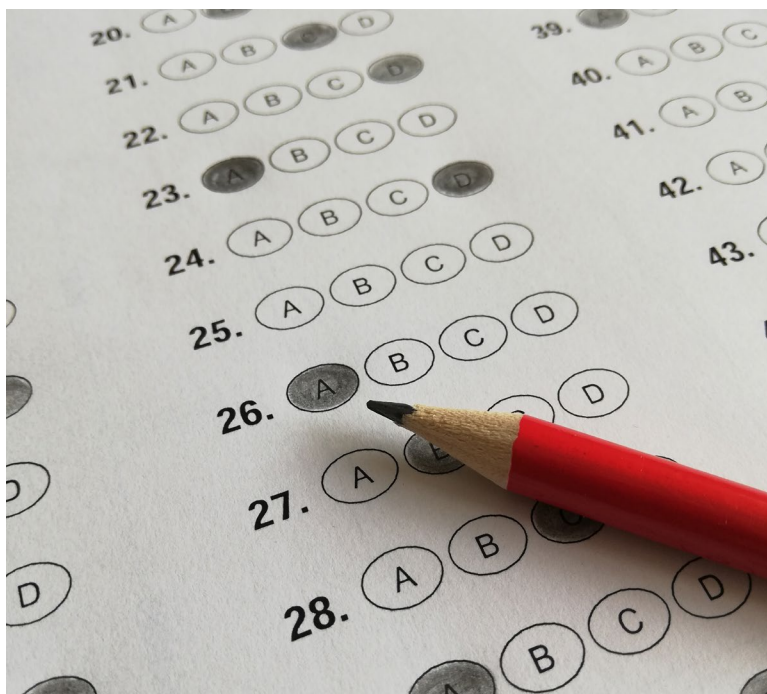


TESTOWANIE I CERTYFIKACJA W GLOTTODYDAKTYCE JĘZYKÓW OGÓLNYCH I SPECJALISTYCZNYCH

Agnieszka M. Sendur



**TESTOWANIE
I CERTYFIKACJA**
W GLOTTODYDAKTYCE
JĘZYKÓW OGÓLNYCH
I SPECJALISTYCZNYCH

Agnieszka M. Sendur

Kraków 2020

Rada Wydawnicza Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego:
Klemens Budzowski, Maria Kapiszewska, Zbigniew Maciąg, Jacek M. Majchrowski

Recenzja: dr hab. Marta Kaliska

Projekt okładki: Oleg Aleksejczuk

Adiustacja: Agnieszka Boniatowska

ISBN 978-83-66007-62-8

Copyright© by Agnieszka M. Sendur &
Krakowska Akademia im. Andrzeja Frycza Modrzewskiego
Kraków 2020

Żadna część tej publikacji nie może być powielana ani magazynowana
w sposób umożliwiający ponowne wykorzystanie,
ani też rozpowszechniana w jakiegokolwiek formie
za pomocą środków elektronicznych, mechanicznych, kopiujących,
nagrywających i innych, bez uprzedniej pisemnej zgody właściciela praw autorskich.

Na zlecenie:



Krakowskiej Akademii
im. Andrzeja Frycza Modrzewskiego
www.ka.edu.pl

Wydawca: Oficyna Wydawnicza AFM, Kraków 2020

Skład: Oleg Aleksejczuk

Spis treści

Wykaz użytych skrótów	7	
Wstęp	11	
Rozdział I. Ocenianie i egzaminowanie w nauczaniu języków obcych dla celów ogólnych i specjalistycznych – zagadnienia teoretyczne		13
1. Przedmiot, cele i podział glottodydaktyki	14	
2. Język ogólny a języki specjalistyczne	15	
3. Terminologia związana z testowaniem	17	
4. Rodzaje testów	19	
5. Kryteria poprawności testu	28	
6. Zasady konstruowania testów	33	
7. Rodzaje zadań testowych i ich charakterystyka	39	
8. Ocenianie i egzaminowanie w glottodydaktyce ogólnej i specjalistycznej – podobieństwa i różnice	77	
9. Podsumowanie	92	
Rozdział II. Testowanie języków obcych na przestrzeni wieków		95
1. Główne okresy w historii testowania języków obcych	95	
2. Testowanie języków obcych – spojrzenie historyczne	107	
3. Podsumowanie: zmiany w podejściu do testowania	128	
Rozdział III. Testowanie i certyfikacja języków obcych dzisiaj		135
1. Określenie poziomu biegłości językowej	136	
2. Certyfikacja języków ogólnych	148	
3. Certyfikacja języków specjalistycznych	158	
4. Stopień specjalizacji egzaminów przeprowadzanych na uczelniach wyższych	179	
5. Stopień sprofilowania zawodowego w egzaminach z języka specjalistycznego	185	

6. Główne różnice pomiędzy egzaminami w glottodydaktyce ogólnej oraz specjalistycznej na podstawie analizy przykładowych egzaminów	188
Zakończenie: bilans i perspektywy certyfikacji językowej	191
Bibliografia.....	197
Załączniki.....	215

Wykaz użytych skrótów

AI	<i>artificial intelligence</i> (sztuczna inteligencja)
ALPT	<i>The Army Language Proficiency Tests</i>
ALTE	Association of Language Testers in Europe
BEC	<i>Business English Certificate</i>
BILC	Bureau for International Language Coordination
BULATS	<i>Business Language Testing Service</i>
CAE	<i>Certificate in Advanced English</i>
CEFR	<i>Common European Framework of Reference for Languages: Learning, Teaching, Assessment</i>
CKE	Centralna Komisja Egzaminacyjna w Warszawie
CLES	<i>Certificat de Compétences en Langues de l'Enseignement Supérieur</i>
CPE	<i>Cambridge Proficiency in English</i>
CS	<i>Clinical Skills</i>
D	zadanie na dobieranie
EAP	<i>English for Academic Purposes</i>
ECML	European Centre for Modern Languages
EfA	<i>English for Accounting</i>
EfT	<i>English for Tourism</i>
ELPAC	<i>English Language Proficiency for Aeronautical Communication</i>
ELTS	English Language Testing Service
EOP	<i>English for Occupational Purposes</i>
EPTB	<i>English Proficiency Test Battery</i>
ESFMG	Evaluation Service for Foreign Medical Graduates
ESOKJ	Europejski System Opisu Kształcenia Językowego

ESP	<i>English for Specific Purposes</i>
ETS	Educational Testing Service
FCE	<i>First Certificate in English</i>
FSI	<i>Foreign Service Institute</i>
GE	<i>General English</i>
GSE	<i>Global Scale of English</i>
GULT	<i>Guidelines for University Language Testing</i>
IELTS	<i>International English Language Testing System</i>
ICAO	International Civil Aviation Organization
IDP	International Development Program of Australian Universities and Colleges
ILR	<i>Interagency Language Roundtable Skill Level Descriptions</i>
IPEC	<i>International Professional English Certificates</i>
KO	zadanie typu krótkiej odpowiedzi
KRASP	Konferencja Rektorów Akademickich Szkół Polskich
KSEJ	Krajowy System Egzaminów Językowych
L	zadanie z luką
LCCI IQ	<i>London Chamber of Commerce and Industry – International Qualifications</i>
LCE	<i>Lower Certificate in English</i>
LGP	<i>Languages for General Purposes</i>
LSP	<i>Languages for Specific Purposes</i>
MEQ	<i>Modified Essay Question</i>
MOTET	<i>MONDIALE Technical English Test</i>
OET	<i>Occupational English Test</i>
OKE	Okręgowa Komisja Egzaminacyjna
OSCE	<i>Objective Structured Clinical Examination</i>
PBT	<i>Paper-based Test</i>
PF	zadanie typu prawda–fałsz
P/F/BI	zadanie typu prawda–fałsz–brak informacji
PLAB	Professional and Linguistics Assessments Board
PTE	<i>Pearson Test of English</i>
RO	zadanie rozszerzonej odpowiedzi
STANAG 6001	Porozumienie Standaryzacyjne NATO nr 6001
sTANDEM	<i>Standardised Language Certificate for Medical Purposes</i>

TBA	<i>Task-based Assessment</i> – ocenianie w podejściu zadaniowym
TSE	<i>Test of Spoken English</i>
TOEIC	<i>Test of English for International Communication</i>
TOEFL	<i>Test of English as a Foreign Language</i>
TOEFL CBT	<i>Test of English as a Foreign Language – Computer-based Test</i>
TOEFL iBT	<i>Test of English as a Foreign Language – Internet-based Test</i>
TOEFL PBT	<i>Test of English as a Foreign Language – Paper-based Test</i>
TRAB	<i>Temporary Registration Assessment Board</i>
TWE	<i>Test of Written English</i>
UCLES	University of Cambridge Local Examinations Syndicate
WEFT	<i>Written English for Tourism</i>
WW	zadanie wyboru wielokrotnego

Wstęp

Sprawdzanie osiągnięć edukacyjnych i ocenianie są nieodzownymi elementami kształcenia. W przypadku kształcenia językowego bardzo istotna, zwłaszcza w ostatnich latach, jest również certyfikacja. Przez certyfikację biegłości językowej wg Jolanty Urbanikowej (2011: 118) należy rozumieć „potwierdzanie nabytych przez uczącego się kompetencji językowych w odniesieniu do uznanego, powszechnie rozpoznawalnego standardu”.

Wymogi rynku pracy czy też potrzeba walidacji efektów uczenia się narzucają kandydatowi konieczność posiadania odpowiednich dokumentów poświadczających biegłość językową na określonym poziomie i w zakresie właściwego rodzaju języka obcego. Treść tych dokumentów, zdaniem Jolanty Sujeckiej-Zajac i Ludmiły Sobolew (2015: 11–12), powinna jednoznacznie wskazywać na jakość danej kompetencji oraz odwoływać się do „rygoryzmu postępowania, niezbędnego do takiego stwierdzenia”. Przez ów „rygoryzm” autorki mają na myśli możliwie jak najbardziej rzetelny i trafny dla wymaganej oceny sposób postępowania, w którym ocena ta opiera się na jasnych i odwołujących się do uznanych powszechnie standardów ewaluacji. Certyfikaty takie można uzyskać, przystępując do egzaminów oferowanych przez odpowiednie instytucje.

Jak twierdzi Dan Douglas (2000: 1), nie istnieje dychotomia: testy (egzaminy) językowe dla celów ogólnych oraz testy dla celów specyficznych, gdyż każdy test stworzony jest dla jakiegoś konkretnego celu. Według badacza różnica tkwi w stopniu specyficzności poszczególnych testów, które da się umieścić na osi kontinuum – od tych przeznaczonych dla celów ogólnych do tych bardziej specyficznych czy też specjalistycznych. Zgodnie z powyższą zasadą nie można postawić wyraźnej granicy pomiędzy tymi dwoma rodzajami egzaminów. A jednak nauczamy języka dla celów ogólnych i przeciwstawiamy temu rodzajowi dydaktyki naucza-

nie języków specjalistycznych. Na rynku edukacyjnym istnieją egzaminy z zakresu języków obcych ogólnych oraz cała gama egzaminów językowych skierowanych do przedstawicieli różnych branż i zawodów, a więc egzaminów do celów specyficznych lub specjalistycznych.

Intencją niniejszej monografii jest zestawienie teorii i praktyki dotyczących testowania i certyfikacji w glottodydaktyce ogólnej i specjalistycznej. W tym celu w rozdziale I dokonany został przegląd teorii odnoszącej się do testowania w ogóle oraz do testowania dla potrzeb zawodowych, rozdział II zawiera spojrzenie na testowanie języków obcych w perspektywie historycznej, natomiast w rozdziale III zaprezentowane oraz porównane zostały egzaminy certyfikujące z tzw. języka ogólnego oraz z języków specjalistycznych o różnym stopniu specyficzności. Analiza przedstawionych przykładów ma na celu skonfrontowanie, czy teoretyczne postulaty znajdują odzwierciedlenie w certyfikacji języków obcych, a więc sprawdzenie, czy i w jakim stopniu funkcjonujące obecnie na rynku egzaminy realizują teoretyczne założenia i wskazówki dotyczące odpowiednio egzaminów w glottodydaktyce ogólnej oraz specjalistycznej.

Rozdział I.

Ocenianie i egzaminowanie w nauczaniu języków obcych dla celów ogólnych i specjalistycznych – zagadnienia teoretyczne

Rozdział traktuje o ocenianiu kompetencji i egzaminowaniu zarówno w nauczaniu języka obcego ogólnego, jak i nauczaniu języków dla celów specjalistycznych. Pierwszym niezbędnym krokiem jest zestawienie terminologii związanej z testowaniem i certyfikacją w interesującym obszarze. Dalej omówione zostaną rodzaje testów z uwzględnieniem ich charakterystycznych cech i funkcji. Zaprezentowane będą również cechy testów stanowiących o ich poprawności, takie jak trafność, rzetelność, praktyczność oraz wartość dydaktyczna, a także wskazane zasady dotyczące poprawnego konstruowania testów.

W dalszej części zostaną opisane rodzaje i typy zadań w testach piśmiennych, w ustnych formach mierzenia kompetencji językowych oraz zadania egzaminacyjne specyficzne dla podejścia zadaniowego. Omówienie poszczególnych rodzajów i typów wzbogacono o przykłady egzaminów, w których dane zadanie znalazło zastosowanie. Na podstawie przedstawionych przykładów podjęta zostanie próba porównania zadań wykorzystywanych w egzaminach z języka ogólnego oraz w ocenianiu dla celów specjalistycznych.

Ostatnia część rozdziału poświęcona jest cechom testowania we wspomnianych dwóch glottodydaktykach. Przyjrzymy się tutaj atrybutom

oceny w dydaktyce języka ogólnego i w nauczaniu języków obcych specjalistycznych ze wskazaniem na ich podobieństwa oraz cechom testów i zadań w egzaminach z języków specjalistycznych, które odróżniają te dwa rodzaje certyfikacji.

By jednak mówić o ocenianiu w glottodydaktyce języków ogólnych i specjalistycznych, należy zacząć od zdefiniowania przedmiotu badań glottodydaktyki, jak również rozróżnienia pomiędzy tymi dwoma rodzajami języka obcego.

1. Przedmiot, cele i podział glottodydaktyki

Glottodydaktyka to bardzo młoda dyscyplina wyodrębniona z lingwistyki. Jest ona dziedziną interdyscyplinarną, posiadającą silne związki zarówno z językoznawstwem, jak i dydaktyką ogólną. Glottodydaktyka, do której odnosi się i ogranicza niniejsza publikacja, to glottodydaktyka języków obcych, która funkcjonuje obok tzw. glottodydaktyki ojczysto-języcznej (Jaroszewska 2014: 53). Przedmiotem badań interesującej nas dyscypliny naukowej są procesy nauczania i uczenia się języków obcych, celem badań glottodydaktycznych natomiast jest „poznanie tych procesów i wypracowanie optymalnych systemów nauczania i uczenia się” (Pfeiffer 2001: 17).

Glottodydaktyka dzieli się na ogólną oraz szczegółową (Komorowska 1982: 19; Pfeiffer 2001: 17). Ta pierwsza zajmuje się nauczaniem języków obcych w ogóle, druga zaś nauczaniem konkretnego języka. Zatem zarówno dydaktyka języków obcych ogólnych, jak i dydaktyka języków obcych specjalistycznych są przykładami dydaktyk szczegółowych. Inną klasyfikacją jest podział na glottodydaktykę czystą i stosowaną. Glottodydaktyka czysta zajmuje się „opisem i wyjaśnieniem przedmiotu swych badań, tj. procesu akwizycji języka obcego”, glottodydaktyka stosowana z kolei „wiąże się z opracowaniem i formułowaniem dyrektyw dla praktyki nauczania i uczenia się, które to dyrektywy wynikają z opisu i wyjaśnień dostarczonych przez glottodydaktykę czystą” (Komorowska 1982: 18). W wyniku tych podziałów wyróżnić można czystą i stosowaną glottodydaktykę ogólną oraz czystą i stosowaną glottodydaktykę szczegółową (Pfeiffer 2001: 15). Tematyka niniejszej monografii skupia się zatem w obrębie czystej i stosowanej glottodydaktyki szczegółowej.

Chociaż termin glottodydaktyka ogólna odnosi się do nauczania języków obcych w ogóle, to ze względów praktycznych jest on również używany przez wielu badaczy (np. Gajewska 2008; Gajewska i Sowa 2014; Gębał 2016; Sowa 2012; Szerszeń 2014) w zestawieniu z tzw. glottodydaktyką specjalistyczną, a więc nauczaniem języków obcych specjalistycznych. Będzie on również używany w tym znaczeniu w tym opracowaniu.

2. Język ogólny a języki specjalistyczne

Istnieje wiele podejść do definicji języków specjalistycznych, opartych na różnych kryteriach. Najprostsze chyba rozróżnienie pomiędzy tzw. językiem ogólnym i językami specjalistycznymi oferuje Dan Douglas (2014: 70), odnosząc to pierwsze do sytuacji, w których cele nauczania nie dają się jasno sprecyzować, te drugie zaś do tych, w których cele te dają się określić z dużą dokładnością. Magdalena Sowa (2016: 5–7) zestawia natomiast szereg różnych czynników, które leżą u podstaw rozlicznych definicji. Jednym z nich jest kryterium dziedzinowe, odnoszące się do konkretnego obszaru komunikacji zawodowej – dziedziny czy dyscypliny – w którym to dany rodzaj języka jest stosowany. Jest ono widoczne w nazewnictwie stosowanym w odniesieniu do kursów, podręczników do nauki języków zawodowych, a także certyfikatów z języków specjalistycznych (np. *Business English*, *English for Tourism*, *Medical English* itp.).

Kolejnym kryterium jest według Sowy związek języka z jego użytkownikami; to odniesienie języka do określonej grupy zawodowej pozwala w prosty sposób wskazać odbiorców danego kursu, podręcznika czy egzaminu językowego. I tak w przypadku języka angielskiego mamy do czynienia z *English for Air Traffic Controllers*, *English for Paramedics* czy *English for Lawyers*.

Jeszcze innym czynnikiem stosowanym do definiowania języka specjalistycznego jest zestawianie go z językiem ogólnym. Jak pisze Sowa, „rozbieżności pomiędzy badaczami stosującymi to kryterium dotyczą rozdzielności i zależności systemów języka specjalistycznego i ogólnego” (2016: 6). Niektórzy (np. Hoffmann 1979) widzą w językach specjalistycznych odrębną, autonomiczną odmianę języka o własnych regułach i charakterystycznych dla niego elementach, inni obserwują silne powiązanie języków specjalistycznych z językiem naturalnym (np. Grucza

2002), a jeszcze inni badacze pozycjonują się pomiędzy dwoma powyższymi opcjami i podchodzą do opisu języków specjalistycznych w sposób pośredni – twierdzą, że wszystkie języki specjalistyczne dzielą z językiem ogólnym jego podstawowe zasoby, takie jak alfabet, system morfologiczny i fonologiczny, natomiast różnią się od niego w zakresie leksyki i składni (Sowa 2016: 6).

Na języki specjalistyczne można spojrzeć również z perspektywy dydaktyki. W odróżnieniu od nauczania języka ogólnego glottodydaktyka specjalistyczna koncentruje się na wąskim celu nauczania oraz na ograniczonym zakresie tematów i treści. Charakteryzuje się również często ograniczeniami czasowymi (kursy specjalistyczne są zazwyczaj krótkie, w przeciwieństwie do kursów języka ogólnego) oraz ograniczeniami w zakresie dostępności materiałów dydaktycznych. Inną jej cechą jest przewaga ucznia względem nauczyciela w zakresie wiedzy fachowej i kompetencji dotyczących odnośnego zawodu.

I wreszcie ostatnie kryterium definiowania języka specjalistycznego wymieniane przez Sowę to kontekst, w jakim język jest używany, oraz znaczenie i zastosowanie terminów w zależności od dyscypliny czy dziedziny. W tym sensie język traktowany jest jako „nośnik wiedzy specjalistycznej” (Sowa 2016: 6).

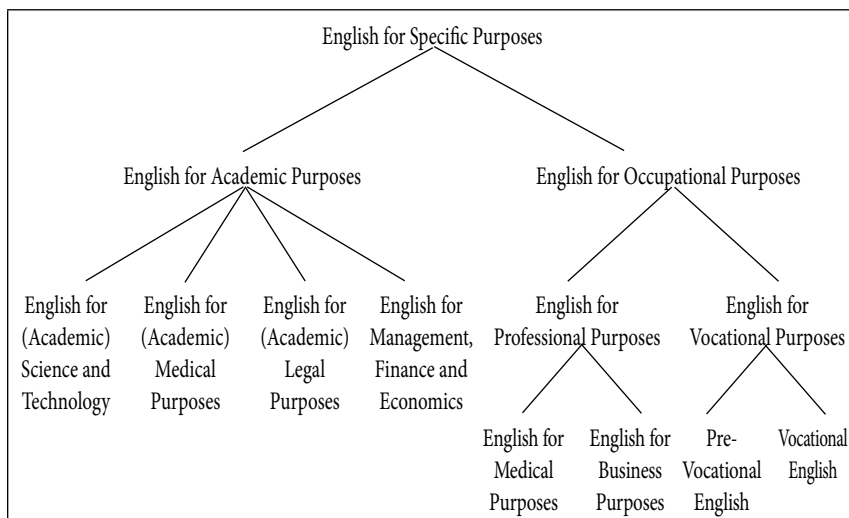
W kontekście wielości definicji i kryteriów podziału pojawia się również wiele klasyfikacji języków obcych dla celów specyficznych, jak również różnice w nazewnictwie (cf. Basturkmen 2010: 6; Belcher 2009: 2–3; Dudley-Evans i St John 1998: 6; Gajewska i Sowa 2014: 17–47; Grucza 2004, 2013: 27–46; Hutchinson i Waters 1991: 16–19). W polskiej literaturze fachowej funkcjonuje wiele terminów odnoszących się do tej kategorii języków obcych: *język specjalistyczny*, *zawodowy*, *specyficzny*, *fachowy*, *branżowy* czy wreszcie – *język dla potrzeb/celów zawodowych*. Wszystkie te nazwy funkcjonują w literaturze przedmiotu jako rodzaj dyskursu przeciwstawny do tzw. języka ogólnego. Ich funkcjonalnymi odpowiednikami są w języku angielskim *Languages for Specific¹ Purposes* (LSP; w odniesieniu do języka angielskiego: *English for Specific Purposes* – ESP), skontrastowane z *Languages for General Purposes* (LGP; w odniesieniu do języka angielskiego: *General English* – GE).

¹ Określenie LSP przyjęło się w anglojęzycznej literaturze przedmiotu w efekcie zjazdu British Council w 1968 roku, zatytułowanego *Languages for Special Purposes*. „*Special*” zostało zastąpione przez „*specific*” w 1977 roku, kiedy to British Council nazwał kolejny zjazd poświęcony językom specjalistycznym *Languages for Specific Purposes* (Gajewska i Sowa 2014: 69).

Ze względów praktycznych w niniejszej publikacji używane będą naprzemiennie polskie terminy: *języki specjalistyczne, języki zawodowe, języki dla celów specjalistycznych/zawodowych*, traktowane tu synonimicznie. Stosowane również będą angielskie skróty LSP i ESP, które wydają się bardziej czytelne i rozpoznawalne niż ewentualny skrót utworzony z nazwy polskiej.

Jak wspomniano wcześniej, pod zbiorczym określeniem języków specjalistycznych lub LSP kryją się rozliczne odmiany specyficznych języków, do których należą zarówno języki przeznaczone dla celów akademickich, jak i zawodowych. Często przywoływaną klasyfikację języka angielskiego dla celów specyficznych zaproponowali Tony Dudley-Evans i Maggie-Jo St John (1998: 6) – por. rysunek 1.

Rysunek 1. Klasyfikacja języka angielskiego dla celów specyficznych za: Dudley-Evans i St John (1998: 6)



3. Terminologia związana z testowaniem

Podstawowymi pojęciami w niniejszym opracowaniu są: *ocenianie i ewaluacja, test i sprawdzian, testowanie oraz certyfikacja*.

Ocenianie według autorów Europejskiego Systemu Opisu Kształcenia Językowego (ESOKJ) (Rada Europy 2003: 153) to ocena biegłości językowej danej osoby, a wszystkie testy językowe stanowią jakąś jej formę.

Terminem szerszym niż ocenianie jest *ewaluacja* – każdy rodzaj oceniania jest formą ewaluacji. Gdy mówimy o nauczaniu języka, kontroli podlega nie tylko biegłość językowa, ale również takie elementy, jak efektywność poszczególnych metod czy materiałów dydaktycznych, a także praca nauczyciela (Rada Europy 2003: 153). Jako że *ewaluacja* zawiera w sobie pojęcie *oceniania*, a publikacja ta dotyczy wyłącznie tego elementu ewaluacji, określenia te będą używane tutaj jako synonimy.

Kolejnymi terminami wymagającymi uściślenia są pojęcia *testu* i *sprawdzianu*. W literaturze funkcjonuje wiele definicji opisujących te określenia, jednak dla celów niniejszego tekstu formułę sprawdzianu i testu przyjęto za Hanną Komorowską:

Sprawdzian to [...] wszelka forma kontroli, zmierzająca do uzyskania przez nauczyciela informacji o tym, ile uczeń wie i co umie zdziałać w języku obcym – może to być forma ustna lub pisemna, wykorzystująca dodatkowe materiały (teksty, obrazki, plansze, arkusze odpowiedzi) lub odbywająca się bez nich.

Test to po prostu wszelki sprawdzian – ustny lub pisemny, prowadzony z dziećmi lub z dorosłymi, krótki lub długi, oficjalny lub prowadzony w domu, na lekcji z prywatnym uczniem (Komorowska 2007: 13–14).

Oznacza to, że określenia te również można uznać za synonimiczne. W innej publikacji Komorowska zawęża pojęcie testu językowego:

Test językowy jest to taki sprawdzian wiadomości i umiejętności uczniów, który ma charakter punktowy i obiektywny, a więc taki, gdzie punktowanie nie zależy od nastroju nauczyciela i jego nastawienia do ucznia (Komorowska 2005: 242).

W takim właśnie znaczeniu termin ten stosowany jest w niniejszej pracy.

Testowanie, wg Wincentego Okonia (2007: 424), to „sprawdzanie osiągnięć szkolnych za pomocą testu osiągnięć szkolnych”. Oczywiście użycie przymiotnika „szkolnych” nie ogranicza nas do niższych etapów edukacyjnych; termin ten dotyczy każdego rodzaju szkolnictwa, w tym również ustawicznego.

Terminem ściśle powiązaniem z ewaluacją jest *certyfikacja*. Podczas gdy ewaluacja polega na zmierzeniu przez nauczyciela pracy uczącego się oraz określeniu jej wartości w danej chwili nauki, certyfikacja jest oficjalnym zatwierdzeniem uzyskanych wyników, wydawanym w celu formalnego uznania ocenianej kompetencji (Sowa 2015: 46). Certyfikacja nie

może zatem odbyć się bez ewaluacji, ale ewaluacja nie musi być zwieńczona certyfikatem (Gajewska i Sowa 2014: 214). W ostatnich latach, m.in. za sprawą ESOKJ i wyróżnionych w nim deskryptorów biegłości językowej, popularna stała się ewaluacja w postaci certyfikacji posiadanych umiejętności językowych i komunikacyjnych, w tym certyfikacja biegłości językowej w zakresie języków specjalistycznych. Certyfikacja biegłości językowej, zgodnie z definicją przywołaną we Wstępie do niniejszej publikacji, to „potwierdzanie nabytych przez uczącego się kompetencji językowych w odniesieniu do uznanego, powszechnie rozpoznawalnego standardu” (Urbanikowa 2011: 118).

4. Rodzaje testów

Różne są cele testowania i różne w związku z tym rodzaje testów. W zależności od tego, jaką informację chcemy uzyskać, będziemy korzystać z innego narzędzia. Rodzaje testów i zadań testowych bez rozróżnienia na testowane przedmioty szeroko omawia Bolesław Niemierko (1999). O różnych rodzajach testów i sposobach testowania oraz oceniania języków obcych piszą m.in.: Arthur Hughes (1995), Hanna Komorowska (1984 i 2007), Nathan T. Carr (2011) oraz autorzy ESOKJ (Rada Europy 2003).

Testy mogą być klasyfikowane na różne sposoby, a różnice pomiędzy poszczególnymi podziałami wynikają z kryteriów klasyfikacji. W tej części rozdziału zebrano wszystkie podziały uwzględnione w opracowaniach cytowanych powyżej autorów.

Podział testów ze względu na formę i sposób udzielania odpowiedzi

Testy można dzielić ze względu na ich formę. Według tego kryterium mamy do czynienia z testami **pisanymi**, w których zdający udziela odpowiedzi w formie pisemnej (*paper and pencil tests*), zapisując swoje rozwiązania zadań lub wskazując gotowe odpowiedzi, oraz z testami **ustnymi** (*oral tests*), gdzie zdający wypowiada słowa i zdania stanowiące rozwiązanie zadań. Większość egzaminów językowych składa się z testów pisanych, sprawdzających umiejętność czytania, rozumienia ze słuchu i pisanie, oraz z testu ustnego – badającego sprawność mówienia.

Istnieje też rozróżnienie pomiędzy testami **słownymi**, w których zadania oparte są na słowie czytany lub mówionym, oraz **obrazkowymi**,

gdzie bazą do zadania jest obrazek czy fotografia, a nie tekst (Komorowska 2007: 15).

Odnosząc się do testów osiągnięć szkolnych, a więc do sprawdzianów wiedzy i umiejętności z różnych przedmiotów, Niemierko dzieli je m.in. ze względu na sposób udzielania odpowiedzi. Według tego podziału, oprócz omówionych powyżej testów pisanych i ustnych, występuje jeszcze **test praktyczny**. W teście praktycznym „uczeń demonstruje sposób wykonywania działań lub wytwór fizyczny” (Niemierko 1999: 55). Klasyfikacja ta nie odnosi się wyłącznie do osiągnięć szkolnych uczniów; może w równym stopniu dotyczyć ewaluacji umiejętności osób dorosłych, w tym umiejętności związanych z wykonywanym zawodem.

Podział testów ze względu na przedmiot testowania

Powszechny jest też podział testów ze względu na przedmiot testowania. W zależności od tego, co sprawdzają, testy językowe mogą być **testami elementów języka** (*discrete-point tests*), badającymi znajomość poszczególnych form językowych, bądź też testami **sprawnościowymi** (*language skills tests*), mierzącymi umiejętność posługiwania się językiem obcym. Testy elementów języka możemy dalej podzielić na: fonologiczne, ortograficzne, leksykalne i gramatyczne. Zadania w takich testach odbiegają zdecydowanie od zadań komunikacyjnych, zbliżonych do działań, które użytkownik języka wykonuje w codziennych sytuacjach życiowych lub zawodowych. Dają one niekompletny obraz umiejętności językowych zdającego, umożliwiają natomiast sprawdzenie opanowania konkretnych zagadnień gramatycznych czy leksykalnych. Wśród testów sprawnościowych znajdziemy testy sprawności mówienia, rozumienia ze słuchu, czytania ze zrozumieniem oraz testy badające umiejętności w ramach sprawności pisania. Możemy też mówić o testach **integracyjnych**, zwanych również globalnymi (*integrated tests*) (Komorowska 2007: 15–16), których zadaniem jest zbadanie jednocześnie kilku różnych sprawności.

Podział oparty na relacji testu do programu nauczania

Inny możliwy podział wynika z tego, czy dany test oparto na programie nauczania, czy jest od takiego programu niezależny. Sprawdziany oparte na programach nauczania to testy plasujące, diagnostyczne, testy osiągnięć i mierzące postępy. Sprawdziany językowe niezależne od programów nauczania to testy biegłości oraz przesiewowe (Carr 2011: 6–9).

Rolą testów **plasujących** (*placement tests*) – albo **poziomujących** – jest określenie poziomu umiejętności uczącego się celem przydzielenia go do odpowiedniej grupy zaawansowania na szeroko rozumianym kursie językowym. Testy **diagnostyczne** (*diagnostic tests*) mają na celu zidentyfikowanie jego mocnych i słabych stron. Dokonana za ich pomocą diagnoza umożliwia nauczycielowi ustalić i zaplanować program kursu dla konkretnej grupy czy konkretnego uczącego się. W celu sprawdzenia, jakie postępy czynią uczniowie w trakcie kursu w odniesieniu do przeobionego materiału i celów kursu, stosujemy **testy postępów** (*progress tests*), natomiast na zakończenie kursu (lub modułu), chcąc sprawdzić, w jakim stopniu uczący się osiągnęli założone cele i opanowali materiał kursu, zastosowany powinien być **test osiągnięć** (*achievement test*). Testy osiągnięć używane są w celu wystawienia oceny końcowej lub też w celu promowania uczącego się do dalszego etapu programu (np. klasy, kolejnego poziomu kursu, zaliczenia przedmiotu w przypadku uczelni wyższych) i sprawdzają stopień osiągnięcia konkretnych celów nauczania, a więc tego, co było nauczane na kursie.

Sprawdziany **biegłości** (*proficiency tests*) mają za zadanie ocenić ogólny poziom nabytych umiejętności zdającego niezależnie od programów, na podstawie których był nauczany. Takie określenie poziomu za pomocą testu biegłości często wiąże się z uzyskaniem określonego certyfikatu, jednak wyniki tego testu mogą również posłużyć do selekcji kandydatów na odpowiednim poziomie umiejętności językowych, na przykład w toku naboru na studia. W takiej sytuacji test biegłości przyjmuje funkcję testu **przesiewowego** (*screening test*).

W zależności od celu oraz powiązanych z nim treści egzaminy wstępne mogą należeć zarówno do kategorii testów opartych na programach nauczania, jak i być od nich niezależnymi. W sytuacji, gdy odbiegają one od wspomnianych programów, nabierają cech testów biegłości. Mogą też zostać oparte na podstawie programowej, uzależniając przyjęcie kandydata na kurs lub studia od stopnia opanowania zawartych w niej treści. Przykład taki stanowi egzamin maturalny z języka obcego.

Rodzajem testu, który łączy w sobie cechy testów biegłości oraz testów osiągnięć, jest tzw. **prochievement test**. Termin ten został utworzony przez połączenie angielskich nazw dwóch rodzajów testów: *achievement* – osiągnięć i *proficiency* – biegłości. Testy te są powiązane z konkretnym programem nauczania, a więc stanowią sprawdzian osiągnięć, ale wykraczają poza ten program. Testują biegłość językową, jednak tylko w kontekstach

i sytuacjach, które były nauczane podczas kursu (Fischer, Chouissa, Dugovičová, Virkkunen-Fullenwider 2011: 46).

Testy mogą także uwzględniać kilka aspektów oceny jednocześnie. Przeprowadzony w trakcie trwania kursu dany test może być testem postępów; może jednak odgrywać jednocześnie rolę testu diagnostycznego, jeśli nauczyciel wykorzysta go do zidentyfikowania słabych stron kursantów i na jego podstawie zdecyduje, które elementy materiału należy powtórzyć i przeciwiczyć (Carr 2011: 7–9).

Szczególnym przypadkiem odbiegającym w swoich celach i formie od wszystkich opisanych powyżej jest test **prognostyczny** (*prognostic test*), związany nie z obecnymi umiejętnościami językowymi uczącego się, a z jego kształceniem językowym w przyszłości. Celem takiego testu jest określenie przewidywanych sukcesów ucznia w procesie opanowania języka obcego, który ma dopiero nastąpić. Może to mieć miejsce nawet w przypadku ucznia, który w ogóle danego języka się nie uczył. Możemy także badać za pomocą testu prognostycznego, czy osoba, która zna język na pewnym poziomie, posiada predyspozycje związane z uczeniem się języka obcego, które predestynują ją na przykład do udziału w jakimś szczególnie atrakcyjnym kursie, na który przyjęci mogą zostać tylko wyselekcjonowani kandydaci. Z sytuacją taką mamy do czynienia w przypadku naboru do klas dwujęzycznych w polskich szkołach czy też kwalifikacji do określonych jednostek w armii amerykańskiej (Smith i Stansfield 2016).

Testy prognostyczne bazują na sztucznym, nieistniejącym w rzeczywistości języku, sprawdzając na przykład zdolność zapamiętywania wyrazów ze słuchu, zapamiętywania wyrazów z tekstu pisanego, zdolność rozumienia znaczeń wyrazów, wnioskowania, wrażliwość gramatyczną oraz zdolność kojarzenia formy dźwiękowej z graficzną (Komorowska 2007: 18–20).

Podział oparty na sposobie interpretacji wyników

Kolejnym podziałem, tym razem bazującym na sposobie interpretacji wyników, są testy **sprawdzające** (*criterion-referenced*) i **różnicujące** (*norm-referenced*). Ocenianie za pomocą testów sprawdzających zwane jest również oceną według kryteriów; ocenianie różnicujące natomiast określane jest także jako ocena według normy (Rada Europy 2003: 159). Wyniki testów mają sens tylko wtedy, kiedy dokonuje się ich porównania z jakimś punktem odniesienia, z pewną normą bądź też w odniesieniu do

jakiegoś kryterium (Carr 2011: 9). W sytuacji, kiedy dokonujemy porównania wyników zdających w ramach jednej grupy, kiedy ustalamy ranking uczących się w danym zespole według poziomu ich znajomości języka, mamy do czynienia z testami różnicującymi. O testach sprawdzających mówimy, gdy rozpatrujemy ich wyniki wobec określonych standardów, nie zaś w porównaniu z wynikami innych uczestników testu. W przypadku tego rodzaju sprawdzianu interesuje nas, jak umiejętności zdającego mają się do ustalonych zewnętrznych kryteriów, czy dana osoba osiągnęła standard wymagań czy nie. Kryteria te będą inne w zależności od celu danego testu (Douglas 2014: 68–69). Wszystkie egzaminy certyfikujące są testami sprawdzającymi.

Podział oparty na celu testowania

Ocenianie **sumujące** (*summative*) i **kształtujące** (*formative*) to kolejny podział, tym razem wynikający z celu testowania. Testy sumujące albo po prostu ocena końcowa (Rada Europy 2003: 161) przeprowadzane są najczęściej na zakończenie rozdziału podręcznika, modułu czy całego kursu i pokazują, w jakim stopniu uczący się opanował materiał. Na podstawie takiej formy ewaluacji wystawiana jest ocena końcowa. Większość testów osiągnięć reprezentuje ocenianie sumujące. Ocenianie przeprowadzane w trakcie nauczania, którego rolą jest monitorowanie bieżących postępów w nauce w celu modyfikowania procesu nauczania i uczenia się, to ocenianie kształtujące, formujące lub – za ESOKJ – ocena bieżąca (Rada Europy 2003: 160). Ocena bieżąca to ciągły proces gromadzenia wiedzy o mocnych i słabych stronach procesu uczenia się, to dane, które nauczyciel powinien wykorzystać zarówno jako pomoc w planowaniu i prowadzeniu kursu oraz jako informację zwrotną dla uczącego się.

Ocenianie obiektywne i subiektywne

Klasyfikacją budzącą pewne kontrowersje jest podział testowania na **obiektywne** (*objective*) i **subiektywne** (*subjective*). Mówiąc o obiektywnym ocenianiu i obiektywnych testach, mamy na myśli testy składające się z zadań wyboru wielokrotnego, na dobieranie oraz typu prawda–fałsz. Zadania te zaliczane są do obiektywnych ze względu na sposób ich oceniania, niewymagający od ocenianego podejmowania własnych decyzji dotyczących poprawności, braku poprawności czy też stopnia poprawności danej odpowiedzi. Zadania w ten sposób skonstruowane (zadania

zamknięte) mają z góry określony klucz odpowiedzi, a ich sprawdzanie może odbywać się bez bezpośredniego udziału człowieka. Zadania subiektywne, a więc i testy, które takie zadania zawierają, wymagają od osoby oceniającej podjęcia decyzji (oceny) według określonych kryteriów. Decyzja ta może być odmienna w przypadku, gdy to samo zadanie zostanie ocenione przez innego egzaminatora.

Dyskusyjność omawianej tutaj klasyfikacji i nazewnictwa bywa kwestionowana przez badaczy jako nietrafna (Carr 2011: 13). Tzw. obiektywne zadania nigdy do końca obiektywne nie są, ponieważ autor testu subiektywnie podejmuje decyzję, co w danym zadaniu umieścić i jakie pytania zadać. Z kolei zadania zaliczane do kategorii subiektywnych mogą być ocenione w bardzo obiektywny sposób, zakładając, że jasno sprecyzowano kryteria oceny, oceniający są do swojej roli dobrze przygotowani, a samo ocenianie przebiega zgodnie z procedurami, jest monitorowane i odpowiednio weryfikowane.

Testowanie bezpośrednie i pośrednie

Kolejną kategorią są testy **bezpośrednie** (*direct*) i **pośrednie** (*indirect*). Ogólnie rzecz ujmując, testy bezpośrednie wymagają od zdającego użycia tej umiejętności, która ma zostać zmierzona. W bezpośrednim sprawdzaniu umiejętności tworzenia wypowiedzi pisemnej oczekujemy od zdającego, aby skonstruował tekst pisany. Jeżeli natomiast sprawność pisania sprawdzana jest za pomocą zadań niewymagających od zdającego korzystania z odnośnej sprawności, a jedynie wykorzystania innych umiejętności, które tylko pośrednio wiążą się z umiejętnością przez nas pożądaną, wówczas mamy do czynienia z testami pośrednimi. W przypadku wspomnianego oceniania pisania testem pośrednim będzie sprawdzenie tej umiejętności za pomocą zadań zamkniętych mierzących znajomość gramatyki czy też poprzez dokonanie przez zdającego korekty błędów.

W momencie pojawienia się egzaminów przeprowadzanych za pomocą komputera, w których sprawność mówienia badana jest nie podczas bezpośredniej rozmowy z egzaminatorem, ale poprzez nagranie wypowiedzi zdającego, mówimy też o testach **pół-bezpośrednich** (*semi-direct*) (Carr 2011: 13–14).

Testy wykonania i ocenianie wiedzy

Testy można także skategoryzować w zależności od tego, czy sprawdzają umiejętność wykonania zadania, czy też oceniają posiadaną wiedzę. Testy umiejętności **wykonania** zadania (*performance tests / performance-referenced tests*) wymagają od zdającego wykazania się użyciem języka w mowie lub w piśmie. Ocenie podlega wypowiedź osadzona w quasi-autentycznym kontekście, zbliżonym do rzeczywistej sytuacji życiowej lub zawodowej. Ocenianie posiadanej **wiedzy** (*knowledge assessment / system-referenced tests*) wymaga od uczącego się wykazania opanowania wiedzy dotyczącej podsystemów językowych – fonologicznego, leksykalnego i gramatycznego (Nunan 2004: 141; Rada Europy 2003: 162).

Podział rodzajów oceny ze względu na miejsce ewaluacji

Na jeszcze jeden podział ewaluacji, odnoszący się w sposób szczególny do testowania specjalistycznych umiejętności zawodowych, zwracają uwagę Elżbieta Gajewska i Magdalena Sowa (2014: 204). Ze względu na miejsce, w którym ocena się odbywa, możemy mówić o **ewaluacji w terenie** oraz o **ewaluacji w środowisku instytucjonalnym**. Ewaluacja w terenie ma miejsce po zakończeniu szkolenia, kiedy to uczący się w praktyce wykonuje zadania, do których przygotowywał się podczas kursu. Ocena takiego działania odbywać się będzie na zewnątrz instytucji kształcącej w zakresie języka specjalistycznego. W momencie, kiedy ewaluacja jest integralną częścią takiego kształcenia, mamy do czynienia z ewaluacją w środowisku instytucjonalnym.

Podział testów ze względu na rangę egzaminu

Testy i egzaminy mają różne funkcje, różne jest także ich znaczenie w procesie dydaktycznym. Z ważnym rozróżnieniem mamy do czynienia, mówiąc o egzaminach powszednich oraz doniosłych (Niemierko 2006; Zawadowska-Kittel 2011). **Egzamin powszedni**, zwany także egzaminem **niskich stawek** (*low-stakes examination*), to test, „w którym znaczenie komentarza dydaktycznego jest większe niż znaczenie informacji o wyniku uczenia się, a więc kształtowanie osiągnięć przeważa nad ich sumowaniem” (Niemierko 2002: 252). Egzaminem powszednim będzie każdy sprawdzian osiągnięć, test śródroczny, diagnostyczny, którego rolą jest wskazanie np. problemów dydaktycznych, czy nawet test końcowy decydujący o promocji do kolejnej klasy bądź na kolejny stopień kursu.

Znaczenie informacji o wyniku **egzaminu doniosłego** – egzaminu **wysokich stawek** (*high-stakes examination*) – jest dużo większe. Na jego podstawie podejmowane są ważne decyzje dotyczące uczącego się: decyzja o przyjęciu do szkoły wyższego szczebla, o przyznaniu uprawnień lub dyplomu. Egzaminem doniosłym jest niewątpliwie polski egzamin maturalny, ale jest nim także np. egzamin dla pilotów i kontrolerów ruchu lotniczego, od wyniku którego zależy uzyskanie lub utrata pewnych uprawnień do wykonywania zawodu. Egzamin certyfikujący z języka obcego może dla zdającego mieć rangę egzaminu powszedniego, gdy będzie stanowił jedynie sprawdzenie i potwierdzenie własnych umiejętności; często będzie miał jednak znaczenie doniosłe, bowiem powodem przystąpienia do takiego egzaminu jest zwykle chęć uzyskania poświadczenia niezbędnego do jakiegoś dalszego celu, jak na przykład zdobycia pracy czy uzyskania awansu zawodowego.

