistyczny *Barbara*) oraz 4 (na mocy zależności wynikających z kwadratu logicznego) – w sumie 3 zestawy.

Wniosek ogólnoprzeczący:

- (1) $\neg PiS$
- (2) SaM
- (3) PeS
- (4) MeP
- (5) $\frac{\neg SiP}{SeP}$

Rozwiązaniem są kombinacje: 1 (na mocy zależności wynikających z kwadratu logicznego oraz operacji konwersji), 3 (na mocy operacji konwersji), 5 (na mocy zależności wynikających z kwadratu logicznego) oraz 2 i 4 (obie przesłanki wraz z wnioskiem tworzą tryb sylogistyczny *Celarent*) – w sumie 4 zestawy.

Wniosek szczegółotwierdzący:

- (1) PaM
- (2) MaS
- (3) MaP
- (4) PiM
- (5) $\frac{MiS}{SiP}$

Rozwiązaniem są kombinacje: 1 i 2 (wraz z wnioskiem przesłanki te tryb sylogistyczny *Bramantip*); 3 i 2 (wraz z wnioskiem przesłanki te tworzą tryb sylogistyczny *Darapti*), 4 i 2 (wraz z wnioskiem przesłanki te tworzą tryb sylogistyczny *Dimaris*) oraz 3 i 5 (wraz z wnioskiem przesłanki te tworzą tryb sylogistyczny *Datisi*) – w sumie 4 zestawy.

Wniosek szczegółowoprzeczący:

- (1) SaP
- (2) $P'oS'$
- (3) PiS
- (4) SaM
- (5) $\frac{PiM}{SoP}$

Rozwiązaniem jest kombinacja: 2 (na mocy operacji kontrapozycji).

Te przesłanki, które nie powinny zostać wykorzystywane przez badanego starano się dobrać w sposób nieprzypadkowy, tj. biorąc pod uwagę najczęstsze błędy popełniane przez ludzi w trakcie rozwiązywania zadań z wykorzystaniem klasycznych zdań kategorycznych.

2.3 WSTĘPNY DOBÓR POZYCJI TESTOWYCH DO PILOTAŻOWEJ WERSJI NARZĘDZIA

Pośród przygotowanych pozycji testowych dokonano wyboru 10, kierując się następującymi kryteriami:

- dobór zadań z wnioskami wszystkich 4 typów;
- dobór takich pozycji testowych, które różniłyby się liczbą poprawnych zestawów przesłanek, w tym wykorzystanie zadania z brakiem prawidłowych zestawów;
- wykorzystanie zadań z negacjami zdaniowymi i nazwowymi w przesłankach, wykorzystanie zadań bez negacji w przesłankach, wykorzystanie zadań z jednym typem negacji w przesłankach;
- wykorzystanie zadania z zanegowanym (zdaniowo) wnioskiem;
- dobór pozycji odmiennych pod względem szczegółowych typów wnioskowań opisujących relacje prawdziwościowe zachodzące pomiędzy przesłankami a wnioskiem (np. dobór pozycji, w których poprawne zestawy przesłanek tworzą z wnioskiem wyłącznie tryby sylogistyczne; dobór pozycji, w których wykorzystywane są tryby i oraz operacje konwersji na tych samych przesłankach; dobór pozycji opartych na zastosowaniu tylko schematów wnioskowania bezpośredniego itd.).

Założono, że 10 będzie odpowiednią liczbą zadań, których własności można sprawdzić w pilotażu. Zestaw 10 zadań był zestawem niejako nadmiarowym – liczba ta i tak miała ulec zmniejszeniu w toku opracowania finalnej wersji narzędzia (por. podrozdział 4). Uznano 10 za górną granicę liczności zbioru zadań – przy większej liczbie pozycji testowych badani mogliby być mniej zmotywowani do kończenia testu bądź być zbyt poznawczo wyczerpani uzupełnianiem TGD, aby mogli rozwiązać zadania na miarę swoich możliwości.

