

Testowanie umiejętności czytania: zadania typu „krótkiej odpowiedzi” i „wielokrotnego wyboru”

mgr Monika Sobejko

Jagiellońskie Centrum językowe

8 września 2017

Aby tworzyć dobre zadania (jednostki testu), autorzy zadań powinni wiedzieć (Schedl & Malloy, 2014):

- jak wygląda teoretyczny **model** danej umiejętności;
- jakie są **ramy**, do których odnosi się dany test:
 - ESOKJ → efekty kształcenia przyjęte w JCJ;
- jak wygląda **proces konstruowania** testów;
- jaki jest **plan** danego testu (tj., specyfikacji)

Umiejętność rozumienia tekstu pisanego – złożoność konstruktu

Szereg podumiejętności

- sprawności i strategię (Grabe, 2013);
- nie wszystkie możemy łatwo testować przy pomocy testów standaryzowanych, czasem obserwacja i ocena kształtująca lektora może być bardziej trafna niż wynik testu standaryzowanego;
- to co zmierzmy (konstrukt) w dużej mierze będzie zależeć od tego jak dobierzemy techniki testowania i jak dobierzemy teksty (Alderson, 2000; Grabe, 2013; Hughes, 2003; Weir, 1993), a więc różne techniki i różnorodne teksty zapewnią większą trafność wyników testu i podejmowanych na podstawie tych wyników decyzji.

Językowe kryteria doboru tekstu

Zależności między zmiennymi takimi jak słownictwo, struktury gramatyczne, sposobem organizacji (*discourse*) i tematyką tekstu są na tyle skomplikowane, że niezmiernie trudno jest wskazać która z tych zmiennych odgrywa największą rolę (Alderson, 2000).

Językowe kryteria doboru tekstu

Wskaźniki łatwości rozumienia tekstu i inne przydatne źródła:

- Coh-metrix
- Text Inspector
- VocabKitchen Profiler
- Compleat Lextutor

- English Vocabulary Profile (EVP)
- British Council | EAQUALS Core Inventory for General English