Testowanie w podejściu zadaniowym

Testowanie powiązane z tradycyjnymi programami nauczania ma do wyboru cały wachlarz różnego rodzaju zadań egzaminacyjnych sprawdzających znajomość elementów gramatyki, leksyki, wybranych umiejętności w zakresie rozumienia tekstów oraz oceniających sprawności produktywnie. Zadania takie można również wykorzystać do testowania umiejętności zdobytych w ramach nauczania w podejściu zadaniowym. Jednak ocenianie powinno być powiązane z procesem nauczania, powinno odzwierciedlać to, co było nauczane. Jeżeli zatem nauka języka odbywała się z użyciem metodologii zadaniowej, takie właśnie podejście należy wykorzystać w testowaniu (Nunan 2004: 138).

Podejście zadaniowe w nauczaniu i ocenianiu umiejętności językowych opiera się, jak sama nazwa wskazuje, na *zadaniu*. Jest ono definiowane przez ESOKJ „jako każde celowe działanie, uważane za konieczne, by rozwiązać problem, wypełnić zobowiązanie lub realizować dążenie” (Rada Europy 2003: 21). W ocenianiu w podejściu zadaniowym (*Task-based Learning Assessment*, dalej: TBA) zdający otrzymuje zadanie, które powinien wykonać, aby osiągnąć konkretny wynik końcowy. Celem działania jest stworzenie pewnego „produktu”, którym będzie ustny lub pisemny tekst osadzony w określonym kontekście (dyskurs), np. pisemny raport wyjaśniający problem i jego możliwe rozwiązania lub ustna prezentacja tych rozwiązań przed określoną widownią w specyficznej sytuacji.

Aby ten cel osiągnąć, zdający musi wykonać szereg działań i powiązanych ze sobą różnych aktywności, np. musi przeczytać materiał pisemny lub wysłuchać wypowiedzi ustnej, bez zrozumienia których wykonanie danego zadania nie będzie możliwe (Fischer, Chouissa, Dugovičová, Virkkunen-Fullenwider 2011: 16).

Zadania wykorzystane w nauczaniu i testowaniu w podejściu zadaniowym powinny być bardzo bliskie zadaniom komunikacyjnym, z którymi zdający stykają się w rzeczywistym świecie lub z którymi zetkną się w przyszłości poza salą lekcyjną. Powinno ono symulować działanie, jakie zdający będzie musiał podjąć poza sytuacją związaną z nauką i testowaniem. W życiu codziennym pracownik nie jest rozliczany z umiejętności czytania czy rozumienia tekstu słuchanego, tylko z efektów własnej pracy wykonanej na podstawie przeanalizowanych materiałów (przeczytanych lub usłyszanych). Tak więc aby egzamin był możliwie jak najbardziej autentyczny, powinno się oceniać tylko sprawności produktywne, obudowane tekstami pisanymi i słuchanymi (Fischer, Chouissa, Dugovičová, Virkkunen-Fullenwider 2011: 27).

Według Davida Nunana (2004: 164) charakterystyczne cechy TBA, bez względu na to, w jakim kontekście odbywa się ocena zadań, to:

- ocenianie bezpośrednie (a nie pośrednie),
- ocena sprawdzająca (a nie różnicująca),
- koncentracja na osiągnięciu specyficznych celów (a nie ocena ogólnej biegłości),
- ocenianie kształtujące (a nie sumujące).

Zadaniami egzaminacyjnymi, które spełniają te kryteria i w związku z tym najlepiej wpisują się w ocenianie w podejściu zadaniowym, jest popularne w podejściu komunikacyjnym odgrywanie ról (*role play*), ale i praca projektowa (*project work*), studium przypadku (*case study*) oraz symulacja globalna (*global simulation*) (Fischer, Chouissa, Dugovičová, Virkkunen-Fullenwider 2011: 21–24).

Mimo że ESOKJ często odwołuje się do zadań i działania jako do szczególnie istotnych elementów nauki języka, to jednak nie odnosi się na swych łamach do oceniania w podejściu zadaniowym; TBA nie jest nawet wymieniony w publikacji wśród przykładowych rodzajów testowania (Fischer, Chouissa, Dugovičová, Virkkunen-Fullenwider 2011: 38).

5. Kryteria poprawności testu

Bez względu na jego rodzaj i cel każdy sprawdzian powinien spełniać szereg ogólnych zasad dotyczących prawidłowego konstruowania testów. Cztery główne kryteria poprawności testu to kryterium trafności, rzetelności, praktyczności (Douglas 2000; Komorowska 2005a, 2007; Rada Europy 2003; Weir 1993) oraz kryterium wartości dydaktycznej (Bachman i Palmer 2002; Hughes 1995; Komorowska 2007).

Trafność

Trafność testu oznacza, że sprawdza on faktycznie te umiejętności, które chcieliśmy zmierzyć, oraz że może być z powodzeniem użyty do celu, do którego został przeznaczony.

Rozróżnia się kilka rodzajów trafności. Główny podział obejmuje **trafność wewnętrzną** (*internal validity*) oraz **zewnętrzną** (*external validity*), zwaną również kryterialną (*criterion-based/criterion-related validity*). Trafność wewnętrzna związana jest z cechami testu jako takiego, zewnętrzna natomiast z relacjami tego testu w stosunku do innych testów lub sposobów pomiaru (Poszytek 2005b: 30). Trafność wewnętrzną dzielimy na:

- trafność treści,
- trafność teoretyczną, zwaną także konstrukcyjną,
- trafność fasadową, inaczej zwaną powierzchowną.

Trafność zewnętrzna dzieli się na:

- trafność prognostyczną,
- trafność diagnostyczną, zwaną też równoczesną (Komorowska 2007: 23–24).

Trafność treści (*content validity*) (Hughes 1995: 22–23) to zgodność treści testu z treściami nauczania. Test charakteryzuje się trafnością treści, jeśli sprawdza stopień osiągnięcia celów nauczania i opiera się na wprowadzonym na zajęciach materiale nauczania, nie pomija żadnych jego istotnych elementów i nie wykracza poza jego ramy.

Trafność teoretyczna lub też **konstrukcyjna** (*construct validity*) powinna być interpretowana na dwa sposoby. Po pierwsze, jest to zgodność testu z założeniami wybranej metody nauczania. Po drugie, należy rozumieć ją jako zgodność wyników w zadaniach testowych badających to samo zagadnienie w różny sposób. W celu zapewnienia trafności kon-

strukcyjnej powinniśmy zadbać, aby dana umiejętność językowa sprawdzana była za pomocą więcej niż jednego zadania i aby zadania te były budowane w różny sposób, dając w efekcie podobne wyniki (Hughes 1995: 26–27; Komorowska 2007: 24).

Test charakteryzuje się **trafnością fasadową**, zwaną także powierzchowną (*face validity*), jeżeli na osobach zainteresowanych, a więc na studentach, nauczycielach czy sponsorach, sprawia wrażenie, że bada to, co badać powinien (Gaszyńska-Magiera i Seretny 2004: 37–38; Heaton 1988: 159–160; Hughes 1995: 27). Wielu badaczy uważa ten rodzaj trafności za nieistotny z punktu widzenia konstruowania testów, przede wszystkim z powodu braku jednoznacznej procedury sprawdzania, czy test ten wymóg spełnia, czy nie (Bachman 2003: 285–289; Bachman i Palmer 2002: 42; Carr 2011: 160).

Mówiąc o **trafności prognostycznej** (*predictive/prognostic validity*) (Komorowska 2007: 24–25), mamy na myśli współzależność pomiędzy wynikami osiąganymi przez zdającego w danym teście, a wynikami, jakie uzyska w niedalekiej przyszłości w innych sprawdzianach tego samego typu. Trafność prognostyczna odgrywa istotną rolę na przykład w przypadku testów plasujących. Określając na podstawie takiego testu, w jakiej grupie zaawansowania należy umieścić danego ucznia, zależy nam, by trafność prognostyczna takiego testu była jak najwyższa (Poszytek 2005b: 31).

Trafność diagnostyczna lub **równoczesna** (*concurrent validity*) odnosi się do relacji wyników danego testu w odniesieniu do wyników innego testu przeprowadzonego mniej więcej w tym samym czasie na tej samej grupie uczniów. Ustala się ją na podstawie kryteriów zewnętrznych, leżących poza badanym testem. Można w ten sposób porównywać wyniki tych samych uczących się na testach pisanych w klasie oraz z egzaminu certyfikującego z języka obcego czy też egzaminu maturalnego z języka i egzaminu certyfikującego na takim samym poziomie (Komorowska 2007: 25; Poszytek 2005b: 31).

Chcąc mieć pewność, że testując, mierzymy te umiejętności, które były zaplanowane do zbadania, musimy podjąć pewne kroki (Komorowska 2005a: 245). Pierwszym etapem w celu zapewnienia trafności testu powinno być sporządzenie listy umiejętności, które chcemy sprawdzić. W obrębie tych umiejętności należy wyselekcjonować materiał językowy przeznaczony do zbadania, konieczne jest zatem sporządzenie listy wyrazów i struktur gramatycznych, które będą sprawdzane za pomocą

testu. Każda umiejętność z powyższej listy powinna być ujęta w teście i każdej z nich winna odpowiadać grupa zadań konkretnie na nią ukierunkowanych. Trzeba również dopilnować, aby żadne zadanie nie badało umiejętności nieujętych we wspomnianym wykazie. Następnym etapem jest przygotowanie zadań testowych. W zależności od wagi danej umiejętności powinna jej być przyporządkowana odpowiednia liczba zadań. Najważniejszym umiejętnościom powinna odpowiadać największa liczba zadań. Umiejętności mniej istotne powinny zajmować odpowiednio mniejszą proporcję testu. Zgodnie z zasadami konstruowania testów każde oddzielne zadanie powinno sprawdzać inną umiejętność. Jeżeli ktoś z nich sprawdza więcej niż jedną umiejętność, należy zadbać, by inne umiejętności sprawdzane przez dane zadanie nie okazały się trudniejsze od tych wytypowanych do zmierzenia. Na ostatnim etapie powinniśmy sprawdzić, czy test obejmuje cały zaplanowany materiał językowy i czy nie wprowadza materiału spoza listy, zbyt trudnego lub nieadekwatnego dla uczniów na danym poziomie. Po zapewnieniu powyższych cech możemy sądzić, że test będzie spełniał kryterium trafności.

Rzetelność

Rzetelność odnosi się do dokładności pomiaru i niezawodności testu. Jest to termin techniczny, określający, w jakim stopniu można polegać na wynikach testu. O rzetelnym teście możemy mówić, jeżeli w przypadku powtórnego jego przeprowadzenia uszeregowanie wyników od najwyższych do najniższych nie zmienia się, a więc zdający, którzy przy pierwszym podejściu uzyskali najlepsze wyniki, osiągną najwyższe wyniki przy drugiej próbie i na odwrót – zdający z najniższymi wynikami przy pierwszym podejściu będą mieć najslabsze wyniki w tej samej grupie zdających podczas wykonywania testu ponownie.

O rzetelności mówimy także w przypadku oceny testu. W tym podejściu chodzi o ocenę tego samego testu przez kilku egzaminatorów. Chcąc uznać test za spełniający kryterium rzetelności szacowanej na podstawie zgodności oceny sędziów, oczekujemy, że będzie on oceniony w taki sposób, by ewentualna różnica w ocenie pomiędzy różnymi egzaminatorami była jedynie minimalna albo by nie było jej w ogóle.

Test nie może być trafny, jeśli nie jest rzetelny. Jeżeli pomiar nie jest spójny, nie może być równocześnie dokładny. Z drugiej strony, możliwe jest, by rzetelny test nie spełniał w tym samym czasie warunków trafności. Sprawdzan może dawać rzetelne wyniki, takie same przy każdej kolejnej

próbie, jednocześnie może być całkowicie nietrafny, jeśli nie mierzy tego, co w założeniu miał badać. Tak więc, należy pamiętać, że choć rzetelność jest niezbędnym czynnikiem, by test był trafny, to jednak jest to zależność niewystarczająca. Zdarza się, że w celu zmaksymalizowania rzetelności testu musimy pogodzić się z obniżeniem poziomu jego trafności. Przykładem takiego kompromisu są zadania typu wybór wielokrotny, które z reguły mają wysoki poziom rzetelności, jednak nie mierzą użycia języka w realnych życiowych sytuacjach, co czyni je mniej trafnymi (Alderson, Clapham, Wall 1995: 187; Hughes 1995: 42).

W celu zachowania maksymalnego stopnia rzetelności testu, Anthony (2018) sugeruje przestrzeganie kilku zasad. Należy więc:

- przeprowadzać równorzędne czynności ewaluacyjne, zachowując takie same warunki (np. czas trwania, warunki lokalowe),
- używać czytelnego i jednoznacznego formatu i układu testu,
- formułować instrukcje i polecenia w sposób jasny i precyzyjny,
- kodować prace zdających w celu uniknięcia tendencyjności w ocenianiu,
- używać jasnych i zrozumiałych kryteriów oceny,
- upewnić się, że zdający znają formę testu, typy zadań testowych oraz kryteria oceny,
- zawrzeć w arkuszu urozmaicone jednostki testowe, które w rozmaity sposób zmierzają różne aspekty sprawdzanej wiedzy i umiejętności,
- ograniczyć dowolność interpretacji sposobu odpowiedzi przez zdających,
- spowodować, by zdający rozumieli znaczenie i wartość danej oceny i podeszli do testu z adekwatną powagą,
- przeprowadzić proces oceny odpowiedzi zdających w jednolity sposób (Anthony 2018: 127).

Praktyczność

Praktyczność testu to cecha związana z nakładem czasu, sił i kosztów koniecznych do jego przeprowadzenia, które powinny być współmierne do rezultatów, jakie dany test ma przynieść. Cecha ta może zdecydować, czy dany test znajdzie zastosowanie w praktyce, czy nie. Trafność i rzetelność testu przy braku jego praktyczności nie są wiele warte. W konsekwencji „w starciu [...] kryterium praktyczności z kryterium trafności w egzaminach językowych ta ostatnia niestety przegrywa” (Komorowska 2005b: 48).

Bachman i Palmer (2002: 36) ujmują praktyczność w formule:

$$\text{praktyczność} = \frac{\text{środki dostępne}}{\text{środki potrzebne}}$$

W przypadku gdy wynik takiego równania jest równy lub wyższy od 1, test można uznać za praktyczny; jeśli wartość ta jest mniejsza od 1, wówczas mówimy o braku praktyczności. Środki uwzględnione w tej formule można podzielić na trzy rodzaje:

- zasoby ludzkie,
- środki materialne,
- czas.

Zasoby ludzkie zaangażowane w przygotowanie i przeprowadzenie testu to na przykład konstruktorzy testów, osoby organizujące egzamin, egzaminatorzy, osoby nadzorujące prawidłowy przebieg egzaminu oraz osoby oceniające. Środki materialne odnoszą się do miejsca, w którym egzamin miałby się odbywać, ale także miejsca przeznaczonego na prace związane z przygotowaniem i oceną testu; to także sprzęty i wyposażenie (komputery, odtwarzacze nagrań audio, kamery do nagrywania egzaminu lub dyktafony itp.) oraz materiały, czyli na przykład papier do wydruku arkuszy, zdjęcia/obrazki itp. Ważnym środkiem jest również czas włożony w przygotowanie, przeprowadzenie oraz ocenę i analizę testu (Bachman i Palmer 2002: 37). Mówiąc o czasie jako o czynniku mającym wpływ na to, czy dany test jest praktyczny, Cyril J. Weir (1990: 34) dodaje jeszcze potencjalne zmęczenie zdających, wynikające z długiego czasu trwania ewaluacji.

Wartość dydaktyczna

Wartość dydaktyczna testu (Komorowska 2007: 30–31), zwana też efektem zwrotnym (*backwash/washback effect*) (Carr 2011:54–55; Hughes 1995) czy oddziaływaniem (*impact*) (Bachman i Palmer 2002: 29–35), to wpływ, jaki dany test ma na proces nauczania i uczenia się. Sposób kontroli i oceny efektów nauczania nigdy nie pozostaje bez znaczenia dla tych procesów. Oczywiście jest, że chcielibyśmy, aby każdy test miał pozytywny wpływ na naukę i umiejętności zdającego, jednak efekt zwrotny może również być negatywny.

Według Bachmana i Palmera test oddziałuje na dwóch płaszczyznach: na mikroplaszczyźnie egzaminy mają wpływ na poszczególne osoby biorące udział w testowaniu, czyli na uczniów i na nauczycieli;

na makroplaszczyźnie natomiast oddziałują one na system edukacji i społeczeństwo.

Wpływ na ucznia wiąże się z przygotowywaniem się do testu, z doświadczeniem związanym z jego zdawaniem, z informacją zwrotną dotyczącą osiągnięć zdającego w postaci wyniku egzaminu i decyzjami podejmowanymi na podstawie osiągniętych wyników (Bachman i Palmer 2002: 31). W przypadku nauczycieli oddziaływanie uwidacznia się w tzw. nauczaniu pod test, czyli podporządkowaniu działań dydaktycznych wymaganiom konkretnego testu, nawet jeśli nie jest to do końca zgodne z preferencjami nauczyciela dotyczącymi sposobów nauczania czy też z samym programem nauczania. Dzieje się tak bardzo często w przypadku nauczania szkolnego, które kończy się egzaminem zewnętrznym, lub nauki na kursie przygotowującym do konkretnego egzaminu certyfikującego.

Jednak wpływ testu nie ogranicza się tylko do osób bezpośrednio zainteresowanych. Ma on również szeroki charakter społeczny. Oddziaływanie testu na makroplaszczyźnie ma ogromne znaczenie zwłaszcza w przypadku tzw. egzaminów doniosłych (*high-stakes tests/examinations*), których wyniki stanowią podstawę do podejmowania istotnych decyzji (np. dotyczących dalszego kształcenia lub pracy). Przykładem takich egzaminów jest matura albo egzaminy certyfikujące z języków obcych: oddziałują one nie tylko na uczniów i nauczycieli, ale na wymagania, treści oraz metodykę programów nauczania, a także na cały system edukacji.

6. Zasady konstruowania testów

O konstruowaniu testów językowych pisze wielu autorów, w tym Carr (2011), Hughes (1995), Komorowska (2007), Weir (1990, 1993). Konstruowanie testów LSP omawia szczegółowo Douglas (2000: 92–282). Dwa pierwsze etapy tworzenia testu wyróżnione przez wszystkich wspomnianych autorów to planowanie testu oraz budowa zadań. Kolejne kroki (różnie nazwane i skategoryzowane) wiążą się z punktacją i ocenianiem, pilotowaniem pierwszych wersji testu oraz monitorowaniem wyników jego kolejnych wersji. W tej części omówione zostaną poszczególne etapy tworzenia testów.

Projektowanie testu

Pierwszym krokiem w planowaniu testu jest ustalenie, co chcemy zbadać i w jakim celu. Zakładając, że test ma być sprawdzianem komunikacyjnym, a takim powinien być każdy test LSP, zanim zaczniemy planować szczegóły dotyczące konstrukcji testu, rodzaju zadań i mierzonych umiejętności, musimy ustalić, w jakich docelowych sytuacjach komunikacyjnych adresat egzaminu będzie potrzebował używać języka obcego, jak również – jakie konkretne umiejętności językowe będą mu do tego potrzebne. W tym celu należy zebrać informacje dotyczące:

- czynności – zadań, w jakich zdający będą uczestniczyć w docelowych sytuacjach komunikacyjnych,
- kontekstu sytuacyjnego, czyli fizycznego i psychofizycznego kontekstu, w którym zdarzenia komunikacyjne mają miejsce,
- interakcji, a więc roli i relacji społecznych, w których komunikacja się odbywa,
- instrumentów, na które składają się środki, sposoby i kanały komunikacji,
- umiejętności szczegółowych, które umożliwią użytkownikowi języka skuteczne funkcjonowanie podczas wykonywania wspomnianych czynności (Weir 1990: 36–37).

Aby odpowiedzieć sobie na pytanie, co chcemy zbadać i po co, należy najpierw wykonać **analizę potrzeb** adresatów testu. Analiza potrzeb jest krokiem niezbędnym niezależnie od rodzaju planowanego testu, jest ona jednak szczególnie istotna w glottodydaktyce specjalistycznej. Na analizę potrzeb składa się szereg różnych czynności związanych ze zbieraniem informacji, a następnie badanie oraz interpretacja uzyskanych danych. Celem takiej analizy jest ustalenie, czego powinno się nauczać na specjalistycznym kursie językowym, a w rezultacie – co i ewentualnie jak powinno być testowane w ramach badania biegłości językowej i oceny umiejętności docelowej grupy zawodowej. Element ten jest konieczny, aby pomóc osobom zaangażowanym w proces kształcenia językowego zrozumieć, w jakim celu oraz w jakich okolicznościach uczący się języka będzie go używał. Chodzi o to, by dać twórcom programów i materiałów dydaktycznych oraz autorom egzaminów obraz docelowych sytuacji zawodowych ucznia (ALTE 2018: 7–8; Dudley Evans i St John, 1998: 140–144; Flowerdew 2013; Frendo 2005; Gajewska 2013; Gajewska i Sowa 2014: 115–137; Hutchinson i Waters, 1991: 53–63; Janowska 2011: 24–26; Ligara i Szupelak 2012: 107–114; Long 2005a, 2005b; Sendur 2017).

Gdy wiemy już, w jakim celu potencjalni zdający będą poddawać swoje umiejętności badaniu za pośrednictwem naszego testu i jakie są ich specyficzne potrzeby użycia języka obcego w sferze zawodowej, możemy przejść do opracowania **specyfikacji dla testu** albo **koncepcji testu** (Niemierko 1997: 48–49), która w przypadku testów szerokiego zasięgu wymaga formy dokumentu pisemnego.

Cel testu, opis jego adresatów i opis docelowych sytuacji użycia języka to jedne z podstawowych informacji, które powinny się w takiej specyfikacji znaleźć. Należy zdefiniować konstrukt, który ma zostać poddany zmierzeniu, i dokonać opisu treści testu, jego formatu, rodzajów i typów zadań testowych wraz z przykładami, kryterium zaliczenia i procedurą punktowania. Zebrane informacje stworzą **plan**, zwany też **kartoteką testu**.

Cel sprawdzianu, jego adresaci i docelowe sytuacje użycia języka są ze sobą ściśle powiązane i wzajemnie z siebie wynikają. Cel testu determinuje jego odbiorcę, który z kolei usadowiony jest w określonych sytuacjach zawodowych. Efektem, do którego dążymy, może być na przykład certyfikacja językowa konkretnej grupy zawodowej. Mówiąc o adresatach egzaminu, pomocny jest bardziej precyzyjny opis niż tylko podanie branży zawodowej, do której należą: mogą się tu znaleźć informacje dotyczące przeciętnego wieku zdających, ich narodowości, języka ojczystego oraz poziomu wykształcenia. Dane te pomogą ustalić typ osób, z którymi egzaminowany może wchodzić w interakcje językowe. Opis docelowych sytuacji użycia języka może dotyczyć miejsca, w którym potencjalna interakcja miałaby się odbywać, materiałów, narzędzi i urządzeń wykorzystywanych w pracy oraz rodzajów zadań komunikacyjnych, w których zdający mogą się znaleźć w sytuacjach zawodowych.

Mówiąc o **treściach** testu (Douglas 2000: 111; Hughes 1995: 49), mamy na myśli wszystkie potencjalne treści wszystkich arkuszy danego testu, nie zaś treść konkretnego, pojedynczego arkusza czy też pojedynczej wersji testu. W zależności od rodzaju testu oraz tego, co chcemy za jego pośrednictwem sprawdzić, inne będą jego treści. Na treści testu składają się operacje, rodzaje tekstów, adresaci tekstu, tematyka, a także organizacja arkusza, w tym liczba zadań, krótka jego charakterystyka, przewidywany na wykonanie zadań czas oraz długość tekstów.

Operacje (Hughes 1995: 49) to konkretne zadania egzaminacyjne, które zdający będzie musiał wykonać. Może to być na przykład odszukanie informacji w tekście, odgadnięcie znaczenia słów z kontekstu czy

zrozumienie ogólnego sensu całego tekstu. Zadania testowe powinny wynikać bezpośrednio z analizy docelowych sytuacji zawodowych oraz z definicji konstruktów. Z różnymi rodzajami tekstów mamy do czynienia zarówno w zadaniach mierzących sprawność czytania, słuchania, jak i pisania. W przypadku sprawności receptywnych będą to teksty, z którymi zdający muszą się zapoznać, które muszą zrozumieć i przetworzyć podczas wykonywania zadania. W zadaniach sprawdzających rozumienie czytania może to być tekst publicystyczny bądź naukowy, natomiast w zadaniu na rozumienie ze słuchu możemy spotkać komunikat czy też fragment wiadomości radiowych. W testach sprawdzających umiejętność pisania przykładowym gatunkiem może być raport, notatka bądź artykuł akademicki. Na egzaminie ustnym także pojawiają się różne rodzaje wypowiedzi, a więc różne rodzaje tekstu, przykładem niech będzie ustna prezentacja lub udział w dialogu.

Adresatami tekstu są osoby, z którymi kandydat powinien umieć się porozumieć – czy to w formie ustnej, czy pisemnej – lub też osoby, do których kierowany jest dany rodzaj tekstu pisanego bądź słuchanego. Tematyka tekstów i zadań powinna być dobrana odpowiednio do charakteru danego testu oraz jego odbiorcy. Na tym etapie należy również dokonać wyboru tych jednostek leksykalnych, elementów gramatyki i funkcji językowych, które wiążą się ze specyficznymi potrzebami potencjalnego adresata egzaminu.

Kolejnym elementem, który należy wziąć pod uwagę pisząc specyfikację, powinien być **format** testu. Określona winna być jego struktura wraz z czasem przeznaczonym na wykonanie całego arkusza oraz poszczególnych jego części, jak również rodzaje i przykłady zadań egzaminacyjnych. Specyfikacja powinna także uwzględniać liczbę i długość tekstów, na których będą oparte zadania sprawdzające umiejętność czytania i słuchania, oraz liczbę i rodzaje wypowiedzi, które zdający będzie musiał stworzyć w części sprawdzającej umiejętność pisania, jak również liczbę poszczególnych jednostek w każdej wiązce zadań.

Należy także ustalić pułap, który zdający powinien przekroczyć, aby można było uznać test za zdany. Może to być minimalna liczba punktów lub wynik procentowy, który należy uzyskać, albo opisowe kryteria, które praca zdającego powinna spełnić, aby mogła otrzymać pozytywną ocenę. Ścisłe związany z kryteriami oceny jest sposób punktowania zadań. Plan testu powinien określić, jaki procent wyniku punktowego całego egzaminu zajmuje dana sprawność i ile punktów można uzyskać za każde poje-

dyncze zadanie. Ustalamy w ten sposób wagę poszczególnych części testu i konkretnych sprawdzanych umiejętności.

Stworzony w efekcie dokument – specyfikacja dla testu – posłuży jako przewodnik do tworzenia arkusza egzaminacyjnego, wyposaży jego autorów w parametry niezbędne do opracowania zadań oraz zapewni egzaminatorom i oceniającym wytyczne potrzebne do przeprowadzenia oceny.

Tworzenie arkusza testu

Przechodząc do kolejnego etapu, jakim jest budowanie arkusza testowego, należy przeprowadzić **selekcję treści**, które chcemy uwzględnić w konkretnym egzaminie. Mając świadomość tego, że nie da się zamieścić w jednym arkuszu wszystkich treści wytypowanych na etapie tworzenia specyfikacji, niezbędne jest dokonanie przemyślanego wyboru poszczególnych jej elementów. Podczas selekcionowania treści powinno się wziąć pod uwagę szereg istotnych zagadnień. Według Weira (1990: 38) charakterystyczne cechy testów komunikacyjnych, które winny być uwzględnione przy dokonywaniu selekcji, to:

- realistyczny kontekst,
- odpowiednia luka informacyjna,
- intersubiektywność,
- zakres rozwinięcia aktywności przez zdającego,
- umożliwienie zdającemu monitorowania własnej efektywności komunikacyjnej,
- przetwarzanie odpowiednio zaaplikowanych danych wejściowych,
- odpowiednie ramy czasowe.

Realistyczny kontekst, o którym pisze Weir, to umiejscowienie zadań testowych w sytuacji, w której zdający może się znaleźć poza sytuacją ewaluacyjną. Odpowiednia luka informacyjna odnosi się do konieczności przetworzenia nowych informacji w taki sposób, jak by to miało miejsce w sytuacji rzeczywistej. Intersubiektywność zadań w testach komunikacyjnych oznacza, że zdający powinien być zaangażowany w zadanie zarówno jako odbiorca komunikatu, jak i jego nadawca. Mówiąc o zakresie rozwinięcia aktywności przez zdającego, Weir ma na myśli fakt, że zadania powinny dawać zdającemu możliwość wykazania się komunikacyjną samodzielnością, a zasady punktowania zadań uwzględniać kreatywną nieprzewidywalność zdającego. Zadania powinny umożliwić zdającemu używanie strategii dyskursu w celu oceny własnej efektywności i pozwolić na dokonanie niezbędnych modyfikacji w akcie komunika-

cji. Zakres i wielość danych do przetworzenia przez zdającego powinna być porównywalna z tą, którą zdający mógłby przetworzyć w rzeczywistej sytuacji użycia języka. Taka sama zasada dotyczy czasu przeznaczanego na wykonanie zadania.

Po dokonaniu selekcji treści można przejść do **budowania zadań testowych**. Czynność ta jest ogromnym wyzwaniem dla autorów testu. Hughes (1995) zaleca, aby odbywała się w formie pracy grupowej. Spojrzenie na dane zadanie z różnego punktu widzenia przez różne osoby pozwala dostrzec błędy i problemy, które autor może przeoczyć lub sobie ich w ogóle nie uświadamiać. Zaangażowanie w pracę nad arkuszem całego zespołu osób nie jest jedynym wymogiem: opracowane zadania powinny zostać poddane weryfikacji poprzez zlecenie wykonania go rodzimym użytkownikom języka oraz osobom, które w przyszłości będą podobny test rozwiązywać. Wynik takiej próby powinien pokazać, które z zadań wymagają udoskonalenia lub całkowitej eliminacji. „Przetestowanie testu” na takiej grupie osób pomoże także w opracowaniu **klucza** odpowiedzi do zadań, który właśnie na tym etapie powinien powstać.

Wspomniana powyżej weryfikacja zadań nazywana jest przez Hughesa (1995) **moderacją**, w przeciwieństwie do **pilotażu**, będącego kolejnym, już prawie ostatnim etapem tworzenia arkusza testu. Wprawdzie przeprowadzanie pilotażu testu na odpowiedniej liczbie osób często bywa niemożliwe lub niewykonalne, stanowi jednak bardzo pożądany krok przed wprowadzeniem arkusza egzaminacyjnego do formalnego użytku, może bowiem wyeliminować przeoczone lub nieuświadomione wcześniej usterki, jak również pozwolić na wykonanie analiz statystycznych, których wyniki powinny być wzięte pod uwagę przy ostatecznym kształcie arkusza.

Wdrożenie i monitorowanie testu

Następne dwa etapy tworzenia testu wg Weira (1990: 40–41) to jego wdrożenie i monitorowanie. Wdrożenie wiąże się z graficznym i technicznym przygotowaniem arkusza, jego drukiem, dystrybucją, informacją dla kandydatów, instytucji, egzaminatorów i osób nadzorujących egzamin. Monitorowanie natomiast dotyczy gromadzenia i analizowania wyników testu oraz kolejnych jego wersji, analizy odbioru produktu przez rynek edukacyjny – przez instytucje, których egzamin dotyczy, jak i przez samych zdających, czy też wpływu testu na nauczanie (*backwash effect*) itp. Oczywiście czynności te dotyczą jedynie egzaminów przeprowadzanych

na większą skalę. Ponieważ jednak ten etap pracy nad testem leży poza obrębem tematyki niniejszej publikacji, nie będzie tutaj szerzej rozwijany.

7. Rodzaje zadań testowych i ich charakterystyka

Według Wincentego Okonia (2007: 481) zadanie testowe to „najmniej-
sza, względnie niezależna, częśćka testu, wymagająca od ucznia udziele-
nia odpowiedzi; może mieć postać polecenia, pytania, wypowiedzi nie-
kompletnej lub twierdzenia poddawanego w wątpliwość”.

W przypadku testowania języków obcych mówimy najczęściej o zada-
niach pisemnych i ustnych. Dzieląc testy ze względu na sposób udzielania
odpowiedzi, Niemierko (1999: 55) wyróżnia testy ustne, pisemne i prak-
tyczne. Na podstawie tej klasyfikacji dokonuje również podziału zadań
testowych na pisemne, ustne i praktyczne (1999: 56 i 78–129). W odnie-
sieniu do języka obcego z praktycznymi zadaniami testowymi mamy do
czynienia w omówionym dalej podejściu zadaniowym.

W dalszej części tego rozdziału przedstawiono rodzaje zadań testo-
wych z podziałem na różne formy wypowiedzi, opisano ich charaktery-
styczne cechy oraz zaprezentowano przykłady konkretnych typów zadań
wykorzystanych w egzaminach certyfikujących z języka ogólnego oraz
z egzaminów językowych dla celów specjalistycznych.

W analizie uwzględniono szereg egzaminów, które dla celów niniej-
szego opracowania zostały poddane głębszemu badaniu – wszystkie
można zaliczyć do kategorii egzaminów doniosłych. Jest to sześć testów
z ogólnego języka angielskiego oraz szereg egzaminów zawodowych.
W przypadku, gdy w w/w arkuszach nie znaleziono przykładów na okre-
ślony rodzaj zadania, odwołano się również do polskich egzaminów ze-
wnętrznych, w których takie zadania są obecne.

Wybór testów do analizy jest całkowicie autorski. Dokonano go prze-
siewowo, metodą selekcji spośród całej gamy egzaminów dostępnych na
polskim rynku edukacyjnym, ale uwzględniając również egzaminy prze-
znaczone dla różnych odbiorców, które wydawały się godne podniesienia
w kontekście badanego tematu. Starano się wybrać przykłady reprezen-
tatywne dla rozmaitych dziedzin. Wzięto pod uwagę również takie kryteria,
jak popularność oraz pewne cechy szczególne danego egzaminu. W do-
borze sugerowano się także różnym stopniem specjalistyczności, tak aby
zapewnić przykłady od bardzo ogólnych do skrajnie specjalistycznych.

Wybierając certyfikaty z języka ogólnego, starano się dokonać selekcji popularnych egzaminów różniących się między sobą. Wśród analizowanych egzaminów znajduje się tradycyjny test z długoletnią tradycją (*Cambridge B2 First*), stworzony przez instytucję zajmującą się testowaniem od ponad stu lat (Cambridge English Language Assessment), jak również bardzo nowoczesny adaptacyjny test komputerowy oparty na technologii sztucznej inteligencji (*Linguaskill*).

Egzaminy dla celów zawodowych zostały dobrane tak, by reprezentować rozmaite branże i różny stopień specjalistyczności. Język medyczny reprezentowany jest podwójnie – przez egzamin z tradycjami sięgającymi połowy lat 80. XX wieku, stworzony na zlecenie rządu Australii, by zaspokoić potrzebę potwierdzania umiejętności językowych imigrantów legitymujących się dyplomami wybranych zawodów medycznych uzyskanymi w kraju nieanglojęzycznym (OET), oraz przez dużo młodszy efekt międzynarodowego projektu opracowanego pod auspicjami i z finansowaniem Komisji Europejskiej (sTANDEM). Pojawiają się również dwa różne egzaminy z języka technicznego (MOTET oraz *telc B2 Technical*). Dla innych branż zawodowych wybrano po jednym egzaminie. W niektórych przypadkach test kierowany jest do całej branży, a nie do konkretnego zawodu (*telc B2 Technical*, EfT, STANAG 6001, *Linguaskill Business*), innym razem w ramach jednego egzaminu mamy do czynienia z oddzielnymi wersjami dla poszczególnych specjalizacji wchodzących w skład danej branży (sTANDEM, OET, ELPAC).

W analizie uwzględniono następujące egzaminy certyfikujące z ogólnego języka angielskiego:

- *Cambridge B2 First* (dawniej: *First Certificate in English*, obecnie używana także skrócona nazwa: *First*; Cambridge English Language Assessment 2015b);
- *telc Dual-Level* na poziomach B1–B2 (*telc Language Tests 2012a* i *2012b*; *telc Language Tests 2014*);
- *telc English B2* (*telc Language Tests 2019*);
- *PTE General – Pearson Test of English* (*PTE General 2018*);
- IELTS (IELTS 2019);
- *Aptis General* (British Council 2020; *Prepare for Aptis*²);
- *Linguaskill General* (Cambridge Assessment English 2019c; Ludlow 2020).

² Zob. <https://www.britishcouncil.org/exam/aptis/practice-materials> [dostęp: 16.06.2020].

Poddane analizie testy certyfikujące z języków do celów specjalistycznych to:

- egzamin z języka angielskiego medycznego sTANDEM (*Standardised Language Certificate for Medical Purposes*; The sTANDEM Consortium 2013 i www.standem.eu);
- *Occupational English Test* (OET) dla branż medycznych (OET 2009, 2014a i 2014b; www.occupationalenglishtest.org)³;
- egzamin dla wojska STANAG 6001 poziom 3⁴;
- egzaminy dla pilotów i kontrolerów ruchu lotniczego *English Language Proficiency for Aeronautical Communication* (ELPAC; Eurocontrol 2013 i Decyzja Nr 40⁵);
- z technicznego języka angielskiego – *MONDIALE Technical English Test* (MOTET)⁶;
- z technicznego języka angielskiego – *telc English B2 Technical* (telc Language Tests 2016);
- z języka angielskiego dla księgowych – *English for Accounting*⁷ (EfA);
- z biznesowego języka angielskiego – *Linguaskill Business* (Cambridge Assessment English 2017; Cambridge Assessment English 2019c; Ludlow 2020);
- z biznesowego języka angielskiego na poziomie C1 – *Business Higher* (BEC Higher; Business English Certificate Higher 2015; Cambridge Assessment English 2019b; www.cambridgeenglish.org)⁸;

³ Occupational English Test (online), <https://www.occupationalenglishtest.org/test-information> [dostęp: 16.05.2020].

⁴ NATO Standardization Agency (2009) *Standardization Agreement. STANAG 6001 NTG (edition 3) – Language Proficiency Levels* (online), http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/Porozumienie_standaryzacyjne_STANAG_6001_edycja_3_angielski.pdf [dostęp: 16.05.2020] oraz *Porozumienie Standaryzacyjne NATO nr 6001 – STANAG 6001* (online), http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/Porozumienie_standaryzacyjne_STANAG_6001_edycja1.pdf [dostęp: 16.05.2020].

⁵ *Decyzja Nr 40 Prezesa Urzędu Lotnictwa Cywilnego z dn. 12 listopada 2014 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla pilotów i kandydatów na pilotów* (online), http://www.ulc.gov.pl/_download/personel_lotniczy/lkeang/Materials/decyzja_nr_40_prezesa_ulc.pdf [dostęp: 16.05.2020].

⁶ Mondiale Testing (online), <http://www.mondiale-testing.de/site/content/view/90/72/lang.en/> [dostęp: 23.02.2017].

⁷ Pearson. *English for Accounting* (online), <http://qualifications.pearson.com/en/qualifications/lcci/english-language/english-for-accounting.coursematerials.html#filter-Query=category:Pearson-UK:Category%2FSpecification-and-sample-assessments> [dostęp: 23.02.2017].

⁸ Zob. <http://www.cambridgeenglish.org/exams/business-certificates/business-higher/exam-format/> [dostęp: 5.05.2020].

- z języka angielskiego w branży turystycznej – *English for Tourism*⁹ (EfT);
- seria egzaminów z języka angielskiego skierowanych do wielu różnych zawodów – *Vocational English Certificate* (VEC)¹⁰.

Posiłkowano się również informacjami na temat egzaminów lektoratowych przeprowadzanych w polskich wyższych uczelniach (Sendur 2016b). W kilku przypadkach – tam, gdzie dany rodzaj zadania nie jest reprezentowany w międzynarodowym egzaminie GE lub ESP, przytoczono także przykłady polskich państwowych egzaminów zewnętrznych z języków obcych (egzamin ósmoklasisty, egzamin gimnazjalny, matura, egzaminy eksternistyczne z zakresu szkoły podstawowej, gimnazjum i liceum ogólnokształcącego¹¹). Niektóre z wymienionych tutaj egzaminów zostały bardziej szczegółowo omówione w rozdziale III.

Pisemne zadania testowe

Zadanie testowe, wg Niemierki (1975: 28), jest „najmniejszą niezależną częstką testu wymagającą od ucznia udzielenia odpowiedzi”. Ten sam autor (1975: 30–43; 1999: 56) dzieli pisemne zadania testowe ze względu na sposób komunikowania odpowiedzi przez ucznia lub zdającego egzamin. Rodzaje i formy zadań zestawia w następujący sposób:

Tabela 1. Typologia pisemnych zadań testowych za: Niemierko (1999)

Rodzaj	Forma
Zamknięte	na dobieranie (D)
	wyboru wielokrotnego (WW)
	prawda–fałsz (PF)
Otwarte	rozszerzonej odpowiedzi (RO)
	krótkiej odpowiedzi (KO)
	z luką (L)

⁹ *LCCI International Qualifications. Level 2 Certificate in Written English for Tourism. Syllabus – Effective for examinations to be held from Series 2, 2010* (online), <http://qualifications.pearson.com/content/dam/pdf/LCCI/Spoken-English-for-Tourism-WEFT/2010/specification-and-sample-assessments/WEFT-Syllabus-Level-2-April-2012.pdf> [dostęp: 4.10.2016].

¹⁰ Zob. <https://www.interveng.com/i-want-to-find-exam-information> [dostęp: 14.06.2020].

¹¹ Centralna Komisja Egzaminacyjna (online), www.cke.edu.pl.

Zadania zamknięte to zadania wymagające wybrania jednej bądź kilku z gotowych odpowiedzi. Zadanie **wyboru wielokrotnego**¹² (WW) to rodzaj zadania zamkniętego wymagającego wybrania jednej lub większej liczby odpowiedzi spośród kilku podanych w zadaniu. Odpowiedziami mogą być wyrazy, zdania, zwroty, wyrażenia matematyczne, symbole, rysunki itp. Zadanie **na dobieranie** (D) jest rodzajem zadania zamkniętego wymagającego poprawnego zestawienia ze sobą dwóch informacji (najczęściej pojęć, definicji, liczb itp.). Rozwiązanie zadania **typu prawda-falsz** (PF) polega natomiast na rozstrzygnięciu, czy zawarte w nim twierdzenie jest prawdziwe, czy fałszywe lub czy spełnia ono określone kryterium, np. poprawności (Centralna Komisja Egzaminacyjna 2008).

Zadania otwarte wymagają samodzielnego sformułowania odpowiedzi przez zdającego. Zadanie **rozszerzonej odpowiedzi** (RO) jest rodzajem zadania otwartego wymagającego rozwiniętej, wieloelementowej i odpowiednio uporządkowanej odpowiedzi w postaci słownej (np. wypracowania), złożonych wyrażeń matematyczno-fizycznych, reakcji chemicznych, schematów, procesów itp. (Do tej grupy zadań zalicza się zarówno zadania pisemne, jak i wypowiedzi ustne.) Zadanie **krótkiej odpowiedzi** (KO) to rodzaj zadania otwartego wymagającego udzielenia zwięzłej odpowiedzi w postaci jednego słowa, liczby, symbolu, zdania lub kilku zdań, wyrażenia matematycznego, prostego rysunku, wyliczenia elementów itp. Zadanie **z luką** (L) wymaga od zdającego wstawienia w pozostawioną przez konstruktora lukę wyrazu, pojęcia, związku wyrazowego, liczby itp. (Centralna Komisja Egzaminacyjna 2008).

Typologia Niemierki odnosi się do wszystkich pisemnych zadań testowych niezależnie od przedmiotu testowania. Znajduje ona zastosowanie również w odniesieniu do zadań w testach językowych. Terminologia ta stosowana jest bardzo często do klasyfikacji zadań przez instytucje zajmujące się testowaniem w Polsce oraz w polskich publikacjach dotyczących tego zagadnienia. Bardziej szczegółową klasyfikację zadań wykorzystywanych w badaniu umiejętności językowych przedstawia stowarzyszenie Association of Language Testers in Europe (ALTE) w opracowaniu *Materials for the Guidance of Test Item Writers* (ALTE 2005). Poniżej

¹² W języku powszechnie używanym zadania tego typu nazywane są *zadaniami wielokrotnego wyboru*, jednak zarówno Niemierko (1975, 1997, 1999), jak i Gaszyńska-Magiera i Seretny (2004) stosują w odniesieniu do tych zadań nazwę *wybór wielokrotny*. Na konieczność stosowania prawidłowej terminologii zwracają również uwagę pracownicy CKE i OKE w szkoleniach i innych pracach dot. egzaminów.

przedstawiono typy zadań zaprezentowanych w klasyfikacji stowarzyszenia ALTE¹³ wpisane w typologię Niemierki.

Zamknięte jednostki testu (*selection items*):

- wybór wielokrotny
 - zadania typu wybór wielokrotny (*multiple choice*)
 - zadania z lukami z wyborem wielokrotnym (*gap-filling/cloze passage with multiple choice options*)
 - wykrywanie błędów/usuwanie zbędnych wyrazów (*extra word error detection*)
- na dobieranie
 - dobieranie/dopasowywanie (*matching*)
 - wypełnianie luk na podstawie selekcji z banku możliwości (*gap-filling with selection from bank*)
 - zadanie na dobieranie wielokrotne (*multiple matching*)
 - uzupełnianie luk na poziomie akapitu (*gap-filling at paragraph level*)
- prawda–fałsz
 - zadania typu prawda/fałsz (*true/false item*).