Wybrane zadania były w istocie wyłącznie schematami finalnych pozycji – w dalszym kroku przystąpiono do zastąpienia występujących w nich zmiennych nazwowych S , P , M

pseudosłowami (konstrukcja pseudosłów przebiegła w sposób analogiczny do opisanego w raporcie Paluszkiewicz [2014]) oraz zastąpienia symboli *a*, *o*, *e*, *i* odpowiednimi wyrażeniami. Zdecydowano się na użycie pseudosłów, a nie słów faktycznie funkcjonujących w języku polskim, ponieważ posiadanie przez dane słowo znaczenia mogło znacznym stopniu interferować z rozwiązywaniem zadania. Aby narzędzie mogło mierzyć to, co w zamiarze powinno, treść danej pozycji testowej musiała „ustalać” relacje między pewnymi obiektami nieznanymi uczestnikowi badania, a zarazem treść zadania powinna brzmieć stosunkowo naturalnie (na co pozwoliło użycie pseudosłów). Przykładowa pozycja z zastosowaniem pseudosłów brzmiała:

- (1) Niektóre czuły są mutkami.
- (2) Niektóre mutki są zymkami
- (3) Każda czufa jest mutką.
- (4) Każda czufa jest zymkiem.
- (5) Nie jest tak, że żaden zymek nie jest mutką.
Niektóre zymki są mutkami.

Konstruując pozycje testowe zdecydowano się na zapis zdań szczegółowych z zastosowaniem wyrażenia „Niektóre są...”, „Niektóre nie są...”. Aby zapobiec konfuzji wynikającej z rozumienia słowa „niektóre” wedle jego interpretacji pragmatycznej (jako „tylko niektóre”), rozważano posłużenie się wyrażeniem „Na pewno co najmniej jeden...” jako najlepiej oddającym sens słowa „niektóre” w duchu interpretacji logicznej [Paluszkiewicz, 2014], niemniej jednak zastosowanie takiego sformułowania czyniło zdania składowe pozycji testowej nadmiernie skomplikowanymi (por. np. zdanie „Nie jest tak, że na pewno co najmniej jeden zymek nie jest mutką”). W celu uniknięcia nieporozumień wynikających z rozbieżności interpretacyjnej problematycznego wyrażenia, zdecydowano się na dodanie stosownych wyjaśnień w instrukcji – w oparciu o stosowny przykład nadmieniono w niej, w jakim znaczeniu posłużono się słowem „niektóre” w teście.

2.4 OCENA SĘDZIÓW KOMPETENTNYCH – OSTATECZNY WYBÓR POZYCJI DO WERSJI PILOTAŻOWEJ NARZĘDZIA

Wybrane 10 pozycji zostało poddanych dwustopniowej ocenie sędziów kompetentnych. Pierwsza grupa sędziów składała się z dwóch osób – doktoranta kognitywistyki oraz filozofa-logika w stopniu doktora habilitowanego. Wybrani sędziowie mieli za zadanie rozwiązanie 10 zadań, a następnie ocenę, czy wybrane pozycje są adekwatnie zróżnicowane, odpowiednio trudne oraz odpowiednio zbudowane (czy właściwa jest ich kolejność w teście oraz czy właściwa jest kolejność przesłanek w obrębie jednej pozycji). Wśród uwag, jakie mieli sędziowie znalazły się głównie kwestie dotyczące kolejności zdań nad kreską. Zestaw wybranych 10 zadań został zmieniony zgodnie z sugestiami sędziów. Na tym etapie sędziowie nie byli zapoznawani z instrukcją testu, tym niemniej zgłosili swoje uwagi dotyczące jej ewentualnej treści.