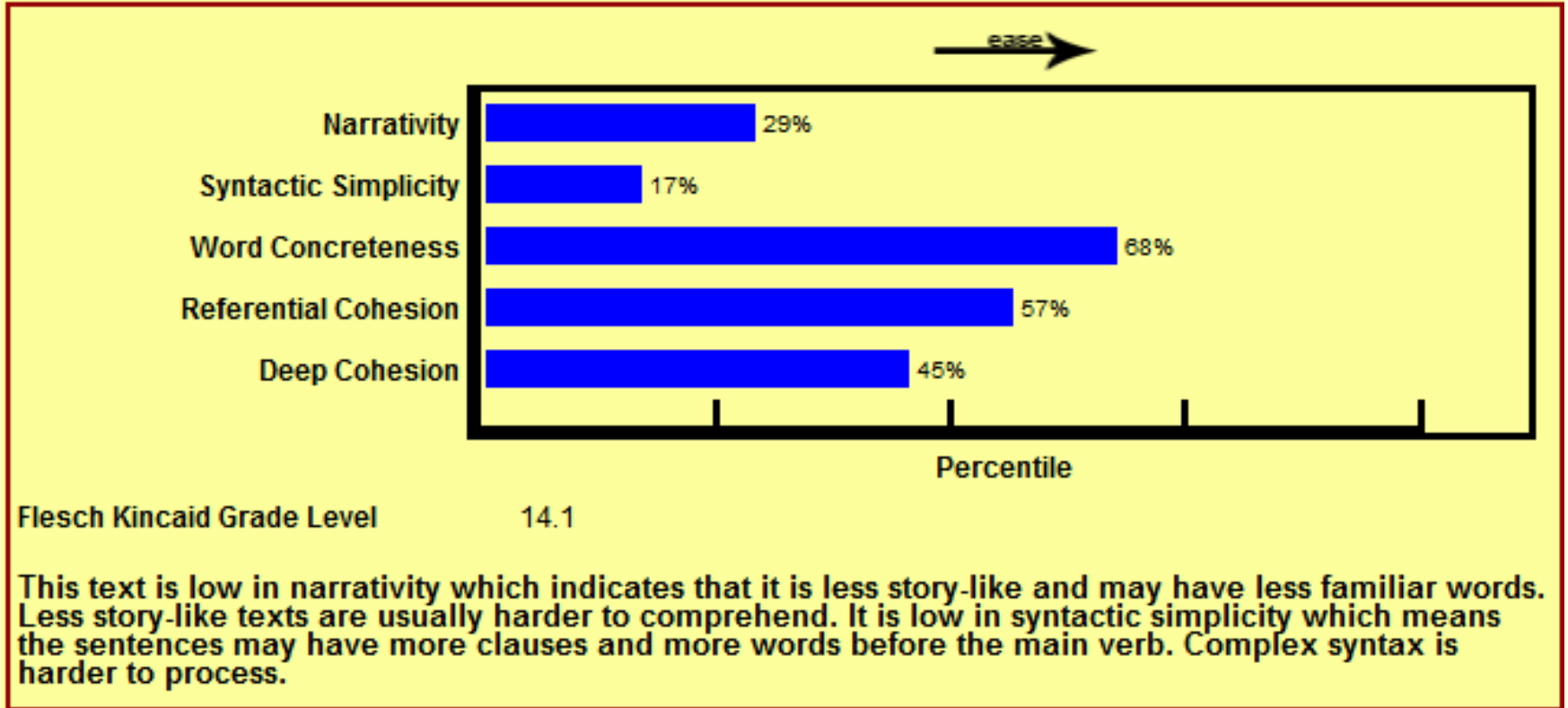
Text Inspector

Readability Scores

Flesch Reading Ease	34.10
Flesch-Kincaid Grade	16.87
Gunning Fog index	20.09

Coh-Metrix

Text Easibility Assessor



VocabKitchen

Words in the text by CEFR Level

Sort A1: 57%	Sort A2: 11%	Sort B1: 8%	Sort B2: 9%	Sort C1: 0%	Sort C2: 0%
[9] a	[1] add	[1] accommodation	[1] assuming	[1] intended	
[3] about	[1] already	[1] according	[1] based	[1] labour	
[4] after	[1] area	[1] aim	[1] conduct	[1] sector	
[1] also	[1] by	[1] aimed	[1] contribute		
[2] an	[1] case	[1] attention	[1] contribution		
[13] and	[10] cent	[1] attracts	[2] current		
[3] are	[1] centre	[1] benefit	[3] data		
[3] as	[1] collect	[1] certain	[1] debate		
[1] asked	[2] completing	[1] completely	[1] detailed		
[2] at	[2] could	[1] considerable	[1] economy		
[1] be	[1] even	[1] continue	[1] estimated		
[1] behind	[1] final	[1] directly	[1] estimates		
[1] best	[2] further	[2] education	[1] estimating		
[1] between	[1] however	[1] ending	[1] evidence		
[1] body	[1] if	[1] figure	[1] exports		
[1] change	[1] immediately	[1] former	[1] force		
[1] close	[2] including	[2] gap	[1] forced		
[1] come	[1] information	[2] government	[1] found		
[1] coming	[12] international	[3] indeed	[1] harming		
[1] countries	[1] jumping	[1] intend	[1] highlighting		
[1] country	[1] just	[1] issues	[1] highly		
[1] course	[1] large	[1] period	[2] impact		

Najlepszą miarą przy doborze tekstów...

....„combined expert judgement” (Alderson,
2000, p. 72), czyli kolektywny osąd panelu
expertów...

**METODY TESTOWANIA
UMIEJĘTNOŚCI ROZUMIENIA
TEKSTU PISANEGO – ICH WADY I
ZALETY**

Metoda „krótkiej odpowiedzi”, czyli metoda należąca do tzw. „zadań otwartych”

Wady:

wymaga interpretacji wypowiedzi kandydata, a więc:

- problem z przygotowaniem klucza, uwzględniającym wszystkie możliwości
- pytania w zasadzie muszą ograniczyć się do powierzchownego rozumienia tekstu, czyli do eksplicytnie wyrażonych treści (Buck, 2001): „what constitutes a reasonable interpretation of the text, (...) what constitutes a sufficient response to the question” (str. 140)
- problematyczna rzetelność oceny (*inter- and intrarater reliability*)
- sprawdzanie testu może być czasochłonne
- opracowanie klucza jest czasochłonne

Metoda „krótkiej odpowiedzi”, czyli metoda należąca do tzw. „zadań otwartych”

Zalety:

- przy dłuższej wypowiedzi możemy prześledzić tok myślenia kandydata
- wypowiedź dłuższa (2-3 zdania) wymaga korzystania ze skali ocen (*rating scale*)
- realistyczna technika (zarówno jeśli jest wykorzystana w formie pytań, uzupełniania zdań lub „transferu informacji” (walor autentyczności)
- bardzo elastyczna technika może być użyta do testowania czytania w celu zrozumienia głównych myśli (*skimming*) i globalnego rozumienia tekstu (transfer informacji) (Weir, 1993). W kontekstach akademickich, Weir(1993) zaleca tą technikę w połączeniu z dłuższymi tekstami.
- nie jest podatna na zgadywanie

Jak sobie poradzić?