Otwarte jednostki testu (*candidate-supplied response item types*):

- z luką
 - zadanie typu wypełnianie luk otwartych (*open gap-filling/cloze*)
 - przekształcenia fleksyjne (*transformations*)
 - zadanie typu przekształcenia z lukami (*transformation cloze*)
 - przekształcenia słowotwórcze (*word formation*)
 - poprawa błędów/czytanie w celu dokonania korekty (*error correction/proof reading*)
- krótkiej odpowiedzi
 - jednostka testu typu krótka odpowiedź na pytanie (*short answer item*)
 - uzupełnianie zdań (*sentence completion*)
 - zadanie typu *note expansion*
- rozszerzonej odpowiedzi
 - transfer informacji (*information transfer*)
 - rozwinięta wypowiedź pisemna (*extended writing questions*).

¹³ Nazewnictwo typów zadań zaprezentowanych przez ALTE (2005) w języku polskim podano za: Gaszyńska-Magiera i Seretny (2004).

Decyzja o tym, jakim rodzajem i typem jednostek testu należy się w danej sytuacji posłużyć, uzależniona jest od tego, co i w jakim celu ma być testowane. Obydwa rodzaje zadań – otwarte i zamknięte – mają zarówno zalety, jak i wady.

a) Zadania zamknięte

Zadania należące do grupy zadań zamkniętych wymagają od zdającego dokonania wyboru odpowiedzi spośród kilku podanych wariantów. Ten rodzaj zadań ma szereg zalet wynikających z ich konstrukcji, niepozbawiony jest on jednak również wad (Niemierko 1999: 129–130; Centralna Komisja Egzaminacyjna 2008: 34; strona internetowa Okręgowej Komisji Egzaminacyjnej w Krakowie¹⁴).

Niewątpliwą zaletą zadań zamkniętych jest szeroki zakres ich zastosowania. W testowaniu języków obcych mogą być wykorzystane w mierzeniu sprawności receptywnych – czytania i słuchania, jak również znajomości gramatyki i leksyki. Z powodzeniem nadają się również do sprawdzania wiedzy zdającego. Wprawdzie w przypadku testowania języka badamy głównie umiejętności zdającego, a nie jego wiedzę, ale w szczególnych przypadkach, jakimi są między innymi testy językowe do celów specjalistycznych, i ona bywa testowana. Z racji swojej ograniczonej długości zadania zamknięte mogą obejmować większy zakres materiału. Udzielanie odpowiedzi zajmuje mało czasu, gdyż zdający tylko wskazuje odpowiednią z podanych propozycji, a nie tworzy jej samodzielnie. Obydwa czynniki składają się na dużą sprawność pomiarową zadań zamkniętych. Jeżeli zadanie skonstruowane jest prawidłowo, to sporządzenie klucza punktowania nie powinno stanowić problemu – jest szybkie i proste, nieporównywalnie łatwiejsze niż tworzenie klucza punktowania do zadań otwartych. Samo sprawdzanie odpowiedzi również zajmuje mało czasu i nie wymaga szczególnych kwalifikacji od osób wykonujących tę czynność; zadania te mogą być sprawdzane elektronicznie, co ma ogromne znaczenie zwłaszcza w testach szerokiego użytku, do których przystępuje duża liczba zdających. Przypisywanie punktów poszczególnym odpowiedziom w zadaniach zamkniętych jest obiektywne, gdyż decyzja osoby sprawdzającej test nie jest uzależniona od jej subiektywnych przekonań i interpretacji; możliwe jest sporządzenie

¹⁴ Zob. http://www.oke.krakow.pl/sprawdzian/biuletyn_12_03/biuletyn_12_2003_3.html [dostęp: 23.02.2017].

i konsekwentne stosowanie stałego klucza punktowania. Oczywiście warunkiem *sine qua non* jest bezbłędna konstrukcja zadania, gwarantująca jednoznaczność odpowiedzi. Prostsza analiza wyników wpływa na sprawność pomiarową zadań zamkniętych.

Pośród wad tej formy testowania znajdujemy m.in. brak możliwości tworzenia syntez przez uczniów, samodzielnego formułowania przez nich hipotez oraz wyrażania własnych opinii. Odpowiedzi ograniczone są do treści wykorzystanych w zadaniu i umiejętności zaplanowanych do zmierzenia przez autorów testu, bez możliwości zaprezentowania innych umiejętności zdającego oraz przedstawienia przezeń własnego toku rozumowania. Kolejną wadą zadań zamkniętych jest brak autentyczności. W codziennym życiu czy też podczas wykonywania czynności zawodowych nikt nie oczekuje od nas rozwiązywania zadań wyboru wielokrotnego bądź dopasowywania gotowych tytułów do fragmentów tekstu, a więc zadania takie dają fałszywy obraz świata i wiedzy ludzkiej jako zamkniętych systemów o stałych, czytelnych prawidłowościach. W przypadku zadań zamkniętych mamy również do czynienia z przewagą formy zadań nad treścią kształcenia – testować za pomocą tego rodzaju zadań możemy tylko to, co daje się w takiej formie zmierzyć. Zadania zamknięte pozwalają sprawdzać ograniczony, ściśle określony zestaw umiejętności. Kolejną, bardzo istotną negatywną cechą omawianego rodzaju zadań jest trudność ich konstruowania. Aby zadania wyboru wielokrotnego spełniły swoją rolę, muszą być dostosowane do całego szeregu wymogów. Niejednokrotnie zadania na dobieranie dają więcej niż jedną możliwość odpowiedzi, a zadania typu prawda–fałsz umożliwiają udzielenie odpowiedzi bez konieczności zapoznawania się z tekstem. Wynika to z faktu, że konstrukcja zadań zamkniętych jest bardzo skomplikowana i wymaga od konstruktorów testu specjalnego przygotowania. I wreszcie cecha, której świadomi są wszyscy zainteresowani – nauczyciele, egzaminatorzy, ale także egzaminowani, a mianowicie podatność na zgadywanie. Zadania zamknięte stwarzają możliwość domyślnego typowania poprawnej odpowiedzi, co zdecydowanie zaburza obraz faktycznych umiejętności zdającego.

• **Zadania typu wybór wielokrotny**

Zadanie wyboru wielokrotnego (WW) to rodzaj zadania zamkniętego wymagającego wybrania jednej lub większej liczby prawidłowych odpowiedzi lub opcji rozwinięcia zdania spośród kilku podanych.

Zadania WW w typologii Niemierki obejmują, oprócz tradycyjnego *multiple choice question* (ALTE 2005), również zadanie z lukami z wyborem wielokrotnym (*gap-filling/cloze passage with multiple choice options*).

Zadanie WW typu **multiple choice question** składa się z tekstu wprowadzającego w formie pytania lub trzonu, będącego początkiem zdania, i z kilku (najczęściej 3–4) możliwych odpowiedzi lub wariantów jego dokończenia. Odpowiedzi nieprawidłowe zwane są dystraktorami, poprawna odpowiedź natomiast to werstraktor. Zadania te mogą być wykorzystane do sprawdzenia sprawności czytania i słuchania, znajomości struktur, leksyki oraz budowy dyskursu. **Zadanie z lukami z wyborem wielokrotnym** (*gap-filling/cloze passage with multiple choice options*), zwane też testem luk sterowanych, składa się z tekstu z lukami oraz listy odpowiedzi z trzema lub czterema wariantami dla każdej luki. Jest ono zazwyczaj używane do testowania gramatyki, słownictwa oraz obydwu tych elementów w powiązaniu ze sprawnością czytania.

Zadania wyboru wielokrotnego gwarantują szybkość, efektywność oraz obiektywność oceniania. Są formą testowania najbardziej zaawansowaną konstrukcyjnie, jak również najbardziej podatną na analizy statystyczne. Jako zadania zamknięte mogą być sprawdzane maszynowo. Jednak tworzenie tego typu zadań jest procesem bardzo trudnym i czasochłonnym, wymaga ściśle określonych umiejętności oraz wiąże się z koniecznością przeprowadzenia testów próbnych (Niemierko 1999: 115–116; Weir 1990: 43).

Zadania WW stosowane są szeroko w ocenianiu ogólnego języka obcego. W egzaminach wykorzystuje się je przede wszystkim do mierzenia sprawności czytania i słuchania oraz do sprawdzania znajomości gramatyki i leksyki. W egzaminie GE *Cambridge B2 First* za pomocą zadań WW sprawdzane jest rozumienie tekstów czytanych i słuchanych. Sprawność czytania badana jest w tym teście wyłącznie za pomocą zadań zamkniętych – wyboru wielokrotnego i na dobieranie. W części sprawdzającej rozumienie ze słuchu egzaminu *First* pojawiają się zarówno zadania otwarte, jak i zamknięte, w tym jedno zadanie WW.

Innym analizowanym egzaminem GE jest *telc Dual-Level* na łączonych poziomach od B1 do B2. W tym teście część badająca umiejętność rozumienia tekstu czytanego składa się z czterech zadań zamkniętych, z czego jedno to zadanie WW z trzema opcjami do wyboru. Rozumienie ze słuchu składa się wyłącznie z zadań WW, każde z trzema wariantami odpowiedzi.

Zadanie z lukami z wyborem wielokrotnym występuje w kilku analizowanych egzaminach GE – w *Cambridge English First* w części *Use of English* oraz w części egzaminu *telc* zatytułowanej *Language Elements*. Obydwa testy sprawdzają za pomocą tego zadania znajomość słownictwa i struktur gramatycznych oraz elementów dyskursu. W przykładowym arkuszu egzaminu *Aptis General* pojawia się ono w części nazwanej *Reading Comprehension*. Należy jednakże zauważyć, że choć zadanie to oparte jest na tekście, a nie na oderwanych zdaniach, to sprawdzane są za jego pomocą umiejętności w zakresie znajomości leksyki i gramatyki, a nie (jak wskazywałaby nazwa części egzaminu) rozumienia tekstu.

Technika WW używana jest również w certyfikacji dla potrzeb zawodowych. Spotykamy ją w szerokiej palecie egzaminów, nie tylko tych o niskim stopniu specjalizacji, ale i w testach nawiązujących do rzeczywistych sytuacji użycia języka w kontekście wykonywanej pracy. W egzaminie z języka angielskiego medycznego *STANDEM* dwa z trzech zadań opartych na tekstach pisanych to WW. Umiejętność rozumienia ze słuchu sprawdzana jest tutaj za pomocą trzech różnych zadań zamkniętych, z czego jedno to zadanie WW z czterema opcjami odpowiedzi do wyboru.

Occupational English Test (OET) to inny egzamin z języka angielskiego przeznaczony dla zawodów medycznych. Zadania WW znajdują się tutaj zarówno w części *Reading* (jedno zadanie z dwóch), jak i *Listening*, gdzie kilka pojedynczych zadań wchodzących w skład większego zadania opartego na tekście słuchanym to właśnie wybór wielokrotny.

W egzaminie dla wojska – *STANAG 6001* poziom 3 – cały test sprawdzający sprawność czytania skonstruowany jest z zadań WW. Dwa z trzech zadań w części badającej sprawność słuchania to również zadania tego typu.

Egzamin dla pilotów i kontrolerów ruchu lotniczego *ELPAC* jest przykładem wysokospecjalistycznego testu maksymalnie zbliżonego do rzeczywistej praktyki zawodowej, niemniej jedno z zadań sprawdzających rozumienie tekstu słuchanego składa się z zadań WW. Co wyróżnia ten egzamin spośród innych testów językowych, to fakt, że wspomniane zadanie WW może mieć jedną lub dwie prawidłowe odpowiedzi. Takie rozwiązanie spotykane jest w egzaminach innego rodzaju, w egzaminach językowych jednak zdarza się bardzo rzadko.

Jak można zauważyć, technika WW wykorzystywana jest do testowania zarówno języka ogólnego, jak i języków nisko i wysoko

specjalistycznych. W większości badanych egzaminów jest ona jednym z kilku stosowanych rodzajów zadań testowych, ale nie jedynym. Istnieją jednak egzaminy, które w całości bądź też których całe części skonstruowane są z zadań WW. Moduł ogólny egzaminu z technicznego języka angielskiego MOTET (*General Technical English Module – GTEM*), obowiązkowy dla wszystkich kandydatów przystępujących do tego egzaminu, jest w całości testem wyboru wielokrotnego. Składają się na niego zadania gramatyczno-leksykalne (50 jednostek WW), zadania sprawdzające sprawność słuchania (10 jednostek WW) oraz zadania w sekcji czytanie ze zrozumieniem (10 jednostek WW). W części pisemnej egzaminu MOTET tylko test sprawdzający umiejętność pisania z przyczyn oczywistych nie jest testem WW.

Innym przykładem takiej praktyki jest egzamin z języka angielskiego dla księgowych – *English for Accounting*. On również w całości składa się z zadań WW. Za ich pomocą sprawdzane jest głównie fachowe słownictwo, ale również umiejętność rozpoznawania poprawnych struktur gramatycznych, wskazywania błędnych form gramatycznych oraz błędnego zapisu ortograficznego. W niewielkim zakresie badane jest także rozumienie czytania na bazie bardzo krótkich tekstów związanych tematycznie z księgowością.

W żadnym z analizowanych egzaminów z języka specjalistycznego nie zaobserwowano natomiast obecności zadania z lukami z wyborem wielokrotnym.

Jak widać z powyższej analizy, zadania WW znajdują bardzo szerokie zastosowanie w egzaminach certyfikujących. Mimo ich wad oraz dość częstej krytyki zarówno pośród laików, jak i osób parających się zawodowo testowaniem, zadania WW są najpowszechniej stosowanym typem zadań testowych, dającym wysoką rzetelność pomiaru (Pomykało 1993: 847). Co więcej, w związku z rosnącą popularnością materiałów dydaktycznych i testowych w formie elektronicznej, jak również tendencją do przechodzenia z testowania w wersji papierowej na testowanie komputerowe, w którym zadania zamknięte, a w szczególności WW oraz typu prawda–fałsz, sprawdzają się dużo lepiej niż inne, należy sądzić, że zadania te długo jeszcze pozostaną obecne w testowaniu języków obcych.

• **Zadania na dobieranie**

Zadanie na **dobieranie**, zwane także dopasowywaniem, w nomenklaturze ALTE (2005) – *matching* – nie jest tożsame z zadaniem o tej samej

nazwie w klasyfikacji Niemierki. Dobieranie (D) w typologii Niemierki ma znaczenie szersze: obejmuje ono zadanie o tej samej nazwie w klasyfikacji ALTE, ale mieści w sobie również omówiony dalej rodzaj zadania na dobieranie wielokrotne, wypełnianie luk na podstawie selekcji z banku możliwości oraz uzupełnianie luk na poziomie akapitu. Wszystkie te zadania polegają na zestawieniu ze sobą elementów przedstawionych w postaci dwóch odrębnych zbiorów. Różnią się natomiast szczegółami swojej konstrukcji.

Dobieranie typu *matching* jest najprostszą formą tego rodzaju zadania. Polega na zestawieniu elementów z dwóch oddzielnych list słów, wyrażeń, części zdań lub fragmentów tekstu. Każdą z opcji można wykorzystać tylko raz; do odpowiedzi zazwyczaj dodane są dystraktory. Zadanie to może sprawdzić kompetencję strukturalną¹⁵, znajomość leksyki, umiejętność reagowania, rozumienie tekstu czytanego bądź słuchanego. Zadanie takie może polegać na przykład na dopasowaniu początków i zakończeń zdań, pytań i odpowiedzi, reakcji do pewnych zachowań językowych, na dobieraniu definicji do słów lub wyrażeń; może też wymagać od zdającego dopasowania dwóch serii opisów w zadaniu badającym sprawność czytania lub rozumienia ze słuchu (np. dopasowanie opisu osoby do opisu aktywności/zainteresowań) bądź też przyporządkowania osób wypowiadających się w nagraniu do stwierdzeń na temat poszczególnych rozmówców. Zadanie to może mieć więc szeroki zakres zastosowania w egzaminach językowych – nadaje się do testów mierzących sprawności czytania i słuchania, umiejętność stosowania funkcji językowych oraz do sprawdzianów gramatyczno-leksykalnych.

Wypełnianie luk na podstawie selekcji z banku możliwości (*gap-filling with selection from bank*) jest nieco bardziej zaawansowanym wariantem zadania na dobieranie. Składa się z tekstu lub zdań z lukami oraz listy (banku) zawierającej wszystkie potrzebne do uzupełnienia tych luk słowa, jak również kilka dystraktorów, które nie pasują do żadnej z luk. Uzupełnienie banku dystraktorami powoduje, że ostatnia luka nie zostanie uzupełniona automatycznie w przypadku, gdy pozostałe wypełniono poprawnie. Zadanie to nadaje się do sprawdzania znajomości leksyki, ale także struktur i funkcji językowych, jak również sprawności czytania.

¹⁵ Termin *kompetencja strukturalna (structural competence)* używany jest w materiałach ALTE (2009, 2011) dotyczących testowania. CEFR (European Commission 2001: 108–115) na swoich stronach odnosi się oddzielnie do kompetencji leksykalnej i gramatycznej (*lexical competence, grammatical competence*).

Zadanie na dobieranie wielokrotne (*multiple matching*) jest wykorzystywane głównie do badania rozumienia tekstu czytanego. Konstrukcja zadania polega na przygotowaniu pewnej liczby pytań lub jednostek tekstu do wyboru, a odpowiedzi podaje się w formie banku słów lub fraz, z których każde może być użyte dowolną liczbą razy. Zadanie na dobieranie wielokrotne jest trudniejsze niż klasyczne zadanie na dobieranie, gdyż brak ograniczenia w użyciu każdej opcji do wyboru powoduje, że poszczególne warianty odpowiedzi nie odpadają w miarę wykonywania zadania, a tym samym nie zmniejsza się liczba możliwych odpowiedzi.

Uzupełnianie luk na poziomie akapitu (*gap-filling at paragraph level*) polega na uzupełnieniu tekstu wyjętymi z niego uprzednio fragmentami – zdaniem bądź kilkoma – w celu odtworzenia poprawnego i logicznego tekstu. Ten rodzaj zadania może dotyczyć wyłącznie testu sprawdzającego umiejętność czytania ze zrozumieniem.

W badanych egzaminach zadanie typu *matching* wykorzystywane jest głównie do mierzenia sprawności słuchania. Jest ono obecne np. w egzaminie z języka medycznego *STANDEM* – spośród trzech zadań w części egzaminu badającej rozumienie ze słuchu jedno jest właśnie zadaniem na dobieranie. Również jedno z zadań na słuchanie w egzaminie *Cambridge B2 First* jest zadaniem tego typu, rolą zdającego jest w tym wypadku dopasowanie osób do wypowiedzi. *Telc Dual-Level* także używa tej formy testowania w teście mierzącym sprawność słuchania: jedno z czterech zadań jest tutaj zadaniem na dobieranie, w którym rolą zdającego jest dopasowanie osób wypowiadających się w nagraniu do informacji na ich temat lub do opinii przez nich przedstawionych.

Uzupełnianie luk na podstawie selekcji z banku możliwości dobrze sprawdza umiejętność zastosowania struktur gramatycznych i leksykalnych oraz elementów dyskursu. Pośród analizowanych materiałów zadania tego typu znaleziono w egzaminach z ogólnego języka angielskiego *telc Dual-Level*, gdzie w części *Language Elements* jedno z dwóch zadań sprawdza wspomniane umiejętności właśnie w taki sposób. Uzupełnianie luk na podstawie selekcji z banku możliwości obecne jest również w części egzaminów kończących lektorat na niektórych polskich wyższych uczelniach (Sendur 2016a, 2016b).

Zadanie na dobieranie wielokrotne w analizowanych egzaminach pojawia się w testach badających sprawność czytania. W *Cambridge B2 First* rolą zdającego jest wskazanie, w którym akapicie tekstu znajduje się żądana informacja. Podobny wariant tego zadania można znaleźć

pośród możliwych zadań na egzaminie maturalnym (Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne 2013) oraz w egzaminie ósmoklasisty z języka obcego (Centralna Komisja Egzaminacyjna 2017). Było ono również obecne w nieistniejącym już egzaminie gimnazjalnym (Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne 2010). We wszystkich tych przypadkach rolą zdającego jest oznaczenie, w którym fragmencie tekstu (lub w którym tekście) znajduje się dana informacja. Za każdym razem do wyszukania są cztery informacje, natomiast teksty, w których należy je odnaleźć, są tylko trzy, tak więc jeden z nich jest wykorzystany dwukrotnie.

Jak wspomniano wcześniej, uzupełnianie luk na poziomie akapitu jest zadaniem badającym sprawność czytania. W egzaminie z języka obcego medycznego sTANDEM jedno z trzech zadań opartych na tekstach pisanych jest klasycznym przykładem tej formy. Zadanie polegające na uzupełnieniu tekstu wyjętymi z niego wcześniej zdaniami znajduje się również w egzaminie z języka angielskiego ogólnego *Cambridge B2 First*. Nieco innym wariantem tego rodzaju zadania jest dobieranie nagłówków do akapitów lub dopasowanie pytań do akapitów stanowiących odpowiedzi na te pytania. W teście języka ogólnego *telc Dual-Level* dwa zadania z czterech stanowią przykłady takiego zadania. Zadania na dobieranie są zadaniami popularnymi w egzaminach językowych. Są one obecne zarówno w testach z języka ogólnego, jak i dla celów specjalistycznych. W przypadku egzaminów specjalistycznych znajdujemy je w arkuszach sprawnościowych – badających czytanie i słuchanie. W testach z języka ogólnego można zaobserwować je także w zadaniach gramatyczno-leksykalnych, które są częściej spotykane w egzaminach z języka ogólnego niż w specjalistycznych, jak również w egzaminach lektoratowych, które są bliższe egzaminom ogólnym.

• **Zadania zamknięte rzadziej spotykane w certyfikacji językowej**

Dobrze znaną jednostką testu jest zadanie **typu prawda–fałsz** (PF; *true/false item*). Rolą zdającego jest zdecydowanie o prawdziwości danego stwierdzenia. Tego rodzaju zadanie stosuje się zazwyczaj w testach badających sprawność czytania lub rozumienia ze słuchu.

Innym wariantem tego zadania jest typ **prawda/fałsz/brak informacji** (P/F/BI), który można zakwalifikować również jako zadanie typu wybór wielokrotny z trzema opcjami odpowiedzi.

W porównaniu do omówionych wcześniej zadań WW jednostki PF nie są tak częste w egzaminach certyfikujących. Spośród analizowanych

tutaj egzaminów technikę tę spotykamy w egzaminach *telc* – zarówno tych z zakresu języka ogólnego, jak i z języka specjalistycznego. Obydwa egzaminy z języka ogólnego – *telc Dual-Level* i *telc English B2*, jak również egzamin z języka specjalistycznego *telc English B2 Technical* – wykorzystują ten rodzaj zadania w części sprawdzającej umiejętność rozumienia tekstu czytanego oraz w części badającej rozumienie ze słuchu. Zadania typu PF są także dosyć częste w egzaminach uczelnianych. Przykłady takich zadań były też przez lata obecne w egzaminie maturalnym z języka obcego na poziomie podstawowym oraz w egzaminie gimnazjalnym. Obecnie Centralna Komisja Egzaminacyjna odchodzi w swoich egzaminach od tego rodzaju zadań jako nazbyt podatnych na zgadywanie.

Zadanie typu P/F/BI zaobserwowane zostało w części *Reading* egzaminu *IELTS General Training* oraz w tej samej części egzaminów VEC.

Wykrywanie błędów lub inaczej **technika eliminacji zbędnych wyrazów** (*extra word error detection*¹⁶) jest zadaniem opartym na tekście pisanym. W większości wierszy tekstu występuje po jednym zbędnym, niepoprawnie użytym słowie. Rolą zdającego jest zidentyfikowanie i wskazanie tego wyrazu. Zadanie to wymaga od zdających zastosowania świadomej wiedzy o strukturach językowych.

Technika eliminacji zbędnych wyrazów była przez lata popularna w niektórych testach egzaminów certyfikujących z ogólnego języka angielskiego (np. FCE). Jest ona jednak rzadko wykorzystywana w nowych wersjach egzaminów GE przeprowadzanych przez Cambridge English i niespotykana w badanych do celów niniejszej publikacji egzaminach specjalistycznych. W egzaminie *Business Higher* na poziomie C1 pośród sześciu zadań mierzących sprawność czytania ze zrozumieniem znajdujemy jedno zadanie określone jako *proofreading*, czyli *korekta tekstu* (Business English Certificate Higher 2015; Cambridge Assessment English 2019b; www.cambridgeenglish.org¹⁷). Nazwa ta, użyta przez organizatora w opisie egzaminu, jest identyczna, jak w przypadku omówionego dalej jednego z wariantów zadania otwartego, polegającego nie tylko na wskazaniu błędnego wyrazu, ale również na naniesieniu korekty błędu. Zadanie występujące w omawianym tutaj egzaminie jest jednak zadaniem zamkniętym.

¹⁶ Zadania tego typu nazywane są także *cloze-elide* (Manning 1987).

¹⁷ Zob. <http://www.cambridgeenglish.org/exams/business-certificates/business-higher/exam-format/> [dostęp: 5.05.2020].

b) Zadania otwarte

Otwarte jednostki testu to takie, w których odpowiedzi są formułowane i zapisywane samodzielnie przez zdającego. Podobnie jak zadania zamknięte mają one zarówno zalety, jak i wady (strona internetowa Okręgowej Komisji Egzaminacyjnej w Krakowie¹⁸ oraz Gaszyńska-Magiera i Seretny 2004: 139–140).

W porównaniu z jednostkami testu opartymi na wyborze zadania otwarte dają znacznie szersze możliwości i zdającemu, i egzaminującemu. Dostarczają większą ilość materiału do testowania, gdyż nie ograniczają się do umiejętności, które daje się zmierzyć za pomocą konkretnego rodzaju zadania, a jednocześnie pozwalają zdającemu wykazać się całym wachlarzem umiejętności. Zadania otwarte sprawdzają kreatywność, pokazują tok myślenia zdającego, pozwalając zarazem na samodzielność pracy i swobodę wypowiedzi. W przypadku tego rodzaju zadań zminimalizowany jest efekt zgadywania odpowiedzi, nieunikniony w przypadku zadań zamkniętych.

Jednak zadania otwarte, mimo licznych zalet, mają swoje ograniczenia. Liczba zadań otwartych w arkuszu egzaminacyjnym, z racji ich długości, jest zwykle niewielka, co powoduje, że arkusz obejmuje mniejszy zakres treści kształcenia. Skonstruowanie tego rodzaju zadań w taki sposób, aby istniała tylko określona liczba oczekiwanych odpowiedzi, nie jest łatwe. W efekcie często występuje wiele różnych możliwych alternatywnych odpowiedzi, które należy uznać. To z kolei sprawia, że bardzo problematyczna jest konstrukcja modelu oceniania i niełatwo jest ocenić te zadania rzetelnie. Trudność w obiektywnej ocenie zadań z kolei skutkuje tym, że ocena wymaga udziału przeszkolonego egzaminatora, a proces sprawdzania i oceniania jest długotrwały i złożony. Czasochłonne jest również udzielanie odpowiedzi przez zdającego oraz analiza i interpretacja wyników.

Poniżej omówione zostaną pokrótce zadania otwarte wymienione przez ALTE (2005) jako przykłady zadań stosowanych do badania umiejętności językowych oraz przykłady wykorzystania takich jednostek w konkretnych egzaminach doniosłych (*high-stakes tests / examinations*), o ile takowe wystąpiły w badanych testach.

• Zadania z luką

Zadania z luką (L) w taksonomii Niemierki obejmują kilka różnych typów zadań według klasyfikacji ALTE. Do tego rodzaju należy zadanie

¹⁸ Zob. http://www.oke.krakow.pl/sprawdzian/biuletyn_12_03/biuletyn_12_2003_3.html [dostęp: 5.11.2017].

typu wypełnianie luk otwartych, przekształcenia fleksyjne, przekształcenia z lukami, przekształcenia słowotwórcze oraz poprawa błędów.

Wypełnianie luk otwartych (*open gap filling*) znane jest wśród anglistów również jako *cloze test*. Tworzenie zadania polega na usunięciu z tekstu słów, które zdający ma następnie uzupełnić. Zgodnie z normą wypracowaną dla języka angielskiego luki w tym zadaniu powinny powtarzać się mniej więcej co siódme słowo (Gaszyńska-Magiera i Seretny 2004: 142). W pierwszych latach wykorzystania tej techniki w testach językowych faktycznie usuwano regularnie co n-ty wyraz, jednak dzisiaj słowa, które mają zostać usunięte z tekstu, częściej selekcjonuje się pod kątem sprawdzania znajomości konkretnych struktur (*rational cloze*) (ALTE 1998: 138). W praktyce luki powstałe wskutek usuwania wyrazów nie powinny pojawiać się częściej niż co 7–10 wyrazów (ALTE 2005: 128–129). Jednostka ta wykorzystywana jest często w testowaniu kompetencji strukturalnej, a więc znajomości gramatyki i leksyki, ale tak naprawdę wymaga współwystępowania wielu różnych umiejętności, takich jak rozumienie i opanowanie leksyki, rozpoznawanie i rozumienie znaczenia struktur gramatycznych oraz ich czynnego opanowania, czytanie ze zrozumieniem, opanowanie ortografii, pisanie i domysłu językowego, a więc jest techniką testowania integracyjnego i komunikacyjnego (Komorowska 2007: 152). Zadania typu *cloze* występują w egzaminach, które zawierają oddzielną część gramatyczno-leksykalną, a więc raczej w testach z języka ogólnego. W przeanalizowanych egzaminach certyfikujących zadanie tego typu pojawia się jedynie w części *Use of English* egzaminu *Cambridge B2 First*.

Zadanie typu **przekształcenia fleksyjne** (*transformation*) bada kompetencję strukturalną, a w tym przypadku głównie znajomość gramatyki. Rolą zdającego jest parafraza zdania lub jego części w taki sposób, by przekazana została ta sama informacja, lecz przy zastosowaniu innej struktury.

W części *Use of English* egzaminu *Cambridge B2 First*, obok innych zadań sprawdzających znajomość słownictwa i umiejętność stosowania zasad gramatycznych, pojawia się zadanie typu transformacji ze słowem kluczem, w którym podane słowo winno być użyte w przetworzonym zdaniu. Zadania takie są również obecne w egzaminach na lektoratach na wyższych uczelniach.

W zadaniach typu **przekształcenia z lukami** (*transformation cloze*) materiałem wyjściowym jest ciągły tekst, z którego usunięto po jednym słowie z każdego wersu. Podana jest inna forma gramatyczna wyjątku

słowa, a rolą zdającego jest ustalenie miejsca, w którym dane słowo należy umieścić, i podanie jego prawidłowej formy. Zadanie tego typu nie występuje w żadnym z badanych egzaminów z języka ogólnego ani ze specjalistycznego języka angielskiego. Prawdopodobnie z powodu braku tego typu zadań w egzaminach jest ono również bardzo rzadko spotykane jako ćwiczenie językowe w podręcznikach kursowych, a mogłoby przecież być ciekawym urozmaiceniem ćwiczeń.

Przekształcenia słowotwórcze (*word formation*) służą do testowania znajomości słownictwa, ale za ich pomocą sprawdzana jest w pewnym stopniu również znajomość gramatyki oraz rozumienie tekstu czytanego. W zadaniach, w których rolą zdającego jest dokonanie przekształcenia słowotwórczego, ze zdania usuwa się jedno słowo, a jego forma pokrewna podana jest jako materiał wyjściowy. Zdający przekształca podaną formę na pasującą do luki w zdaniu.

I znowu – spośród przeanalizowanych zadań egzaminacyjnych tylko *Cambridge B2 First* oraz część egzaminów kończących lektorat na wyższych uczelniach decyduje się sprawdzać umiejętności zdających w takiej formie. W przypadku egzaminu *First* słowotwórstwo sprawdzane jest na bazie tekstu. W egzaminach uczelnianych bazą dla tego zadania są często samodzielne zdania. Nie zaobserwowano tego typu zadań w egzaminach ESP, chociaż mogłyby one być przydatne w przypadku specyficznej terminologii niektórych języków specjalistycznych.

Zadanie typu **poprawa błędów** lub **czytanie w celu dokonania korekty** (*error correction/proof reading*) używane jest w testach kompetencji strukturalnej. Rolą zdającego jest zidentyfikowanie słowa w niepoprawnej formie i dokonanie jego korekty. Błędnie zastosowane wyrazy umieszczone są w każdym wierszu tekstu. Zadanie tego typu w ogóle nie występuje w przeanalizowanych tutaj egzaminach, pojawia się natomiast jako ćwiczenie językowe w podręcznikach języka ogólnego¹⁹. Brak tego typu zadania na egzaminach może wynikać z wywodzącego się jeszcze z metody audiolingwalnej i behawioryzmu przekonania, że ekspozycja na niewłaściwą formę powoduje jej utrwalenie, a więc powstanie niewłaściwego nawyku językowego.

Wart przytoczenia jest pewien interesujący przypadek egzaminu *English for Accounting* (EfA). Test ten, będąc w całości skonstruowany

¹⁹ Np. podręczniki GE wydawnictwa Pearson: *Total English Intermediate* (Clare i Wilson 2006), *Total English Upper Intermediate* (Acklam i Crace 2009), *Speakout Intermediate* (Clare i Wilson 2011), *Speakout Upper Intermediate* (Eales i Oakes 2011).

z zadań WW, zawiera kilka podzadań, w których zdający mają wskazać słowa niepoprawne gramatycznie (kompetencja strukturalna) bądź też stwierdzić, ile wyrazów w krótkim tekście jest błędnie zapisanych (kompetencja ortograficzna). Nie są to przykłady zadań konstrukcyjnie wpisujących się zadanie typu *error correction*, które wg klasyfikacji ALTE jest zadaniem otwartym i wymaga od zdającego dokonania poprawy błędu. Wspomniane zadania w egzaminie EfA oczekują od zdającego jedynie zidentyfikowania błędów/ów.

• Zadania typu krótkiej odpowiedzi

Jednostka **typu krótkiej odpowiedzi** w nomenklaturze ALTE (2005) – **short answer item** – nie jest tożsama z zadaniem krótkiej odpowiedzi w typologii Niemierki (1999). *Short answer item* odnosi się faktycznie do odpowiedzi na zadane pytanie – za pomocą słowa, liczby, krótkiego wyrażenia lub zdania. Zadanie krótkiej odpowiedzi (KO) Niemierki mieści w sobie zadanie o tej samej nazwie w języku angielskim (*short answer item*), jednak ma znaczenie szersze, obejmuje bowiem również inne (opisane dalej zadania), np. uzupełnianie zdań i zadanie typu *note expansion*.

Zadanie typu krótkiej odpowiedzi wg terminologii ALTE jest odpowiednio do mierzenia sprawności czytania i słuchania, jak również pisania na niskich poziomach zaawansowania. Nie jest ono jednak obecnie często używane w egzaminach szerokiego zasięgu na wyższych poziomach.

Zadania tego typu pojawiają się w egzaminach GE na niższych poziomach zaawansowania przeprowadzanych na mniejszą skalę niż pozostałe tutaj opisane. Przykładem mogą być eksternistyczne egzaminy dla dorosłych z zakresu szkoły podstawowej i liceum ogólnokształcącego²⁰. W egzaminie eksternistycznym z zakresu szkoły podstawowej (Centralna Komisja Egzaminacyjna 2015a) sprawdzane jest w ten sposób rozumienie prostego tekstu czytanego w języku obcym, natomiast w egzaminie eksternistycznym z języka angielskiego z zakresu liceum ogólnokształcącego (Centralna Komisja Egzaminacyjna 2015b) bada się za jego pomocą umiejętność pisania – zdający udzielają jedno- lub dwuzdaniowej odpowiedzi na pytania dotyczące przedstawionego w arkuszu egzaminacyjnym obrazka. Umiejętność pisania sprawdzana jest w ten sposób również w wielopoziomowym, komputerowym egzaminie certyfikującym *Aptis General*. W pierwszym, najprostszym zadaniu w tej sekcji egzaminu zadaniem zdającego jest uzupełnienie prostego formularza.

²⁰ Zob. <https://www.cke.edu.pl/egzamin-eksternistyczny/> [dostęp 5.11.2017].

W egzaminach certyfikujących LSP spotykamy ten rodzaj zadania w teście dla pilotów i kontrolerów ruchu lotniczego ELPAC, w części sprawdzającej rozumienie ze słuchu. Pośród różnego rodzaju zadań zamkniętych i otwartych badających sprawność słuchania znajdują się również pytania, które wymagają udzielenia odpowiedzi w postaci krótkiego zdania lub frazy. Zadanie typu *short answer item* pojawia się także w części *Listening* w egzaminie dla branży medycznej OET. W teście tym jednostka *short answer* nie jest skonstruowana tradycyjnie w formie pytanie–odpowiedź, ale jako uzupełnianie notatek z usłyszanego wywiadu medycznego, czyli w formie hasło–notatka.

W zadaniu typu **uzupełnianie zdań** (*sentence completion*) podana jest część zdania, a rolą zdającego jest uzupełnienie go na podstawie przeczytanego bądź też wysłuchanego tekstu. Zasadą dotyczącą konstrukcji tego typu zadania testowego jest sformułowanie go w taki sposób, aby istniała tylko jedna poprawna i jednoznaczna odpowiedź albo ograniczona liczba akceptowanych odpowiedzi.

W analizowanych dla celów niniejszego tekstu egzaminach zadanie takie pojawia się w arkuszu badającym sprawność słuchania egzaminu z języka ogólnego *Cambridge B2 First*. Znajdujemy je także w części rozumienia tekstu słuchanego w egzaminach dla pilotów i kontrolerów ruchu lotniczego ELPAC, w egzaminach wojskowych STANAG 6001, dla branży medycznej OET i sTANDEM. Nietypowym przykładem jest wykorzystanie tego rodzaju zadania w części rozumienia tekstu czytanego w egzaminie OET. W tym przypadku polega ono na odszukaniu w kilku krótkich tekstach medycznych informacji potrzebnych do uzupełnienia innego tekstu. Tekst z lukami, który zdający uzupełniają, jest streszczeniem informacji zawartych w przedstawionych tekstach do przeczytania.

Angielska nazwa zadania użyta w publikacji ALTE (2005) – **note expansion** – odnosi się do zadania, w którym rolą zdającego jest zbudowanie poprawnych zdań z podanych elementów leksykalnych przypominających notatki. W tym celu zdający musi odpowiednio zmienić formy gramatyczne podanych wyrazów, niejednokrotnie także ich szyk, oraz dodać pewne elementy, jak na przykład przyimki, przedimki czy zaimki. Tego rodzaju zadanie może pojawić się w teście sprawdzającym kompetencję strukturalną, bądź też umiejętność pisania (*cf.* przykład 1). Termin **rozsypanka wyrazowa**, użyty jako odpowiednik angielskiego **note expansion** przez autorki *Przewodnika dla autorów zadań do testów*

językowych (Gaszyńska-Magiera i Seretny 2004: 146–147), nie do końca jest zgodny z jego użyciem w polskiej literaturze dotyczącej testowania. Termin ten w polskich publikacjach związanych z tematem (Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne 2010; Komorowska 2007: 87–88, 139, 148) odnosi się raczej do uproszczonej wersji zadania typu *note expansion*, w której zdający nie musi zmieniać form gramatycznych podanych wyrazów, ani też nie powinien niczego dopisywać, a zadanie polega wyłącznie na uporządkowaniu podanych („rozsypanych”) wyrazów celem stworzenia poprawnego zdania (cf. przykład 2).

Zadanie używane od jakiegoś czasu w polskich egzaminach zewnętrznych z języków obcych (egzamin gimnazjalny; Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne 2010), które polega na budowaniu fragmentów zdań z podanych elementów leksykalnych, nosi nazwę **gramatykalizacji** (cf. przykład 3). Jest ono również modyfikacją zadania typu *note expansion*, jako że dotyczy fragmentów, a nie całych zdań, jak to ma miejsce w opisie ALTE.

Przykłady poszczególnych rodzajów zadań testowych, wpisujących się częściowo w charakterystykę zadania typu rozsypanka wyrazowa, zaprezentowano poniżej:

Przykład 1. Zadanie typu *note expansion* za: ALTE (2005: 132).

Dear Mr Harris

- a) I be very pleased / meet you / teachers' conference / London last year.
- b) It be kind / you / invite me / come and see you while I be / England / this summer.
- c) I hope / pay a visit / your school / 26th and 27th June if / not be inconvenient.
- d) Please / not rearrange / programme / me.
- e) I be very happy / fit in / whatever you / do at that time.
- f) I like / stay overnight / 26th June and hope / arrange accommodation / me.
- g) I telephone you once / reach London / confirm / exact time / arrival / school.
- h) I look forward / meet / again.

Yours sincerely

Przykład 2. Rozsypanka wyrazowa wg terminologii Komorowskiej i CKE w podręczniku *English for Safety, Security and Law Enforcement* (Sendur i Warecka 2012: 21).

POLICE	THE	A	DRINK-DRIVING	FOR	MAN	ARRESTED
4. _____						

Przykład 3. Zadanie typu gramatykalizacja w arkuszu z języka angielskiego egzaminu gimnazjalnego w roku 2013 (Centralna Komisja Egzaminacyjna 2013).

7.2. It was cold and there (be/many/cloud) _____ in the sky.
--

Omówiony tutaj typ zadania nie występuje w przeanalizowanych międzynarodowych egzaminach GE ani LSP w żadnej z przedstawionych powyżej wersji. Być może wynika to z faktu, że rozsypanka wyrazowa według nazewnictwa Komorowskiej i CKE to zadanie sprawdzające bierną znajomość języka obcego na niższych poziomach zaawansowania, a dla celów niniejszego opracowania badane były przede wszystkim testy na poziomie B1 i wyższych. Jeśli zaś chodzi o trudniejszą wersję zadania, jaką jest gramatykalizacja, to przyczyną braku jej użycia w egzaminach LSP może z kolei być duża popularność innych form sprawdzania znajomości gramatyki, jak na przykład przekształcenia fleksyjne, czyli inaczej transformacje, które sprawdzają podobne umiejętności.

• Zadania rozszerzonej odpowiedzi

Do zadań rozszerzonej odpowiedzi (RO; Niemierko 1999) zaliczamy zadanie typu transfer informacji oraz rozwinięta odpowiedź pisemna.

Zadanie typu **transfer informacji** (*information transfer*) polega na wykorzystaniu przez zdającego informacji podanej w pewnej formie w celu wykonania innego zadania. Może ono na przykład polegać na wykorzystaniu informacji przedstawionej w formie tekstu pisanego w celu stworzenia pisemnego raportu albo narysowania diagramu. Zadanie to sprawdza umiejętności związane z tworzeniem tekstu pisanego, rozumieniem tekstu pisanego lub słuchanego oraz kompetencją strukturalną (ALTE 2005: 134; Gaszyńska-Magiera i Seretny 2004: 148).

Przykładem testu, który wykorzystuje transfer informacji w swoich zadaniach, jest część pisemna egzaminu *English for Tourism* (Eft). Zadania te oparte są na materiale tekstowym, z którego zdający musi wyselekcjonować informacje potrzebne do wykonania kolejnych zadań polegających na stworzeniu kilku własnych tekstów. W tym celu musi wykazać się umiejętnością rozumienia tekstów pisanych, a następnie sprawnością pisania. Podobnie wygląda sytuacja w części *Writing* egzaminu dla branży medycznej sTANDEM. Jedno z zadań w tym teście polega na przeanalizowaniu i pisemnym podsumowaniu danych przedstawionych w formie graficznej (np. wykresu). W części badającej sprawność pisania egzaminu z języka medycznego OET zdający przetwarza wyselekcjonowane informacje z karty pacjenta, stanowiącej dane wejściowe do zadania, celem stworzenia formalnego skierowania do innego lekarza specjalisty, w którym przekazuje on informacje dotyczące pacjenta oraz historii jego choroby. Celem wszystkich tych zadań jest zbliżenie zadania egzaminacyjnego do autentycznej działalności zawodowej zdającego.

Zadanie typu **rozwiniętej odpowiedzi pisemnej** (*extended writing question*) może mieć różne postacie. Może ono być skonstruowane w formie krótkiego tematu, czy tytułu pracy, którą zdający będzie musiał napisać; może też być oparte na dłuższym, złożonym materiale wejściowym (*input*), który egzaminowany będzie musiał najpierw przetworzyć, aby następnie skonstruować na jego podstawie swój własny tekst.

Wszystkie egzaminy certyfikujące GE oraz większość LSP zawierają w swoich testach zadania rozwiniętej wypowiedzi pisemnej. W niektórych egzaminach uczelnianych taka forma nie jest w ogóle sprawdzana, zapewne ze względu na trudność i czasochłonność procesu oceniania oraz dużą liczbę zdających przystępujących do egzaminu podczas sesji. Część egzaminów LSP daje kandydatom możliwość wyboru poszczególnych sprawności, więc nie każdy zdający musi przystąpić do testu umiejętności pisania, w którym znajduje się omawiany tutaj rodzaj zadania (np. egzamin z angielskiego języka technicznego MOTET, dla zawodów medycznych sTANDEM, dla branży turystycznej Eft). Jeden z analizowanych egzaminów z założenia nie bada sprawności pisania, gdyż umiejętność ta nie jest niezbędna do wykonywania zawodu, którego przedstawiciele egzamin ten zdają. Mowa tutaj o teście dla pilotów i kontrolerów ruchu lotniczego ELPAC.