W kolejnym kroku 10 pozycji testowych wraz z opracowaną już instrukcją rozwiązywania testu zaprezentowano grupie studentów kognitywistyki trzeciego roku (7 osób), którzy

w ramach swoich studiów odbyli kilka kursów z logiki (w tym zapoznali się z podstawami sylogistyki). Zadaniem studentów było rozwiązanie narzędzia, ocena jasności sformułowania instrukcji oraz odniesienie się do budowy pozycji testowych oraz budowy narzędzia w ogóle. Na bazie uwag sędziów z drugiego etapu zmieniono kolejność pozycji testowych, dokonano dodatkowych modyfikacji w zakresie kolejności przesłanek w niektórych zadaniach, poprawiono błędy językowe występujące w treści pozycji testowych oraz znacząco zmodyfikowano treść instrukcji (w szczególności skrócono ją). Z udziałem sędziów kompetentnych oszacowano również przybliżony czas uzupełniania narzędzia. Uzyskując wyniki zadań rozwiązanych przez studentów opracowano metodę oceny poprawności wykonania poszczególnych pozycji testowych (por. podrozdział 2.6).

2.5 OPRACOWANIE TREŚCI INSTRUKCJI

Treść ostatecznej instrukcji skomponowano w taki sposób, aby nie przytłaczała ilością treści, będąc przy tym dostatecznie jasno sformułowaną. Po pierwsze, zadbano o to, aby uzasadnić posłużenie się pseudosłowami w teście – przystępujący do badania proszony był o wyobrażenie sobie, że czyta książkę, której bohaterami są fantastyczne stworzenia żyjące na planecie XYZ, takie jak flewy, pafki, nozaki, malety itp. Uprzedzono, że zdania stanowiące treść pozycji testowych TGD dotyczyć mają związków pomiędzy tymi stworzeniami.

W treści instrukcji nie posługiwano się określeniami takimi jak „wynikanie” czy „dedukcyjny” – zadanie sprowadzające się do odszukiwania zdań, pomiędzy którymi zachodzi wynikanie sformułowano bardziej opisowo. Badanemu polecano, aby stwierdził, które ze zdań znajdujących się nad kreską musiałyby być prawdziwe, aby można było uznać zdanie pod kreską. Wybór dokonany miał być poprzez wskazanie numeru zdania lub numerów zdań. Uprzedzono, że w skład pojedynczego zestawu może wchodzić jedno zdanie lub więcej oraz, że może zdarzyć się tak, że taki zestaw nie istnieje. Podkreślono również wyraźnie, że zadaniem badanego jest wybór takiego zdania bądź zdań, których przyjęcie wystarcza do uznania zdania pod kreską (tj. dany zestaw nie zawiera zdań zbędnych) oraz, że jedno zdanie może być wykorzystane w różnych zestawach.

W instrukcji zaprezentowane przykładowe zadanie wraz z rozwiązaniem oraz wyczerpującym wyjaśnieniem. Instrukcja kończyła się koniecznymi wyjaśnieniami dotyczącymi rozumienia słowa „niektóre” oraz negacji nazwowych.

Pełna wersja narzędzia użytego w pilotażu (wraz z kluczem) znajduje się w załącznikach do raportu.

2.6 SPOSÓB OCENY POPRAWNOŚCI ROZWIĄZANIA ZADANIA

Rozwiązanie każdego z zadań oceniano uwzględniając liczbę wszystkich odpowiedzi udzielonych przez badanego, liczbę odpowiedzi poprawnych badanego oraz liczbę odpowiedzi poprawnych w zadaniu. Należy w tym miejscu wyraźnie podkreślić, że jako odpowiedź uznawano również brak wskazania jakiegokolwiek zestawu (głównie z uwagi na fakt, że w jednym z zadań uwzględnionych w TGD poprawną odpowiedzią jest właśnie brak wska-

zania jakiegokolwiek kombinacji).

Przez Z oznaczmy zbiór wszystkich zdań nad kreską w danym zadaniu; $Z = \{z_1, z_2, z_3, z_4, z_5\}$. Oznaczmy teraz przez R zbiór rozwiązań zadania, przez R_p zbiór rozwiązań badanego oraz przez n liczbę rozwiązań zadania ($|R|$). Zarówno R , jak i R_p są podzbiorami zbioru potęgowego Z ($R \subseteq \mathcal{P}(Z)$; $R_p \subseteq \mathcal{P}(Z)$). Operowanie na zbiorach rozwiązań pozwoliło w naturalny sposób poradzić sobie z problemem niewskazania żadnego zestawu przesłanek w danym zadaniu przez gracza. Gdy gracz wskazał jako rozwiązanie zestawu zdań o numerach 1, 2+3 oraz 4+5, $R_p = \{\{1\}, \{2, 3\}, \{4, 5\}\}$, natomiast gdy nie wskazał żadnego zestawu, zbiór jego odpowiedzi nie był pusty, a jednoelementowy, złożony ze zbioru pustego: $R_p = \{\emptyset\}$.