- ograniczyć wypowiedź kandydatów do 3-4 słów;
- opracować klucz poszerzony o wszelkie akceptowalne interpretacje:
 - poprawiając zadania dzielimy odpowiedzi na „poprawne”, „graniczne” (*borderline*) i „niepoprawne”. Po jakimś czasie, wśród tych granicznych jesteśmy w stanie prześledzić tok rozumowania kandydatów na tyle, żeby zdecydować, które ostatecznie są akceptowalne, a które nie. Stopniowo, otrzymamy mniej więcej spójny system oceniania odpowiedzi (Buck, 2001).
- moderować zadania: inni lektorzy, mała grupa studentów (*trialling*);
- pilotaż, pilotaż, pilotaż...

Metoda „wielokrotnego wyboru”, czyli metoda należąca do tzw. „zadań zamkniętych”

Wady:

- tak naprawdę nie jesteśmy w stanie prześledzić toku myślenia kandydata
- w przypadku testowania gramatyki czy słownictwa, trzeba pamiętać, że testują tylko rozpoznawanie (*recognition*)
- technika jest podatna na zgadywanie
- brak autentyzmu: niektórzy badacze uważają, że jest to umiejętność sama w sobie, odmienna od umiejętności rozumienia tekstu pisanego i umiejętności rozumienia ze słuchu (Alderson, 2000)
- silny „efekt metody” – zadania MCQ mogą skłonić kandydata do ponownej interpretacji tekstu i wybrania opcji, o której bez podanych rozwiązań nie pomyślał by (Alderson, 2000; Buck, 2001)
- przygotowywanie zadań jest czasochłonne, a przygotowywanie dobrych zadań jest trudne
- mogą mieć negatywny efekt zwrotny (*washback*)

Metoda „wielokrotnego wyboru”, czyli metoda należąca do tzw. „zadań zamkniętych”

Zalety:

- zadania mają wysoką rzetelność (*inter- and intrarater reliability*)
- poprawianie zadań jest szybkie (szablony, sprzęt OCR)
- jeśli dokonujemy pilotażu, bardzo łatwo ocenić ich stopień trudności, usunąć te zadania, które nie spełniły swojej roli, itd.
- nie zależą od umiejętności pisania
- w większości, zadania są znane kandydatom (zadania „krótkiej odpowiedzi” czasem dezorientują niektórych kandydatów, gdy w sposobie sformułowania pytań pojawią się jakieś dwuznaczności)

Wskazówki dla autorów zadań wielokrotnego wyboru

1. Każde zadanie powinno mierzyć jeden ważny problem. Unikamy zadań dotyczących informacji „trywialnych”. Są to na ogół zadania, które obniżają wskaźnik rzetelności testu (tzw. *Cronbach alpha*) (Buck, 2001).
2. Zadania nie mogą być współzależne (tj. odpowiedź na jedno zadanie nie może być powiązana z odpowiedzią na drugie).
3. Wybranie poprawnej odpowiedzi nie powinno być możliwe na podstawie wiedzy ogólnej (lub specjalistycznej!) kandydatów (lub części kandydatów), bez przeczytania lub wysłuchania tekstu (*passage dependency*).
4. Zadania powinny pojawiać się w takiej kolejności w jakiej informacje są zamieszczone w tekście (wyjątkiem jest sprawdzanie umiejętności typu *scanning*).

Wskazówki dla autorów zadań wielokrotnego wyboru

5. Poprawna odpowiedź musi naprawdę być „poprawna” (musimy unikać wariantów, które są sobie zbyt bliskie znaczeniowo).
6. Każdy wariant powinien być „atrakcyjny” dla niektórych kandydatów („żelazna reguła” – każdy dystraktor powinien „przyciągnąć” minimum 8% kandydatów). We wszystkich wariantach powinny pojawiać się elementy z tekstu (tzw. „zahaczki”), przyciągające uwagę kandydatów.
7. Poprawna odpowiedź nie może różnić się od dystraktorów (na przykład, długością lub stylem). Wszystkie warianty powinny być podobne pod względem stopnia skomplikowania.
8. Jeśli jakieś elementy powtarzają się, należy zamieścić je w trzonie (a nie powtarzać w wariantach).