Pośród szeregu różnych form wypowiedzi pisemnej sprawdzanych na egzaminach za pomocą zadania typu rozwiniętej odpowiedzi pisemnej występują różne rodzaje listów i e-maili, rozmaite typy rozprawki, artykuł, blog, opis, instrukcje, raport, recenzja, opowiadanie, sprawozdanie i formalna notatka służbowa.

W *Cambridge B2 First* (egzamin GE) obowiązkowe dla wszystkich zdających jest napisanie wypracowania z wykorzystaniem danych wejściowych – informacji podanych w poleceniu. Drugie zadanie w egzaminie to jeden z trzech tematów do wyboru. Rodzaje tekstu, które mogą pojawić się w tej części, to artykuł, formalny lub nieformalny list bądź e-mail, raport albo recenzja. Wypowiedź zdającego ma być oparta na opisie sytuacji zaprezentowanym w poleceniu.

W egzaminie MOTET zdający pisze dwa teksty spośród podanych pięciu form do wyboru (np. artykuł, blog, opis, list, instrukcje). Zadania w części *Writing* egzaminu dla branży medycznej sTANDEM mają na celu sprawdzić umiejętność konstruowania listów formalnych oraz umiejętność pisania analiz badań naukowych. STANAG 6001 dla wojska wymaga od zdających napisania trzech tekstów: listu formalnego z wykorzystaniem podanych informacji, formalnej notatki służbowej związanej tematycznie ze wspomnianym listem oraz sprawozdania.

Rozwinięta wypowiedź pisemna stanowi również element omówionych wcześniej zadań typu transfer informacji w egzaminach EfT, OET oraz sTANDEM. Efektem końcowym innych złożonych czynności, które zdający ma w tych zadaniach wykonać, są różnego rodzaju teksty pisane.

Zadania te wydają się dość autentyczne, zarówno jeśli chodzi o egzamin z języka ogólnego *First*, jak i te w opisanych egzaminach ESP. Większość z nich odpowiada rzeczywistym działaniom językowym w życiu codziennym lub w pracy bądź też możliwym działaniom zawodowym w przyszłości.

Specyficzne zadania ustne testujące sprawność mówienia

Sprawność mówienia może być badana w różny sposób. Wybór zadania będzie zależał m.in. od tego, czy zdający ma być egzaminowany indywidualnie, w parze bądź w większej grupie, czy kandydaci będą odgrywać zadane role, czy wypowiadać się na zadany temat oraz jakie materiały stymulujące zostaną użyte w celu zbadania kompetencji zdających. W grę

wchodzi również nagrywanie odpowiedzi lub wypowiedzi zdającego w reakcji na polecenie podczas egzaminu przeprowadzanego w wersji komputerowej.

W ramach mierzenia umiejętności mówienia możliwy do wykorzystania jest cały szereg różnych zadań, jak wypowiedź na zadany temat, opisywanie zdjęcia / obrazka, opowiadanie / streszczenie wydarzenia / opowiadania, zadanie typu pytanie–odpowieź, dyskusja, zadanie z luką informacyjną i odgrywanie ról.

ALTE (2005: 142–147) i Komorowska (2007: 103–119) jako najpopularniejsze rodzaje zadań testujących sprawność mówienia podają:

- technikę wywiadu,
- prezentację,
- użycie ikonicznych materiałów stymulujących,
- użycie pisemnych materiałów stymulujących,
- zadanie z luką informacyjną,
- technikę symulacji i odgrywania ról zawodowych.

a) Przykłady zadań testujących sprawność mówienia

Technika **wywiadu** (*interview*) to jedna z najprostszych form testowania sprawności mówienia. Polega na swobodnej rozmowie z uczniem czy zdającym w celu zorientowania się co do poziomu umiejętności budowania wypowiedzi ustnej zdającego oraz uczestniczenia w konwersacji. Wywiad może być swobodny lub kontrolowany. Różnica pomiędzy tymi dwoma typami zadania ustnego polega na stopniu ograniczeń w zadawaniu pytań przez osobę egzaminującą i udzielaniu odpowiedzi przez zdającego. W wywiadzie swobodnym brak jest precyzyjnych uściśleń dotyczących długości odpowiedzi, ujednoczenia poleceń i innych sformułowań wypowiedzianych przez egzaminującego. W przypadku wywiadu kontrolowanego elementy te są poddane precyzyjnym kryteriom. W egzaminach certyfikujących zadanie tego typu bywa wykorzystywane jako nieformalna rozgrzewka przed kolejnymi zadaniami w części ustnej. W testach, w których takie zadanie wstępne jest wpisane w scenariusz egzaminu, często nie jest ono w ogóle punktowane, określone są natomiast sztywne ramy dotyczące jego formy i przeznaczonego czasu.

Z analizowanych egzaminów tylko *Cambridge B2 First* wprowadził tę formę badania umiejętności interakcji w ramach określonych tematów

jako zadanie egzaminacyjne podlegające ocenie. Część ustna egzaminu *English for Tourism* rozpoczyna się od dwuminutowej rozmowy wstępnej, w której egzaminator zadaje kandydatowi pytania dotyczące podstawowych danych osobowych, pochodzenia, pracy lub studiów, planów na przyszłość itp. Zadanie to nie jest oceniane. Rozmowa wstępna w formie wywiadu ma taką samą funkcję w teście z technicznego języka angielskiego MOTET oraz w egzaminach *Linguaskill General* i *Business*, gdzie również nie jest punktowana.

Rozmowa tego typu – na tematy dotyczące życia codziennego – jest wpisana w użycie języka do celów ogólnych, nie odpowiada natomiast żadnemu *stricte* zawodowemu zadaniu językowemu. Zasadna wydaje się zatem ocena tej umiejętności na egzaminie GE oraz zaniechanie takiej oceny w przypadku testu LSP, gdzie rozmowa ta stanowić będzie jedynie rodzaj rozgrzewki przed dalszą paraautentyczną symulacją. Warte jednak rozważenia jest uwzględnienie tego typu ewaluacji w egzaminach LSP dla zawodów, w których tzw. *small talk* może być istotnym elementem dyskursu zawodowego. Przykładem na potrzebę posiadania umiejętności prowadzenia takiej pozornie niezobowiązującej konwersacji jest język biznesowy, którego elementem mogą być także nieformalne rozmowy na spotkaniach biznesowo-towarzyskich (Frendo 2005: 64–65). Rozmowy takie, w warunkach zawodowych, powinni umieć prowadzić również przedstawiciele takich zawodów jak kosmetyczka (Beinstein 1975: 91) czy pokojówka hotelowa (Jasso-Aguilar 2005).

W zadaniu typu **prezentacja** (*presentation*) rolą zdającego jest kilkuminutowa wypowiedź na temat, który albo został wcześniej przygotowany, albo przedstawiony zdającemu w trakcie egzaminu. Po prezentacji powinna nastąpić dyskusja na jej temat, co pozwoli wyeliminować lub zmniejszyć błąd oceny wynikający z uprzedniego przygotowania wypowiedzi bądź też pamięciowego nauczenia się treści przygotowanego materiału.

Prezentacja jest obecna w bardzo wielu egzaminach językowych zarówno sprawdzających GE, jak i ESP. W części ustnej egzaminu *telc Dual-Level* jedno z trzech zadań ma formę prezentacji. Zdający otrzymuje dwa tematy, z których wybiera jeden i prezentuje go po pięciominutowym przygotowaniu. Wprawdzie do tematów dołączona jest seria zdjęć z nimi związanych, mają one jednak służyć jedynie jako inspiracja, pomoc w zaplanowaniu wypowiedzi, a nie baza do prezentacji, toteż zdający nie musi

się do nich odnosić. Po prezentacji następuje związana z nią dyskusja. W omawianym egzaminie tematy do prezentacji są tematami uniwersalnymi, występującymi często w podręcznikach do języka ogólnego, które powinny być dość bliskie dla przeciętnego dorosłego człowieka, jak również dla nastolatka aspirującego do uzyskania certyfikatu językowego na poziomie B1–B2. Mogą one dotyczyć np. relacji pomiędzy pracą a życiem osobistym czy ochrony środowiska z punktu widzenia pojedynczego człowieka (cf. przykład 4).

W części ustnej egzaminu STANAG pierwszą część stanowi tzw. *briefing*, a więc 3–4-minutowy monolog z wykorzystaniem podanych informacji, po którym następuje dyskusja z drugim kandydatem polegająca na wymianie pytań i odpowiedzi oraz komentowaniu treści wypowiedzi. Przed przystąpieniem do egzaminu zdający mają 15 minut na przygotowanie wypowiedzi ustnej. Zadanie wydaje się dość autentycznie naśladować działania zawodowe pracownika wojska (cf. przykład 5).

W egzaminach MOTET prezentacje odbywają się spontanicznie, zdający nie ma czasu na przygotowanie się do dyskusji. W przypadku tego egzaminu trudno jest ocenić przydatność zawodową omawianej formy na podstawie jedyne dostępnego zadania przykładowego. Tematyka przykładowego zadania (korzystanie z portalu społecznościowego Facebook; cf. przykład 6) może kojarzyć się z docelową profesją jego odbiorcy, jaką jest zawód informatyka (*Information Technology*), znajomość tematu nie jest jednak rzeczywiście związana z faktem jego wykonywania. Osoba nie mająca nic wspólnego z w/w dziedziną może dużo lepiej orientować się w temacie niż informatyk niekorzystający z portali społecznościowych. Zaprezentowane przez instytucję egzaminującą przykładowe zadanie wydaje się być bardziej adekwatne dla egzaminu skierowanego do młodzieży szkolnej lub akademickiej, która we wspomnianym temacie jest prawdopodobnie doskonale zorientowana.

Oprócz wymienionych testów prezentacja jest również obecna w części ustnej wielu egzaminów uczelnianych (Sendur 2016b: 176–177).

Przykład 4. Zadanie typu prezentacja w części ustnej egzaminu *telc Dual-Level B1-B2* (telc Language Tests 2012b: 24).

Topic 1: The environment

What do you think individual people can do to help protect the environment?
How do you contribute and why is it important to you?



Topic 2: Work-life balance

How do you balance your work with the rest of your life?
How can you achieve success at work and have a happy personal life?



Przykład 5. Zadanie z przykładowego egzaminu ustnego STANAG 6001 poziom 3 (arkusz L3 054)²¹

Briefing 2

You serve in a small Polish force on a mission abroad. One of your patrols has been cut off in hostile terrain. Give a briefing on two possible courses of action to rescue the soldiers, describing their advantages and disadvantages. Recommend one and justify your choice.

Possible options:

- wait up to 12 hours for support from allied forces
- send a Polish QRF platoon immediately
- other

²¹ Zob. <http://wsnjo.wp.mil.pl/pl/77.html> [dostęp: 6.11.2017].

Przykład 6. Zadanie z przykładowego ustnego egzaminu MOTET²²

One of the most popular, if not the most popular social media is *facebook*. Could you explain for a layman how *facebook* works, what the benefits and particularly the risks are, and how these risks can be avoided.

Rozmowa na podstawie materiału stymulującego (użycie ikonicznych materiałów stymulujących; *speaking based on picture prompts*) to zadanie polegające na ustnym opisie i dyskusji na temat zdjęcia, zestawu zdjęć bądź innych materiałów graficznych. W egzaminie GE Cambridge B2 First oraz ELPAC dla pilotów i kontrolerów lotniczych materiałem wyjściowym do tego zadania jest zdjęcie. W teście z języka medycznego sTANDEM jest to materiał graficzny w postaci wykresu, tabelki lub rysunku. We wszystkich powyższych przypadkach zdający prezentuje opis obrazka lub informacje przedstawione za pomocą materiału graficznego, a następnie odbywa się rozmowa na zaprezentowany temat.

W egzaminie Cambridge B2 First tematyka wypowiedzi opartej na zdjęciach dotyczy życia codziennego, relacji międzyludzkich itp. Część ustna odbywa się w parach, a więc po zakończeniu monologowej wypowiedzi opartej na materiale zdający zadają sobie nawzajem pytania i komentują swoje wypowiedzi.

W przypadku egzaminu dla pilotów obrazek nie stanowi materiału zawodowego wymagającego opracowania, jest jedynie pretekstem do mającej nastąpić po omówieniu fotografii rozmowy na ogólne tematy dotyczące kwestii lotniczych. Rozmowa ta ma na celu zademonstrowanie przez kandydata pełnego zakresu swojej biegłości językowej.

Najbardziej autentyczne wydaje się zadanie tego typu w egzaminie sTANDEM. Rolą zdającego jest tu przeanalizowanie materiału graficznego w formie wykresu, a następnie zaprezentowanie i zinterpretowanie informacji w nim zawartych w kilkuminutowej wypowiedzi ustnej, po której następują pytania egzaminatora.

Zadanie oparte na **pisemnych materiałach stymulujących** różni się od opisanego wcześniej rodzaju materiału stymulującego, który stanowi podstawę do wypowiedzi ustnej. Badanie sprawności mówienia może być oparte na materiale graficznym, jak w opisanych powyżej egzaminach, lub na tekście (*written prompts*). Materiałem do przetworzenia może być

²² Zanotowane na podstawie nagrania wideo przykładowego egzaminu ustnego udostępnionego przez instytucję przeprowadzającą egzaminy MONDIALE Testing.

krótkie zdanie, hasło, bądź też dłuższy tekst. Rolą zdającego może być prezentacja materiału, odpowiedź na pytania jego dotyczące, przedstawienie własnej opinii na temat wynikający z materiału oraz dyskusja.

Pisemne materiały stymulujące wykorzystane w części sprawdzającej sprawność mówienia w badanych egzaminach są często bardzo krótkie (kilka krótkich zdań lub haseł), podobnie jak materiał przedstawiony zdającemu w omówionych wcześniej przykładach zadań typu prezentacja. Jednak we wspomnianych przykładach materiały służą wyłącznie jako inspiracja, jako pomoc w przygotowaniu prezentacji, zdający w swojej wypowiedzi nie musi wykorzystywać zawartych w nich informacji. Tekstowe materiały stymulujące mogą być wykorzystane dużo bardziej efektywnie przy założeniu, że zdający muszą w swojej wypowiedzi odnieść się do informacji w nich zawartych.

W jednym z zadań w części ustnej *Cambridge B2 First* arkusz podaje zdającym temat w formie pytania oraz listy haseł – możliwych odpowiedzi, do których zdający odpowiadający w parach mają się odnieść we wspólnej dyskusji. Egzamin dla wojska, STANAG 6001, przedstawia kilkuzdaniowe pytanie–polecenie uzupełnione o trzy hasła (słowa lub wyrażenia), które również należy uwzględnić w wypowiedzi. Bardziej rozbudowana forma tego zadania wykorzystana jest w egzaminie *English for Tourism* (EfT): w bilecie egzaminacyjnym oprócz jednego głównego tematu, do którego zdający ma się odnieść w swojej wypowiedzi, znajduje się pięć pytań i zdań naprowadzających, które mają służyć jako materiał pomocniczy porządkujący wypowiedź.

W egzaminie *First*, który jest testem GE, tematyka omawianego zadania dotyczy zagadnień związanych z codziennym życiem, m.in. społeczności, których zdający jest częścią (np. spędzanie czasu wolnego przez mieszkańca miasta/wsi, podejście do turystyki z punktu widzenia mieszkańca miasta turystycznego itp.), a więc może być tematem dyskusji towarzyskiej lub debaty społecznej. W egzaminie ESP, jakim jest EfT, tematyka zadania jest ściśle związana z odnośną branżą, samo zadanie jednak nie jest wprost zadaniem zawodowym. Podobnie jak w egzaminie GE wymaga ono wyrażenia swojej opinii na zadany temat, a nie wykonania działania charakterystycznego dla wykonywanego zawodu. Omawiane zadanie w egzaminie STANAG 6001 jest natomiast przykładem zadania niespełniającego w ogóle wymogów testowania LSP. Zagadnienia, do których odnoszą się zadania oparte na pisemnych materiałach stymulujących w tym egzaminie, dotyczą m.in. świata zwierząt, kultury, nauki

języków, opieki zdrowotnej, czytania i zakupów, a więc są to tematy wpisujące się bardzo mocno w tematykę certyfikacji ogólnej. Należy jednak zwrócić uwagę, że wskaźniki biegłości językowej opisane w *Tabeli stopni znajomości języków w Porozumieniu Standaryzacyjnym NATO nr 6001 – STANAG 6001*²³, które stanowią punkt wyjścia dla egzaminów STANAG 6001, w swoim założeniu w bardzo niewielkim stopniu odnoszą się do konkretnych zagadnień militarnych. Są one skonstruowane tak, by definiować ogólną biegłość językową, a więc umiejętność komunikowania się w szerokim zakresie tematów z różnymi odbiorcami i w różnych okolicznościach (Monaghan 2012: 29). I właśnie taką rolę zadanie to spełnia.

Zadanie z luką informacyjną (*information gap task*) opiera się na autentycznej potrzebie komunikacji wynikającej z różnicy w zakresie posiadanych informacji na dany temat przez odbiorcę komunikatu oraz jego nadawcę. Rolą zdającego jest uzyskanie potrzebnych mu informacji od interlokutora w celu wypełnienia owej luki. Rozmówcą może być drugi zdający w egzaminach zdawanych w parach lub egzaminator. Luka informacyjna wykorzystywana jest często w podręcznikach jako ustne ćwiczenie językowe²⁴. Nie jest ona wykorzystana w żadnym z opisanych tutaj egzaminów.

Godną uwagi jest technika kontroli i oceny sprawności mówienia w formie **symulacji i odgrywania ról zawodowych** (*role play*). Zdający otrzymuje kartę z opisem sytuacji i swojej roli w mającej nastąpić interakcji oraz informacją na temat tego, co ma w danej rozmowie osiągnąć lub jakie poglądy ma wyrażać. Technika ta bardzo dobrze sprawdza się jako ćwiczenie językowe na kursach języka specjalistycznego dla celów zawodowych, jak również jako zadanie egzaminacyjne w testach ukierunkowanych na taki cel. Dzieje się tak przede wszystkim dlatego, że w przypadku oceniania kompetencji w zakresie języka zawodowego ważne jest przede wszystkim jego operacyjne użycie, a nie same treści językowe. Jak zauważa Magdalena Sowa (2015: 43), sam komponent językowy nie jest wystarczający w ewaluacji znajomości języka, dlatego w ocenie należy wziąć pod uwagę rzeczywistą skuteczność zawodową w języku obcym. Ową skuteczność daje się ocenić właśnie w zadaniach typu *role play*.

²³ Zob. http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/Porozumienie_standaryzacyjne_STANAG_6001_edycja1.pdf [dostęp: 6.11.2017].

²⁴ Np. podręczniki GE wydawnictwa Pearson: *Total English Intermediate* (Clare i Wilson 2006), *Total English Upper Intermediate* (Acklam i Crace 2009), *Speakout Intermediate* (Clare i Wilson 2011), *Speakout Upper Intermediate* (Eales i Oakes 2011).

We wszystkich przeanalizowanych egzaminach, w których pojawia się ten rodzaj zadania, jedna rola w symulacji odgrywana jest przez zdającego, drugą natomiast przyjmuje na siebie egzaminator. W egzaminie ELPAC odgrywane są role pilota i kontrolera ruchu lotniczego przez zdającego i egzaminatora, zgodnie z wykonywanym zawodem egzaminowanym. W następującym kolejnym zadaniu egzaminator przyjmuje rolę przełożonego, a kandydat zdaje szczegółową relację z wydarzeń z zadania poprzedzającego. W egzaminie z języka medycznego *STANDEM* zadanie to oparte jest na szczegółowym scenariuszu. W bilecie egzaminacyjnym określone jest miejsce rozmowy, rola zdającego i egzaminatora, opisana jest sytuacja i problem, z jakim zgłasza się pacjent, oraz tekst, z którego informacje powinny być wykorzystane w rozmowie. Podobnie skonstruowane są zadania w części sprawdzającej sprawność mówienia w egzaminie *English for Tourism* oraz *Occupational English Test* dla branży medycznej. Bilet egzaminacyjny zawiera szczegółowy opis sytuacji, na podstawie którego zdający wraz z egzaminującym odgrywają scenkę imitującą autentyczną sytuację zawodową związaną odpowiednio z branżą turystyczną oraz medyczną.

Technika ta w egzaminach ESP sprawdza się doskonale, dostarczając możliwości wykonywania zadań bardzo zbliżonych do zadań zawodowych, z którymi zdający spotykają się lub mogą spotkać się w pracy. Nie musi ona jednak ograniczać się do testów LSP. Odgrywanie ról może z równym powodzeniem dotyczyć codziennych życiowych sytuacji, jak pytanie o drogę, kupowanie biletu czy rezerwowanie pokoju w hotelu. Nie występuje ona jednak w żadnym z badanych tutaj egzaminów GE.

Różne rodzaje ustnych zadań testowych sprawdzają inne umiejętności szczegółowe, a więc decyzja o ich zastosowaniu lub nie, uzależniona będzie od potrzeb danego egzaminu. Warto zwrócić uwagę, że niektóre z zadań badają także więcej niż tylko sprawność mówienia. Wywiad, dyskusja po prezentacji, zadanie z luką informacyjną i odgrywanie ról to zadania, które wymuszają interakcję ustną pomiędzy zdającym a egzaminującym bądź też pomiędzy dwoma lub trzema egzaminowanymi jednocześnie. Zadania te sprawdzają zatem ustną interakcję, a więc łączą sprawność mówienia z rozumieniem ze słuchu.

Zadania egzaminacyjne specyficzne dla podejścia zadaniowego

Trzecim rodzajem zadań według podziału ze względu na sposób udzielania odpowiedzi są zadania praktyczne.

Zadanie praktyczne jest zadaniem testowym wymagającym od ucznia celowego oddziaływania na otaczający świat materialny. Przynosi wynik, czyli *produkt*, który może być oceniany osobno lub, w przypadku bardziej złożonych czynności ucznia, łącznie z zarejestrowanym przebiegiem, to jest *procesem* działania praktycznego. Produkt pozbawiony wartości rynkowej nazywamy *wytworem* działania. (Niemierko 1999: 79)

Zadania praktyczne, choć naturalne w codziennym życiu, w uczeniu się i wykonywaniu zawodu z racji swoich cech nie są tak popularne w egzaminach, jak zadania pisemne i ustne. Charakteryzuje je bowiem wysoki koszt, czasochłonność, trudność i w niektórych przypadkach narażenie na niebezpieczeństwo (Niemierko 1999: 79 za: Valencia 1991).

W nauczaniu języków obcych z zadaniami praktycznymi mamy do czynienia przede wszystkim w przypadku języka do celów specjalistycznych, w szczególności w podejściu do nauczania języka opartym na działaniu, czyli w podejściu zadaniowym. W ocenianiu w podejściu zadaniowym ważne jest odzwierciedlenie aktualnych i przyszłych komunikacyjnych celów użytkownika języka, a więc zadanie egzaminacyjne powinno jak najbardziej przypominać zadanie zawodowe. W przypadku języka obcego zadanie takie będzie integrować różne sprawności, tak jak to się dzieje w życiu codziennym, a ów „wytwór działania” może mieć formę ustną, pisemną, może też łączyć różne formy wypowiedzi.

Prawie każde zadanie dydaktyczne może być użyte do celów oceny postępów uczącego się. Różnica pomiędzy zadaniem wykorzystanym w procesie nauczania, a użytym w procesie oceniania polega na odpowiednim dostosowaniu sposobu jego przeprowadzenia, formy zapisu pracy osoby zdającej oraz sposobu analizy wyników (Nunan 2004: 154). W przypadku języków obcych w ocenianiu w podejściu zadaniowym (*Task-based Assessment* – TBA) najlepiej wpisuje się:

- odgrywanie ról zawodowych (*role play*),
- praca projektowa (*project work*),
- studium przypadku (*case study*),
- symulacja globalna (*global simulation*). (Fischer *et al.* 2011: 21–24)

a) Przykłady zadań w podejściu zadaniowym

Zadanie typu **odgrywanie ról zawodowych** nadaje się do mierzenia sprawności mówienia i było już wcześniej omówione w kontekście zadań ustnych. Może ono być także częścią innego, większego zadania, jak na przykład praca projektowa czy symulacja globalna. W podejściu

zadaniowym technika ta nie różni się niczym pod względem konstrukcyjnym od zadania typu *role-play* opisanego w części dotyczącej zadań testujących sprawność mówienia. Na co warto jednak zwrócić uwagę w przypadku podejścia zadaniowego, to autentyczność zadań, która ma tutaj dużo większe znaczenie. Podczas gdy w podejściu komunikacyjnym w zadaniu takim zdający może przyjmować różne role, w których może, ale nie musi się znaleźć w przyszłości, to w podejściu zadaniowym role i sytuacje ograniczają się *stricte* do takich, z którymi zdający powinien zetknąć się w pracy zawodowej lub w życiu codziennym w najbliższej przyszłości. Zasadę tę bardziej obrazowo przedstawia Fischer *et al.* (2011: 21), podając jako przykład rolę lekarza, którą w zadaniu typu *role play* w podejściu zadaniowym zdający może otrzymać tylko, gdy taki zawód wykonuje lub do wykonywania tego zawodu się przygotowuje. Podejście komunikacyjne w takiej sytuacji nie jest aż tak restrykcyjne. Wszystkie omówione wcześniej – w części poświęconej testowaniu sprawności mówienia – przykłady zadań egzaminacyjnych typu odgrywanie ról (ELPAC dla pilotów, sTANDEM dla branży medycznej oraz *English for Tourism*) spełniają wspomniany warunek dotyczący autentyczności zadań w ocenianiu w podejściu zadaniowym.

David Nunan (2004: 133) definiuje **pracę projektową** jako „maksymalizację”, jako serię zintegrowanych ze sobą zadań, które wykonane po kolei stworzą końcowy projekt. Praca projektowa opiera się na autentycznym działaniu i współpracy uczących się, stwarza szansę rozwoju i ewaluacji kompetencji ogólnych oraz kompetencji socjolingwistycznej (Janowska 2011: 158).

Technika ta polega na realizacji wspólnego przedsięwzięcia (projektu) na podstawie przyjętych założeń. Pozwala ona na integrację różnych sprawności oraz kompetencji z wykorzystaniem zróżnicowanych technik i nośników informacji. Uczący się/egzaminowany podczas pracy projektowej wykorzystuje różne typy dokumentów, posługuje się różnymi odmianami dyskursu, aktywnie uczestniczy w różnych sytuacjach komunikacyjnych (Gajewska i Sowa 2014: 197–199).

Wg Janowskiej (2011: 156) projekt pedagogiczny powinien:

- być ukierunkowany na konkretny produkt,
- rozciągać się w czasie,
- mobilizować cały zespół (choć może też być realizowany indywidualnie),
- dotyczyć wielu dziedzin wiedzy.

Podobnie jak pozostałe rodzaje zadań wykorzystywanych w podejściu zadaniowym technika ta stworzona została z myślą o uczeniu się i nauczaniu języka, ale może być również wykorzystana w celu ewaluacji.

Próbę wykorzystania tej techniki oceniania działań językowych studentów podjął zespół GULT opracowujący kształt egzaminu językowego w podejściu zadaniowym – *Guidelines for University Language Testing* (Fischer *et al.* 2011). Na chwilę obecną praca projektowa nie znalazła jednak szerokiego zastosowania w ocenianiu biegłości językowej i certyfikacji; nie jest ona także uwzględniona w strukturze żadnego z analizowanych tutaj egzaminów certyfikujących, ani też w egzaminach kończących lektorat w badanych wyższych uczelniach w Polsce (*cf.* Sendur 2016b).

„**Studium przypadku** to prezentacja konkretnej sytuacji z życia codziennego lub zawodowego przedstawiona za pośrednictwem specyficznych faktów, postaw i opinii, na podstawie których ma zostać podjęta decyzja”²⁵ (Fischer *et al.* 2011: 23 za: Kaiser 1983). Studium lub analiza przypadku (*case study*) pozwala na integrację wielu elementów kształtujących docelowe kompetencje związane z użyciem języka obcego. Podczas analizy przypadku stosuje się szereg różnych form pracy indywidualnej, pracy w parach lub grupowej. Uczestnicy zadania otrzymują dane, poszukują rozwiązań, wcielają się w różne role i wchodzą w interakcje z innymi uczestnikami (Gajewska i Sowa 2014: 196–197).

Podczas gdy wykorzystanie studium przypadku w nauczaniu języka obcego posiada wiele zalet (Breszka-Jędrzejewska 2019: 73), to jednak zadanie to nie jest szeroko wykorzystywane w językowych egzaminach międzynarodowych; nie jest obecne w żadnym z analizowanych do celów niniejszej publikacji egzaminów, oprócz projektu GULT, w którym to studium przypadku stanowi jeden z etapów pracy projektowej.

Symulacja globalna (*global simulation*) polega na symulowanym działaniu w wyznaczonych jej tytułem–tematem ramach, opartym na scenariuszu opisującym miejsce akcji, głównych bohaterów, materiały pomocnicze oraz oczekiwany produkt końcowy (Gajewska i Sowa 2014: 193). Symulacja angażuje w działanie całą grupę osób (nawet trzydziestoosobową klasę). Uczestnicy, na kanwie otrzymanego scenariusza, tworzą określoną społeczność, którą łączą wzajemne relacje (np. wcielają się w role mieszkańców bloku lub pracowników hotelu), a język obcy staje się narzędziem do realizacji określonych celów komunikacyjnych. Chociaż symulacja została pierwotnie zaprojektowana jako ćwiczenie

²⁵ Tłumaczenie własne – A.S.

lekcyjne, to według Nunana (2004: 156) może być wykorzystana w celach pomiaru. Podobnie jednak jak praca projektowa i studium przypadku nie znajduje ona zastosowania w analizowanych tutaj egzaminach.

Z omówionych czterech rodzajów zadań charakterystycznych dla podejścia zadaniowego tylko jedno znajduje szerokie zastosowanie w zawodowych egzaminach językowych – odgrywanie ról zawodowych. Pozostałe trzy – praca projektowa, studium przypadku i symulacja – nie występują w badanych dla celów niniejszej pracy egzaminach. Brak zastosowania tych technik w ocenianiu na szeroką skalę wynika z ich niskiego stopnia praktyczności. Zadania tego typu powinny być rozłożone w czasie, ponadto wymagają one współpracy kilku osób, co z kolei utrudnia ocenę indywidualnych umiejętności uczestników działania. Przeszkodą jest także potrzeba porównywalności oceny z innymi egzaminami, które mierzą poszczególne sprawności w sposób tradycyjny. W przypadku egzaminów przeprowadzanych na dużą skalę wykorzystanie tego typu zadań angażowałoby zbyt wiele sił i środków, niewspółmiernych do oczekiwanego rezultatu, co sprawia, że nie najlepiej nadają się do certyfikacji. Mogą one jednak z pełnym powodzeniem być wykorzystane w trakcie kursu, jako ćwiczenie językowe, oraz w bieżącej ocenie umiejętności.

Zadania w egzaminach GE i ESP – podsumowanie

W powyższym podrozdziale analizie poddano zadania testowe w wybranych egzaminach z zakresu ogólnego języka angielskiego oraz z egzaminów językowych dla celów specjalistycznych. W wyniku dokonanego przeglądu rysuje się kilka ogólnych wniosków dotyczących rodzajów zadań w egzaminach tych dwóch typów.

Pośród zadań wybranych do sprawdzania umiejętności językowych w formie pisemnej zdecydowanie najbardziej popularną techniką egzaminacyjną w badanych egzaminach okazuje się zadanie typu wybór wielokrotny. Jest ono stosowane powszechnie zarówno w testach języka angielskiego ogólnego, jak i w egzaminach ESP, do sprawdzania umiejętności w zakresie sprawności receptywnych oraz znajomości gramatyki i leksyki. Z wszystkich przeanalizowanych egzaminów tylko *English for Tourism* nie wykorzystuje w żadnym ze swoich testów zadań tego typu. Nie bez znaczenia jest tutaj obserwacja, że egzamin ten bada umiejętności językowe na podstawie zadań bardzo bliskich zadaniom zawodowym, do których WW nie należy.

Spośród form zamkniętych również dużą popularnością cieszą się różne rodzaje zadań na dobieranie. Wykorzystywane są one do sprawdzania sprawności słuchania (sTANDEM, *telc Dual-Level* i *First*) oraz do mierzenia sprawności czytania (*telc Dual-Level*). Uzupełnianie luk na poziomie akapitu jest obecne tylko w jednym z badanych egzaminów ESP (sTANDEM) oraz w egzaminie GE (*First*), natomiast uzupełnianie luk na podstawie selekcji z banku możliwości spotkać można tylko w egzaminie GE (*First*) oraz w egzaminach uczelnianych, które są bliższe egzaminom GE niż LSP (Sendur 2016b: 176).

Spośród zadań otwartych zdecydowaną przewagę nad innymi rodzajami ma zadanie typu krótkiej odpowiedzi oraz uzupełnianie zdań. Są one szeroko eksploatowane zarówno przez egzaminy GE, jak i ESP do sprawdzania umiejętności słuchania, czytania, a także w niewielkim stopniu pisanie. Zadanie typu rozwiniętej odpowiedzi jest obecne w różnej formie w większości omawianych testów, pod warunkiem, że sprawność pisanie znajduje się wśród badanych przez dany egzamin aspektów języka.

Istnieją także zadania, które w ogóle nie pojawiają się w badanych testach ESP, są natomiast obecne w testach GE. Są to głównie jednostki sprawdzające kompetencję strukturalną, takie jak zadanie z lukami z wyborem wielokrotnym, *cloze* (wypełnianie luk otwartych), uzupełnianie luk na podstawie selekcji z banku, transformacje i słowotwórstwo. Wymienione typy zadań pojawiają się w egzaminach *First*, *telc* oraz niektórych egzaminach uczelnianych.

Istnieje też szereg zadań wymienianych przez ALTE (2005) jako zadania testowe, które jednak w ogóle nie występują w analizowanych arkuszach egzaminacyjnych, pojawiają się natomiast jako ćwiczenia językowe w podręcznikach do języków obcych. Niektóre z nich znajdujemy także w egzaminach o mniejszym zasięgu, jak na przykład egzaminy eksternistyczne CKE (Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne 2015a). Przykłady takich zadań to przekształcenia z lukami, rozsypanka wyrazowa, *note expansion*, wykrywanie błędów (technika eliminacji zbędnych wyrazów) i ich poprawianie.

Analiza egzaminów ustnych pokazuje, że wykorzystywane w nich zadania mają rozmaite formy, w tym łączące różne rodzaje zadań ustnych. Na przykład prezentacja może być oparta na ikonicznych bądź tekstowych materiałach stymulujących; odgrywanie ról może zawierać elementy zadania z luką informacyjną. Wszystkie z omówionych typów zadań badających sprawność mówienia znajdują zastosowanie w badanych do

celów niniejszej pracy egzaminach. Jednak zdecydowanie najczęściej spotykaną formą testowania języka mówionego w przeanalizowanych testach jest odgrywanie ról (PTE, ELPAC, sTANDEM, EfT, OET) oraz prezentacja (telc, STANAG, MOTET i egzaminy uczelniane).

Mimo że w ostatnich latach dużo mówi się o zadaniowym podejściu do nauczania języków obcych, zwłaszcza w kontekście języków zawodowych, takie podejście do egzaminowania (*Task-based Assessment* – TBA) nie jest szeroko rozpowszechnione. Świadczą o tym między innymi rodzaje zadań w badanych egzaminach ESP. W wielu z nich zadania egzaminacyjne są zadaniami zamkniętymi, co wyraźnie kłóci się z teorią podejścia zadaniowego, która mówi, że nauczanie, a więc i testowanie, powinno być bliższe życiu, realistyczne, winno bazować na specyficznych rzeczywistych sytuacjach, opartych na faktach i autentycznych dokumentach. Testowe zadania zamknięte nie są autentyczne. Również sprawdzane umiejętności dość często nie przekładają się na działanie, na wykonywanie zadań zawodowych. W kilku egzaminach ESP (np. MOTET, EfA) oraz egzaminach uczelnianych aspirujących często do miana testów ESP/LSP kompetencja strukturalna sprawdzana jest bezpośrednio przez zamknięte i otwarte zadania gramatyczno-leksykalne, podobnie jak ma to miejsce w testach GE. Są jednak także egzaminy ESP, które odchodzą od bezpośredniego testowania znajomości leksyki oraz poprawności gramatycznej na rzecz mierzenia tych umiejętności na podstawie symulowanych działań zawodowych, jak na przykład tworzenie tekstu pisanego czy interakcja w formie ustnej. W tych przypadkach arkusz egzaminacyjny nie zawiera zadań gramatyczno-leksykalnych, natomiast zakres bogactwa językowego i poprawności jest brany pod uwagę w kryteriach oceny wypowiedzi pisemnej i ustnej (np. OET, STANAG, EfT, sTANDEM, ELPAC).

Złożone zadania językowe wykorzystywane w nauczaniu w podejściu zadaniowym, jak symulacja globalna, praca projektowa czy studium przypadku, nie znajdują zastosowania w analizowanych egzaminach przeprowadzanych na szeroką skalę. Powodem może być z jednej strony czasochłonność wykonania zadania i trudność oceny, z drugiej zaś potrzeba porównywalności wyników z innymi egzaminami, które mierzą poszczególne sprawności w sposób tradycyjny. Tak więc, nawet w egzaminach identyfikujących się jako TBA, część umiejętności sprawdzana jest w sposób tradycyjny, między innymi za pomocą zadań zamkniętych.

8. Ocenianie i egzaminowanie w glottodydaktyce ogólnej i specjalistycznej – podobieństwa i różnice

Głównym celem nauczania zarówno języków ogólnych, jak i specjalistycznych jest przygotowanie uczących się do kontaktów z użytkownikami danego języka obcego. W obydwu przypadkach, oprócz rozwijania kompetencji językowej, nacisk stawia się na rozwijanie kompetencji komunikacyjnej. W omawianych dwóch glottodydaktykach wskazane jest posługiwanie się autentycznymi materiałami. Zarówno w nauczaniu języka ogólnego, jak i języka dla celów specjalistycznych rolą nauczyciela jest stworzenie w klasie warunków podobnych do naturalnego środowiska komunikacyjnego, co ma sprzyjać pragmatycznemu aspektowi nauczania. Różnica polega na kontekście kontaktów, w których używany jest język obcy – w zależności od sytuacji jest to kontekst ogólny albo zawodowy.

Wśród zasadniczych cech odróżniających nauczanie języka dla celów specjalistycznych od nauczania języka ogólnego Mangiante i Parpette (Gajewska i Sowa 2014: 29) wymieniają dokładnie określony przez odbiorcę wąski i konkretny cel nauczania/uczenia się, często pilną potrzebę uczącego się, a w konsekwencji – stosunkowo krótki czas nauki, skupienie się – głównie albo wyłącznie – na wybranych sytuacjach oraz na kluczowych dla odbiorców kompetencjach. Nauczyciela takiego kursu niejednokrotnie charakteryzuje nieznanomość bądź jedynie częściowa znajomość nauczanych zagadnień fachowych. Często brakuje także gotowych materiałów dydaktycznych. W konsekwencji niezbędna okazuje się współpraca nauczyciela ze specjalistami z danej dziedziny lub korzystanie z ich pomocy.

W glottodydaktyce języka ogólnego egzamin opracowują specjaliści w dziedzinie językoznawczej. W przypadku egzaminu z języka specjalistycznego nad kształtem, treściami oraz kryteriami oceny pracować powinni wspólnie przedstawiciele trzech grup eksperckich: po pierwsze – ekspert w dziedzinie oceniania; po drugie – osoba posiadająca wiedzę z zakresu danej dziedziny, która może być źródłem autentycznych materiałów oraz wiedzy na temat codziennej komunikacji w określonej dziedzinie; po trzecie – nauczyciel, który jest specjalistą od treści i zawartości merytorycznej tworzonych tekstów (Bartosik 2019: 61).

Biorąc pod uwagę specyfikę nauczania języków specjalistycznych, tworzenie programu nauczania dla tej dziedziny będzie też nieco odmienne niż tworzenie programów dla języków ogólnych. Według Strevensa (Gajewska i Sowa: 28–29), tworzenie programu nauczania języka specjalistycznego dokonuje się poprzez wybór i skupienie się na sprawnościach językowych bezpośrednio związanych ze specyficznymi potrzebami ucznia (np. kursy jednosprawnościowe) oraz selekcję wyłącznie tych jednostek leksykalnych, elementów gramatyki (tzw. kursy skośne – Komorowska 2005b: 42) i funkcji językowych oraz dobór tematów i zagadnień, które wiążą się z tymi potrzebami.

O testowaniu dla celów specjalistycznych pisze Douglas (2000). Dzieląc testy na takie, które mierzą umiejętności w języku ogólnym, oraz takie, które sprawdzają biegłość w zakresie języka specjalistycznego, należy wziąć pod uwagę spostrzeżenie tego autora, który mówi, że:

[...] testy nie są albo testami dla celów ogólnych, albo testami dla celów specyficznych – wszystkie testy są stworzone dla jakiegoś celu; istnieje pewne kontinuum poziomu specyficzności od bardzo ogólnego do bardzo specyficznego, a każdy test może plasować się w dowolnym punkcie tej osi²⁶. (Douglas 2000: 1)

Stwierdzenie to jest bardziej czytelne w przypadku terminologii dotyczącej tych dwóch odmian języka w języku angielskim niż polskim. Wynika to z różnicy w terminie przeciwstawnym do *języka ogólnego* w języku polskim i angielskim, co zostało szerzej omówione w innym miejscu. Powszechnie stosowany polski termin *język specjalistyczny* nie oddaje znaczenia, które nosi wyrażenie z języka angielskiego – *języki do celów specyficznych* (*languages for specific purposes* – LSP), a właśnie do stopnia specyficzności języka związanego z różnymi zawodami i potrzebami odnosi się obserwacja Douglasa.

Zatem skoro nie ma jasno wytyczonej granicy pomiędzy testami do celów ogólnych oraz do celów specyficznych, to czy istnieją różnice pomiędzy testowaniem w tych dwóch glottodydaktykach? Jeśli tak, to z czego te różnice wynikają i na czym polegają? Odpowiedzi na te pytania zostały zawarte w poniższym podrozdziale.

²⁶ Tłumaczenie własne – A.S.; oryginalny fragment tekstu: „[...] tests are not either general purpose or specific purpose – all tests are developed for some purpose – but that there is a continuum of specificity from very general to very specific, and a given test may fall at any point on the continuum”.

Ocenianie i egzaminowanie w glottodydaktyce języków ogólnych i LSP – podobieństwa

Ocenianie umiejętności językowych w dydaktyce specjalistycznej w swoich założeniach nie różni się od oceniania w innych obszarach nauczania języka (Sendur 2020: 50). Osoby tworzące test muszą wziąć pod uwagę jego cel, charakterystykę zdającego oraz docelowe sytuacje użycia języka. Każdy test, bez względu na to, czy dotyczy języka ogólnego czy LSP, powinien stosować się do omówionych wcześniej podstawowych zasad jakości badania, jakimi są rzetelność, trafność, praktyczność i wpływ zwrotny testu (Bachman i Palmer 2002: 19–36; Douglas 2013: 367; Komorowska 2007: 22–31). Zarówno test z języka ogólnego, jak i egzamin kierowany do fachowców określonej branży wymaga od jego twórcy zdefiniowania celu, przeprowadzenia analizy potrzeb, zebrania danych dotyczących użycia języka w kontekście, analizy docelowych zadań komunikacyjnych i stworzenia takich zadań egzaminacyjnych, które będą odzwierciedlały zadania docelowe (Douglas 2013: 367–368).

Omówione w niniejszym rozdziale rodzaje testów i zadań egzaminacyjnych dotyczą zarówno testowania i egzaminowania w dydaktyce tzw. języków ogólnych, jak i języków do celów specjalistycznych czy też specyficznych. W obydwu glottodydaktykach korzystamy z różnych rodzajów testów w zależności od celu, dla którego danego narzędzia potrzebujemy. Testy te mogą mieć formę pisemną lub ustną; mogą sprawdzać znajomość elementów języka bądź badać poszczególne sprawności; mogą przyporządkować uczniów do grup kursowych albo diagnozować dalsze potrzeby rozwoju umiejętności. Podczas trwania kursu LSP, analogicznie do kursów z języka ogólnego, nauczyciel wykorzysta testy postępów i osiągnięć, a uczący się będzie w pewnym momencie nauki chciał lub musiał sprawdzić swoją biegłość w zakresie czy to języka ogólnego, czy też specjalistycznego.

Jak wykazano w podrozdziale 1.6., zarówno w testach z języka ogólnego, jak i egzaminach LSP wykorzystywany jest cały wachlarz różnego rodzaju zadań testowych, choć daje się zauważyć, że pewne typy zadań są zdecydowanie częściej spotykane w testach GE niż w egzaminach LSP, a kilka z nich w ogóle w badanych testach LSP się nie pojawia.

Zasady konstruowania testu są takie same, bez względu na to, czy jest to test z języka ogólnego czy LSP. Chcąc przygotować profesjonalny egzamin, należy najpierw stworzyć jego projekt, a więc na podstawie analizy potrzeb opracować specyfikację testu, a następnie sporządzić jego plan.

Konstruując arkusz egzaminacyjny, trzeba wpieryw dokonać selekcji treści, następnie zbudować pojedyncze zadania, opracować klucz do zadań i przeprowadzić pilotaż. Wreszcie, opracowany egzamin i arkusz testu można wdrożyć, a następnie go monitorować.