Wskaźnik poprawności rozwiązania pojedynczego zadania obliczano wedle wzoru: $Ev = \frac{2 \cdot |R \cap R_p|}{n + |R_p|}$ (przez pionowe kreski oznaczono moc zbioru). Dla każdego zadania badany mógł otrzymać wskaźnik z przedziału od 0 do 1. Jak można się domyślać, $Ev = 1$, gdy $R = R_p$. Dla przykładu rozważmy kilka sytuacji:

- w zadaniu, w którym należało wskazać 4 zestawy przesłanek badany wskazał 1 zestaw, będący zarazem zestawem prawidłowym, wówczas $Ev = 0,4$; w tym samym zadaniu inny badany wskazał 3 zestawy, przy czym tylko jeden z nich był poprawny, w tym przypadku $Ev = 0,29$ (w przybliżeniu);
- w zadaniu, w którym należało wskazać 5 zestawów przesłanek badany wskazał 5, z czego 3 były poprawne, wówczas $Ev = 0,6$; w tym samym zadaniu inny badany wskazał 4 zestawy – wszystkie poprawne, w tej sytuacji $Ev = 0,89$ (w przybliżeniu);
- w zadaniu, w którym należało wskazać 2 zestawy badany wskazał 3 niepoprawne kombinacje zdań, wówczas $Ev = 0$; w tym samym zadaniu badany nie wskazał żadnego zestawu, w tej sytuacji również $Ev = 0$;
- w zadaniu, w którym prawidłowym rozwiązaniem było niewskazanie żadnego zestawu zdań badany udzielił odpowiedzi złożonej z 2 zestawów, wówczas $Ev = 0$; inny badany zareagował prawidłowo – nie wskazał żadnego zestawu, w tym przypadku $Ev = 1$.

Wynikiem ogólnym w teście była średnia ze wszystkich uzyskanych wskaźników poprawności (zatem również mogła przyjmować wartości od 0 do 1).

3 BADANIE PILOTAŻOWE

3.1 STATYSTYKI OPISOWE

Wszystkie uzyskane wyniki analizowano za pomocą oprogramowania IBM SPSS Statistics w wersji 23 (PS IMAGO 3).

W badaniu pilotażowym uczestniczyło 26 badanych, w tym 14 kobiet. Badani byli studentami – najwięcej uczestników studiowało kognitywistykę ($N = 4$), psychologię ($N = 4$) oraz automatykę i robotykę ($N = 3$). Dwie osoby nie zadeklarowały studiowanego kierunku.

Średni wiek osób badanych wynosił 22,65 ($SD = 1,44$).

Analizy rozkładu wyników ogólnych uzyskanych w teście dokonano za pomocą testu Shapiro-Wilka. Wynik testu pozwolił na wniosek, że rozkład ten jest zgodny z rozkładem normalnym ($W = 0,977$; $p = 0,794$). Pozostałe statystyki opisowe zostały zaprezentowane w tabeli 3.1.

Tabela 3.1: Statystyki opisowe dla wyniku ogólnego TGD w wersji pilotażowej

	<i>M</i>	<i>SD</i>	<i>Me</i>	<i>Min</i>	<i>Max</i>	<i>W</i>	<i>p</i>
<i>MeanEv</i>	0,49	0,20	0,46	0,07	0,88	0,98	0,79

3.2 WŁASNOŚCI PSYCHOMETRYCZNE NARZĘDZIA

Założono, że test TGD bada jeden homogeniczny konstrukt – giętkość dedukcyjną. Trudność zadań składających się na TGD jest jednak zróżnicowana. Tabela 3.2 zestawia wskaźniki trudności obliczone dla wszystkich 10 zadań wchodzących w skład testu. Wskaźnik trudności dla danego zadania obliczono jako średni wskaźnik jego poprawności w całej próbie wyrażony procentowo.