Wskazówki dla autorów zadań wielokrotnego wyboru

9. Trzon może mieć formę pytania lub zdania do uzupełnienia przez jeden z wariantów.
10. Trzon i warianty muszą uzupełniać się w sposób poprawny gramatycznie.
11. Należy unikać wariantów wzajemnie się wykluczających i oraz zawierających słowa takie jak „zawsze” i „nigdy”.
12. Należy unikać konstrukcji przeczących. Jeśli się pojawią, słowo przeczące należy zaznaczyć **łustym drukiem**, a wszystkie warianty powinny mieć formę twierdzącą.
13. Trzon nie powinien zawierać podpowiedzi słownych, które naprowadzają kandydata na poprawną odpowiedź.
14. Pozycja poprawnej odpowiedzi powinna zmieniać się losowo, a każda litera (A, B, C) powinna zostać użyta z podobną częstotliwością.

Metoda pracy z tekstem

1. Najpierw, musimy odpowiedzieć na pytanie – na czym polega „zrozumienie danego tekstu”?

What would constitute an understanding of [that] text? (Weir, 1993)

2. Następnie, należy opracować “mapę” głównych idei i istotnych szczegółów zawartych w tekście. Taka mapa może mieć formę diagramu (tak jak mapy myśli Tony Buzana, 1974) lub linearnej sekwencji. Następnie porównujemy taką mapę z tym, co przygotowali inni lektorzy – i, próbujemy uzgodnić nasze rozumienie istotnych treści zawartych w tekście i ich wzajemnych relacji. (Interesujące badanie dotyczące takiej procedury opisał Sarig, 1989). Weir (1993) traktuje to jako pierwszy ważny krok zapewniający *trafność (validity)* naszego testu.

Na tej podstawie, możemy przygotować zadania otwarte (na przykład, streszczenie, zadanie złożone z jednostek typu „krótka odpowiedź”) lub zadania zamknięte (na przykład, typu „wybór wielokrotny”).

Rodzaj tekstu	1) popularno-naukowy tekst zawierający elementy argumentacji, napisany stylem formalnym lub pół-formalnym; 2) list do redakcji, esej, artykuł wstępny, felieton, recenzja filmu lub książki 3) literatura piękna
Długość	650-750 słów
Metoda	wielokrotny wybór (<i>multiple choice</i>) 2 dystraktory+1 odpowiedź poprawna
Zadanie 1	rozumienie głównej myśli całego tekstu
Zadanie 2	rozumienie informacji szczegółowej w tekście
Zadanie 3	rozumienie zawartej w tekście argumentacji
Zadanie 4	rozumienie na podstawie kontekstu nieznanego słowa lub zwrotu
Zadanie 5	rozumienie głównej myśli w wybranym paragrafie
Zadanie 6	rozumienie stanowiska lub intencji autora tekstu/tonu autora tekstu/celu w jakim dany tekst powstał

Źródła

- Alderson, C. (2000). *Assessing Reading*. Cambridge, UK: Cambridge University Press.
- Alderson, Clapham C, & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge, UK: Cambridge University Press.
- Buck, G. (2001). *Assessing Listening*. Cambridge, UK: Cambridge University Press.
-
- Cobb, T. VocabProfile Home. <https://www.lexutor.ca/vp/> [2. września 2017]
-
- Coh-Metrix Text Easibility Assessor. (n.d.). <http://tea.cohmetrix.com/> [4 września 2017]
-
- English Vocabulary Profile. <http://vocabulary.englishprofile.org/staticfiles/about.html> [4. września 2017]
- Garner, J. *Vocabkitchen Profiler*. <http://vocabkitchen.com/profiler> [2. września 2017]
-
- Grabe, W., & Jiang, X. (2013). Assessing Reading. In A. J. Kunnan (Ed.) *The Companion to Language Assessment* (pp. 185–200) DOI: 10.1002/9781118411360
-
- Hughes, A.(2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- North, B., Ortega, A., Sheehan, S. (2010) *British Council | EAQUALS Core Inventory for General English*. British Council/EAQUALS <http://englishagenda.britishcouncil.org/continuing-professional-development/cpd-teacher-trainers/british-council-eaquals-core-inventory-general-english> [4 września 2017]
- *Przewodnik dla autorów zadań do testów językowych* (2004). (Ed.) W. Martyniuk. Kraków: TAIWPN Universitas.
-
- Sarig, G. (1989). Testing meaning construction: can we do it fairly? *Language Testing* , 6, 77-94.
-
- Schedl, M., & Malloy, J. (2014). Writing Items and Tasks. In A. J. Kunnan (Ed.) *The Companion to Language Assessment* (pp. 185–200)
-
- Text Inspector: The professional web tool for analysing texts. <http://textinspector.com/> [2. września 2017]
-
- Weir, C.J. (1993). *Understanding and developing language tests*. London: Prentice-Hall.