Podsumowując, istnieje cały szereg cech, które nie odróżniają testowania w zakresie języków ogólnych od testowania języków do celów specjalistycznych, a które są istotne w testowaniu w obydwu glottodydaktykach.

Ocenianie i egzaminowanie w glottodydaktyce języków ogólnych i LSP – różnice

Mimo wielu cech wspólnych istnieją jednak różnice w szczegółach dotyczących testowania języka ogólnego oraz języków dla celów specjalistycznych. Jako źródło tych różnic Douglas (2000: 6–8; 2013: 368–369) wskazuje charakterystyczne cechy języków dla celów specyficznych. Po pierwsze, **użycie języka zmienia się w zależności od kontekstu**. Po drugie – **język dla celów specyficznych jest precyzyjny**.

Przykładem na to, że użycie języka zmienia się zależnie od kontekstu, jest użycie dyskursu specjalistycznego przez lekarzy. Inny będzie język użyty przez medyka w relacji z pacjentem, inny zaś w przypadku, gdy rozmówcą jest drugi lekarz, nawet gdy temat rozmowy będzie zbieżny. Mimo to w obydwu przypadkach rodzaj dyskursu zakwalifikujemy do ogólnej kategorii języka medycznego (Douglas 2013: 368).

Charakterystyką użycia języka w kontekście zawodowym jest również jego wysoka precyzja. Mówimy tu o tzw. żargonie zawodowym, który musi zostać opanowany przez fachowców z danej dziedziny w ramach przygotowania do zawodu. To, co dla niefachowca będzie *małym palcem lewej nogi*, dla specjalisty z dziedziny medycyny jest *piątym palcem lewej stopy* (nb. słowo *noga* w ogóle nie istnieje w polskim żargonie medycznym); gdy pacjent mówi, że ma *dziurę w górnej szóstce*, dla stomatologa jest to *ubytek w pierwszym trzonowcu górnym prawym/lewym*.

a) Cechy testów dla potrzeb specyficznych

Wymienione powyżej cechy języków specjalistycznych czy specyficznych generują podstawowe różnice pomiędzy testowaniem dla celów ogólnych i dla celów LSP. Dwie zasadnicze cechy testowania LSP, które odróżniają ten rodzaj testów od sprawdzianów języka ogólnego według Douglasa (2000: 2–18; 2013: 369–371), to **autentyczność zadań** oraz **związek pomiędzy wiedzą językową i wiedzą specjalistyczną**.

Wprawdzie zadania egzaminacyjne nigdy nie będą w pełni autentyczne, gdyż sam fakt, że odbywają się w sytuacji egzaminacyjnej, na to nie pozwala (Douglas 2013: 371), jednak w każdym teście, niezależnie od tego, czy sprawdza język raczej ogólny, czy bardziej fachowy, zadania egzaminacyjne powinny być maksymalnie zbliżone do zadań, które zdający może w przyszłości wykonywać z użyciem języka obcego. Zadania testowe powinny naśladować autentyczne sytuacje. Nie jest zatem zasadne na przykład przydzielenie uczniowi na ustnym teście szkolnym roli, w której ma prowadzić symulowaną rozmowę z własnym rodzicem w języku obcym, gdyż taka sytuacja w rzeczywistości ma nikłe szanse zaistnienia (Sendur 2020: 51). Nie jest to sytuacja autentyczna i tym samym nie powinna znaleźć się w egzaminie z języka obcego. Kwestia autentyczności zadań ma jeszcze istotniejsze znaczenie w przypadku egzaminów LSP. W tego rodzaju sprawdzianach zadania konstruowane na potrzeby ewaluacji powinny być całkowicie zbieżne z zadaniami, które zdający spotyka lub spotka w przyszłości w swojej praktyce zawodowej. To zbliżenie zadań egzaminacyjnych do zadań nieegzaminacyjnych (zawodowych) ma na celu zwiększenie prawdopodobieństwa, że zdający wykona zadanie testowe w taki sam sposób, jakby je wykonał w autentycznej sytuacji zawodowej (Sendur 2020).

Widdowson (1984: 258) i Douglas (2000: 16–18; 2014: 24–26) podkreślają różnicę pomiędzy pojęciami autentyczności użycia języka, a oryginalności zadań, które bywają mylone. Zadanie testowe może być oparte na oryginalnych materiałach, ale niekoniecznie będzie ono spełniało kryterium autentyczności. O autentyczności zadania nie świadczy bowiem użyty w nim materiał, tylko sposób jego wykorzystania w celu wykonania zadania. Innymi słowy, w egzaminie LSP, tekst powinien zostać wykorzystany do rozwiązania problemu zawodowego, a nie na przykład do testowania znajomości gramatyki czy słownictwa. Sprawdzanie rozumienia oryginalnego tekstu za pomocą zadań wyboru wielokrotnego jest również przykładem zadania egzaminacyjnego, które nie jest autentyczne, gdyż prawdopodobieństwo, że zdający w pracy zawodowej będzie rozwiązywał zadania zamknięte oparte na tekście, jest znikome.

W podręczniku Stowarzyszenia ALTE – *Manual for Language Test Development*, stworzonym na zlecenie Rady Europy, autorzy wprowadzają rozróżnienie pomiędzy autentycznością sytuacyjną a interakcyjną (2011: 12).

Autentyczność sytuacyjna dotyczy stopnia, w jakim zadania egzaminacyjne odzwierciedlają działania językowe podejmowane w rzeczywistej sytuacji komunikacyjnej. **Autentyczność interakcyjna** z kolei odnosi się do stopnia naturalności interakcji pomiędzy zdającym a zadaniem oraz procesy mentalne, które temu towarzyszą. Przykładem zadania autentycznego sytuacyjnie może być sprawdzian rozumienia ze słuchu w teście dla pracowników branży turystycznej, w którym zdający muszą wysłuchać długoterminowej prognozy pogody. Aby dać zdającemu rzeczywisty powód do wsłuchania się w informacje i sprawić, że zadanie egzaminacyjne będzie cechowało się autentycznością interakcyjną, rolą zdającego może być opracowanie planu wycieczki z uwzględnieniem informacji dotyczących pogody.

Mimo że często teksty egzaminacyjne będą musiały być adaptowane ze względu na badany poziom biegłości, a tym samym przestaną być autentyczne z punktu widzenia językowego, to sytuacja, w której zdający będzie używał języka i jego interakcja z tekstem bądź z innym zdającym czy egzaminatorem, może spełniać warunki autentyczności.

Nie należy zapominać o autentyczności również w kontekście oceniania. Kryteria oceny wykonania zadania powinny być oparte na opiniach ekspertów z branży dotyczących zasady postępowania w autentycznych sytuacjach zawodowych (Douglas, 2000: 67–71; Hamp-Lyons i Lumley, 2001: 131).

Jeszcze bardziej wyrazistą cechą egzaminów LSP jest interakcja pomiędzy znajomością języka i wiedzą specjalistyczną. W testach sprawdzających ogólną biegłość językową wiedza ogólna nie powinna mieć wpływu na poprawność wykonywanych zadań; jest elementem, który może mieć negatywny wpływ na wyniki pomiaru. W tym przypadku jej wpływ na poprawne rozwiązanie zadania powinien być ograniczony do minimum. Tymczasem w testowaniu LSP znajomość zagadnień specjalistycznych jest niezbędna i odgrywa zasadniczą rolę w badaniu umiejętności językowych z danej dziedziny (Douglas 2000: 29–32). Mimo że celem testowania LSP nie jest ocena kompetencji zawodowych, to kompetencje te i wiedza zawodowa są nierozzerwalnie związane z wykonaniem językowym w danym obszarze zawodowym (Douglas 2013: 369).

b) Kompetencje oceniane w ewaluacji dla potrzeb specjalistycznych

W glottodydaktyce ogólnej kształcenie i ocenianie zakłada równomierne opanowanie wszystkich sprawności, bez względu na zainteresowania

i potrzeby uczącego się (Komorowska 2019: 159). I tak w egzaminach z języków ogólnych prawie zawsze mamy do czynienia z testami sprawdzającymi wszystkie cztery główne sprawności językowe, a ich waga w obrębie całego egzaminu jest podobna lub taka sama. W przypadku języków specjalistycznych poszczególne sprawności nie muszą być opanowane na tym samym poziomie (Gajewska 2008: 79–80; Komorowska 2005b: 42) i w związku z tym różne może być ich znaczenie w egzaminach LSP. W zależności od potrzeb i wymogów danej branży zawodowej w nauczaniu i egzaminowaniu w zakresie języka obcego mówimy o **kompetencjach i sprawnościach priorytetowych**, które mogą być ograniczone lub cząstkowe, co oznacza, że ich waga w poszczególnych egzaminach będzie różna, a w niektórych przypadkach pewne sprawności mogą w ogóle nie występować. Wprawdzie Janine Courtillon (Gajewska i Sowa 2014: 207) uważa rozumienie ze słuchu i sprawność czytania za podstawowe w procesie nabywania każdego języka obcego i w związku z tym za podlegające ocenie niezależnie od celu podejmowanej nauki, to jednak istnieją przypadki egzaminów LSP, które zrezygnowały nawet z tych sprawności. Znamienitymi przykładami są omówiony w niniejszej publikacji egzamin ELPAC dla pilotów i kontrolerów ruchu lotniczego oraz jego polski odpowiednik – egzamin Krajowego Systemu Egzaminów Językowych (KSEJ; Sendur 2020), które w ogóle nie zawierają elementu sprawdzającego sprawności czytania. W kilku innych egzaminach LSP wybór poszczególnych części egzaminu, a tym samym badanych sprawności, oddany został do decyzji zdających. W przypadku egzaminu sTANDEM dla lekarzy, pielęgniarek i farmaceutów zdający wybiera pomiędzy testem sprawdzającym wyłącznie sprawności czytania i pisanie oraz testem badającym mówienie i słuchanie. Może oczywiście również zdecydować o przystąpieniu do obydwu testów i uzyskać dwa niezależne certyfikaty (*Reading and Writing* oraz *Speaking and Listening*). Podobnie wygląda sytuacja w przypadku egzaminu z technicznego języka angielskiego MOTET, w którym każdy zdający musi zaliczyć obowiązkowy moduł ogólny, natomiast pozostałe elementy testu są opcjonalne. W zależności od potrzeb lub też oceny własnych umiejętności kandydat wybiera tylko test sprawdzający umiejętność pisanie, oddzielny egzamin sprawdzający sprawność mówienia lub też obydwu z wymienionych komponentów.

W testowaniu LSP mamy również do czynienia z zadaniami ewaluacyjnymi, które łączą w sobie różne sprawności. Zadania takie uwzględ-

niąją **integrację kilku sprawności** – istotnych dla danej dziedziny (Koromorowska 2005b: 43) – mobilizowanych na rzecz działania zawodowego, które razem warunkują wykonanie zadania i podlegają wspólnej ocenie (Dudley-Evans i St John 1998: 226; Gajewska i Sowa 2014: 208–209; Sendur 2020: 52).

W przypadku języków specjalistycznych ocena samej kompetencji językowej nie jest wystarczająca. Przy ewaluacji umiejętności językowych należy wziąć też pod uwagę kompetencję zawodową, w skład której wchodzi kompetencja interkulturowa i wiedza zawodowa. Mówimy w tym wypadku o ocenie **podwójnej kompetencji**. W praktyce oznacza to, że poprawna językowo wypowiedź nie zawsze będzie oceniona jako prawidłowa w danym kontekście zawodowym, musi ona bowiem uwzględniać również przyjęte w danym środowisku zawodowym sposoby postępowania (Gajewska i Sowa 2014: 208–209.).

c) Cechy zadań w testach dla potrzeb specjalistycznych

Bachman (2003: 118–152) przedstawia listę kategorii metod testowania komunikacyjnego. Zgodnie z jego taksonomią są to: środowisko testu (*testing environment*), instrukcja (*rubric*), dane wejściowe (*input*), oczekiwana odpowiedź (*expected response*) oraz związek pomiędzy danymi wejściowymi i spodziewaną odpowiedzią (*relationship between input and response*). Za Bachmanem swoją charakterystykę zadań w testach języka obcego zawodowego formułuje Douglas (2000: 50–74). Opisane przez niego elementy budowy zadań w testach LSP to: instrukcja, dane wejściowe, oczekiwana odpowiedź, interakcja pomiędzy danymi wejściowymi i spodziewaną odpowiedzią oraz zasady oceny. Poniżej przedstawiono i pokrótce opisano cechy testów LSP według tego autora.

Używane w języku angielskim słowo *rubric* w odniesieniu do zadań ewaluacyjnych to więcej niż w języku polskim *instrukcja* czy *polecenie do zadania*. W skład tego elementu zadania (Bachman i Palmer 2002: 50–52) wchodzi nie tylko sama instrukcja dotycząca tego, co zdający ma wykonać. Składa się na nią również opis testu, informacja dotycząca czasu przeznaczanego na wykonanie testu, jego części lub pojedynczego zadania, celu zadania, sposobu i formy udzielenia odpowiedzi, struktury testu lub zadania oraz kryteriów oceny. Dla celów niniejszego opracowania, w odniesieniu do angielskiego słowa *rubric*, stosowanego w opisanym tu kontekście, używany jest termin *instrukcja*. Należy jednak zwrócić uwagę na inne znaczenie i użycie angielskiego terminu *rubric*, które stosowa-

ne jest również w odniesieniu do *kryteriów oceny* (cf. Dawson 2017; The Glossary of Educational Reform²⁷).

Sposób udzielania odpowiedzi to informacja, w jaki sposób zdający powinien wykonać zadanie, np. „zakreśl prawidłową odpowiedź”, „zaznacz odpowiednie kratki”, „wypełnij tabelę” czy „wciśnij odpowiedni klawisz” w przypadku zadania wykonywanego na komputerze (cf. przykład 7 i 8). Informacja dotycząca struktury zadania powinna zawierać takie dane jak liczba podzadań czy też ich względna waga (cf. przykład 7). Przeznaczony czas może dotyczyć zarówno konkretnego zadania, jak też całej części egzaminu (cf. przykład 7 i 9). Instrukcja powinna również zawierać informacje dotyczące oceniania i punktowania poszczególnych zadań (cf. przykład 7). Zdający powinien się z niej dowiedzieć, co będzie brane pod uwagę przy ocenie, jaka jest punktacja danego zadania itp.

Przykład 7. Instrukcja na stronie tytułowej części *Reading and Use of English* egzaminu *Cambridge English First* (Cambridge English Language Assessment 2015a: 1)

Time 1 hour 15 minutes

INSTRUCTIONS TO CANDIDATES

Do not open this question paper until you are told to do so.

Write your name, centre number and candidate number on your answer sheet if they are not already there.

Read the instructions for each part of the paper carefully.

Answer all the questions.

Read the instructions on the answer sheet.

Write your answers on the answer sheet. Use a pencil.

You **must** complete the answer sheet within the time limit.

At the end of the test, hand in both this question paper and your answer sheet.

INFORMATION FOR CANDIDATES

There are 52 questions in this paper.

Questions **1 – 24** and **43 – 52** carry one mark.

Questions **25 – 30** carry up to two marks.

Questions **31 – 42** carry two marks.

²⁷ Zob. <https://www.edglossary.org/rubric> [dostęp: 16.12.2020].

Przykład 8. Instrukcja do zadania w części *Use of English* egzaminu *Cambridge English First* (Cambridge English Language Assessment 2015a: 2)

Part 1

For questions 1 – 8, read the text below and decide which answer (A, B, C or D) best fits each gap. There is an example at the beginning (0).


Mark your answers **on the separate answer sheet**.

Example:


0 A band B set C branch D series

0	A	B	C	D
<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Przykład 9. Instrukcja do części sprawdzającej sprawność słuchania próbnego testu dla pilotów ELPAC²⁸



English Language Proficiency for Aeronautical Communication



Instructions Sample test - part 1

You will hear a number of short controller messages to pilots. Answer the questions in the spaces provided.

There is an example at the beginning.

There will be a 10 second pause at the end of each communication for you to complete your answers.

You now have 15 seconds to look at the questions before the recording starts.

W autentycznej sytuacji zawodowej informacje dotyczące sposobu wykonania zadania zawodowego są często implicytne, niewymagające sprecyzowania ich wprost. Pracownik sam wie, w jaki sposób powinien przekazać współpracownikom czy klientowi określoną informację. W przypadku zadań testowych informacje te muszą być jasno określone i niepozostawiające żadnych wątpliwości dotyczących sposobu wykonania zadań oraz innych wymagań z nimi zawiązanych.

²⁸ Próbny arkusz egzaminacyjny dostępny na stronie Eurocontrol: <http://www.eurocontrol.int/services/english-language-proficiency-aeronautical-communication-elpac> [dostęp: 20.02.2015].

Dane wejściowe²⁹ (*input*) to materiał związany z użyciem języka w docelowych zawodowych sytuacjach komunikacyjnych, który ma być przetworzony przez użytkowników języka i który powinni oni uwzględnić podczas wykonywania zadania. W przeciwieństwie do instrukcji, która jest niezależna od kontekstu zawodowego, dane wejściowe w teście LSP są ściśle związane z kontekstem specjalistycznym. Dane te składają się z opisu tła sytuacyjnego (*prompt*) oraz z materiału stymulującego (*input data*).

Tło sytuacyjne to specyficzne składowe sytuacji komunikacyjnej. Informacje dotyczące kontekstu związanego ze specyficznymi celami są niezbędne, aby użytkownik języka mógł zaangażować się w zadanie komunikacyjne. Obejmują one np. określenie miejsca i czasu sytuacji, jej uczestników, celu oraz inne cechy zdarzenia komunikacyjnego, jak forma, ton wypowiedzi, normy interakcji czy gatunek tekstu.

Materiał stymulujący to autentyczny materiał w formie wizualnej lub dźwiękowej, który użytkownik języka musi przetworzyć podczas wykonywania zadania. Mogą to być teksty, wykresy, tabele, zdjęcia, obrazki, nagrania, gesty interlokutora czy też autentyczne przedmioty, jak narzędzia, urządzenia lub sprzęt laboratoryjny.

Przykład danych wejściowych składających się z tła sytuacyjnego (*Situation 1*) i materiału stymulującego w egzaminie *Written English for Tourism (WEfT) Level 1* przedstawia przykład 10. Z opisu sytuacji (*Situation 1*) dowiadujemy się, gdzie ma ona mieć miejsce (recepcja Hotelu „Four Seasons” w Hong Kongu), jakiego terminu dotyczy (goście przyjadą za dwa dni), wiemy, że uczestnikami będzie dwudziestoposobowa grupa biznesmenów z Londynu, którzy uczestniczą w zjeździe przedsiębiorców w Hong Kongu. Celem postawionym przed zdającym jest napisanie listu powitalnego skierowanego do spodziewanych gości, uwzględniającego szereg konkretnych informacji (usługi dostępne dla gości, w tym restauracja i zaplecze rekreacyjne).

Materiał stymulujący w prezentowanym przykładzie zawiera informacje tekstowe i zdjęciowe dotyczące oferty hotelu i restauracji hotelowej, wyposażenia sal konferencyjnych oraz zaplecza rekreacyjnego.

²⁹ Dla celów niniejszego tekstu tłumaczenie słowa *input* w kontekście struktury zadań egzaminacyjnych zostało przyjęte za Gajewską i Sową (2014) jako *dane wejściowe*; w polskiej literaturze tematu funkcjonuje również termin *materiał wyjściowy* (Gaszynska-Magiera i Seretny 2004: 115), który odnosi się do tej samej cechy zadania testowego.

Przykład 10. Przykład danych wejściowych egzaminu *Written English for Tourism (WEFT) Level 1* (LCCI International Qualifications 2010b: 2–4)

SITUATION 1

You work on the reception desk of the Four Seasons Hotel in Hong Kong, China. A group of 20 businessmen from London are attending a trade convention in Hong Kong and will arrive in your hotel 2 days from now on Saturday 31 Jan 2011.

Your manager would like you to write a welcome letter for the group. He has asked you to include essential information on the services available to them during their stay including restaurant and leisure facilities.

TASK 1

Read the information below and write the welcome letter with a summary of the services, which you feel would be of interest to your guests and information on restaurant choices and leisure facilities.

**Restaurant Facilities at
The Four Seasons**

The Morning Room: *Hot and cold buffet breakfast 7-10 am*

Location: *lower ground floor*

Caprice Restaurant: *a la Carte and table d'hote lunch and dinner*



Lunch **12:00 noon - 2:30 pm**

Dinner **6:00 pm - 10:30 pm***

***Closing time for last order**

With superb views over Victoria Harbour and the Kowloon Peninsula from its location on the sixth floor, Caprice's stunning surroundings feature Chinese-inspired Czech crystal chandeliers. With an open kitchen that provides a show of chefs at work, the restaurant offers contemporary French cuisine accompanied by the finest French wines, as well as other European and New World wines from renowned vineyards. Another highlight is the exclusive Caprice cellar, with Hong Kong's widest selection of French cheeses. Caprice also offers an adjoining bar and lounge area for cocktails, and three private dining rooms with seating for up to 16 guests.

24 hour room service is also available. Please see our special menu.

Business amenities and services

Around the clock, Four Seasons can assist you with business, entertainment or travel arrangements through the Business Centre and our resourceful, multilingual Concierge.

- **24-hour business centre**
- **Business equipment**
- **Computers**
- **Executive Club**
- **Meeting and function space**
- **Secretarial service**
- **Wired or wireless Internet access in meeting rooms**
- **Newspapers delivered to room or collect at reception**
- **Free phone for taxis in reception area**
- **Safety deposit boxes available**



Fitness Centre

The 24-hour Fitness Centre overlooks spectacular Victoria Harbour from floor-to-ceiling windows. It offers the latest in cardiovascular equipment – with each machine featuring an individual LCD touch-screen television and headphones – as well as an extensive selection of weight-training machines and free weights.

Tai chi, yoga and pilates classes are also available. Please enquire at the Fitness Centre for details.

Tennis

Victoria Park Tennis Centre

Victoria Park in Causeway Bay is the largest park on Hong Kong island, with six standard tennis courts available for rent to the public. The Tennis Centre Court, with a spectator seating capacity of 3,611, has been host to numerous international tournaments since 1981.

Details

Estimated travel time: 15 minutes drive

Speak to Concierge if interested

You may write your letter here.

**The Four Seasons Hotel
8 Finance Street
Hong Kong Central
China
Tel: 800 819505 3
Fax: 852 3196 8888
Website: www.fourseasons.com**

Oczekiwana odpowiedź to założenia osób tworzących test dotyczące czynności, które zdający winni wykonać w reakcji na opisaną w instrukcji i obudowaną danymi wejściowymi sytuację docelowego użycia języka. Odpowiedź może mieć formę ustną lub pisemną, nagrania audio lub wideo, wydruku komputerowego czy prezentacji multimedialnej, a nawet gestu, czynności lub szeregu różnych czynności i wymagać użycia sprawności receptywnych i/lub produktywnych. Zdający może mieć do wykonania zadanie związane z dokonaniem wyboru lub też zadanie krótkiej albo rozbudowanej odpowiedzi. W zaprezentowanym przykładzie rolę zdającego jest napisanie listu z wykorzystaniem odpowiednich informacji ujętych w danych wejściowych.

Mówiąc o **interakcji pomiędzy danymi wejściowymi i spodziewaną odpowiedzią**, Douglas (2000: 63–67) podaje, w jaki sposób dane wejściowe i spodziewana odpowiedź nawzajem na siebie wpływają. Odbywa się to w trzech obszarach: wzajemnej reakcyjności, zakresu i bezpośredniości.

Reakcyjność (*reactivity*) odnosi się do informacji zwrotnej otrzymanej podczas takiej interakcji. Jest to stopień, w jakim dane wejściowe mogą ulec zmianie na skutek odpowiedzi udzielonych przez użytkownika języka. Obejmuje ona zakres od odwzajemnionego do nieodwzajemnionego. W naturalnej rozmowie interakcja pomiędzy informacjami wejściowymi a odpowiedzią jest zazwyczaj wysoce odwzajemniona. Interlokutorzy, sygnałami werbalnymi i niewerbalnymi, przekazują sobie nawzajem informację, czy i w jakim stopniu przekazany komunikat został zrozumiany, czy wymaga wyjaśnienia lub doprecyzowania. Na skutek uzyskanych w ten sposób sygnałów zmienia się dalszy przekaz. Jeśli informacja jest dla odbiorcy niezrozumiała, następuje modyfikacja pewnych elementów wypowiedzi.

W zadaniach testowych stopień reakcyjności może być zróżnicowany, podobnie jak ma to miejsce w sytuacjach zawodowych. W przypadku sprawdzianu pisania zdający nie otrzymuje w trakcie konstruowania tekstu żadnej informacji zwrotnej dotyczącej zaprezentowanych przez siebie informacji, a więc reakcyjność jest nieodwzajemniona. O odwzajemnionej reakcyjności w odniesieniu do egzaminu możemy ewentualnie mówić w przypadku jego części ustnej, w której rozmówca w jakiś sposób reaguje na wypowiedź zdającego. Jednak i w tym przypadku reakcyjność będzie mniejsza niż w rzeczywistej sytuacji życiowej. W większości formalnych egzaminów osoba egzaminująca nie powinna dawać po sobie poznać stopnia satysfakcji z udzielonych odpowiedzi, więc zdający może błędnie interpretować brak ewentualnych sygnałów ze strony egzaminatora – jako potwierdzenie poprawności udzielonej odpowiedzi. W takiej sytuacji reakcyjność jest raczej nieodwzajemniona.

Zakres (*scope*) interakcji pomiędzy danymi wejściowymi i spodziewaną odpowiedzią odnosi się do ilości lub różnorodności danych wejściowych, które zdający musi przetworzyć przed wykonaniem zadania. W testach LSP zasięg jest zazwyczaj dość wąski ze względu na ograniczenia czasowe oraz mnogość różnych rodzajów zadań koniecznych do dokonania prawidłowej oceny. Jednak ze względu na potrzebę zapewnienia bogatej kontekstualizacji w testowaniu LSP wskazany jest w miarę szeroki zakres danych wejściowych, jak również oczekiwanych odpowiedzi.

Odpowiedź zdającego jest mniej lub bardziej uzależniona od danych wejściowych. **Bezpośredniość** (*directness*) jest elementem, który pozwala ocenić stopień, w jakim odpowiedzi zdającego są uzależnione od danych wejściowych, a nie od specjalistycznej wiedzy użytkownika języka. Zada-

nie, w którym zdający nie wykorzystuje wiedzy specjalistycznej albo wykorzystuje ją jedynie w minimalnym zakresie, to zadanie o dużym stopniu bezpośredniości. W sytuacji, w której zadanie można wykonać wyłącznie posiadając szeroki zakres wiedzy specjalistycznej, będziemy mówić o niskim stopniu bezpośredniości. Nie znaczy to jednak, że w przypadku egzaminów LSP oczekujemy, aby zadania charakteryzowały się właśnie niskim stopniem bezpośredniości, a więc aby wymagały szerokiego zakresu wiedzy specjalistycznej. Ważne jest, by zadanie egzaminacyjne zawierało odpowiednią ilość elementów plasujących go w kontekście zawodowym.

Kryteria oceny według Douglasa (2000: 67–71) powinny jasno określać, jakie umiejętności językowe mają podlegać ocenie, czyli precyzować, które aspekty tych umiejętności użytkownika języka będą konieczne, by wykonać zadanie. Kryteria oceny powinny także przedstawiać sposób punktowania poszczególnych odpowiedzi. Głównym zadaniem oceny jest sprawdzenie, czy uczący się osiągnęli na zakończenie nauki zamierzone cele. W przypadku ewaluacji w zakresie języka specjalistycznego ocena samych umiejętności językowych nie jest wystarczająca. Pożądane jest uzupełnienie kryteriów oceny zorientowanych na język zadań egzaminacyjnych wskaźnikami stosowanymi przez fachowców przy ocenie zadań zawodowych w autentycznych sytuacjach zawodowych (Douglas 2001: 173–175). McNamara (1996: 43–44), przyglądając się różnym kryteriom wykonania zadania, wyróżnił silne i słabe testy wykonania językowego (*strong/weak performance tests*). W wariantach silnych ocena wykonania zadania opiera się przede wszystkim na kryteriach pozajęzykowych. W tym przypadku jakość języka jest oceniana w takim zakresie, w jakim może ona mieć wpływ na poprawność i efektywność wykonania zadania. Z kolei w słabych testach kryteria oceny oparte są głównie na jakości języka zademonstrowanej podczas wykonania zadania.

9. Podsumowanie

Na kartach tego rozdziału omówiona została podstawowa terminologia związana z testowaniem i egzaminowaniem języków obcych. Dokonano przeglądu różnych klasyfikacji testów językowych oraz poszczególnych rodzajów testów, które znajdują zastosowanie w nauczaniu języków obcych. Odniesiono się tutaj także do głównych kryteriów stanowiących o poprawności testu, jakimi są trafność, rzetelność, praktyczność

i wartość dydaktyczna, oraz omówiono zasady i kolejne etapy konstruowania testów. Opisane zostały rodzaje zadań testowych wykorzystywanych w nauczaniu i testowaniu języków obcych. Omawiane typy zadań egzaminacyjnych zostały poparte przykładami jednostek występujących w analizowanych do celów niniejszego opracowania egzaminach certyfikujących z języka angielskiego i innych testach językowych oraz w egzaminach certyfikujących z języków specjalistycznych.

Celem analizy egzaminów w tym rozdziale było przedstawienie przykładów zadań lub wskazanie testów, w których takie zadania są wykorzystywane, nie zaś analiza arkuszy egzaminacyjnych. Zestawienie budowy poszczególnych arkuszy szczegółowo omawianych w tej publikacji egzaminów (Rozdział III), z uwzględnieniem wykorzystywanych w nich rodzajów zadań, przedstawia Załącznik 3.

Na podstawie analizy zadań wypracowano szereg wniosków dotyczących występowania poszczególnych rodzajów zadań w egzaminach GE i ESP. Przyglądnięto się także podstawowym cechom, które są wspólne dla testów w zakresie języka ogólnego oraz testów LSP. Wreszcie, wskazano na różnice pomiędzy językiem ogólnym oraz językami specjalistycznymi, a w rezultacie – na dystynktywne cechy rozróżniające omawiane dwa rodzaje testowania.

Istotą oceniania w LSP, odróżniającą ten rodzaj ewaluacji od testowania w glottodydaktyce ogólnej, okazuje się mierzenie tzw. podwójnej kompetencji. Celem testowania języków obcych dla potrzeb zawodowych jest sprawdzenie, czy użytkownik języka będzie w stanie poradzić sobie z realizacją zadań językowych w typowych zawodowych sytuacjach komunikacyjnych. A zatem w egzaminach tego typu, obok umiejętności językowych, pomiarowi poddane winny być również wiedza i kompetencje zawodowe, gdyż w rzeczywistym działaniu zawodowym są one ze sobą ściśle powiązane. W związku z tym egzamin, którego celem jest sprawdzenie, czy zdający jest w stanie operacyjnie korzystać ze swoich umiejętności językowych, powinien być oparty na zadaniach naśladowujących autentyczne działanie w docelowych sytuacjach zawodowych, co z kolei będzie miało wpływ na techniki oraz kryteria oceny. Wskazane różnice pomiędzy testowaniem języka ogólnego oraz języków dla celów specjalistycznych świadczą o potrzebie wdrażania oceniania specjalistycznego w zawodach i branżach, w których umiejętność posługiwania się językiem obcym jest istotna i w których, z jednej strony, znajomość języka ogólnego może być niewystarczająca, z drugiej zaś, testowanie sprawn-

ści i elementów języka zbędnych w komunikacji zawodowej, może być bezcelowe.

Czy faktycznie twórcy egzaminów LSP uwzględniają postulaty dotyczące zasad testowania specjalistycznego w swoich testach, sprawdzimy w rozdziale III; tymczasem w rozdziale II omówione zostaną zmiany zachodzące w podejściu do testowania na przestrzeni wieków, co pozwoli lepiej zrozumieć kształt i formę egzaminów językowych obecnych na współczesnym rynku edukacyjnym.

Rozdział II.

Testowanie języków obcych na przestrzeni wieków

Mówiąc o testowaniu języków obcych, warto przyjrzeć się zmianom, jakie przez lata zachodziły w tej kwestii. Celem niniejszego rozdziału jest przybliżenie charakterystycznych cech oceniania w różnych okresach historycznych oraz ewolucji, jakiej podlegało testowanie. Przedstawione zostaną główne epoki w historii testowania, wynikające ze zmieniającego się podejścia do nauczania języków obcych. Zaprezentowany zostanie także – w spojrzeniu diachronicznym – przegląd pierwszych egzaminów w glottodydaktyce specjalistycznej (zwłaszcza w nauczaniu specjalistycznego języka angielskiego), z odniesieniem do omówionych wcześniej cech okresów, w których dane testy się pojawiły, oraz zmian, którym zostały poddane na przestrzeni lat.

1. Główne okresy w historii testowania języków obcych

Dowodem na to, że języki obce były nauczane już w średniowieczu, są zachowane podręczniki z tego okresu. Co ciekawe, te pierwsze skrypty dotyczą rodzaju języka, który dzisiaj zakwalifikowalibyśmy jako język do celów specyficznych. *La manière de langage qui enseigne à parler et à écrire le français* – wydany w Anglii w 1396 roku podręcznik do nauki konwersacji w języku francuskim przedstawia użycie języka w sytuacjach, z którymi jego użytkownik może się zetknąć podczas podróży. Kolejny przykład to dwujęzyczny podręcznik opublikowany ok. 1483 roku

przez Williama Caxtona – *Tres bonne doctrine pour aprendre briefment fransoys et engloys* – *Right good lemyng for to lerne shortly frenssh and englyss*. Jest to bardzo praktyczny zbiór nauk skierowany głównie do angielskich kupców, którzy nie mieli możliwości komunikowania się w języku angielskim poza Wyspami Brytyjskimi. Książka nie zawiera żadnych informacji *stricte* na temat języka, prezentuje natomiast praktyczne wyrażenia przydatne w codziennej komunikacji, takie jak zwroty powitalne i pożegnalne, użyteczne słowa określające przedmioty gospodarstwa domowego, relacje rodzinne, nazewnictwo członków służby, wreszcie krótkie dialogi dotyczące kupowania i sprzedawania towarów. Podręcznik zawiera także listę przydatnych słów i wyrażeń określających produkty spożywcze oraz przykładową rozmowę dotyczącą handlu tkaninami – wełną, skórą i innymi materiałami. Inne elementy podręcznika to krótkie opisy postaci związanych z kupiectwem i innymi rzemiosłami oraz dialog dotyczący poszukiwania zakwaterowania i płacenia za niego. W roku 1553, również dla osób zajmujących się handlem, została wydana w Antwerpii *A Treatise for to Learn to Speak French and English*. Autorem tej książki był francuski nauczyciel języka angielskiego (Howatt 1997: 3–8).

Obecność pochodzących z tego okresu podręczników językowych, skierowanych do konkretnych profesji, dowodzi, że pewien rodzaj języków specjalistycznych był nauczany już w średniowieczu. Nie istniało jednak w tym czasie w Europie formalne testowanie językowe; nie odnotowujemy również opisu żadnych innych form oceny umiejętności w zakresie języka obcego zawodowego. Nie ulega wszakże wątpliwości, że najlepszym sprawdzianem w zakresie umiejętności językowych było samo życie i konieczność przetrwania w obcym kraju, zorganizowania sobie i współtowarzyszom podróży noclegu i stawy oraz załatwienia spraw zawodowych związanych np. z handlem.

Inaczej prezentuje się kwestia języka obcego nauczanego w szkole – do połowy XVIII wieku była to prawie wyłącznie łacina oraz greka. Języki nowożytne nauczane były głównie w kształceniu indywidualnym z prywatnym nauczycielem. Nauka języków innych niż starożytne opierała się na dialogach, głównie w formie pytanie–odpowiedź (Howatt 1997: 5–6) oraz uczeniu się całych tekstów na pamięć. Zatem tutaj sprawdzanie umiejętności odbywało się poprzez odpytywanie ucznia z wyuczonych tekstów. Jak pisze Bernard Spolsky (1975), początek formalnego testowania w Europie to połowa XIX wieku, kiedy to wraz z powstaniem

w 1858 roku The University of Cambridge Local Examinations Syndicate (UCLES) pojawił się system egzaminów publicznych.

W różnych okresach odmiennie rozumiano cele, do których zmierza nauka języka, zróżnicowane były metody nauczania języków obcych, inne były także metody sprawdzania umiejętności uczących się. W zależności od dominującej w danym okresie metody nauczania zmianie ulegało podejście do testowania. W historii testowania biegłości w zakresie posługiwania się językami obcymi dają się wyróżnić cztery główne okresy. Trzy pierwsze, zgodnie z terminologią zaproponowaną przez Spolsky'ego (1975), to okres przednaukowy lub tradycyjny, okres psychometryczno-strukturalistyczny i psycholingwistyczno-socjolingwistyczny. Czwarty okres w testowaniu biegłości językowej, dodany przez badaczy w późniejszym czasie (m.in. Heaton 1988: 19–24; Komorowska 2007: 11; Weir 1990), określaný jest jako okres komunikacyjny. W każdej z tych epok obowiązywały inne wartości i przekonania dotyczące tego, jaka wiedza i umiejętności winny być sprawdzane oraz w jaki sposób.

Przedstawiona klasyfikacja jest podziałem bardzo uproszczonym. Inaczej bowiem rozwijało się testowanie w Stanach Zjednoczonych, innymi torami przebiegał ten proces w Europie; różnice zauważalne były również pomiędzy poszczególnymi krajami europejskimi. Uproszczenie to daje jednak dość wyrazisty obraz kierunku zmian zachodzących w podejściu do oceniania biegłości językowej, pozwalając lepiej je zrozumieć.

Poniżej przedstawiono krótką charakterystykę głównych metod nauczania języka i wynikających z nich metod pomiaru dydaktycznego. Jeżeli nie zaznaczono inaczej, przy redakcji tego rozdziału wykorzystano szereg różnych źródeł omawiających zagadnienia związane z historią nauczania (Howatt 1997; Komorowska 2005a; Larsen-Freeman 2008) oraz z historią oceniania (Gajewska i Sendur 2015; Heaton 1988; Hughes 1995; Hüttner *et al.* 2011; Komorowska 1984, 2007; O'Sullivan 2011; Poszytek 2005a, 2008; Spolsky 1995; Weir 1990).

Okres przednaukowy

Według klasyfikacji Spolsky'ego (1975) okres przednaukowy (tradycyjny) w testowaniu to okres począwszy od połowy XIX wieku do II wojny światowej. Nie znaczy to jednak, że w okresie poprzedzającym nie było żadnych form sprawdzania umiejętności językowych. W cesarskich Chinach element językowy był obecny w sprawnie działającym systemie państwowych egzaminów dla urzędników władz centralnych. Sformalizowane,

przeprowadzane według ściśle określonych, bardzo rygorystycznych zasad, w warunkach uniemożliwiających oszukiwanie, organizowane na ogromną skalę standaryzowane egzaminy państwowe funkcjonowały w Chinach przez ponad 1300 lat, aż do roku 1905 (Elman 2000; Feng 1994: 1–6; Miyazaki 1981; Wilbrink 1997: 10–11).

W Europie, przed omawianym okresem tradycyjnym, ocenianie odbywało się bez jakiegokolwiek podłoża metodycznego. Do XVII wieku nauczanie nowożytnych języków obcych przebiegało prawie całkowicie nieformalnie, w małych grupach lub indywidualnie – z prywatnym nauczycielem. Szkoły były głównie zainteresowane nauczaniem łaciny, a także, w pewnym stopniu, greki. Uczniowie uczyli się na pamięć tekstów oraz zasad gramatyki. Kontrola nauczania polegała przede wszystkim na recytowaniu, niejednokrotnie całkowicie bezmyślnym, wyuczonych na pamięć reguł gramatycznych oraz tłumaczeniu tekstów, nader często pod groźbą kary chłosty. Potrzeba utylitarnej nauki nowożytnego języka obcego stała się bardziej zauważalna zarówno w Wielkiej Brytanii, jak i innych krajach europejskich, dopiero w XIX wieku. Wreszcie, ok. roku 1900, część szkół brytyjskich wprowadziła nauczanie jednego, a niektóre nawet kilku z głównych europejskich języków do programu szkolnego.

W latach pięćdziesiątych XIX wieku w Anglii pojawił się system egzaminów publicznych przeprowadzanych pod nadzorem uniwersytetów. Przypomnijmy: w 1858 roku powstał University of Cambridge Local Examinations Syndicate (UCLES), którego rolą było przeprowadzanie egzaminów dla studentów spoza Uniwersytetu w Cambridge celem podniesienia standardów edukacji. Był to przełomowy moment w systemie szkolnictwa i miał bardzo duże znaczenie dla nauczania języków obcych. To właśnie wydarzenie stanowi początek okresu przednaukowego w nowoczesnym ocenianiu.

W okresie tradycyjnym metody pomiaru dydaktycznego nie miały podstaw naukowych, zaś ocenianie odbywało się całkowicie intuicyjnie. W nauczaniu języków obcych mamy w tym czasie do czynienia z dwoma głównymi metodami: gramatyczno-tłumaczeniową, którą w Stanach Zjednoczonych nazywano ówczesnie metodą pruską, oraz metodą bezpośrednią, zwaną również konwersacyjną. W przypadku tej drugiej trudno mówić o wyodrębnionej fazie kontroli, gdyż sama metoda polegała przede wszystkim na uczestniczeniu w naturalnych rozmowach w języku obcym, najlepiej z rodzimym użytkownikiem danego języka.

Jak pisze Hanna Komorowska (1984: 20–22), o kontroli i ocenie (choć jeszcze nie o testowaniu) można mówić już w odniesieniu do metody

gramatyczno-tłumaczeniowej. Głównym celem nauczania przy użyciu tej metody było czytanie i rozumienie tekstów literackich, umiejętność ich tłumaczenia oraz umiejętność analizowania struktur gramatycznych na podstawie tłumaczonych tekstów. Nieco później pojawiły się też dłuższe wypowiedzi pisemne. Zgodnie z założeniami metody uczeń powinien być świadomy zasad gramatycznych, zaś umiejętność komunikowania się nie była celem nauki języka obcego. Tak więc sprawności słuchania i mówienia były całkowicie zaniedbywane, a nacisk kładziono na poprawność. Nauczanie języka obcego odbywało się przy użyciu języka ojczystego.

Cele metody gramatyczno-tłumaczeniowej przekładały się na sposoby kontroli opanowania języka. W odniesieniu do tego okresu mówimy o **esejowo-tłumaczeniowej metodzie oceniania** (Heaton 1988: 15). Przedmiotem kontroli i oceny było rozumienie tekstów literackich i umiejętność tłumaczenia ich na język ojczysty, pisanie wypracowań oraz testy gramatyczne sprawdzające umiejętność stosowania zasad gramatycznych. Jak zauważa Komorowska (2019: 157–158), funkcja translacji była wówczas zupełnie inna, niż rozumiemy ją dzisiaj. Za pomocą tłumaczenia sprawdzano umiejętność czytania ze zrozumieniem, jak również umiejętność pisania w języku ojczystym. Kolejną cechą kontroli w metodzie gramatyczno-tłumaczeniowej był jej globalny charakter: w nauczaniu nie wyróżniano żadnych umiejętności, które należałoby wykształcić u ucznia. Nie były one również wyróżnione w fazie kontroli i oceny. Od nauczycieli nie wymagano kwalifikacji związanych z testowaniem, nie istniały wytyczne dla nauczycieli języków obcych dotyczące tworzenia testów ani dla egzaminatorów w zakresie obiektywnego oceniania. Proces oceniania był wysoce subiektywny i nieprzewidywalny, a wiedza na jego temat jedynie intuicyjna. Powszechnie sądzono, że wykształcenie nauczycielskie jest wystarczające, aby w pełni usankcjonować oceny nauczycielskie. Nie stosowano też żadnych narzędzi statystycznych do interpretowania wyników testów. Trudno zatem mówić o jakiegokolwiek rzetelności czy obiektywności testowania w tym okresie.

Pierwsze próby obiektywizacji kontroli nauczania dają się zaobserwować po I wojnie światowej. W latach dwudziestych XX wieku w Stanach Zjednoczonych pojawiają się zorganizowane prace nad konstrukcją testów językowych. Tworzone w tym okresie testy są w dalszym ciągu sprawdzianami pisemnymi ukierunkowanymi na kontrolę zagadnień gramatycznych, ale w życie wchodzi nowe rodzaje pomiaru charakterystyczne dla kolejnego okresu: testy elementów języka.

W latach trzydziestych rozwija się umiejętność standaryzowania zadań egzaminacyjnych i analizowania ich pod względem statystycznym. W dalszym ciągu jednak testowanie umiejętności językowych nastawione jest na tekst pisany o charakterze literackim.

Okres psychometryczno-strukturalistyczny

Wraz ze zmianą zapatrywań na naturę uczenia się języka – pod wpływem strukturalizmu w językoznawstwie i behawioryzmu w psychologii – w latach czterdziestych XX wieku rodzi się w Stanach Zjednoczonych metoda audiolingwalna, zwana też strukturalną albo metodą Mim-mem (Komorowska 2019: 158). Po raz pierwszy zaczyna się mówić o czterech sprawnościach językowych, a celem nauczania staje się opanowanie kolejno sprawności słuchania i mówienia oraz czytania i pisania.