Tabela 3.2: Wskaźniki trudności zadań w wersji pilotażowej TGD

Zadanie	Wskaźnik trudności
1.	42,27%
2.	63,52%
3.	36,85%
4.	11,53%
5.	64,83%
6.	36,94%
7.	76,66%
8.	67,17%
9.	53,53%
10.	33,27%
średnia:	48,96%

Przyjmuje się, że do ostatecznej wersji testu należy zakwalifikować pozycje najlepiej różnicujące badanych na całym kontinuum zmienności cechy, tj. pozycje ze wskaźnikiem trudności wynoszącym około 50% [Hornowska, 2007]. Pozycje TGD są jednak ze sobą wysoce skorelowane, co wyklucza zastosowanie powyższej zasady. Zaleceniem w tym wypadku jest, aby średnia trudność oscylowała w okolicach 50%, natomiast poszczególne zadania posiadały zróżnicowaną trudność – od najtrudniejszych do najłatwiejszych [Anastasi i Urbina, 1999]. Jak sugeruje tabela 3.2, w przypadku pilotażowej wersji TGD udało się spełnić oba warunki, tj. średnia trudność oscyluje wokół pożądanej wartości oraz zadania są zróżnicowane pod względem trudności.

Rzetelność narzędzia obliczono opierając się na modelu rzetelności Guttmana. Jako wskaźnik rzetelności narzędzia TGD wybrano jeden z 6 wskaźników (λ) zaproponowanych przez autora modelu tj. λ nr 2, której wybór rekomendowany jest w sytuacji, gdy analizujemy narzędzie złożone z zadań o zróżnicowanej trudności [Guttman, 1945]. Wartość λ nr 2 wyniosła 0,893, co świadczy o wysokiej rzetelności narzędzia.

Przeprowadzenie analizy czynnikowej metodą analiz głównych składowych pozwoliło na wniosek, że narzędzie ma strukturę jednoczynnikową, co dodatkowo uprawomocnia wniosek o jego homogenicznym charakterze (wzięto pod uwagę kryterium Cattela, tj. analizę wykresu osypiska [Cattell, 1966]).

3.3 POPEŁNIANE BŁĘDY

Aspektem, na który warto zwrócić uwagę w toku analiz narzędzia jakim jest TGD są typy błędów popełnianych przez rozwiązujących ten test.

Błędami możliwymi do popełnienia w TGD są:

- pominięcia (niewskazania właściwych zestawów);
- rozwiązania nadmiarowe (poprawny zestaw, zwykle jednozdaniowy, z dołączonymi dodatkowymi zdaniem/zdaniem):
 - nadmiarowe negatywne (poprawny zestaw połączony z zestawem niepoprawnym),
 - nadmiarowe pozytywne (poprawne zestawy połączone),
- rozwiązania całkowicie niepoprawne (każdy element zestawu jest niepoprawny; do tego typu rozwiązań zalicza się także „fragmenty” zestawów poprawnych).

Każdy z typów błędów (po podsumowaniu w ramach całego narzędzia) potencjalnie mógłby zostać potraktowany w toku pomiaru TGD jako osobna kategoria zmiennej. Warto byłoby również rozważyć opracowanie takiego wskaźnika poprawności rozwiązania zadania, który z osobna uwzględniałby wszystkie przytoczone typy błędów (a nie traktował je zbiorczo; wskaźnik poprawności w obecnym kształcie nie „rozróżnia” rozwiązań nadmiarowych od całkowicie niepoprawnych).