W Stanach Zjednoczonych pojawiają się sprawdziany wzorowane na testach psychometrycznych używanych przez psychologów. Z czasem zostają one przyjęte również w Europie. Po raz pierwszy testy językowe opierają się na zasadach naukowych i po raz pierwszy w testowaniu uwzględnia się analizy naukowe. W miejsce oceniania globalnego, charakterystycznego dla poprzedniego okresu, wprowadzone zostało **testowanie atomistyczne** zaproponowane przez Roberta Lado (1961). Dążenie do obiektywizacji oceny testowej zostało ukierunkowane na mierzalność. Zadania otwarte ustąpiły miejsca testom elementów języka, takich jak słownictwo, struktury gramatyczne, poprawna pisownia, a nawet fonetyka, intonacja czy sprawność pisania. Uważano, że w celu prawidłowej oceny i obiektywnego punktowania poszczególnych sprawności czy umiejętności językowych każdy element języka powinien być testowany oddzielnie. Testy składały się z dużej liczby zadań sprawdzających opanowanie tych indywidualnych elementów, a każde zadanie badało opanowanie jednego elementu języka. Poszczególne zadania egzaminacyjne konstruowane były na bazie pojedynczych zdań, bez osadzenia przykładu w jakimkolwiek konkretnym kontekście językowym. Testowanie poszczególnych sprawności językowych również odbywało się rozdzielnie.

Do powszechnego użytku wprowadzono w bateriach testowych najpierw zadania typu prawda–fałsz, a następnie technikę wyboru, w której zdający wybierał spośród kilku podanych odpowiedzi tę, którą uznał za poprawną, odrzucając jednocześnie wszystkie błędne. Wprowadzenie techniki wyboru zaowocowało analityczną procedurą oceniania. Takie

podejście do kontroli i oceny sprawiło, że testy stały się bardziej obiektywizowane, o wysokich wskaźnikach trafności i rzetelności, a wyniki mierzalne, porównywalne i łatwe do obróbki statystycznej.

Ocenianie sprawności mówienia, które było ważnym elementem testowania w okresie tradycyjnym, stało się dużym wyzwaniem dla podejścia psychometrycznego. Przez wiele lat, mimo podejmowania prób wypracowania technik obiektywnego mierzenia tej sprawności, do nowoczesnych egzaminów w Stanach Zjednoczonych w ogóle nie wprowadzano zadań sprawdzających bezpośrednio umiejętność posługiwania się językiem mówionym. Kontroli podlegały tylko niektóre umiejętności związane ze sprawnością mówienia, jak na przykład wymowa i intonacja. Były one sprawdzane za pomocą zadań zamkniętych, na przykład poprzez rozpoznanie par wyrazów, które się rymują (Hughes 1995: 15).

Nauczanie i testowanie w okresie psychometryczno-strukturalistycznym były traktowane jako rozłączne, odrębne aktywności. Nauczyciele mieli zajmować się nauczaniem, a testowanie należało do instytucji osobno do tego powołanych, zatrudniających wykwalifikowanych ekspertów.

Znaczenie okresu strukturalno-psychometrycznego dla dziedziny testowania znajomości języka jest bardzo istotne. Wypracowano w tym czasie metody analizy statystycznej, sposoby badania trafności i rzetelności testu, zasady planowania jego zawartości w odniesieniu do zestawu struktur językowych oraz techniki konstruowania samodzielnych jednostek testu typu wybór wielokrotny (ALTE 2005: 10).

Okres psycholingwistyczno-socjolingwistyczny

Na początku lat siedemdziesiątych XX wieku testy elementów języka zaczęły być uważane za niewystarczające do mierzenia umiejętności językowych. Uznano, że kompetencja językowa to szereg wpływających na siebie wzajemnie umiejętności, których nie można badać rozłącznie, a skoro kompetencja językowa to połączenie wszystkich umiejętności językowych, testować ją należy również łącznie. Wiedza dotycząca elementów języka nic nie znaczy, jeśli użytkownik nie potrafi powiązać tych elementów w taki sposób, aby sprawnie komunikować się w określonej sytuacji.

Nastała nowa era testowania – okres psycholingwistyczno-socjolingwistyczny i **testowanie integracyjne** (Heaton 1988: 16–19), w którym mierzy się kompetencję językową jako całość. Nastąpiło odejście od testów elementów języka na rzecz technik globalnych, sprawdzających

między innymi umiejętność produkowania dłuższych wypowiedzi – zarówno ustnych, jak i pisemnych. Zwrócono uwagę na konieczność jednoczesnego badania rozwoju kilku sprawności językowych, które w sytuacji komunikacyjnej występują zazwyczaj łącznie. Mierzeniu poddawano globalną biegłość językową uczącego się, a nie poszczególne elementy składające się na ową biegłość.

W tym okresie w nauczaniu mamy do czynienia z metodą kognitywną. Programy nauczania i podręczniki hołdowały zasadom nauczania funkcjonalnego i komunikacyjnego. Większy nacisk zaczęto kłaść na sprawność rozumienia wypowiedzi i samodzielności ich budowania. Na nowo odkryta została metoda gramatyczno-tłumaczeniowa, zwrócono uwagę na pewne aspekty metody bezpośredniej i elementy tych metod wprowadzono ponownie do nauczania.

Zadania typu *cloze* i dyktando, krytykowane w poprzednim okresie, zostały przywrócone do łask i stały się typowymi dla okresu psycholingwistyczno-socjolingwistycznego technikami kontroli. Testy typu *cloze* były tworzone poprzez usunięcie co szóstego lub siódmego słowa, a rolą uczącego się było właściwe wypełnienie luk. Czynność ta wymagała od zdającego wykorzystania jednocześnie umiejętności i strategii rozumienia czytania oraz znajomości leksyki, struktur gramatycznych i struktury dyskursu. Również dyktando dawało dużo lepszą miarę umiejętności łączenia sprawności językowych niż testowanie elementów języka. Poprawne wykonanie dyktanda wymaga uważnego słuchania, odtworzenia w formie pisemnej usłyszanego tekstu oraz wykorzystania pamięci krótkotrwałej. Obydwa typy zadań testowych sprawdzają umiejętność posługiwania się językiem na bazie dłuższego tekstu, nie zaś na zestawach oderwanych od siebie zdań, co nadawało im charakter większej autentyczności.

Wkrótce jednak wysunięto zarzut, że takie rodzaje zadań nie mierzą umiejętności posługiwania się językiem obcym w realnych sytuacjach życiowych. Wykonując zadania typu *cloze* oraz pisząc dyktando, uczący pokazuje jedynie umiejętność przetwarzania i stosowania struktur i leksyki w konkretnym znaczeniowo kontekście. Za pomocą tych dwóch technik nie da się natomiast zmierzyć, na ile kompetencja językowa przekłada się na praktyczne umiejętności w czytaniu, słuchaniu, pisaniu i mówieniu. I tak stopniowo nastąpiło przesunięcie zainteresowania z mierzenia kompetencji językowej w kierunku mierzenia komunikacyjnej kompetencji językowej.

Okres komunikacyjny

Z czasem dostrzeżono, że kompetentne posługiwanie się językiem nie zależy wyłącznie od poziomu wiedzy lingwistycznej. Fakt, że uczeń potrafi wyprodukować pewne struktury podczas lekcji, nie znaczy, że będzie to umiał powtórzyć w rzeczywistej komunikacji poza klasą. Komunikacja wymaga, aby uczący się potrafił prawidłowo zastosować funkcje językowe odpowiednie dla określonego kontekstu społecznego. Inaczej mówiąc, komunikowanie się wymaga więcej niż kompetencji językowej – wymaga kompetencji komunikacyjnej (Hymes 1966 i 1972), czyli wiedzy kiedy, jak, co i do kogo powiedzieć, a zatem umiejętności efektywnego porozumiewania się w języku obcym.

a) Podejście komunikacyjne

W latach siedemdziesiątych XX wieku powstało, a w osiemdziesiątych rozpowszechniło się podejście komunikacyjne w nauczaniu języków obcych. Teoria kompetencji komunikacyjnej Della Hymesa (1966) w dużym stopniu rozszerzyła zakres tego, co obejmowało rozumienie języka i umiejętność użycia języka w kontekście, zwłaszcza w odniesieniu do wymogów społecznych w zakresie wykonania. Hymes zauważył, że istnieją specyficzne kulturowo zasady stosowania języka, które uzależniają użycie języka od właściwości kontekstu komunikacyjnego. Do powstania nowego kierunku w dydaktyce przyczyniły się również prace Rady Europy uwieńczone opublikowaniem w 1975 i 1976 roku pierwszych poziomów progowych – *Threshold Level* (van Ek i Alexander 1975) i *Un Niveau-Seuil* (Coste et al. 1976).

Celem nauczania stała się kompetencja komunikacyjna. Nastąpiło przesunięcie z nauczania systemu języka na nauczanie języka jako środka do komunikowania się. Język miał być narzędziem komunikacji, a celem nauczania stało się rozwijanie praktycznych umiejętności posługiwania się nim. Za mniej istotne zaczęto uważać bezbłędne opanowanie obcego systemu językowego, a znaczącą rolę otrzymało skuteczne porozumiewanie się. Mniejszą wagę zaczęto przywiązywać do pełnej poprawności gramatycznej, o ile tylko jej brak nie powodował zablokowania komunikacji. Celem ćwiczeń lekcyjnych przestało być produkowanie absolutnie poprawnych zdań, a stała się nim skuteczność przekazywania informacji i intencji w typowych sytuacjach życia codziennego.

Zmiany w podejściu do nauczania języka ponownie wpłynęły na odmianę spojrzenia na testowanie. W ocenianiu w podejściu komuni-

kacyjnym przeniesiono środek ciężkości z kontroli opanowania określonych struktur na sprawdzenie umiejętności aktywnego posługiwania się językiem (Madsen 1983: 6–7). Nastąpiło przejście z testowania kompetencji do testowania wykonania językowego.

Testy językowe w podejściu komunikacyjnym charakteryzują się dwoma zasadniczymi cechami. Po pierwsze, są testami wykonania, w których ocena ma się odbywać podczas aktywnego udziału zdającego w rozwiniętej aktywności komunikacyjnej – produktywnej lub receptywnej bądź jednocześnie produktywnej i receptywnej. Po drugie, testy te biorą pod uwagę role społeczne, które kandydaci mogą przypuszczalnie przyjąć w realnym życiu, i formułują wymagania dla takich właśnie ról. To ta druga cecha, czyli zwrócenie uwagi na społeczną funkcję języka, odróżnia testowanie komunikacyjne od poprzedzającego go testowania integracyjnego. Innowacyjnymi cechami egzaminowania stało się także użycie autentycznych tekstów i realistycznych zadań.

W **komunikacyjnej metodzie testowania** (Heaton 1988: 19–24) umiejętności językowych obserwujemy połączenie tradycji analitycznej, czyli testów atomistycznych, oraz tradycji syntetycznej, znajdującej odzwierciedlenie w podejściu esejowo-tłumaczeniowym i integracyjnym. Wykorzystywane są tutaj zarówno charakterystyczne dla testowania atomistycznego zadania zamknięte, jak i otwarte. Do testowania wprowadzono sprawdziany umiejętności radzenia sobie z dłuższymi tekstami, takimi jak wypracowanie, streszczenie, dyktando oraz dłuższa wypowiedź monologowa zdającego. Po raz pierwszy pojawiły się zadania z luką informacyjną, stwarzające uczącym się rzeczywistą potrzebę komunikowania się ze sobą (Gaszyńska-Magiera i Seretny 2004: 23).

W okresie komunikacyjnym nastąpiło znaczne zbliżenie procesu kontroli do procesu nauczania. Zadania kontrolne oraz ćwiczenia komunikacyjne i sprawnościowe wykonywane w czasie nauki przyjęły identyczną lub bardzo zbliżoną formę. Repertuar technik testowania i lista ćwiczeń lekcyjnych bardzo się poszerzyły. Rodzaje zadań, które królowały w poszczególnych tradycjach testowania jako jedynie słuszne, są obecnie wykorzystywane w bateriach testów składających się z całego repertuaru różnych rodzajów zadań i różnych technik egzaminacyjnych. W zależności od celu testu i jego rodzaju preferowane są odmienne techniki testowania.

b) Podejście zadaniowe

Po wieloletnich wysiłkach zespołów specjalistów z różnych krajów swej wersji ostatecznej doczekał się projekt standaryzacji nauczania i uczenia się języków obcych w Europie. Zapoczątkowany przez poziomy progowe (*Threshold Level*) (van Ek i Alexander 1975), w roku 2001 został wydany w wersji angielskiej *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* – CEFR (Council of Europe 2001). Wersja polska doczekała się publikacji w roku 2003 pod tytułem *Europejski system opisu kształcenia językowego* (ESOKJ; Rada Europy 2003). Twórcy ESOKJ zaproponowali w niej nowe podejście do nauczania języka, nazwane podejściem zadaniowym, jednak jego początki dawały się zauważyć już w latach dziewięćdziesiątych XX wieku. Podejście zadaniowe w nauczaniu języka, nazywane również „uczeniem się/nauczaniem przez działanie”, „uczeniem się przez doświadczenie” czy „uczeniem się/nauczaniem przez wykonywanie zadań” (Janowska 2011: 11 i 77), zakłada włączenie uczącego się do procesu dydaktycznego, a za najlepszy sposób nabywania wiedzy i umiejętności uważa działanie. Stąd też dyskusja nad prawidłowością tłumaczenia nazwy tegoż podejścia na język polski: Przemysław Gębal (2019: 139–140) kwestionuje zasadność użycia w polskich przekładach (m.in. w polskiej wersji ESOKJ) oraz w polskiej literaturze tematu (m.in. Janowska 2011) terminu „podejście zadaniowe” na rzecz „podejścia działaniowego”. Na rozbieżności w terminologii między „podejściem działaniowym” (*Action-oriented Approach*) a „nauczaniem poprzez zadania” (*Task-based Learning*) zwraca również uwagę Marta Kaliska (2016: 186–188). Jednak, w związku z szerokim zastosowaniem tłumaczenia „podejście zadaniowe” w literaturze, do której odwołujemy się w niniejszej monografii, termin ten został tutaj zachowany.

Większość badaczy widzi w podejściu zadaniowym kontynuację podejścia komunikacyjnego, jego nową, udoskonaloną wersję. Cechy podejścia komunikacyjnego, które są bardzo wyraźne również w podejściu zadaniowym, to nacisk na przekaz informacji, użycie autentycznych materiałów, integrowanie sprawności oraz usytuowanie uczącego się w centrum kształcenia językowego. Według Johanna Fischera *et al.* (2011: 18–19) tym, co podejście zadaniowe wnosi do podejścia komunikacyjnego, jest wprowadzenie holistycznych aktów komunikacji do specyficznego kontekstu i specyficznych sytuacji, przy czym cel takiej komunikacji odzwierciedla bieżące lub przyszłe komunikacyjne cele zdającego. W perspektywie podejścia zadaniowego komunikowanie i uczenie

się sprowadzone są do wykonywania zadań, w których przejawiają się konkretne kompetencje i strategie stosowane przez użytkownika języka w celu zakończenia zadania z powodzeniem. Istota tego podejścia polega na tym, że uczący się aktywnie uczestniczy w rozwiązywaniu problemów, posługując się w tym celu językiem docelowym, przy czym – jak zauważa Magdalena Sowa – „komunikacja staje się tu środkiem do osiągnięcia celu, a nie celem samym w sobie” (Sowa 2010: 92).

Logicznym kolejnym krokiem po wprowadzeniu do dydaktyki języków obcych podejścia zadaniowego stało się stworzenie nowego modelu testowania. **Ocenianie w podejściu zadaniowym** (*Task-based Assessment* – TBA) ma uczynić testy językowe bardziej relewantnymi dla zdającego, takimi, które zmierzają, w jakim stopniu będzie on w stanie komunikować się w specyficznej sytuacji komunikacyjnej w swojej dziedzinie zawodowej. Testy w podejściu zadaniowym mają być bliższe życiu, bardziej realistyczne, testowanie ma bazować na specyficznych rzeczywistych sytuacjach opartych na faktach i autentycznych dokumentach. Testy TBA charakteryzują się większą transparentnością, są również bardziej zrozumiałe dla zdającego, który jest w stanie lepiej samodzielnie ocenić swoje wyniki.

W nauczaniu i testowaniu według podejścia zadaniowego działania powinny być bardzo bliskie rzeczywistym działaniom komunikacyjnym uczącego się/zdającego, zbieżne z zadaniami, przed którymi osoby te staną w rzeczywistym świecie. Zadanie testowe powinno zatem symulować działanie, w jakim zdający będzie musiał uczestniczyć w codziennym życiu poza salą lekcyjną bądź też podczas wykonywania pracy zawodowej. Według Davida Nunana (2004: 164) TBA powinno być ocenianiem bezpośrednim i koncentrować się na osiągnięciu specyficznych celów, a sama ocena powinna mieć charakter kształtujący i posiadać funkcję sprawdzającą, a nie różnicującą.

Zadaniami egzaminacyjnymi najlepiej wpisującymi się w podejście zadaniowe są zadania typu odgrywanie ról (*role play*), praca projektowa (*project work*), studium przypadku (*case study*) i symulacja globalna (*global simulation*) (Fischer *et al.* 2011: 21–24).

2. Testowanie języków obcych – spojrzenie historyczne

W zaprezentowanym powyżej podziale na główne okresy w historii testowania języków obcych przedstawiono obowiązujące w danym czasie teorie testowania i charakterystyczne cechy tworzonych w tym czasie egzaminów, bez względu na to, do jakich celów dany egzamin został stworzony. Nie istnieje odrębna historia testowania języków do celów specjalistycznych. W związku z powyższym na użytek niniejszego opracowania zebrano informacje dotyczące najważniejszych egzaminów pojawiających się na rynku edukacyjnym – brytyjskim i amerykańskim – w poszczególnych okresach historycznych, dokonano ich porównania w odniesieniu do cech testów w kolejnych okresach, jak również prześledzono zmiany w tych egzaminach, jakie istniały i rozwijały się na przestrzeni różnych epok. Do tego celu wybrano egzaminy, które poprzez swoją innowacyjność, szeroki zakres odbiorców bądź po prostu popularność odcisnęły swój ślad na historii testowania języków specjalistycznych.

Do kategorii egzaminów LSP Dan Douglas (2000: 2) zalicza testy z języków akademickich oraz testy do celów zawodowych. Taką samą klasyfikację języków do celów specyficznych w odniesieniu do języka angielskiego (ESP) przyjmują Tony Dudley-Evans i Maggie-Jo St John (1998: 5–7) oraz Tom Hutchinson i Alan Waters (1991: 16–18), pisząc o angielskim języku akademickim (*English for Academic Purposes* – EAP) oraz języku angielskim do celów zawodowych (*English for Occupational Purposes* – EOP). W badaniu uwzględniono testy należące do tych dwóch kategorii, ale znajdziemy tutaj także egzaminy, których rola oraz cel zmieniły się na przestrzeni lat (CPE). Pośród opisanych przykładów znajduje się również jeden z najstarszych egzaminów certyfikujących biegłość w ramach ogólnego języka angielskiego – *Lower Certificate in English* – prekursor znanego do dzisiaj egzaminu przeprowadzanego pod auspicjami Uniwersytetu w Cambridge: *Cambridge B2 First*.

Pierwsze egzaminy certyfikujące z języka angielskiego przeprowadzane przez The University of Cambridge Local Examinations Syndicate (UCLES)

Pierwszym formalnym egzaminem z języka angielskiego, który z czasem zaczął być przeprowadzany na szeroką skalę, był *Cambridge Proficiency in English* (CPE). Pierwotnie przeznaczony do certyfikowania biegłości językowej u obcokrajowców – nauczycieli języka angielskiego, z czasem

zmienił swoje przeznaczenie i stał się egzaminem akademickim. Drugim sztandarowym egzaminem przeprowadzanym przez UCLES był *Lower Certificate in English*. Jego zadaniem było mierzenie biegłości na poziomie niższym niż CPE. Obydwa egzaminy odegrały istotną rolę w historii testowania języka angielskiego, obydwie również w dalszym ciągu są obecne na międzynarodowym rynku certyfikacji. Dlatego też od nich zaczniemy historyczny przegląd egzaminów językowych.

a) *Cambridge Proficiency in English* (CPE)

The University of Cambridge Local Examinations Syndicate (UCLES) po raz pierwszy przeprowadził *Cambridge Proficiency in English* (CPE) w roku 1913. Ten przedstawiciel ery przednaukowej w testowaniu może według Douglasa (2000: 2) aspirować do miana pierwszego testu LSP, gdyż skierowany był do przyszłych nauczycieli języka angielskiego, a więc posiadał jasno sprecyzowany cel zawodowy, chociaż nie spełniał zasad dotyczących konstruowania egzaminów z języka obcego specjalistycznego w dzisiejszym rozumieniu.

W 1913 roku (Weir 2002: 2–3, 2005: 6, 2013: 2–6) egzamin trwał łącznie 12 godzin i składał się z części pisemnej i ustnej (Tabela 2.). Był kierowany do osób, których ojczystym językiem był francuski lub niemiecki, w związku z czym zadania w części pisemnej obejmowały tłumaczenie tekstów z języka angielskiego na francuski lub niemiecki i odwrotnie. Zdający mieli także napisać wypracowanie, którego tematyka zarówno w pierwszej, jak i w późniejszych sesjach oscylowała wokół znajomości historii i kultury Wielkiej Brytanii, a wymagania względem zdających w części dotyczącej literatury były takie same, jak na egzaminach wstępnych na brytyjskie wyższe uczelnie dla rodzimych użytkowników języka angielskiego. Pośrednio sprawdzana była również sprawność słuchania: w części dotyczącej fonetyki języka angielskiego kandydat musiał zaprezentować złożoną wiedzę dotyczącą tego, jak wytwarzane są dźwięki i jak brzmią słowa w języku angielskim, zaproponować sposoby nauczania wymowy określonych dźwięków w języku angielskim, jak również wykazać się umiejętnością użycia transkrypcji fonetycznej. Część ustna składała się z trzech elementów: dyktanda, zadania polegającego na głośnym czytaniu tekstu oraz z rozmowy. Warto zauważyć, że część ustna obecna była na tym egzaminie już w jego pierwszej wersji, co zdecydowanie odróżniało go od innych testów językowych w tym okresie i później, a zwłaszcza od amerykańskiego konkurenta – TOEFL.

Tabela 2. Struktura egzaminu CPE w 1913 roku

Część pismenna	tłumaczenie z j. angielskiego na francuski lub niemiecki	2 godz.
	tłumaczenie z j. francuskiego lub niemieckiego na angielski	2,5 godz.
	wypracowanie	2 godz.
	literatura	3 godz.
	fonetyka	1,5 godz.
Część ustna	dyktando	0,5 godz.
	czytanie i rozmowa	0,5 godz.

Do egzaminu w 1913 roku przystąpiło troje kandydatów i żaden z nich go nie zdał. Jednak dla UCLES był to pierwszy poważny krok w kierunku testowania języka obcego i początek długiej historii działalności w tej dziedzinie. Sam egzamin CPE stał się najdłużej funkcjonującym na rynku edukacyjnym testem językowym, który – w zmienionej formie – jest obecny w ofercie Cambridge English Language Assessment do dnia dzisiejszego.

Z czasem rozszerzył się krąg odbiorców egzaminu. W roku 1932 zniknęły z egzaminu zadania sprawdzające znajomość fonetyki (Weir 2003: 3). Osobami zainteresowanymi certyfikatem przestali być jedynie przyszli nauczyciele języka. W 1937 roku Uniwersytet w Cambridge uznał poziom języka poświadczony za pomocą egzaminu CPE za odpowiedni do przyjęcia kandydata na studia (Spolsky 1995: 64). Rok później w ślad za nim poszedł również Uniwersytet w Oksfordzie (Weir 2003: 5) i tak *Cambridge Proficiency* stał się egzaminem akademickim.

W 1938 roku, jako alternatywa dla części dotyczącej literatury angielskiej, do egzaminu wprowadzono zadanie pisemne o tematyce ekonomiczno-handlowej. Rozszerzono także ofertę języków w części dotyczącej tłumaczenia tekstów (Weir 2003: 6). Egzamin stawał się coraz popularniejszy i zyskiwał coraz więcej kandydatów.

Kolejne istotne zmiany wprowadzone zostały w 1953 roku, kiedy to w miejsce arkusza sprawdzającego umiejętność tłumaczenia tekstów (*Translation*) zdający mógł wybrać zadania *Use of English*, sprawdzające wówczas rozumienie tekstu czytanego, transformacje zdań, tworzenie koherentnych tekstów z podanych zdań, zadanie sprawdzające umiejętność poprawnego stosowania znaków przestankowych, umiejętność redagowania tekstu, tworzenia tekstu typu opis oraz stosowania przedrostków i przyrostków (Weir 2003: 9).

W erę strukturalistyczno-psychometryczną egzamin CPE został wprowadzony dopiero w roku 1966, kiedy to jego formuła została po raz kolejny poddana zmianom. Do części *Use of English* wprowadzono zadania wyboru wielokrotnego z trzema opcjami odpowiedzi (Weir 2013a: 4). Do części sprawdzającej umiejętność pisania dodano tematy dla osób zainteresowanych handlem, z kolei obok testu sprawdzającego znajomość literatury brytyjskiej dodano jako opcje do wyboru teksty naukowe, teksty dotyczące przemysłu i handlu oraz część odnoszącą się do życia i instytucji Wielkiej Brytanii. Po raz pierwszy na egzaminie pojawia się także test badający sprawność czytania (Weir 2002: 3–4).

W roku 1975 egzamin CPE przeszedł kolejną transformację, która z kolei wprowadziła go w erę komunikacyjną. Język, również na egzaminie, przestał być traktowany jako system, a zaczął być postrzegany jako środek do komunikowania się (Weir 2005: 8–9). Większy nacisk położono na użycie języka (*English in Use*). Na egzaminie pojawia się wówczas po raz pierwszy sprawdzian umiejętności rozumienia ze słuchu (*Listening Comprehension*), sprawdzany całkowicie za pomocą testu wyboru wielokrotnego. Wprowadzono również obowiązkowy arkusz rozumienia tekstu czytanego (*Reading Comprehension*) także skonstruowany z zadań zamkniętych WW (wcześniej zadania z tekstem pisany były obecne wyłącznie w opcjonalnej części dotyczącej przemysłu i handlu oraz życia i instytucji Wielkiej Brytanii). Z testu znika dyktando i część tłumaczeniowa, jak również zadania sprawdzające znajomość literatury (Weir 2002: 4, 2013a: 5).

b) *Lower Certificate in English / First Certificate in English*

Wprowadzony w życie w 1939 roku *Lower Certificate in English* (LCE) (Hawkey 2009; Spolsky 1995; www.eltnews.gr³⁰) był drugim po *Certificate of Proficiency in English* (CPE) egzaminem z języka angielskiego dla obcokrajowców przeprowadzanym przez UCLES. Potrzeba stworzenia egzaminu na niższym poziomie niż CPE została spowodowana napływem do Zjednoczonego Królestwa najpierw uchodźców z Hiszpanii, na skutek hiszpańskiej wojny domowej, a następnie z innych krajów okupowanej Europy.

³⁰ Zob. <https://www.eltnews.gr/education/1520-the-history-of-the-cambridge-exams> [dostęp: 13.07.2020].

W swojej pierwszej odsłonie egzamin składał się z trzech części:

- ustnej, skonstruowanej z dyktanda, głośnego czytania oraz rozmowy,
- arkusza zatytułowanego *English Composition and Language*, na który składało się wypracowanie na wybrany temat oraz seria testów sprawdzających umiejętność poprawnego użycia języka,
- arkusza opartego na wybranych lekturach.

W przeciwieństwie do swojego pierwowzoru – CPE, jak również wbrew zasadom obowiązującej w dalszym ciągu metody gramatyczno-tłumaczeniowej, LCE nie zawierał w swojej pierwszej wersji zadań sprawdzających umiejętność tłumaczenia tekstów. Jednak już cztery lata później arkusz został uzupełniony o takie zadanie w wersji opcjonalnej – wymiennej za zadanie oparte na lekturze (Hawkey 2009: 7–9). Wersja z roku 1943 zawierała więc wybór pomiędzy arkuszem opartym na lekturze tekstu literackiego oraz arkusz tłumaczenia na język angielski. Egzamin składał się wyłącznie z zadań otwartych bez opisanych kryteriów obiektywnej oceny. Nie spełniał zatem kryterium rzetelności, o które w tym czasie zadbano już w Stanach Zjednoczonych, wprowadzając do egzaminu językowego *College Entrance Examination Board* m.in. zadania typu prawda–fałsz (Spolsky 1995: 65).

W roku 1944 sprawdzian umiejętności tłumaczenia tekstu na język angielski oferował kandydatom wybór spośród 18 języków, w tym z języka polskiego, arabskiego, hebrajskiego, czeskiego, perskiego i szwedzkiego. Na listach kandydatów przystępujących do egzaminu w grudniu 1943 roku widnieje znaczna liczba polskich żołnierzy. Po zakończeniu II wojny światowej LCE okazał się najpopularniejszym egzaminem przeprowadzanym przez UCLES. W 1947 roku przystąpiło do niego ponad 4 tysiące kandydatów, w porównaniu z ok. 2 tysiącami, które przystąpiły w tym samym czasie do CPE (www.eltnews.gr).

W efekcie rozwijającego się podejścia komunikacyjnego w 1975 roku egzamin przeszedł dość radykalne zmiany. Wprowadzono pięć obowiązkowych części testu: wypracowanie (*Composition*), sprawdzian rozumienia testu czytanego (*Reading Comprehension*), test leksykalno-gramatyczny (*Use of English*), test rozumienia ze słuchu zawierający jednostki wyboru wielokrotnego (*Listening Comprehension*) oraz egzamin ustny (*Interview*). Egzamin zmienił również wtedy swoją nazwę na *First Certificate in English* (FCE). Zmiany zainicjowane w 1975, a następnie pewne modyfikacje wprowadzone w roku 1984, oprócz zasad podejścia komunikacyjnego miały również na uwadze zwiększenie wagi sprawności słu-

chania oraz mówienia względem pozostałych części arkusza, jak również autentyczność tekstów stanowiących bazę dla testów czytania.

W kolejnych zmianach wprowadzonych do egzaminu w roku 1996 zwiększono zakres umiejętności sprawdzanych w testach pisania, słuchania oraz mówienia. Waga wszystkich pięciu części egzaminu została wyrównana i od tego momentu każda z nich stanowi 20% wartości wszystkich punktów. Również od tego momentu w egzaminie ustnym bierze udział dwoje zdających jednocześnie.

c) Egzaminy językowe dla celów akademickich (EAP)

Po I wojnie światowej, na skutek napływu do amerykańskich szkół i uczelni dużej liczby obcokrajowców, których znajomość języka angielskiego pozostawiała wiele do życzenia, wprowadzono za pomocą Aktu Imigracyjnego (*Immigration Act of 1926*) obowiązek sprawdzania poziomu znajomości języka angielskiego kandydatów przed przyjęciem na studia. Uczelnie zgłosiły do College Entrance Examination Board zapotrzebowanie na egzamin językowy, który spełniałby ten wymóg. I tak w 1930 roku odbył się pierwszy egzamin kwalifikowany jako egzamin z języka angielskiego dla celów akademickich (EAP) – *English Competence* (Spolsky 1995: 55). W roku 1964 jego miejsce zajął *Test of English as a Foreign Language* (TOEFL).

Konieczność stworzenia egzaminu sprawdzającego kompetencję językową potrzebną do studiowania w języku angielskim pojawiła się także w Wielkiej Brytanii. Również tutaj powodem było zgłaszanie się na studia dużej liczby kandydatów, dla których język angielski nie był językiem ojczystym. Potrzeba ta pojawiła się jednak w Zjednoczonym Królestwie później niż w Stanach Zjednoczonych. W roku 1963 British Council wraz z Uniwersytetem w Birmingham wprowadził egzamin językowy dla potrzeb akademickich *The English Proficiency Test Battery* (EPTB).

d) English Competence Examination

Według Spolsky'ego (1995: 55) *English Competence Examination* był prawdopodobnie pierwszym egzaminem sprawdzającym znajomość języka angielskiego jako obcego w Stanach Zjednoczonych. Douglas twierdzi, że egzamin ten może nawet aspirować do roli jednego z pierwszych egzaminów LSP, a konkretnie EAP, jako że był to test dla potrzeb akademickich. Podobnie jak opisany wcześniej *Cambridge Proficiency in English*, *English Competence Examination* miał jasno zdefiniowane cele

i odbiorcę. Jednak żaden z tych dwóch egzaminów nie spełniał koniecznych warunków i zasad konstruowania testów LSP ze względu na brak analizy specyficznego użycia języka w docelowych sytuacjach komunikacyjnych, jak również z powodu faktu, że zadania egzaminacyjne nie były związane z zadaniami, z którymi zdający mogli się spotkać w pracy zawodowej lub, jak w tym wypadku, w sytuacji związanej ze studiami na wyższej uczelni (Douglas 2000: 2–3).

English Competence w swojej pierwotnej formie składał się z pięciu części (Bachman *et al.* 1995: 4–6; Spolsky 1993: 6–7, 1995: 55–57). Część pierwsza oparta była na czterech krótkich tekstach: prostym tekście naracyjnym, trudniejszym tekście o tematyce historycznej, dziennikarskiej lub innej, krytycznym tekście na istotny temat społeczny i wreszcie tekście naukowym. Wszystkie zadania w tej części były zadaniami typu „plus/minus”, czyli według dzisiejszej nomenklatury: prawda–fałsz. Część druga składała się z dłuższego tekstu z pytaniami w formie otwartej. Część trzecią stanowiło dyktando bezpośrednio oraz zadanie polegające na odtworzeniu z pamięci dyktowanego tekstu. Czwarta część to egzamin ustny, w którym rolą egzaminatora było ustalenie oceny w trzystopniowej skali: biegły, zadawalający, niezadawalający. Oceniana była płynność wypowiedzi, jakość i szybkość reakcji, wymowa, dykcja, opanowanie struktur, spójników, słownictwa i idiomów; egzaminator miał także stwierdzić, czy zdający był ewentualnie nieśmiały i wykazujący się brakiem pewności siebie. W części piątej zdający pisał wypracowanie o długości 250–300 słów na wybrany temat.

W roku 1933 wprowadzono nową, zmodyfikowaną formułę egzaminu. Rozszerzono rodzaj zadań sprawdzających umiejętność czytania; podczas odsłuchu dyktanda, którego tekst miał być później odtworzony z pamięci, wprowadzono możliwość robienia notatek, a część ustna składała się z 10–15 minutowej dyskusji i konwersacji. W niezmienionej formie pozostało wypracowanie (Bachman *et al.* 1995: 4–6).

English Competence Examination był przeprowadzany przez pięć kolejnych lat. W 1936 roku został przerwany, między innymi z powodu pogłębiającego się światowego kryzysu ekonomicznego.

W 1938 roku, w związku z brakiem funduszy na dalsze przeprowadzanie egzaminu przez College Entrance Examination Board oraz ze względu na pojawiającą się nową potrzebą poświadczania kompetencji językowych imigrujących z Europy do Stanów Zjednoczonych lekarzy i prawników pochodzenia żydowskiego, The New York State Board

of Education dokonała porównania egzaminów *English Competence* oraz *Cambridge Proficiency in English*. Podjęto decyzję o akceptowaniu CPE jako egzaminu odpowiedniego do potrzeb certyfikacji tych grup zawodowych, jednak – przy okazji tej analizy – zwrócono uwagę na różnice pomiędzy porównywanymi testami. Egzamin uniwersytetu w Cambridge trzymał się sztywno formuły egzaminów tradycyjnych i subiektywnego sposobu ich oceniania, podczas gdy jego amerykański odpowiednik już w tym okresie zawierał obiektywne zadania „nowego typu” (Bachman *et al.* 1995: 4–6).

Zaprzestanie przeprowadzania *English Competence Examination* nie wstrzymało prac nad nowym amerykańskim egzaminem. Zadanie to przejął powstały w 1947 roku Educational Testing Service (ETS). W efekcie tych prac narodził się egzamin TOEFL.

e) *Test of English as a Foreign Language* (TOEFL)

W roku 1964 odbył się pierwszy *Test of English as a Foreign Language* (TOEFL) (Alderson 2009; Spolsky 1995: 282–286). W swojej pierwotnej wersji składał się z pięciu części: rozumienia ze słuchu (*Auditory Comprehension*), umiejętności pisania (*Writing Ability*), struktur (*Structure*), słownictwa (*Vocabulary*) i rozumienia tekstu czytanego (*Reading Comprehension*). Był skonstruowany wyłącznie z zadań zamkniętych – z 270 jednostek wyboru wielokrotnego z czterema opcjami do wyboru.

Część odpowiadająca za sprawdzenie rozumienia ze słuchu składała się z różnego rodzaju tekstów, w tym dialogów i krótkich wykładów. Moduł badający rozumienie tekstu pisanego skomponowany był z klasycznych tekstów na poziomie uniwersyteckim i 30 zadań typu wybór wielokrotny. Rodzaje tekstów, na bazie których skonstruowane były zadania, to teksty o tematyce z zakresu nauk społecznych, nauk przyrodniczych oraz tekst literacki. Technika pośrednią sprawdzana była również sprawność mówienia i pisania. Za pomocą zadań zamkniętych badane było „poczucie poprawnego stylu”. Zadaniem zdających było uzupełnić akapity wyjętymi fragmentami, które miały pasować stylistycznie do tekstu, oraz wskazać błędne pod względem stylistycznym elementy.

Davies (2007: 75) zarzuca amerykańskiemu egzaminowi brak jakiegokolwiek rozwoju i zmian w okresie ponad czterdziestu lat, brak wynikający z bardzo silnych tradycji psychometrycznych, które w Stanach Zjednoczonych miały dużo większy wpływ na tworzone tam egzaminy niż na testowanie w Wielkiej Brytanii i Europie. Całkowicie odmienne

zdanie prezentuje Educational Testing Service (ETS) – instytucja odpowiedzialna za test, która twierdzi, że przez ponad czterdzieści lat TOEFL był pionierem innowacji w testowaniu kompetencji językowych, w tym również z wykorzystaniem najnowszych technologii (ETS 2011: 5).

Rzeczywiście, wprowadzane zmiany przez wiele lat dotyczyły w głównej mierze formy przeprowadzania egzaminu, w tym medium, za pośrednictwem którego zdający odpowiadali na pytania. W 1979 roku (ETS 2011: 3–5) do istniejącego wówczas od piętnastu lat egzaminu składającego się wyłącznie z zadań typu wybór wielokrotny dodano nieobowiązkowy, dodatkowy test sprawdzający mówienie – *Test of Spoken English* (TSE), a następnie, w 1986, też nieobowiązkowy test badający sprawność pisania – *Test of Written English* (TWE). W 1998 roku wprowadzono komputerową wersję testu – TOEFL CBT (*Computer-based Test*), który po raz pierwszy zawierał arkusz mierzący sprawność pisania. Wprowadzony w 2005 roku internetowy test TOEFL iBT (*Internet-based Test*) całkowicie wyparł wersję komputerową i w dużym stopniu, choć niezupełnie, wersję papierową (TOEFL PBT – *Paper-based Test*). Dopiero ta wersja doczekała się znaczących zmian w konstrukcji i rodzaju sprawdzanych kompetencji. Po raz pierwszy w długoletniej historii testu wprowadzono zadania integrujące różne sprawności (ETS 2010). W części egzaminu sprawdzającej sprawność mówienia cztery z sześciu zadań są zadaniami integrującymi sprawności słuchania i czytania z mówieniem. Podobnie wygląda jedno zadanie w teście umiejętności pisania: jest oparte na tekście czytany i słuchany, które zdający musi przeanalizować w celu stworzenia wypowiedzi pisemnej, a więc zadanie to łączy trzy różne sprawności.

f) *The English Proficiency Test Battery* (EPTB)

W Wielkiej Brytanii pierwszy egzamin od początku konstruowany jako test dla potrzeb akademickich pojawił się w roku 1963. *The English Proficiency Test Battery* (EPTB) (Davies 2007: 73–79; Read 2015: 121–122; Weir i O’Sullivan 2017: 114–122) został opracowany przez British Council i Uniwersytet w Birmingham na potrzeby weryfikacji umiejętności językowych kandydatów na studia w Wielkiej Brytanii.

Ze względów praktycznych egzamin badał tylko sprawności receptywne, głównie czytanie. Test zawierał zadania sprawdzające elementy języka – intonację, akcent, rozróżnianie fonemów oraz gramatykę, ale zawierał także część mierzącą umiejętności praktyczne, jak czytanie i rozumienie ze słuchu tekstów akademickich oraz test szybkości czytania.

Elementy te świadczyły o dużym na owe czasy stopniu innowacyjności testu. Teksty, na których oparte były zadania badające sprawność czytania, bazowały zarówno na zagadnieniach ogólnych, jak i tematyce związanej z naukami ścisłymi. Typy zadań w tej części to *C-test* (*cloze* z podaną pierwszą literą brakującego wyrazu), a zadanie na szybkość czytania to tzw. *cloze-elide*, w którym do angielskiego tekstu zostały dopisane wyrazy w innym języku (Weir 2013b³¹).

Główną część testu EPTB stanowił test gramatyczny oraz zadania atomistyczne. Z tego powodu wpisuje się on w dużym stopniu w podejście strukturalistyczne. Okres, w którym opisywany egzamin powstał, był w Wielkiej Brytanii czasem rozkwitu testowania atomistycznego. Dlatego też Weir (2013b) plasuje test w okresie psychometryczno-strukturalistycznym. Jednak mimo obecności zadań atomistycznych i braku jednostek sprawdzających sprawności produktywne egzamin był dość innowacyjny. Zostały do niego wprowadzone niektóre pomysły zaproponowane przez Johna B. Carrola w 1961 roku, charakteryzujące kolejną epokę w testowaniu – okres psycholinwistyczno-socjolingwistyczny, jak zadania typu *cloze* oraz test szybkości czytania (Davies 2007: 79; Read 2015: 169–170). Również fakt, że test posiadał komponent sprawdzający rozumienie tekstu czytanego oraz tekstu słuchanego, wyróżniał go w owym czasie na tle innych egzaminów dostępnych w Wielkiej Brytanii. Uniwersytet w Cambridge wprowadził obowiązkowy komponent rozumienia tekstów czytanych do swoich egzaminów dopiero w 1975 roku.

EPTB był używany przez British Council w swoich zagranicznych placówkach od 1965 do 1980 roku, kiedy to jego miejsce zajął egzamin *English Language Testing Service* (ELTS).

g) *English Language Testing Service* (ELTS)

Na skutek rewolucji komunikacyjnej w nauczaniu w roku 1980 *The English Proficiency Test Battery* został zastąpiony nowym egzaminem – *English Language Testing Service* (ELTS) (Clapham 1996: 51–57; Read 2015: 123–125; Weir i O’Sullivan 2017: 136–191). Egzamin został przygotowany we współpracy British Council oraz University of Cambridge Local Examinations Syndicate (UCLES). ELTS podkreślał istotę

³¹ Informacja z wykładu zatytułowanego *Tests, Testers and Testing: The British Council and Language Testing 1941–2013*, zaprezentowanego w listopadzie 2013 roku w Pekinie podczas sympozjum *New Directions English: Language Assessment*. Nagranie umieszczone w Internecie przez British Council China: <https://www.youtube.com/watch?v=Wps9IymFwXA> [dostęp: 30.08.2015].

użycia języka w realnych sytuacjach życiowych w kontekście akademickim (Davies 2007: 79–80; www.ielts.org³²), co było zgodne z nowym podejściem. Celem nauczania języka stała się efektywna komunikacja w różnych sytuacjach. Nastąpiło odejście od nauczania i testowania systemów języka oraz wymagania absolutnej poprawności językowej na rzecz użycia języka jako środka komunikacji.

Rolą ELTS miała być ocena umiejętności językowych ukierunkowanych na konkretny cel. Z tego powodu w ofercie znalazły się moduły do wyboru dla zdającego. Egzamin w swojej pierwszej odsłonie składał się z pięciu części, w tym z dwóch ogólnych, wspólnych dla wszystkich kandydatów, i trzech tzw. modułowych. Części dotyczące ogólnego języka angielskiego to testy umiejętności czytania i słuchania. Moduły do wyboru to: nauki przyrodnicze (*Life Sciences*), nauki społeczne (*Social Studies*), nauki ścisłe (*Physical Sciences*), technika (*Technology*), medycyna (*Medicine*) oraz moduł ogólnoakademicki (*General Academic*). W części modułowej znajdował się test umiejętności uczenia się (*Study Skills*), skonstruowany z zadań opartych na tekście pisany, oraz test sprawdzający umiejętność pisania, jak również rozmowa ustna. Co ciekawe, zadania w części sprawdzającej sprawności pisania i mówienia oparte były na jednym z tekstów z części *Study Skills*. Taka integracja zadań w tym okresie była absolutnym ewenementem. Zadania w części ogólnej sprawdzającej czytanie i słuchanie oraz *Study Skills* skonstruowane były z jednostek wyboru wielokrotnego. Bardzo istotną zmianą wprowadzoną przez nowy egzamin była również obowiązkowa część ustna.

W takiej formie egzamin funkcjonował do 1989 roku, kiedy to do zespołu składającego się z British Council i UCLES dołączył partner z Australii – International Development Program of Australian Universities and Colleges (IDP). Po dokonaniu badań walidacyjnych i rewizji formuły egzaminu ELTS wprowadzono w życie *International English Language Testing System* (IELTS) (Clapham 1996: 57; Weir i O’Sullivan 2017: 184–207).

h) *International English Language Testing System* (IELTS)

International English Language Testing System, łącząc w sobie cechy atomistycznego testowania elementów języka z testowaniem umiejętności komunikacyjnych, na którym bazował ELTS (Davies 2007: 80–81),

³² History of IELTS: http://www.ielts.org/researchers/history_of_ielts.aspx [dostęp: 30.08.2016].

zaferował rozsądny kompromis pomiędzy tymi dwoma szkołami testowania.

W swojej pierwszej wersji w 1989 roku IELTS składał się z dwóch modułów ogólnych sprawdzających umiejętności w zakresie słuchania i mówienia oraz dwóch modułów specjalistycznych mierzących sprawności czytania i pisania. Zadania w częściach specjalistycznych miały na celu sprawdzenie umiejętności kandydatów w ramach wybranych przez nich kierunków studiów. Dostępne wcześniej w egzaminie ELTS sześć modułów ograniczono do trzech dyscyplin:

- modułu technicznego i nauk ścisłych (*Physical Science and Technology*),
- modułu nauk przyrodniczych i medycznych (*Life and Medical Sciences*),
- modułu nauk społecznych i biznesu (*Business Studies and Social Sciences*).

Egzamin można było zdawać również w wersji całkowicie ogólnej, w której części oparte na czytaniu i pisaniu sprawdzały umiejętności w zakresie języka ogólnego.

Podczas rewizji egzaminu w 1995 roku trzy moduły specjalistyczne zostały zredukowane do dwóch: *Academic Reading* oraz *Academic Writing*. Czytanie i pisanie przestały być od tej pory ze sobą połączone tematycznie, jak to miało miejsce w wersji z roku 1989. Celem rozdzielenia tych dwóch sprawności było zapewnienie trafności badania każdej z nich.

Na przestrzeni lat egzamin IELTS poddany był kolejnym zmianom (2001 – rewizja sprawdzianu mówienia i 2005 – nowe kryteria oceny wypowiedzi pisemnej), obecnie nadal funkcjonuje na rynku edukacyjnym, ciesząc się ogromną popularnością i uznaniem (Read 2015: 125–128; Weir i O’Sullivan 2017: 208–254).

Egzaminy dla wojska

II wojna światowa uświadomiła potrzebę odpowiedniej oceny umiejętności językowych żołnierzy. Wcześniej, mierzenie poziomu kompetencji językowej wojskowych odbywało się wyłącznie na zasadzie samooceny, co niejednokrotnie okazywało się metodą zgubną, zwłaszcza w tak wrażliwej branży (Pulliam i Ich 1968). Pierwsze egzaminy językowe z przeznaczeniem dla wojska powstały w Stanach Zjednoczonych. *Army Language School* we współpracy z *Personnel Research Office of the Adjutant General’s Office* stworzyły w 1948 roku *Army Language Tests*. W 1958 roku natomiast wprowadzony został ustny *Foreign Service Institute* (FSI) – pierwszy

amerykański bezpośredni test mówienia, na podstawie którego parę lat później stworzona została skala biegłości dla wszystkich czterech sprawności językowych – *Interagency Language Roundtable Skill Level Descriptions* (ILR). Ta z kolei, stała się bazą dla deskryptorów biegłości językowej wykorzystywanych przez NATO: STANAG 6001.

a) *Army Language Tests*

Pierwsze egzaminy językowe dla wojska powstały w Stanach Zjednoczonych w roku 1948. Zostały wówczas przygotowane arkusze egzaminacyjne dla 25 języków, skonstruowane z zadań badających sprawność czytania, słuchania i pisania oraz znajomość gramatyki (Brooks i Hoffman 2014: 1–2).

Po kilku latach intensywnego wykorzystywania arkuszy egzaminacyjnych stwierdzono konieczność zmiany nie w pełni sprawdzającej się formuły. W 1954 roku stworzone zostały dwa prototypy testów mierzących umiejętność tłumaczenia z języka rosyjskiego i mandaryńskiego (chińskiego). Egzaminy składały się z dwóch części: sprawdzającej umiejętność słuchania oraz testu rozumienia tekstu czytanego. Każda z części składała się z 60 zadań wyboru wielokrotnego. W następnej kolejności dokonano tłumaczenia angielskiej wersji prototypów egzaminów – punkt po punkcie – na inne języki. W efekcie stworzono identyczne arkusze dla 36 języków. Powstałe w ten sposób nowe egzaminy przemianowano na *The Army Language Proficiency Tests* (ALPT; Pulliam i Ich 1968).

Fakt, że treści egzaminu zostały przetłumaczone na inne języki z oryginalnych wersji przygotowanych dla języka rosyjskiego i chińskiego, był przez lingwistów krytykowany. Ich zdaniem procedura ta miała negatywny wpływ na trafność testów (Sinaiko 1966: 28). Krytyce podlegał również fakt, że egzamin miał wyłącznie formę pisemną, a więc nie sprawdzał bardzo ważnych umiejętności mówienia i ustnego reagowania na usłyszane teksty (Sinaiko 1966: 34).

b) Egzamin ustny *Foreign Service Institute* (FSI) oraz skala biegłości *Interagency Language Roundtable Skill Level Descriptions* (ILR)

Równocześnie z pracami nad *Army Language Tests*, a następnie *Army Language Proficiency Tests* (ALPT), *Foreign Service Institute* (FSI) zajmował się opracowaniem testów umiejętności wykonania językowego na potrzeby *US Civil Service Commission*. Ich częścią miał być test sprawdzający umiejętność mówienia.

Podczas gdy egzaminy ustne pozostawały ważną częścią testowania w okresie przednaukowym, w okresie psychometryczno-strukturalistycznym były rzadko obecne w bateriach testów albo też sprawność ta była sprawdzana wyłącznie pośrednio. Wynikało to z faktu, że ocena egzaminu ustnego stanowiła duże wyzwanie dla rzetelności testu (Spolsky 1995: 99). W związku z tym, że nie istniały opracowane kryteria oceny wypowiedzi ustnej, FSI przygotował swoją własną sześciopięciową skalę ze wskaźnikami dla każdego poziomu. I tak w roku 1958 wprowadzony został ustny test FSI (Brooks i Hoffman 2014: 1–2). Ocena każdej wypowiedzi przez kilku egzaminatorów oraz wprowadzenie kryteriów doprowadziły do zwiększenia obiektywności oceniania, oceny przyznawane przez różnych egzaminatorów stały się porównywalne, a sam test stał się powszechnie uznanym narzędziem.

Opracowana skala składała się z sześciu poziomów biegłości zawierających listę elementów, które miały być wzięte pod uwagę podczas oceny egzaminatora. Te elementy to akcent, zrozumiałość wypowiedzi, jej płynność oraz słownictwo i gramatyka. Deskryptory były skonstruowane w taki sposób, że skala mogła być stosowana w odniesieniu do różnych języków, nie tylko do angielskiego. Ocena na poziomie 1 świadczyła o braku umiejętności posługiwania się językiem obcym, natomiast poziom 6 oznaczał biegłość odpowiadającą umiejętnościom reprezentowanym przez natywnego użytkownika języka.

Od samego początku prac nad nowoczesną skalą oceny wypowiedzi ustnej świadomość roli kompetencji językowej i umiejętności komunikacyjnych były bardzo ważnymi kwestiami. Zadanie sprawdzające umiejętność komunikowania się w języku obcym polegało na odgrywaniu ról.

Test FSI został wprowadzony do oceny biegłości językowej pracowników różnych amerykańskich agencji rządowych, a jego skala oceny została zaadaptowana przez agencje Departamentu Obrony Stanów Zjednoczonych – Defense Language Institute, Central Intelligence Agency (CIA) oraz Korpus Pokoju – do mierzenia ustnego posługiwania się językiem angielskim przez swoich ochotników (Fulcher 1993: 12; Herzog b.d.).

Egzamin FSI został uznany przez Spolsky'ego (1995: 99) za pierwszy ustny test biegłości sprawdzający umiejętność mówienia w języku obcym w sposób bezpośredni. W 1968 roku kryteria opracowane na potrzeby mierzenia umiejętności mówienia zostały rozszerzone na poz-

stałe sprawności i tak powstały deskryptory ILR (*Interagency Language Roundtable Skill Level Descriptions*; Brooks i Hoffman 2014: 2–3 i ILR³³). Deskryptory FSI zostały zmodyfikowane i rozszerzone, a następnie uzupełnione o poziomy pośrednie (1+, 2+ itd.). Nowa skala ILR została ponownie wprowadzona do szerokiego użycia w amerykańskich agencjach rządowych.

c) STANAG 6001

Świadomość potrzeby zastosowania kryteriów oceny umiejętności językowych dla celów wojskowych była obecna również w Wielkiej Brytanii. W 1976 roku Bureau for International Language Coordination (BILC) przy brytyjskim Ministerstwie Obrony zaadoptowało, a następnie zaadaptowało amerykańską skalę ILR i zapoczątkowało *Porozumienie o standaryzacji w sprawie stopni znajomości języków – STANAG (Standard Agreement) 6001* (Brooks i Hoffman 2014: 3) obowiązujące do dzisiaj w krajach członkowskich NATO (Gos 2009: 126–127). Skala STANAG 6001 składa się z sześciu stopni – od całkowitego braku umiejętności do poziomu umiejętności reprezentowanego przez natywnego użytkownika języka, i odnosi się do wszystkich czterech sprawności.

Egzaminy w branży medycznej

Potrzeba sprawdzenia umiejętności językowych osób ubiegających się o uznanie kwalifikacji medycznych i możliwość wykonywania zawodów medycznych w innym kraju i w innym języku niż uzyskane kwalifikacje stawia przed instytucjami poszczególnych państw konieczność określenia odpowiednich procedur oraz formy takiej kwalifikacji językowej. W krajach anglojęzycznych potrzeba pozyskania wykwalifikowanych specjalistów z dziedziny nauk medycznych jest bardzo duża. Dowodem na to może być liczba lekarzy praktykujących lub uzupełniających swoje wykształcenie w Stanach Zjednoczonych, którzy kwalifikacje zawodowe uzyskali poza granicami tego kraju. Według informacji na stronie internetowej Educational Commission for Foreign Medical Graduates wynosi ona ok. 25%³⁴.

³³ Interagency Language Roundtable. *History of the ILR Scale*, <http://www.govtilr.org/Skills/IRL%20Scale%20History.htm> [dostęp: 7.11.2019].

³⁴ Informacja ze strony: <http://www.ecfmg.org/about/history.html> [dostęp: 7.02.2020]. Dane na stronie uaktualnione: 8.02.2017 r.

Procedury ubiegania się o uznanie kwalifikacji, w tym sposób potwierdzania umiejętności w zakresie języka angielskiego, podlegały zmianom na przestrzeni lat i różnią się w zależności od kraju. Poniżej przedstawiono przykłady wczesnych egzaminów weryfikujących kompetencje językowe lekarzy oraz zmian, jakie zachodziły w potwierdzaniu kompetencji językowych w branży medycznej na przykładzie egzaminu z medycznego języka francuskiego oraz egzaminów z języka angielskiego dla lekarzy i innych specjalistów nauk medycznych w Wielkiej Brytanii, Australii i Stanach Zjednoczonych.

a) Egzamin językowy dla studentów medycyny w koloniach francuskich

Już w okresie przednaukowym mamy ślady praktycznego egzaminu sprawdzającego znajomość specjalistycznego języka obcego. Metoda bezpośrednia, ze swoim naciskiem na język mówiony i jego praktyczną znajomość, była promowana na kontynencie europejskim w formalnym systemie edukacji (Howatt 2014: 82–84). Można ją również zaobserwować w nauczaniu języka zawodowego. Na początku XX wieku w koloniach francuskich wykład akademicki na studiach medycznych był zastępowany kształceniem przy łóżku chorego. Egzamin językowy stanowił wywiad na temat pacjentów prowadzonych przez danego lekarza na podstawie ich kart historii choroby (Mourlhon-Dalies 2008: 62–63 za: Salinger 2007).

b) Egzamin *Temporary Registration Assessment Board* (TRAB) i *Professional and Linguistic Assessments Board* (PLAB)

Za pierwszy prawdziwy egzamin LSP, spełniający podstawowe zasady konstruowania testów LSP, Douglas (2010: 3–4) uważa *Temporary Registration Assessment Board* (TRAB). Egzamin został wprowadzony w 1975 roku przez British General Medical Council w celu oceny umiejętności językowych i zawodowych lekarzy wykształconych poza Zjednoczonym Królestwem, a ubiegających się o czasową zgodę na wykonywanie zawodu w Wielkiej Brytanii. Jak na owe czasy egzamin był złożonym i zaawansowanym testem zarówno pod względem swojego formatu, jak i zawartości oraz konstrukcji (Taylor i Pill 2014: 3). Badał jednocześnie kompetencje zawodowe oraz umiejętności komunikowania się w języku angielskim. Komponent językowy składał się z testu sprawności rozumienia ze słuchu (nagranie na taśmie), tzw. zmodyfikowanego

eseju (*modified essay question* – MEQ) oraz egzaminu ustnego, w którym sprawdzana była zarówno wiedza medyczna, jak i umiejętności językowe (Hunt 1976: 812–813). Zadanie MEQ skonstruowane było z karty historii choroby pacjenta oraz szeregu pytań, na które zdający udzielał krótkiej otwartej odpowiedzi pisemnej, co w efekcie stanowiło rodzaj sterowanej pracy pisemnej. Zadanie tego typu zostało stworzone na użytek właśnie egzaminów dla lekarzy i do tej pory funkcjonuje zarówno w testach wiedzy medycznej, jak i niektórych egzaminach języka angielskiego medycznego (Boczkowski 2001: 149; Khan i Aljarallah 2011).

W ocenie części ustnej egzaminu TRAB brał udział egzaminator lingwista oraz specjalista z dziedziny medycznej, którzy byli przeszkoleni w zakresie stosowania przyjętych kryteriów oceny (Weir 2005: 154).

Część językowa egzaminu oparta była na analizie języka mówionego i pisanego faktycznie używanego przez lekarzy, pielęgniarki i pacjentów w brytyjskich szpitalach. Podczas tworzenia egzaminu językoznawcy pracowali wspólnie z ekspertami z dziedziny medycyny. Egzamin zawierał cechy poprawnie skonstruowanego testu języka specjalistycznego: oparty był na analizie specyficznego użycia języka w docelowych zawodowych sytuacjach komunikacyjnych, zadania były autentyczne, a pomiędzy umiejętnościami językowymi i wiedzą specjalistyczną zdających miała miejsce interakcja (Douglas 2014: 3–4). Pośród zadań egzaminacyjnych znajdowało się zadanie sprawdzające rozumienie ze słuchu oparte na nagraniu, praca pisemna w formie eseju oraz test ustny (Taylor i Pill 2014: 498). Niestety, informacje dotyczące pierwszych arkuszy egzaminacyjnych dostępne są jedynie w opisach zawartych w literaturze naukowej na ten temat. Oryginalne zestawy egzaminacyjne zostały zniszczone przez instytucję zajmującą się ich przeprowadzaniem, obecnie i w przeszłości³⁵.

Po dokonaniu rewizji egzaminu w roku 1978 TRAB przemianowano na *Professional and Linguistic Assessments Board* (PLAB). Egzamin stał się dwuetapowym testem biegłości w języku angielskim oraz wiedzy medycznej i umiejętności komunikacyjnych. W pierwszej części egzaminu sprawdzane były wiedza i umiejętności medyczne. Część druga – *Objective Structured Clinical Examination* (OSCE) – to test, którego celem jest zbadanie umiejętności klinicznych kandydata, w tym umiejętności efektywnej komunikacji z pacjentami, ich rodziną i innymi pracownikami branży medycznej w języku angielskim (Taylor i Pill 2014: 498).

³⁵ Informacja z własnej korespondencji z General Medical Council z dn. 10.09.2015 r.

Egzamin w takiej formie funkcjonował do 1995 roku, kiedy to w miejsce komponentu *Use of English* wprowadzono możliwość, a później obowiązek dokumentowania umiejętności językowych za pomocą coraz bardziej powszechnego certyfikatu językowego IELTS. Celem wprowadzenia tego elementu przesiewowego było między innymi oszczędzenie kandydatom kosztów związanych z koniecznością przybycia na egzamin do Wielkiej Brytanii w przypadku, gdyby ich umiejętności językowe były niewystarczające do podjęcia egzaminu medycznego (Taylor i Pill 2014: 499–500). Ten sposób dokumentowania kompetencji językowych lekarzy starających się o uznanie kwalifikacji w Wielkiej Brytanii obowiązuje do dnia dzisiejszego. W części drugiej nadal obowiązującego egzaminu PLAB w dalszym ciągu bada się – oprócz umiejętności klinicznych – umiejętność efektywnej komunikacji w środowisku medycznym. Część ta przeprowadzana jest przez dwóch egzaminatorów – przedstawicieli branży medycznej; nie bierze w niej udziału specjalista od języka (McNamara 1996: 41; Taylor i Pill 2014: 498).

c) *Occupational English Test (OET)*

W drugiej połowie lat osiemdziesiątych XX wieku, na zlecenie rządu Australii, opracowano egzamin z języka angielskiego dla przedstawicieli profesji medycznych imigrujących do tego kraju. Proces uznawania kwalifikacji w zawodach medycznych od tej pory składa się z trzech etapów: oceny biegłości w zakresie języka angielskiego za pomocą egzaminu *Occupational English Test*, testu wiedzy medycznej z zakresu konkretnej specjalizacji oraz praktycznego egzaminu z kompetencji klinicznych (Jacoby i McNamara 1999).

Do roku 1987 OET był testem języka ogólnego, opartym w dużym stopniu na znajomości angielskiej literatury (Latchford 1996: 3), i jako taki poddawany był krytyce za brak trafności i rzetelności (McNamara 1990: 18). Pojawiła się wówczas propozycja skonstruowania egzaminu, który miał mierzyć umiejętności kandydatów w zakresie efektywnego komunikowania się w miejscu pracy (McNamara 1996: 98). Odbyły się szeroko zakrojone konsultacje z grupami ekspertów, w tym z osobami kształcącymi na kierunkach medycznych, nauczycielami języka angielskiego jako obcego oraz osobami posiadającymi kwalifikacje w branży medycznej uzyskane poza granicami Australii, które były w trakcie lub już przeszły przez proces potwierdzania kwalifikacji zawodowych w tym kraju. Konsultacje te uzupełniono analizą dostępnej literatury oraz bez-

pośrednimi obserwacjami w miejscu pracy (OET 2009: 3). Nowy kształt testu został oparty na analizie zadań zawodowych w branży medycznej w zakresie mówienia, słuchania, czytania i pisanie oraz wynikających z tej analizy potrzeb językowych (OET 2014a: 3).

Od tej pory OET funkcjonuje jako test przesiewowy dla branży medycznej w Australii, Nowej Zelandii i Singapurze. Regularnie udoskonalany, unowocześniany, badany pod względem trafności i rzetelności, sprawdza kompetencje językowe specjalistów z zakresu stomatologii, medycyny, dietetyki, pielęgniarstwa, farmacji, fizjoterapii, radiologii, logopedii, terapii zajęciowej, optometrii, podiatrii i weterynarii.

Test ma konstrukcję tradycyjnych egzaminów GE, podzielony jest bowiem według badanych sprawności. Egzamin opiera się na tematyce medycznej i zadaniach zawodowych dla poszczególnych specjalności medycznych, nie sprawdza jednak wiedzy medycznej. Na takie rozwiązanie nie zezwalają przepisy australijskie (w przeciwieństwie do przepisów w Wielkiej Brytanii [McNamara 1996: 41] i Stanach Zjednoczonych, gdzie połączenie tych dwóch rodzajów kompetencji było możliwe). Wiedza i umiejętności w zakresie medycyny weryfikowane są za pomocą dwóch kolejnych etapów procedury potwierdzania kwalifikacji, a mianowicie wspomnianych wcześniej – testu wiedzy medycznej oraz praktycznego egzaminu z kompetencji klinicznych (OET 2014b).

d) Egzaminy potwierdzające kwalifikacje językowe lekarzy w Stanach Zjednoczonych

W Stanach Zjednoczonych w 1956 roku powstała prywatna organizacja non-profit Evaluation Service for Foreign Medical Graduates (ESFMG), której celem było poświadczanie kwalifikacji medycznych osób przyjeżdżających do USA lub planujących emigrację do tego kraju³⁶. Początkowo obcokrajowcy ubiegający się o uznanie kwalifikacji zdawali między innymi test językowy opracowany przez tę instytucję, z czasem wprowadzono możliwość poświadczenia kwalifikacji językowych za pomocą egzaminu TOEFL lub jego opcjonalnego elementu – *Test of Spoken English* (TSE; Taylor i Pill 2014: 500–501). W latach 1998–2004 kandydaci spoza Stanów, oprócz egzaminu z ogólnego (lub akademickiego) języka angielskiego zobowiązani byli zaliczyć egzamin *Clinical Skills Assessment*, który badał zarówno wiedzę medyczną, jak i kompetencje językowe w bezpośrednim

³⁶ Informacja ze strony: <http://www.ecfmg.org/about/history.html> [dostęp: 7.02.2020]. Dane na stronie uaktualnione: 8.12.2017 r.

kontakcie z pacjentem. W 2004 roku nastąpiła zdecydowana zmiana dotycząca wymagań stawianych kandydatom nienatywnym. Zniesiono wówczas całkowicie obowiązek zdawania egzaminu językowego, na rzecz zintegrowanego egzaminu *Step 2 Clinical Skills (CS) Examination*, który łączy w sobie test kompetencji zawodowych i biegłości językowej. Test ten jest obowiązkowy dla wszystkich kandydatów ubiegających się o prawo wykonywania zawodu lekarza, niezależnie od tego, czy wykształcenie medyczne uzyskali w Stanach Zjednoczonych, czy w innym kraju. Egzamin odbywa się w symulowanych sytuacjach kontaktu lekarza z pacjentem, a ocena umiejętności językowych przeprowadzana nie jest przez lingwistę ani specjalistę z dziedziny medycyny, a przez odpowiednio przeszkolonych „pacjentów” biorących udział w egzaminie. Pacjenci ci oceniają kandydata pod kątem umiejętności uzyskiwania informacji przez lekarza, komunikacji i kompetencji interpersonalnych oraz biegłości w zakresie języka mówionego (Taylor 2013; Taylor i Pill 2014: 500–501; USMLE 2016: 4).

Język angielski w biznesie i w miejscu pracy

Wraz z coraz większą popularnością nauczania biznesowego języka angielskiego w Azji, a zwłaszcza w Japonii i Chinach, pojawiła się również potrzeba opracowania narzędzia do pomiaru umiejętności w tym zakresie. Prekursorem w dziedzinie testowania *Business English* została instytucja zajmująca się testowaniem w Stanach Zjednoczonych od 1947 roku – Educational Testing Services. Nieco później tematem zajęły się tym również firmy brytyjskie trudniące się badaniem umiejętności językowych, w tym UCLES.

a) *Test of English for International Communication (TOEIC)*

Według Barry’ego O’Sullivana (2006: 1–2) pierwszym egzaminem z biznesowego języka angielskiego jest *Test of English for International Communication (TOEIC)* wprowadzony na rynek edukacyjny w Stanach Zjednoczonych w 1979 roku przez Educational Testing Services (ETS). Pierwszy test, zaprojektowany na użytek rynku japońskiego, opierał się całkowicie na teorii psychometryczno-strukturalistycznej i był skonstruowany wyłącznie z zadań wyboru wielokrotnego. W takiej formie funkcjonuje do dnia dzisiejszego. Z jednej strony egzamin cieszy się ogromnym sukcesem komercyjnym, z drugiej, między innymi ze względu na swoją formę, poddawany jest surowej krytyce, głównie przez brytyjskich specjalistów (O’Sullivan 2006: 19).

b) *Business English Certificate* (BEC)

Na początku lat dziewięćdziesiątych XX wieku szybki rozwój chińskiej ekonomii spowodował zwiększenie świadomości znaczenia języka angielskiego w międzynarodowym handlu i biznesie. Rząd chiński zwrócił się do UCLES z prośbą o stworzenie egzaminu mierzącego umiejętności językowe w zakresie biznesu. Pierwszy egzamin *Business English Certificate* na najniższym planowanym poziomie (BEC 1) odbył się właśnie w Chinach w 1993 roku, a przystąpiło do niego 5 tys. kandydatów. W kolejnym, 1994, roku po raz pierwszy przeprowadzony został egzamin na wyższym poziomie: BEC 2, a w 1996 – BEC 3. Te pierwsze egzaminy biznesowe kierowane były głównie do regionu Azji i Pacyfiku, stąd też ich treści odnosiły się do lokalnych warunków tego obszaru. Potrzeby w zakresie testowania języka biznesu w innych częściach świata zostały uwzględnione dopiero w 1998 roku (Booth 2002: 4).

Według początkowych założeń egzamin miał obejmować dwa poziomy biegłości językowej – A2 i B1 (ESOKJ). Jednak ze względu na bardzo ograniczony zakres języka biznesu, który mógłby być testowany na poziomie A2, BEC 1 zawierał też elementy języka ogólnego oraz biurowego. W efekcie tej pierwszej próby ustalono, że odpowiednim poziomem do sprawdzania języka specjalistycznego biznesowego to poziom B1. Krytyce podlegała także część ustna egzaminu, która uznana została za zbyt ogólną i mało specjalistyczną (Booth 2002: 5–6).

Brytyjski egzamin, w przeciwieństwie do swojego amerykańskiego odpowiednika, wpisywał się w erę psycholingwistyczno-socjolingwistyczną, traktując testowanie języka w sposób bardziej holistyczny i zintegrowany niż testy WW egzaminu TOEIC. Już w swojej pierwszej wersji zawierał zarówno zadania zamknięte, jak i otwarte, i składał się z testów mierzących poszczególne sprawności: czytania, pisania, słuchania i mówienia.

W 1998 roku rozpoczął się proces dostosowywania egzaminów BEC do potrzeb szerszego, międzynarodowego odbiorcy; treści zostały mocniej powiązane z tematyką biznesową oraz bardziej uniwersalne, uwzględniające potrzeby i doświadczenia kandydatów z różnych krajów. Zmieniły się także proporcje pomiędzy poszczególnymi częściami, a tym samym testowanymi sprawnościami. W pierwszych wersjach egzaminu większy nacisk kładziono na sprawności receptywne, a zwłaszcza na umiejętność czytania. Po rewizji egzaminów w 1999 i 2000 roku proporcje pomiędzy czterema sprawnościami zostały zrównane, co spowodowało, że sprawności produktywne stały się bardziej istotne niż wcześniej (O'Sullivan 2006: 123).

c) *Business Language Testing Service* (BULATS)

W 1997 roku został wprowadzony w życie wielopoziomowy egzamin *Business Language Testing Service* (BULATS), sprawdzający umiejętności z zakresu języka obcego w pracy i biurze. Koncepcja nowego egzaminu pojawiła się wraz z pomysłem stworzenia serii egzaminów biznesowych z różnych języków (O'Sullivan 2006: 63). Celem nowego testu miało być mierzenie kompetencji językowych w zakresie związanym z miejscem pracy, a oprócz języka angielskiego był on przeprowadzany dla języka hiszpańskiego, francuskiego i niemieckiego.

W wersji oryginalnej arkusz standardowy testu składał się z trzech sekcji: słuchania, czytania oraz gramatyki i słownictwa. Od 1998 roku kandydaci mogli przystąpić również do części ustnej oraz do testu sprawdzającego sprawność pisania, natomiast w roku 2000 wprowadzona została komputerowa wersja egzaminu (Booth 2002: 7).

BULATS został oficjalnie wycofany z oferty Cambridge Assessment English w 2019 roku i zastąpiony przez *Linguaskill Business*³⁷ – wielopoziomowy test przeprowadzany online, oparty na technologii sztucznej inteligencji.

3. Podsumowanie: zmiany w podejściu do testowania

Historia testowania obejmuje wiele wydarzeń, z których część stanowiła jej kamienie milowe (Tabela 3). Zaproponowany przez Bernarda Spolsky'ego podział historii testowania na trzy okresy został przyjęty przez rzeszę badaczy zajmujących się tą tematyką, a przez niektórych częściowo zmodyfikowany. Cyril J. Weir (1990), który dopisał do podziału dokonanego przez Spolsky'ego okres komunikacyjny, bardzo obrazowo porównuje zmiany zachodzące w zakresie testowania do wahadła, które kołysze się pomiędzy dwoma ekstremalnymi podejściami: testowaniem analitycznym i integracyjnym (Rysunek 2). Nacisk w poszczególnych okresach przesuwa się z testowania umiejętności użycia języka (tłumaczenie, pisanie dłuższych tekstów, rozumienie tekstów literackich, umiejętność streszczania itp.) na testowanie formy (gramatyka, słownictwo, rozróżnianie fonemów itp.) i z powrotem (Davies 2008: 1).

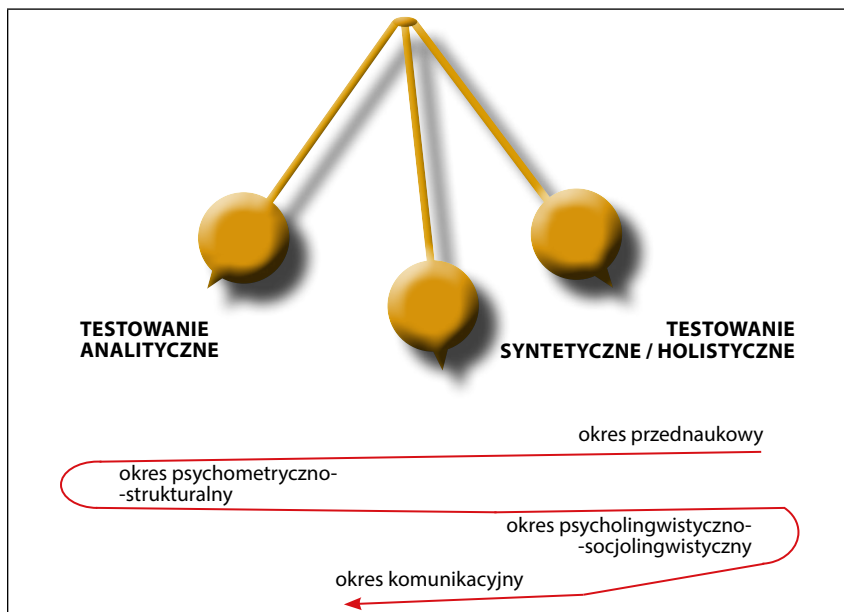
Keith Morrow (1985 za Poszytek 2008: 38–39) dokonał swojej własnej kategoryzacji historycznych okresów w testowaniu, odnosząc się do nich

³⁷ Informacja ze strony: <https://www.cambridgeenglish.org/pl/exams-and-tests/bulats/> [dostęp: 21.05.2020].

nie mniej obrazowo. Mówiąc o okresie przednaukowym, używa metafory „rajskiego ogrodu” (*Garden of Eden*), okresy psychometryczno-strukturalistyczny oraz psycholingwistyczno-socjolingwistyczny to dla niego „dolina łez” (*Vale of Tears*), okres komunikacyjny zaś nazywa „ziemią obiecaną” (*Promised Land*).

Przyglądając się jednak konkretnym przykładom egzaminów, należy stwierdzić, że zmiany w podejściu do testowania nie zachodziły wszędzie w tym samym czasie i nie przebiegały w taki sam sposób w różnych krajach. Zauważył to sam Spolsky (2007), który po latach głębszych badań zaczął mieć wątpliwości co do słuszności zaproponowanej przez siebie przed laty klasyfikacji (Spolsky 2007: 11–13, 2017: 375–376).

Rysunek 2. Metafora wahadła za: Weir (1990)



Podział na trzy lub cztery okresy daje niewątpliwie schematyczny, uproszczony i ograniczony obraz zmian, jakie zachodziły w dziedzinie testowania języków obcych na przestrzeni lat, ale jednocześnie pozwala ją zrozumieć kierunek, powód i cel tych ewolucji. Dlatego też, mimo jej oczywistych mankamentów, badający ten temat lingwiści wciąż odnoszą się do zapoczątkowanej przez Spolsky’ego kategoryzacji.

Spójrzmy, w jaki sposób i w jakim stopniu konkretne przykłady egzaminów odbiegają od zarysowanych ram.

Formalne testowanie umiejętności obcokrajowców w zakresie języka angielskiego w Wielkiej Brytanii rozpoczęło się wcześniej niż w Stanach Zjednoczonych. Uniwersytet w Cambridge zaangażował się w egzaminowanie poza granicami kraju już kilka lat po powstaniu Local Examinations Syndicate (1858). W 1863 roku pierwszych dziesięciu kandydatów zostało przeegzaminowanych w Trynidadzie, a w 1898 istniało już 36 centrów egzaminacyjnych w innych koloniach brytyjskich. W tym samym roku do egzaminów przystąpiło 1220 kandydatów (Roach 1971: 145–146). Jednak prawdziwym kamieniem milowym w historii testowania był egzamin CPE, przeprowadzony po raz pierwszy w 1913 roku. W Stanach Zjednoczonych pierwszy egzamin sprawdzający znajomość języka angielskiego jako obcego pojawia się dopiero w 1930 roku (*College Entrance Examination Board English Competence Examination*).

Prace nad rozwojem testowania języków obcych i zmiany w sposobie testowania następowały w innym tempie w różnych krajach. Wynikało to z odmiennych potrzeb oraz doświadczeń. Przez lata Uniwersytet w Cambridge bardziej zainteresowany był tym, co testować (kompetencje, treści), niż jak to robić (forma testów), odwrotnie niż instytucje zajmujące się testowaniem w USA. Trafność testowania odgrywała zawsze dużo większą rolę w Wielkiej Brytanii niż w Stanach Zjednoczonych, gdzie rzetelność i użyteczność testów były uważane za zdecydowanie ważniejsze. Nie oznacza to, że rzetelność nie miała dla brytyjskich instytucji żadnego znaczenia, jednak nie ona decydowała o tym, jak skonstruowany był test i co było za jego pomocą sprawdzane. Wobec wyboru: trafny test, który nie charakteryzuje się wartościami psychometrycznymi, czy rzetelny test, który niekoniecznie mierzy pożądane umiejętności, ten pierwszy uznawany był przez brytyjskie ośrodki egzaminacyjne za lepszy (Spolsky 1995: 63–66).

W Stanach Zjednoczonych, z ekonomicznej konieczności, pogoń za obiektywnością i praktycznością testów miała decydujące znaczenie. Tam podejście psychometryczno-strukturalistyczne dało o sobie znać dużo wcześniej niż w Wielkiej Brytanii. Początek XX wieku przyniósł konieczność testowania umiejętności językowych tysięcy studentów i pracowników wojska. Testy elementów języka okazały się najszybszym, a więc najbardziej efektywnym sposobem testowania. W połowie lat dwudziestych XX wieku tzw. obiektywne testy albo testy nowego typu (Spolsky 1995: 40–41) były już w USA mocno ugruntowane. W roku 1940 po raz pierwszy do opracowania wyników egzaminów zostały użyte maszyny. W Wielkiej Brytanii natomiast, po latach niechęci wobec psychometrycz-

nego podejścia, w 1966 roku zadania tego rodzaju pojawiły się wreszcie na egzaminie CPE (Spolsky 1995: 84 i 338).

Zmiany w podejściu do testowania nie następowały drastycznie. Przebiegały w różnym tempie w różnych krajach i były wprowadzane stopniowo. Konkretnie egzaminy zwykle nie odzwierciedlały w pełni podejścia obowiązującego w poszczególnych okresach. Elementy charakterystyczne dla jednego okresu były obecne w egzaminach w innych epokach. Zwraca na to uwagę Alan Davies (2008: 1), pisząc, że na wszystkich etapach utrzymywana była pewna równowaga. W okresie tradycyjnym uwagę przywiązywano także do sprawdzania znajomości słownictwa, w okresie strukturalnym pochyłano się również nad zadaniami sprawdzającymi rozumienie tekstów czytanych, a w okresie komunikacyjnym, obok innych umiejętności, sprawdzana jest znajomość gramatyki. Daje się to zaobserwować, przyglądając się dostępnym w literaturze opisom przykładowych egzaminów. I tak na przykład, mimo że CPE w roku 1913 reprezentował okres przednaukowy w testowaniu i gramatyczno-tłumaczeniową metodę nauczania, charakteryzując się „nieobiektywnymi” zadaniami, jednostkami sprawdzającymi umiejętność tłumaczenia tekstów oraz wiedzę językową i kulturową, zawierał on także zadania komunikacyjne, jak mówienie i pisanie (Weir 2013b). Podobnie egzamin dla lekarzy w koloniach francuskich z początku XX wieku – mimo że uplasowany czasowo w okresie przednaukowym, obecne w nim zadania są charakterystyczne dla podejścia komunikacyjnego i zadaniowego. Egzamin *English Proficiency Test Battery* powstał w momencie, kiedy królował jeszcze analityczny system oceniania, co spowodowało, że w dużej mierze składał się z jednostek testujących elementy języka charakterystycznych dla tego okresu, ale pojawiły się w nim także zadania typowe dla kolejnego okresu w historii testowania – psycholingwistyczno-socjolingwistycznego. Tak więc ten konkretny przykład wpisuje się po części w dwa okresy w historii testowania. Na zmiany w testowaniu języków specjalistycznych wpływ miało także uświadamianie sobie odmiennych potrzeb osób przystępujących do takich egzaminów.

John Brian Heaton (1988: 15) uważa, że konstruując testy, nie należy ograniczać się do jednego tylko podejścia, niezależnie od tego, jak doskonałe się wydaje. Nie zawsze też podejścia te nawzajem się wykluczają. Powyższe przykłady dowodzą, że taki stosunek do tworzenia testów mieli w dużym stopniu ich autorzy w różnych okresach historycznych.

Tabela 3. Znaczące wydarzenia w testowaniu języków obcych w układzie chronologicznym

okres przednaukowy	1858	założenie University of Cambridge Local Examinations Syndicate (UCLES)
	1863	pierwsze egzaminy językowe przeprowadzone przez Local Examinations Syndicate w Trynidadzie
	1913	pierwszy egzamin <i>Cambridge Proficiency in English</i> (CPE)
	1930	egzamin <i>English Competence – College Entrance Examination Board</i>
	1938	CPE – wprowadzono arkusz ekonomiczno-handlowy
	1938	porównanie egzaminów <i>English Competence</i> oraz CPE przez The New York State Board of Education
	1939	<i>Lower Certificate in English</i> (LCE; przemianowany w 1975 roku na <i>First Certificate in English</i>)
	1947	powstanie Educational Testing Service (ETS) w USA
	1948	<i>Army Language Tests</i> (ALT)
	1954	<i>Army Language Proficiency Tests</i> (ALPT)
okres psychometryczno-strukturalistyczny	1958	ustny egzamin <i>Foreign Service Institute</i> (FSI)
	1961	Lado – testowanie atomistyczne
	1963	<i>The English Proficiency Test Battery</i> (EPTB)
	1964	pierwszy <i>Test of English as a Foreign Language</i> (TOEFL)
	1966	egzamin CPE – pierwsze zadania zamknięte wyboru wielokrotnego
	1968	skala ILR (<i>Interagency Language Roundtable Skill Level Descriptions</i>)
	1971	początek prac nad <i>Common European Framework of Reference for Languages</i> (CEFR)
	1975	publikacja poziomów progowych <i>Threshold Level</i>
	1975	CPE – po raz pierwszy pojawia się sprawdzian umiejętności rozumienia ze słuchu i rozumienia tekstu czytanego; znika dyktando i część tłumaczeniowa oraz test z literatury
	1975	<i>Lower Certificate in English</i> (LCE) przemianowany na <i>First Certificate in English</i> (FCE)
okres psycholingwistyczno-socjolingwistyczny	1975	Temporary Registration Assessment Board (TRAB)
	1976	STANAG 6001
	1978	Professional and Linguistic Assessments Board (PLAB)
	1979	TOEFL <i>Test of Spoken English</i>
	1979	<i>Test of English for International Communication</i> (TOEIC)
	1980	miejsce EPTB zajmuje <i>English Language Testing Service</i> (ELTS)
	1986	TOEFL <i>Test of Written English</i>
	1987	nowa forma egzaminu <i>Occupational English Test</i> (OET)
	1989	miejsce ELTS zajmuje <i>International English Language Testing System</i> (IELTS)
	1990	powstanie Association of Language Testers in Europe (ALTE)

1991	<i>Certificate in Advanced English (CAE)</i>
1993	<i>Business English Certificate (BEC)</i>
1997	<i>Business Language Testing Service (BULATS)</i>
1998	pierwsza wersja komputerowa egzaminu TOEFL (<i>Computer-based Test</i>)
2001	publikacja <i>Common European Framework of Reference for Languages (CEFR)</i>
2005	pierwsza wersja internetowa egzaminu TOEFL (<i>Internet-based Test</i>)

Rozdział III.

Testowanie i certyfikacja języków obcych dzisiaj

Jak twierdzi Douglas (2014: 70), każdy test stworzony jest dla jakiegoś konkretnego celu, a granice pomiędzy testami językowymi dla celów ogólnych oraz testami dla celów specyficznych dość mocno się zacierają. Mimo to obserwujemy na rynku edukacyjnym ofertę egzaminów certyfikujących język tzw. ogólny, bez żadnego ukierunkowania zawodowego, oraz egzaminy przeznaczone dla specyficznych celów – czy to akademickich, biznesowych, czy też skierowanych do konkretnych zawodów bądź branż.

Głównym celem niniejszego rozdziału jest przegląd i porównanie dostępnych na rynku egzaminów z języka angielskiego ogólnego oraz testów egzaminacyjnych z różnych języków specjalistycznych. Na podstawie analizy wybranych przykładów wskazane zostaną cechy, które odróżniają te dwa rodzaje testów. Do celów opisowych i porównawczych wyselekcjonowane zostały przykłady bądź to najbardziej znanych i popularnych certyfikatów, bądź najbardziej charakterystyczne lub interesujące z perspektywy badacza tematu testowania. Prezentację egzaminów poprzedzają odniesienia do skal i opisów poziomów biegłości językowej wykorzystywanych w certyfikacji ogólnej i specjalistycznej, a w szczególności stosowanych w opisach umiejętności ocenianych za pomocą analizowanych tutaj egzaminów.

Do opisu i porównania egzaminów zastosowano trzy formularze: *Formularz analizy egzaminów GE* (Załącznik 1), *Formularz analizy egzaminów LSP* (Załącznik 2) oraz *Zestawienie analizowanych egzaminów*

GE i ESP (Załącznik 3). Metodologia analizy poszczególnych egzaminów została opisana odpowiednio we wstępach do podrozdziału dotyczącego egzaminów z języka ogólnego oraz egzaminów specjalistycznych.

1. Określenie poziomu biegłości językowej

W przypadku wczesnych standaryzowanych egzaminów językowych trudno jest mówić o porównywalności poziomów. Umiejętności językowe zdającego spełniały albo też nie spełniały założonych wymagań w ramach narzuconego przez konkretny egzamin poziomu. Przyznany certyfikat potwierdzał umiejętności zdającego zbadane danym testem, a nie na danym poziomie. Na rynku edukacyjnym dostępnych było niewiele certyfikatów, więc instytucje, które miały z tymi dokumentami do czynienia, wiedziały, jaki ogólny poziom powinny reprezentować osoby legitymujące się konkretnym świadectwem. Dokonywane były pojedyncze badania porównawcze pomiędzy wybranymi testami, jednak zasadniczym celem takiego zestawienia było ustalenie, czy dane egzaminy sprawdzają umiejętności kandydatów na zbliżonym poziomie zaawansowania (np. porównanie amerykańskiego *English Competence Examination* z brytyjskim *Cambridge Proficiency in English* w 1938; Bachman *et al.* 1995: 4–6). Wyniku danego egzaminu nie odnoszono do konkretnego poziomu biegłości, gdyż nie istniały żadne wspólne klasyfikacje umiejętności językowych. Egzamin sam w sobie stanowił odnośnik poziomu.

Jednym z powodów rozpoczętych w 1971 roku prac nad wspólnym europejskim systemem klasyfikacji umiejętności językowych była chęć ujednoczenia i unormowania skal poziomów biegłości w językach obcych. O pracach nad projektem mówi m.in. John Trim (2011) w wywiadzie nagrany z okazji 10-lecia publikacji CEFR, a także pisze na ten temat Hanna Komorowska (2016). Wydanie *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe 2001) miało miejsce po 30 latach – w roku 2001; od tej pory CEFR na dobre zagościł we wszystkich działaniach dotyczących uczenia się, nauczania i oceniania języków obcych nie tylko w Europie, ale i na świecie. Publikacja polskiego wydania *Europejskiego systemu opisu kształcenia językowego: uczenie się, nauczanie, ocenianie* (Rada Europy 2003) miała miejsce dwa lata później. W 2018 roku z kolei wydana

została rozbudowana wersja dokumentu: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume with New Descriptors* (Council of Europe 2018).

Jak czytamy w informacji wstępnej *Europejskiego systemu opisu kształcenia językowego*, dokument ten

[...] oferuje wspólną dla całej Europy podstawę do opracowywania planów nauczania, zaleceń programowych, egzaminów, podręczników itp. W wyczerpujący sposób opisuje wiedzę, umiejętności i sprawności potrzebne do osiągnięcia biegłości w posługiwaniu się językiem, a także kontekst kulturowy, w którym jest osadzony każdy język. Istotną jego częścią jest określenie poziomów biegłości językowej, pozwalające oceniać postęp w opanowywaniu danego języka – na każdym etapie uczenia się i przez całe życie. (Rada Europy 2003: 13)

Oprócz omówienia procesów uczenia się i nauczania języka, rozważań na temat roli zadań w tych procesach, przedstawienia różnych celów i sposobów oceniania, bardzo znaczącym elementem ESOKJ są zawarte w nim skale biegłości językowej – skala ogólna, skala samooceny, tabela dotycząca poziomów biegłości językowej w zakresie wypowiedzi ustnych oraz szczegółowe skale wskaźników biegłości językowej. Wszystkie wymienione skale odnoszą swoje wskaźniki do sześciu głównych poziomów biegłości językowej, którymi są:

- poziom podstawowy – A1 i A2,
- poziom samodzielności – B1 i B2,
- poziom biegłości – C1 i C2.