4 FINALNA WERSJA TESTU

Jak nadmieniono w podrozdziale 2.2, 10 uznano za górną granicę liczności zbioru tych wymagających poznawczo pozycji testowych. Okazało się jednak, że już 10 zadań wpływało negatywnie na motywację części uczestników do kończenia testu oraz męczyło niektórych badanych. Zmniejszenie liczby zadań pozwoliłoby na wykorzystanie TGD w badaniach, w toku których uczestnicy uzupełnialiby kilka testów podczas jednego spotkania, ale również umożliwiłoby użycie TGD jako samodzielnego narzędzia, którego uzupełnienie zajęłoby mniej czasu i byłoby mniej frustrujące dla badanych.

Do finalnej wersji testu zdecydowano włączyć 8 zadań. Uznano, że 8 pozycji testowych umożliwi spełnienie wszystkich wstępnych kryteriów doboru pozycji do testu (opisanych w podrozdziale 2.3).

Decydując się na eliminację pozycji testowych nie dążono do zwiększania hipotetycznej rzetelności narzędzia, ponieważ niosło to za sobą ryzyko konstruowania testu, którego pozycje testowe byłyby redundantne. Większą uwagę przyłożono do uzyskanych współczynników trudności: konstruując ostateczną wersję testu należało zadbać o utrzymanie średniego poziomu trudności zbliżonego do 50% oraz zróżnicowania wskaźników trudności poszczególnych zadań.

Zdecydowano się na usunięcie pozycji 1 (5 rozwiązań; wniosek ogólnotwierdzący) oraz 10 (6 rozwiązań, wniosek ogólnoprzeczący). Uznano, że w przypadku pozycji nr 10 niezwykle łatwo o uzyskanie punktu w przypadkowy sposób (niemal jakakolwiek kombinacja dwóch zdań była kombinacją poprawną), wszystkie poprawne zestawy zdań tworzyły z wnioskiem tryb sylogistyczny (ew. z zastosowaniem zależności wynikającej z kwadratu logicznego), zatem poprawne zestawy odpowiedzi są z gruntu mało zróżnicowane pod względem „źródła” swojej poprawności; jednocześnie pozycja ta uzyskiwała podobny wskaźnik trudności co inne pozycje w teście, zatem usunięcie jej nie przyniosłoby szkody w postaci zmniejszenia (hipotetycznej) różnorodności trudności pozycji. Dodajmy, że pozycja nr 10 została włączona do pilotażowej wersji testu głównie z uwagi na fakt, że jako jedyna posiadała tak dużą liczbę poprawnych zestawów rozwiązań, od momentu konstrukcji wstępnej wersji testu dostrzegano część z jej słabości, które ujawniły się w trakcie badania empirycznego. Pozycję testową nr 1 usunięto z uwagi na fakt, że w toku analizy czynnikowej na tle innych pozycji nisko korelowała z główną wyodrębnioną składową, ponadto, opracowując treść owej pozycji oparto się wyłącznie na jednym trybie sylogistycznym (*Barbara*) oraz, ewentualnie, zależności wynikających z kwadratu opozycji, co czyniło pozycję tę nieco gorszą od pozostałych zadań z punktu widzenia różnorodności typów wnioskowań wykorzystanych przy ich konstrukcji. Średnia (hipotetyczna) trudność testu po wykluczeniu obu pozycji wynosiła 51,25%.

Końcowa wersja testu wraz z kluczem została umieszczona w załączniku.

5 BIBLIOGRAFIA

- Anastasi, A., Urbina, S. [1999]. *Testy psychologiczne*. A. Matczak (red.). Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Cattell, R. B. [1966]. The scree test for the number of factors. *Multivariate behavioral research*, 1(2) s. 245–276.
- Guttman, L. [1945]. A basis for analysing test-retest reliability. *Psychometrika*, 10, s. 255–282.
- Hornowska, E. [2007]. *Testy psychologiczne: teoria i praktyka..* Warszawa: Wydawnictwo naukowe „Scholar”.
- Paluszkiewicz, K. [2014]. *Polisyllogisms – report*. Raport badawczy, Instytut Psychologii, Uniwersytet im. Adama Mickiewicza w Poznaniu.
- Scott, W. A [1962]. Cognitive complexity and cognitive flexibility. *Sociometry*, 4, s. 405–414.
- Żarnecka-Biały, E. [1995]. *Historia logiki dawniejszej*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.