Wskaźniki biegłości określają umiejętności w zakresie różnych sprawności na poszczególnych poziomach w sposób uniwersalny, niezależny od języka obcego. Skale biegłości ESOKJ odnoszą się przede wszystkim do ogólnych celów kształcenia językowego, a kontekst posługiwania się językiem obcym obejmuje zarówno sferę prywatną, publiczną i edukacyjną, jak również zawodową.

Opis biegłości językowej oferowany przez CEFR/ESOKJ jest dzisiaj niewątpliwie najbardziej znanym i najczęściej cytowanym dokumentem w dziedzinie klasyfikacji umiejętności językowych, a do jego skal odnosi się większość instytucji zajmujących się certyfikacją znajomości języków obcych ogólnych. Nie jest on jednak ani pierwszym, ani jedynym istniejącym i obowiązującym sposobem klasyfikacji tych umiejętności (Jang 2014: 23–34).

Dla celów wojskowych w 1976 roku (a więc na wiele lat przed zaistnieniem ESOKJ) sformułowany został opis poziomów umiejętności językowych w *Porozumieniu STANAG 6001*. Dla pilotów i kontrolerów ruchu lotniczego powstała całkowicie niezależna od ESOKJ *Skala klasyfikacji biegłości językowej ICAO*. Swoją własną skalę od lat stosuje IELTS (*IELTS Band Scales*)³⁸. Wyniki na certyfikacie podawane są w dziewięciostopniowej punktowej skali IELTS (1–9) wraz z odniesieniem do poziomów ESOKJ. Na użytek egzaminu sTANDEM dla lekarzy, pielęgniarek i farmaceutów jego twórcy stworzyli własny opis poziomów na podstawie skal i deskryptorów ESOKJ.

Innego zadania podjęli się autorzy projektu *Global Scale of English* (GSE)³⁹, opracowując cztery nowe, bardzo precyzyjne i rozbudowane skale biegłości dla kompetencji w zakresie języka akademickiego, języka angielskiego dla dorosłych, angielskiego w sferze zawodowej oraz języka angielskiego dla dzieci (Pearson 2015a, 2015b, 2015c, 2015d). Skale te są ściśle powiązane ze skalami ESOKJ – uzupełniają je i dokonują bardziej precyzyjnego podziału wewnątrz głównych poziomów *Europejskiego systemu*. Są one obecnie używane obok skal ESOKJ w opisie profilu biegłości kandydatów przystępujących do egzaminów certyfikujących Pearson Assessment.

Podobne rozwiązanie przyjął Cambridge English Assessment, który od stycznia 2015 roku do prezentowania wyników wszystkich swoich egzaminów stosuje własną skalę *Cambridge English Scale*⁴⁰. Ten sposób przedstawiania wyników egzaminów dostarcza szczegółowych informacji o każdej mierzonej sprawności językowej i pozwala dokonać porównania wyników uzyskanych na różnych egzaminach Cambridge wg jednolitej, precyzyjnej skali. *Cambridge English Scale* również przekłada swoje szczegółowe poziomy na ogólną skalę biegłości ESOKJ.

W tej części rozdziału przyjrzymy się opisom biegłości ESOKJ odnoszącym się do sfery zawodowej, jak również innym skalom wykorzystywanym w certyfikacji specjalistycznej.

³⁸ Zob. <https://www.cambridgeenglish.org/exams-and-tests/ielts/results/> [dostęp: 10.07.2020].

³⁹ Strona projektu *Global Scale of English*: <https://www.english.com/gse> [dostęp: 10.07.2020].

⁴⁰ Zob. <https://www.cambridgeenglish.org/pl/exams-and-tests/cambridge-english-scale/> [dostęp: 10.07.2020].

Skale biegłości językowej Europejskiego systemu opisu kształcenia językowego a język specjalistyczny

O tym, że zaproponowane przez ESOKJ skale biegłości zorientowane są na cele ogólne i że do innych celów należy je odpowiednio modyfikować, piszą sami autorzy publikacji (Rada Europy 2003: 32–33). Potrzebę uzupełnienia kryteriów biegłości w odniesieniu do różnych języków specjalistycznych wyraża wielu innych autorów, w tym Elżbieta Gajewska (2011) i Paul Aubin (2008).

Aby sprawdzić, w jakim stopniu skale biegłości ESOKJ zaspakają potrzeby certyfikacji języków specjalistycznych oraz jaki mają wpływ na kształt poszczególnych egzaminów ESP, na potrzeby niniejszej pracy dokonano analizy skal.

Sfera zawodowa w rozumieniu ESOKJ dotyczy sytuacji komunikacyjnych, w których „dana osoba podejmuje działania związane ze swoim zawodem i/lub wykonywaną pracą” (Rada Europy 2003: 52). Pośród gamy deskryptorów dotyczących umiejętności językowych z zakresu języka ogólnego można znaleźć takie, które odnoszą się do sfery zawodowej:

- „Potrafi zrozumieć jasną informację o faktach z życia codziennego lub na tematy związane z pracą⁴¹, odbierając zarówno ogólne przesłanie, jak i konkretne szczegóły”⁴².
- „Potrafi zrozumieć wykłady lub rozmowy na tematy specjalistyczne, pod warunkiem znajomości tematu oraz prostej prezentacji i jasnej struktury wypowiedzi”⁴³.
- „Potrafi zrozumieć główne treści wykładów, wystąpień i raportów oraz innych form prezentacji akademickich/zawodowych, złożonych pod względem treści, leksyki i struktury”⁴⁴.
- „Potrafi czytać proste, konkretne teksty na tematy związane z własną specjalnością i zainteresowaniami, uzyskując zadawalający stopień ich zrozumienia”⁴⁵.
- „Potrafi zebrać informacje, koncepcje i opinie z wysoko wyspecjalizowanych źródeł w swojej dziedzinie”⁴⁶.

⁴¹ Tu i poniżej: podkreślenia moje – A.S.

⁴² Rozumienie ze słuchu: ogólny opis umiejętności. Poziom B1. (Rada Europy 2003: 68).

⁴³ Słuchanie jako widz/słuchacz. Poziom B1. (Rada Europy 2003: 69).

⁴⁴ Słuchanie jako widz/słuchacz. Poziom B2. (Rada Europy 2003: 69).

⁴⁵ Czytanie ze zrozumieniem: ogólny opis umiejętności. Poziom B1. (Rada Europy 2003: 70).

⁴⁶ Czytanie i rozumienie informacji oraz argumentacji. Poziom B2. (Rada Europy 2003: 71).

- „Rozumie długie i skomplikowane instrukcje dotyczące własnej specjalności”⁴⁷.
- „Potrafi dosyć swobodnie prowadzić rozmowę na typowe i nietypowe tematy związane z własnymi zainteresowaniami i sprawami zawodowymi”⁴⁸.
- „Potrafi nadążać ze zrozumieniem dyskusji na tematy związane ze swoją specjalnością”⁴⁹.
- „Potrafi dokonywać wymiany skomplikowanych informacji i porad odnośnie całej gamy spraw związanych z własną rolą w pracy”⁵⁰.
- „Umie z pewną dozą pewności siebie dokonywać wymiany konkretnych informacji na znane sobie, rutynowe i nietypowe tematy dotyczące własnej specjalności”⁵¹.
- „Dysponuje odpowiednim zakresem słownictwa w sprawach związanych z jego specjalnością”⁵².

Wskaźniki te nie odnoszą się zatem do konkretnej branży, dziedziny czy rodzaju pracy bądź zawodu, ale do pracy ogólnie, właśnie jako do jednej ze sfer życia dorosłego użytkownika języka.

Ujmując pewne zakresy umiejętności w sposób tak ogólny, ESOKJ nie uzurpuje sobie wyłączności w zakresie opisu biegłości językowej; przeciwnie: w założeniach publikacji wielokrotnie podkreślane jest, że jego skale dotyczą ogólnych celów kształcenia językowego. Autorzy ESOKJ wskazują wprost, że dla innych celów można, a nawet należy tworzyć nowe, własne skale bądź też zawężać lub rozszerzać już istniejące:

Ustalenie ogólnej skali poziomów w żadnym stopniu nie ogranicza tworzenia własnych skal poziomów i etapów nauczania przez instytucje należące do różnych sektorów i systemów edukacyjnych. [...]

Celowe wydaje się również przedstawianie poziomów i wskaźników biegłości w różny sposób dla różnych celów. [...]

Ograniczając zakres poziomów i kategorii wyłącznie do tych, które odpowiadają konkretnemu celowi, można dodać więcej elementów: bardziej szczegółowych poziomów i kategorii. Takie uszczegółowienie pozwoli na stworzenie zestawu modułów pozostających we wzajemnych

⁴⁷ Czytanie i rozumienie instrukcji. Poziom B2. (Rada Europy 2003: 71).

⁴⁸ Interakcja ustna: ogólna ocena umiejętności. Poziom B1. (Rada Europy 2003: 74).

⁴⁹ Dyskusja formalna i udział w spotkaniach. Poziom B2. (Rada Europy 2003:77).

⁵⁰ Wymiana informacji. Poziom B2. (Rada Europy 2003:79).

⁵¹ Wymiana informacji. Poziom B1. (Rada Europy 2003:79).

⁵² Zakres słownictwa. Poziom B2. (Rada Europy 2003:101).

zależnościach, a jednocześnie nawiązujących do określonej skali ogólnej *Europejskiego Systemu Opisu*.

(Rada Europy 2003: 32–33)

Deskryptory rekomendowane dla egzaminów sTANDEM

Egzamin sTANDEM dla lekarzy, pielęgniarek i farmaceutów był efektem pracy zespołu nauczycieli akademickich, badaczy i ekspertów z różnych krajów europejskich współpracujących w ramach projektu finansowanego ze środków Unii Europejskiej (Charpy i Carnet 2014: 2). Celem projektu było opracowanie na rynek europejski standardów wymagań dotyczących sprawności i umiejętności w zakresie specjalistycznego języka medycznego oraz samego egzaminu mierzącego biegłość językową przedstawicieli tej branży.

Opisując prace nad tworzeniem egzaminu, Charpy i Carnet (2014: 5–6) podkreślają fakt, iż normatywne wskaźniki ESOKJ nie zawsze wpisują się w cele, jakie stoją przed nauczaniem i ocenianiem języków specjalistycznych. Cytowani autorzy stwierdzają, że skale odnoszące się do języka ogólnego nie zaspakajają potrzeb certyfikacji w zakresie LSP. W celu dostosowania skal ESOKJ do potrzeb egzaminów sTANDEM stworzony został opis wskaźników biegłości odnoszący się do czterech głównych sprawności językowych na trzech poziomach objętych egzaminem: B1, B2 i C1 – *Descriptors Recommended for sTANDEM*⁵³.

Na podstawie analizy potrzeb dla rynku medycznego ustalono profil docelowego kandydata egzaminu. Określono, jakie sprawności winny być testowane, a w konsekwencji – z jakich części powinien składać się test. W oparciu o wskaźniki ESOKJ opracowano listę deskryptorów rekomendowanych dla egzaminu sTANDEM. W praktyce ze skal *Europejskiego systemu opisu* wyselekcjonowano te wskaźniki biegłości, które wpisują się w potrzeby zawodowe pracowników branży medycznej, pomijając równocześnie te, które według twórców egzaminu nie dotyczą ich umiejętności docelowych. Deskryptory wybrane do celów testów sTANDEM nie zostały w żaden sposób zmodyfikowane, nie dodano również innych wskaźników spoza ESOKJ. Deskryptory rekomendowane dla egzaminów sTANDEM są zatem listą wyselekcjonowanych na użytek egzaminu wskaźników ESOKJ.

⁵³ Zob. http://www.standem.eu/wp-content/uploads/2012/07/sTANDEMdescriptors_03.08.2012.pdf [dostęp: 16.11.2015].

Tabela stopni znajomości języków STANAG 6001

W 1976 roku weszło w życie *Porozumienie Standaryzacyjne NATO nr 6001 – STANAG 6001* – dotyczące języków obcych w siłach zbrojnych Organizacji Paktu Północnoatlantyckiego (Organizacja Paktu Północnoatlantyckiego. Wojskowa Agencja Standaryzacji 1976). Jednym z elementów dokumentu jest *Tabela stopni znajomości języków* opisująca poszczególne umiejętności na różnych poziomach biegłości, bez odniesienia do konkretnego języka obcego.

W swojej wersji pierwotnej skala składała się z pięciu stopni/poziomów:

- Stopień 1 – podstawowy,
- Stopień 2 – dostateczny (zdolność wykonywania pracy w ograniczonym zakresie),
- Stopień 3 – dobry (minimum profesjonalne),
- Stopień 4 – bardzo dobry (znajomość profesjonalna),
- Stopień 5 – doskonały (język ojczysty lub dwujęzyczność).

W późniejszym czasie do wymienionych wyżej poziomów podstawowych zostały dodane poziomy pośrednie: 0, 1+, 2+, 3+ i 4+ (NATO Standardization Agency 2009).

Skale obejmują cztery sprawności językowe: słuchanie i czytanie oraz mówienie i pisanie, a opisy biegłości językowej na poszczególnych poziomach odnoszą się osobno do każdej z nich. Wskaźniki poziomów składające się na poszczególne skale w swoim założeniu nie odnoszą się do konkretnych zagadnień militarnych. Są one skonstruowane w taki sposób, aby definiować ogólną biegłość językową – umiejętność komunikowania się w szerokim zakresie tematów, w różnych okolicznościach i z różnymi odbiorcami (Monaghan 2012: 29). Na wyższych poziomach biegłości (od stopnia 2), obok wskaźników dotyczących sfery prywatnej i publicznej, występują również deskryptory odnoszące się do sfery zawodowej. Odwołanie się do sfery zawodowej jest tu jednak znowu dość pobieżne: wskaźniki tylko w niewielkim stopniu dotyczą umiejętności potrzebnych w branży wojskowej, nawiązując jedynie dość ogólnie do „tematów zawodowych”, „dziedziny wiedzy fachowej” czy „zawodowych i specjalistycznych sytuacji”. Rzadziej pojawiają się bardziej precyzyjne wskaźniki dotyczące konkretnych sytuacji związanych z pracą w wojsku, jak na przykład:

- „Potrafi posługiwać się językiem w celu realizacji tak typowych zadań w pracy zawodowej jak odpowiadanie na zastrzeżenia, wyjaśnianie

stanowiska, uzasadnianie decyzji i opinii, reagowanie na pytania, prezentacja i obrona przyjętych kierunków działania⁵⁴.

- „Potrafi wykazać się kompetencją językową podczas prowadzenia spotkań, odpraw lub innych rozszerzonych i uszczegółowionych prezentacji, wysuwać hipotezy i realizować zadania w nieznanym temacie obszarach i sytuacjach⁵⁵.
- „Kontekst obejmuje wiadomości, artykuły informacyjne i redakcyjne ze znanych czasopism, skierowanych do wykształconych rodzimych czytelników, korespondencję osobistą i zawodową, sprawozdania oraz materiały z określonych dziedzin wiedzy fachowej⁵⁶. (Załącznik nr 1 do Decyzji Nr 251/MON Ministra Obrony Narodowej z dnia 26 czerwca 2015 r.)

Możemy zatem stwierdzić, że skala stopni znajomości języków STANAG 6001 nie odnosi się bezpośrednio do zagadnień militarnych i w niewielkim stopniu odwołuje się do umiejętności związanych ze specjalistycznym użyciem języka obcego. Tym niemniej egzaminy resortowe przeprowadzane na podstawie *Porozumienia STANAG 6001* na wyższych poziomach obok elementów dotyczących tematyki ogólnej zawierają znaczący komponent ściśle wojskowy, co wynika z dokumentów narodowych państw członkowskich NATO⁵⁷.

Skale stopni znajomości języka STANAG 6001 wrosły w tradycję dokumentowania kwalifikacji językowych w wojskach NATO tak mocno, że nastanie skal ESOKJ, którym poddało się wiele instytucji zajmujących się nauczaniem i ocenianiem biegłości językowej, nie wpłynęło na zmianę obowiązującej w tym środowisku klasyfikacji poziomów biegłości. Także w żadnych oficjalnych dokumentach NATO nie odnosi swoich poziomów biegłości do poziomów ESOKJ. Fakt ten pociąga za sobą problem w postaci braku bezpośredniego przełożenia poziomów STANAG 6001 na poziomy ESOKJ, co ma znaczenie dla uznawalności certyfikatów, potwierdzania kwalifikacji i walidacji efektów uczenia się. Próby porównania tych dwóch systemów opisu biegłości były dokonywane zarówno przez niezależne instytucje (Linguapeace Europe 2005), jak i indywidualnych badaczy, jednak wyniki tych badań nie są spójne. Tak więc na

⁵⁴ Sprawność mówienia. Poziom 3.

⁵⁵ Sprawność mówienia. Poziom 3.

⁵⁶ Rozumienie tekstu czytanego. Poziom 3.

⁵⁷ Najnowszy polski dokument w tej sprawie to Decyzja Nr 251/MON Ministra Obrony Narodowej z dnia 26 czerwca 2015 r. w sprawie kształcenia i egzaminowania ze znajomości języków obcych w resorcie obrony narodowej z późniejszymi zmianami.

chwile obecna brak jest jednoznacznego określenia relacji pomiędzy tymi dwoma skalami (Green 2012: 87). Niemniej właśnie dla celów uznawalności certyfikatów STANAG 6001 zostały przyjęte pewne odniesienia poziomów skali STANAG względem skal ESOKJ. Zestawienie takie jest dostępne m.in. na stronach wydawnictwa publikującego podręczniki przygotowujące do egzaminów wg STANAG 6001 (Tabela 4).

Tabela 4. Porównanie poziomów CEFR i STANAG 6001. Zestawienie na podstawie strony internetowej podręczników kursowych przygotowujących do egzaminów STANAG 6001: *Campaign. English for the Military*. Online: <http://www.campaignmilitaryenglish.com/Course/teacher.htm> [dostęp: 19.04.2020]

STANAG 6001	COMMON EUROPEAN FRAMEWORK	
5555 NATIVE/BILINGUAL		
4444 FULLY PROFESSIONAL	C2 Mastery	Proficient User
	C1 Effective Operational Proficiency	
3333 MINIMUM PROFESSIONAL	B2+ Vantage+	Independent User
	B2 Vantage	
	B1+ Threshold+	
2222 LIMITED WORKING	B1 Threshold	Basic User
	A2+ Waystage+	
1111 ELEMENTARY	A2 Waystage	Basic User
	A1 Breakthrough	
0 UNSCALED		

Skala klasyfikacji biegłości językowej ICAO

Jeszcze innym przykładem opisu wymaganej biegłości językowej, tym razem z przeznaczeniem całkowicie do celów zawodowych, jest *Skala klasyfikacji biegłości językowej ICAO* (*ICAO Language Proficiency Rating Scale*; International Civil Aviation Organization 2010: A7) ustanowiona w 2003 roku przez Międzynarodową Organizację Lotnictwa Cywilnego (ICAO), która zajmuje się między innymi ustalaniem standardów w zakresie bezpieczeństwa komunikacji lotniczej. Skala biegłości, według której oceniane są umiejętności językowe pilotów i kontrolerów ruchu lotniczego pracujących w środowisku międzynarodowym, została opracowana na skutek zwiększających się obaw dotyczących poziomu biegłości językowej pracowników lotnictwa, od którego może zależeć bezpieczeństwo ruchu lotniczego. Istotnie, analizy przyczyn wypadków lotniczych

przeprowadzone pod koniec lat dziewięćdziesiątych XX wieku, wykazały, że problemy w komunikacji językowej były jednym z poważnych czynników mających wpływ na przebieg zdarzeń (Cookson 2009: 22.1–22.14; International Civil Aviation Organization 2010: 1.1–1.2; Read i Knoch 2009: 21.1–21.2).

ICAO nie narzuca konkretnego egzaminu, który pracownicy lotnictwa mieliby zdać w celu poświadczenia wymaganego poziomu biegłości, pozostawiając decyzję dotyczącą wyboru określonego narzędzia pomiaru krajom członkowskim ICAO. Egzaminem obowiązującym polskich pilotów i kontrolerów był do roku 2016 omówiony w dalszej części niniejszej pracy ELPAC, a obecnie jest nim jego polski odpowiednik: Krajowy System Egzaminów Językowych (KSEJ).

Skala klasyfikacji ICAO definiuje sześć poziomów biegłości językowej w kontekście lotniczym oraz w zakresie języka ogólnego i jest niezależna od języka, którego opis ma dotyczyć. Jednakże, mimo że skalę można odnieść do biegłości w dowolnym języku obcym, w praktyce wymagania ICAO dotyczą języka angielskiego jako języka obowiązującego w międzynarodowej komunikacji lotniczej (International Civil Aviation Organization 2010: vii).

Nazewnictwo poszczególnych poziomów przedstawia się następująco:

- Poziom 1 – poniżej podstawowego (*pre-elementary*),
- Poziom 2 – podstawowy (*elementary*),
- Poziom 3 – poniżej operacyjnego (*pre-operational*),
- Poziom 4 – operacyjny (*operational*),
- Poziom 5 – rozszerzony (*extended*),
- Poziom 6 – biegły (*expert*). (Dyrektywa 2006/23/WE Parlamentu Europejskiego i Rady z dnia 5 kwietnia 2006 r. w sprawie wspólnotowej licencji kontrolera ruchu lotniczego)

Poziomy 1–3 nie uprawniają do wykonywania operacji lotniczych w międzynarodowej przestrzeni powietrznej. Osoby, których umiejętności w zakresie języka angielskiego plasują się w obrębie tych poziomów, mogą wykonywać loty wyłącznie w przestrzeni powietrznej swojego kraju. Aby spełnić wymagania ICAO dotyczące biegłości językowej, kandydat do licencji pilota lub kontrolera ruchu lotniczego albo też posiadacz takiej licencji musi wykazać, że osiągnął co najmniej poziom operacyjny (poziom 4).

Skala klasyfikacji ICAO jest całkowicie niezależna od skali ESOKJ i w żaden sposób się do niej nie odwołuje. Biorąc pod uwagę fakt, iż

obydwa systemy składają się z sześciu podstawowych poziomów, można wyjść z założenia, że stopnie jednej skali przekładają się bezpośrednio na poziomy drugiej, nie jest to jednak hipoteza uprawniona. Przede wszystkim skala ICAO odnosi się jedynie do dwóch sprawności językowych. Umiejętności pracowników lotnictwa sprawdzane są wyłącznie w ramach rozumienia ze słuchu oraz mówienia. Ponadto obok umiejętności odnoszących się do ogólnej biegłości językowej oraz tematyki związanej z wykonywaną pracą wskaźniki w skali ICAO odwołują się również do „nadmierzających i niespodziewanych okoliczności”⁵⁸:

- „Mogą występować błędy, szczególnie w okolicznościach nadzwyczajnych i niespodziewanych, ale rzadko zmieniają znaczenie”⁵⁹.
- „Może często z powodzeniem parafrazować wypowiedzi, gdy w okolicznościach nadzwyczajnych i niespodziewanych brakuje mu słownictwa”⁶⁰.
- „Mówiący w obliczu komplikacji językowej lub sytuacyjnej, względnie nieoczekiwanego obrotu wydarzeń rozumie wolniej lub wymaga procesu wyjaśnień”⁶¹.
- „Inicjuje i utrzymuje wymianę zdań, nawet gdy ma do czynienia z niespodziewanym obrotem wydarzeń”⁶².
- „Rozumienie jest poprawne w sprawach powszechnych, konkretnych i związanych z pracą, i w większości poprawne, gdy mówiący staje w obliczu komplikacji językowej lub sytuacyjnej, lub w obliczu niespodziewanego obrotu wydarzeń”⁶³.

W porównaniu z opisanymi wcześniej skalami ESOKJ, sTANDEM i STANAG 6001, skala ICAO jest dużo bardziej ukierunkowana na konkretne zawodowe potrzeby użytkowników języka, których umiejętności językowe miałyby być mierzone za jej pomocą. Ma ona również bardzo duży wpływ na kształt egzaminów, które powstały wskutek wprowadzonych wymagań.

Od momentu publikacji skali ICAO miał miejsce szereg prób porównania jej poziomów z poziomami różnych popularnych egzaminów

⁵⁸ Skala klasyfikacji biegłości językowej – poziomy: biegły, rozszerzony i operacyjny. W: Dyrektywa 2006/23/WE Parlamentu Europejskiego i Rady z dnia 5 kwietnia 2006 r. w sprawie wspólnotowej licencji kontrolera ruchu lotniczego.

⁵⁹ Struktura. Poziom operacyjny 4 (Dyrektywa 2006/23/WE...: L 114/35).

⁶⁰ Słownictwo. Poziom operacyjny 4 (*ibidem*).

⁶¹ Rozumienie. Poziom operacyjny 4 (*ibidem*).

⁶² Interakcja. Poziom operacyjny 4 (*ibidem*).

⁶³ Rozumienie. Poziom rozszerzony 5 (*ibidem*).

językowych (np. TOEFL, TOEIC, IELTS), jak i innych skal biegłości, w tym ESOKJ. Stwierdzono, iż można zauważyć ogólne relacje pomiędzy pewnymi obszarami użycia języka, nie jest jednak możliwe dokonanie całościowej ich korelacji (International Civil Aviation Organization 2010: 4.5.14).

Skala ESOKJ i skale biegłości dla języków specjalistycznych – podsumowanie

Opisane skale biegłości dotyczące specyficznych kontekstów posługiwania się językiem obcym można podzielić na takie, które wyrastają ze skal ESOKJ, a więc odnoszą się do poziomów ESOKJ i jego deskryptorów, oraz takie, które z tym opisem biegłości mają niewiele wspólnego. Skalę sTANDEM stanowią wybrane na potrzeby konkretnego języka specjalistycznego, a właściwie na potrzeby konkretnego egzaminu deskryptory ESOKJ. Poziomy biegłości opisane wskaźnikami STANAG 6001 oraz ICAO nie przekładają się na poziomy ESOKJ, co może w pewnym stopniu utrudniać proces uznawalności certyfikatów, potwierdzania kwalifikacji oraz walidacji efektów uczenia się.

Wspomniane skale można porównywać również w zakresie, w jakim dotyczą określonego języka zawodowego. Skale ESOKJ nastawione są przede wszystkim na ogólne kompetencje językowe. Pośród swoich deskryptorów, ESOKJ posiada takie, które odnoszą się do tematów związanych z pracą użytkownika języka, z jego „własnymi zainteresowaniami i sprawami zawodowymi”, a więc do pracy jako jednej ze sfer życia. Podobnie skale STANAG 6001, które z założenia mają definiować ogólną biegłość językową, dotyczą zatem języka ogólnego i też tylko okazjonalnie odnoszą się do języka w sferze zawodowej. Skale, które opisują umiejętności bliższe konkretnym potrzebom zawodowym, to opisy biegłości sTANDEM i ICAO. Deskryptory sTANDEM zostały specjalnie wyselekcjonowane z ESOKJ pod kątem potrzeb językowych konkretnych grup zawodowych, chociaż właśnie z racji swojego pochodzenia nie odnoszą się bezpośrednio do zawodów, których opis ma dotyczyć. Skala biegłości ICAO odnosi się zarówno do ogólnych umiejętności komunikacyjnych na poszczególnych poziomach, jak i do umiejętności koniecznych przy wykonywaniu zawodów pilota i kontrolera ruchu lotniczego. Należy podkreślić, że mimo mniejszego lub większego dostosowania wskaźników do potrzeb poszczególnych zawodów, wszystkie opisane skale biegłości, jako bazę dla bardziej specyficznych umiejętności, przyjmują umiejętności związane z ogólną biegłością językową.

2. Certyfikacja języków ogólnych

W niniejszym podrozdziale opisanych zostało sześć egzaminów kwalifikujących się do kategorii egzaminów dla celów ogólnych. Do analizy wytypowano jedno z bardziej obecnie popularnych międzynarodowych egzaminów, które oferowane są jako testy z ogólnego języka angielskiego, a dla celów porównawczych przyjęto najczęściej certyfikowany poziom biegłości, jakim jest B2 wg ESOKJ. Dwa spośród opisywanych egzaminów są testami wielopoziomowymi, które mierzą umiejętności językowe na wszystkich lub kilku poziomach biegłości, w tym na interesującym nas poziomie B2.

Analizy porównawczej dokonano, zestawiając opisywane egzaminy w *Formularzu analizy egzaminów GE* (Załącznik 1) oraz w zbiorczym *Zestawieniu analizowanych egzaminów GE i ESP* (Załącznik 3).

Punktem wyjścia do opracowania pierwszego narzędzia był zastosowany w innym badaniu egzaminów specjalistycznych (Sendur 2020) *Formularz analizy egzaminów*. Został on najpierw dostosowany do potrzeb analizy egzaminów specjalistycznych omawianych w niniejszym opracowaniu, a następnie na jego podstawie stworzona została uproszczona wersja, nieuwzględniająca elementów istotnych dla certyfikacji specjalistycznej. Powstały w ten sposób formularz analizy egzaminów ogólnych zawiera informacje dotyczące formy egzaminu (papierowa/komputerowa), elementów składowych arkusza egzaminacyjnego (rozumienie tekstu czytanego, rozumienie ze słuchu, pisanie, mówienie, znajomość struktur językowych, znajomość leksyki ogólnej, umiejętność stosowania funkcji językowych) oraz tego, czy są one obowiązkowe, czy opcjonalne. Wyszczególniono w formularzu zadania, w których występuje integracja sprawności językowych. Zawarta jest w nim również ogólna informacja dotycząca rodzaju wykorzystywanych zadań egzaminacyjnych (zamknięte/otwarte).

Drugie wspomniane narzędzie zawiera bardziej szczegółowe informacje dotyczące zarówno egzaminów GE, jak i ESP. Wymienione są w nim szczegółowe typy zadań wg klasyfikacji Niemierki (1999) w odniesieniu do poszczególnych części egzaminu, jak również proporcje punktowe poszczególnych arkuszy. Zestawienie zawiera również informację dotyczącą skal biegłości, stanowiących odniesienie dla poziomów biegłości sprawdzanych za pomocą egzaminu. Pewne informacje powtarzają się w oby-

dwu formularzach, jednak dwa różne sposoby ich przedstawienia dają możliwość nieco innego spojrzenia na analizowane testy.

Egzaminy z ogólnego języka angielskiego poddane analizie w tej części rozdziału to: *Aptis General*, *Cambridge B2 First*, *IELTS General Training*, *Linguaskill General*, *PTE General* oraz *telc English B2*.

Cambridge B2 First

Egzamin *Cambridge B2 First* (dawniej: *First Certificate in English* oraz *Cambridge English First*; Cambridge Assessment English 2019a) występuje w wersji papierowej oraz komputerowej. Sprawdza znajomość języka angielskiego w ramach czterech podstawowych sprawności językowych: czytanie, słuchanie, pisanie i mówienie. Test umiejętności czytania rozszerzony jest o zadania sprawdzające znajomość gramatyki i leksyki: *Reading and Use of English*. Według opisu na polskiej stronie Cambridge Assessment English⁶⁴ sprawdza on umiejętności zdającego w zakresie czytania różnego rodzaju tekstów, takich jak książki, gazety i czasopisma oraz znajomość gramatyki i słownictwa. Test sprawności pisania (*Writing*) wymaga od kandydatów umiejętności napisania dwóch różnych wypowiedzi w określonej formie. Sprawdzenie rozumienia tekstów słuchanych (*Listening*) bada umiejętność rozumienia szeregu materiałów mówionych, takich jak programy informacyjne, prezentacje i codzienne rozmowy. Z kolei *Speaking Test* sprawdza umiejętność skutecznego komunikowania się w sytuacjach rozmowy bezpośredniej. Do egzaminu ustnego zdający przystępuje wraz z jednym lub dwoma innymi kandydatami.

Część *Reading and Use of English* składa się z zadań sprawdzających rozumienie tekstów czytanych, słownictwa oraz struktur gramatycznych. Jest ona skonstruowana z zadań takich jak *cloze test*, zadanie na dobieranie wielokrotne, przekształcenia słowotwórcze, przekształcenia fleksyjne (transformacje), wybór wielokrotny (WW) oraz uzupełnianie luk na poziomie akapitu. Na arkusz *Listening* składają się zadania zamknięte typu WW oraz wielokrotne dobieranie, ale także zadania otwarte typu uzupełnianie zdań. *Writing* składa się z dwóch zadań: pierwsze to obowiązkowy tekst w formie rozprawki, w drugim zaś zdający może dokonać wyboru spośród trzech tematów. Rodzaje tekstów, jakie mogą pojawić się w tym zadaniu egzaminacyjnym to artykuł, nieformalny lub formalny email lub

⁶⁴ Zob. <https://www.cambridgeenglish.org/pl/exams-and-tests/first/exam-format/> [dostęp: 5.06.2020].

list, raport oraz recenzja. Egzamin ustny sprawdza umiejętności wyrażania i uzasadniania własnych opinii, wyrażania zgody lub sprzeciwu wobec opinii interlokutora oraz umiejętność udziału w dyskusji. Wszystkie części egzaminu dotyczą życia codziennego oraz tematów ogólnych. Części odpowiadające poszczególnym sprawnościom językowym oraz arkusz gramatyczno-leksykalny mają taką samą wartość punktową.

Uczestnik egzaminu otrzyma wynik w postaci ogólnej oceny z całego egzaminu, odniesienia do poziomu CEFR oraz szczegółowej punktacji w ramach poszczególnych części egzaminu z odniesieniem do skali *Cambridge English Scale*.

Pearson Test of English (PTE)

Pearson Qualifications⁶⁵ oferuje m.in. egzaminy certyfikujące z języka angielskiego ogólnego (*PTE General*), akademickiego (*PTE Academic*), dla dzieci w wieku od 6 do 13 lat (*PTE Young Learners*) oraz z języka biznesu (*English for Business*).

*PTE General*⁶⁶ (*PTE General 2018*), podobnie jak inne testy z języka ogólnego, oparty jest na czterech kluczowych sprawnościach językowych i sprawdza biegłość na sześciu poziomach zaawansowania. Egzamin na poziomie B2 (Level 3) sprawdza umiejętności potrzebne do studiowania, pracy, podróży i innych celów związanych z codziennym funkcjonowaniem, a zdający powinni wykazać się rozumieniem oraz umiejętnością wyrażania w mowie i piśmie informacji, pomysłów, uczuć, opinii oraz podstawowych funkcji językowych odnoszących się do sytuacji z życia codziennego.

Egzamin składa się z dwóch głównych części: pisemnej i ustnej. Arkusz pisemny zbudowany jest z zadań zamkniętych oraz otwartych, sprawdzających umiejętności w zakresie słuchania, czytania i pisania. Posiada też charakterystyczny i niespotykany obecnie w innych egzaminach językowych element łączący dwie sprawności. Jest to rodzaj dyktanda, sprawdzający i oceniający równocześnie umiejętności w zakresie dwóch sprawności: słuchania i pisania. Test nie posiada natomiast

⁶⁵ Zob. <https://qualifications.pearson.com/en/qualifications/pearson-test-of-english.html> [dostęp: 10.06.2020].

⁶⁶ Obecna nazwa obowiązuje do końca 2020 roku, od 2021 certyfikat będzie funkcjonować pod nową nazwą: *Pearson English International Certificate* i będzie dostępny również w wersji elektronicznej (informacja z korespondencji z dn. 18.09.2020 r., przesłana przez Pearson Poland do rektorów uczelni członkowskich KRASP za pośrednictwem Przewodniczącego KRASP).

odrębnej części gramatycznej (*Use of English*); poprawność gramatyczna jest weryfikowana pośrednio, w trakcie oceny pozostałych umiejętności językowych.

Writing składa się z dwóch zadań – dwóch tekstów użytkowych o tematyce ogólnej: krótkiego listu/emaila oraz jednego rodzaju tekstu do wyboru spośród szeregu możliwości, takich jak artykuł, wpis na blogu, ulotka, recenzja, raport czy rozprawka. Jakość języka oceniana jest pod względem bogactwa leksykalno-gramatycznego, poprawności, spójności tekstu oraz poprawności ortograficznej.

Tematy, z jakimi zdający będą się mierzyć podczas egzaminu ustnego są związane z danymi osobowymi zdającego oraz jego zainteresowaniami, konkretnymi lub bardziej abstrakcyjnymi tematami ogólnymi w formie prezentacji, dyskusji i ustnego opisu oraz odgrywania ról.

Wszystkie cztery sprawności językowe oceniane w egzaminie PTE mają taką samą wartość punktową. Wyniki prezentowane są w formie ogólnej oceny z egzaminu wraz z odniesieniem do skali ESOKJ oraz precyzyjnych wyników punktowych uzyskanych za poszczególne sprawności językowe oraz sprawdzane podsystemy języka.

Telc English B2

Telc GmbH oferuje egzaminy z wielu języków obcych na różnych poziomach zaawansowania, od A1 do C2. W celach porównawczych analizie poddany zostanie tutaj egzamin z języka angielskiego na poziomie B2.

Telc English B2 (Telc Language Tests 2019) skonstruowany jest z części pisanej oraz mówionej. Na część pisaną składają się sprawdziany umiejętności w ramach sprawności językowych: *Reading Comprehension*, *Listening Comprehension* oraz *Written Expression*, a także test gramatyczno-leksykalny – *Language Elements*. Wszystkie zadania, z wyjątkiem tych w części badającej umiejętność tworzenia tekstu, są jednostkami zamkniętymi: WW, na dobieranie oraz prawda–fałsz. W części badającej sprawność pisania, zadaniem zdającego jest stworzenie półformalnego emaila z wykorzystaniem podanych w arkuszu danych wejściowych.

W egzaminie ustnym – *Oral Expression* – bierze udział jednocześnie dwoje lub troje zdających. Kandydaci muszą wykazać się umiejętnością opisywania własnych doświadczeń z zakresu tematyki związanej z życiem codziennym, brać udział w dyskusji bazującej na tekście o tematyce ogólnej zamieszczonym w karcie egzaminacyjnej oraz wspólnego planowania i podejmowania decyzji opartej na dyskusji z pozostałymi uczest-

nikami egzaminu ustnego. Umiejętności zdającego w zakresie mówienia oceniane są w odniesieniu do czterech podstawowych kryteriów: ekspresja wypowiedzi (styl, bogactwo języka, skuteczność realizacji celu wypowiedzi), sposób realizacji zadania (aktywność w dyskusji, wykorzystanie odpowiednich strategii językowych, płynność wypowiedzi), poprawność językowa oraz wymowa i intonacja.

Cztery elementy arkusza mają taką samą wagę punktową: czytanie, słuchanie, wypowiedź ustna oraz sprawność pisania połączona z testem gramatyczno-leksykalnym i wszystkie są obowiązkowe dla zdającego.

International English Language Testing System (IELTS)

Egzamin IELTS (IELTS 2019) oferowany jest w wersji papierowej oraz komputerowej. To test wielopoziomowy, sytuujący wyniki zdających na 9-stopniowej skali oceny *IELTS Band Scores* (1–9)⁶⁷. Ogólne wyniki punktowe zdającego są również odniesione do poziomów biegłości ESOKJ. Certyfikat występuje w dwóch wersjach: *Academic* oraz *General Training*. Obydwa testy składają się z czterech części opartych na głównych sprawnościach językowych. Części *Listening* i *Speaking* są takie same dla obydwu wersji egzaminu, natomiast tematyka części *Reading* oraz *Writing* różni się w zależności od wybranego wariantu.

Zagadnienia tematyczne podejmowane w egzaminie *IELTS General Training*, który wpisuje się w zakres tego rozdziału, dotyczą sytuacji związanych z pracą i życiem towarzyskim w praktycznych, codziennych kontekstach. *Listening* składa się głównie z zadań zamkniętych (ale znajdziemy tam także zadanie z luką) i sprawdza umiejętności w zakresie rozumienia głównej myśli tekstu, wyszukiwania informacji, rozumienia opinii rozmówcy, celu wypowiedzi i umiejętności podążania za treścią wypowiedzi.

Reading sprawdza cały szereg różnych umiejętności, jak rozumienie głównej myśli tekstu, wyszukiwanie informacji, rozumienie logiki wyводу, opinii autora tekstu i celu wypowiedzi. Wersja badająca umiejętności w ramach języka ogólnego opiera się na materiałach, z którymi użytkownik języka może spotkać się na co dzień w anglojęzycznym środowisku. Teksty użyte w zadaniach egzaminacyjnych to fragmenty literatury, artykuły prasowe, ogłoszenia, reklamy, teksty z podręczników i wytyczne pracownicze. Sprawdzenie rozumienia tekstu czytanego wykorzystuje cały

⁶⁷ Zob. <https://ieltpolska.pl/pl/skala-ielts/> [dostęp: 10.06.2020].

szereg różnych zadań testowych. Pośród zadań zamkniętych możemy tutaj spotkać wybór wielokrotny, prawda/fałsz/brak informacji oraz całą paletę zadań na dopasowywanie. Występują tu również zadania otwarte, jak uzupełnianie zdań, notatek, streszczenia, informacji w tabelach czy wykresach oraz odpowiedzi na pytania.

Tematyka części *Writing*, zgodnie z informacją podaną w opisie egzaminu, odnosi się do ogólnych zainteresowań oraz do sytuacji z życia codziennego. Zadania sprawdzają umiejętności na podstawie dwóch rodzajów tekstu pisanego: nieformalnego, półformalnego lub formalnego listu oraz rozprawki.

Egzamin ustny składa się z trzech zadań opartych na tematyce związanej z życiem prywatnym zdającego, domem, rodziną, pracą, nauką i zainteresowaniami. W ocenie brane są pod uwagę płynność i koherencja wypowiedzi, zasób leksyki, zasób struktur i ich poprawność oraz wymowa.

Wyniki każdego odrębnego arkusza są odnoszone do kryteriów 9-stopniowej skali IELTS. Wszystkie części egzaminu oparte na głównych sprawnościach językowych mają taką samą wagę w skali całego testu. Na podstawie wyników poszczególnych arkuszy egzaminu określany jest średni poziom umiejętności kandydata, wyrażony w skali IELTS oraz ESOKJ.

Aptis

Aptis jest jednym z najnowszych międzynarodowych egzaminów certyfikujących z języka angielskiego. Jest to wielopoziomowy test komputerowy oferowany przez British Council, dający szybką ocenę umiejętności językowych osoby zdającej, ze wskazaniem poziomu w skali Rady Europy (ESOKJ) (British Council 2020). Certyfikat występuje w czterech wariantach: *Aptis General*, *Aptis Advanced*, *Aptis for Teens* oraz *Aptis for Teachers*. W tej części publikacji interesują nas dwie pierwsze wersje egzaminu.

Aptis General jest „testem z języka angielskiego ogólnego przeznaczonym dla osób dorosłych funkcjonujących w środowisku zawodowym”⁶⁸. Test ten bada umiejętności posługiwania się językiem angielskim w przedziale od A1 do C według skali Rady Europy, bez rozróżnienia na poziomy C1 i C2. *Aptis Advanced* z kolei testuje umiejętności językowe w przedziale od B1 do C2.

⁶⁸ Zob. <https://www.britishcouncil.pl/egzaminy/aptis/czym-jest-aptis> [dostęp: 7.06.2020].

Egzamin w obydwu wersjach składa się z głównej części obowiązkowej dla wszystkich kandydatów – *Aptis Core Test*, zbudowanej z zadań sprawdzających umiejętności w zakresie gramatyki i słownictwa – oraz z oddzielnych, nieobowiązkowych testów mierzących umiejętności w zakresie pisania, mówienia, rozumienia ze słuchu oraz czytania. W zależności od własnych potrzeb bądź też wymagań pracodawcy kandydaci mogą ograniczyć się do wybranych elementów egzaminu lub przystąpić do wszystkich jego części.

Jako że *Aptis* jest egzaminem komputerowym i z założenia ma udostępniać wyniki w trybie prawie ekspresowym (do 48 godzin), części *Grammar & Vocabulary*, *Reading* oraz *Listening* składają się wyłącznie z zadań zamkniętych (WW i na dobieranie). Arkusz sprawdzający umiejętność tworzenia tekstu pisanego złożony jest z czterech zadań o rosnącym poziomie trudności. Gradacja sprawdzanych umiejętności postępuje od krótkich, kilkuwyrazowych, przez kilkudzaniowe odpowiedzi na pytania, po dwa dłuższe teksty w formie emaila – nieformalnego oraz formalnego. Część ustna również odbywa się za pośrednictwem komputera. Zdający nagrywa swoje odpowiedzi na pojawiające się na ekranie pytania. Egzamin ustny składa się z czterech zadań, każde z serią pytań o rosnącym poziomie trudności. Pytania sprawdzają umiejętności (takie, jak opisywanie, wskazywanie wad i zalet, wyrażanie i uzasadnianie opinii) oraz dotyczą zarówno codziennych życiowych sytuacji, jak i tematów abstrakcyjnych. Odpowiedzi do zadań otwartych sprawdzane są przez egzaminatora, natomiast wyniki do zadań zamkniętych generowane są komputerowo.

Wszystkie pięć części egzaminu – cztery testy sprawnościowe oraz test leksykalno-gramatyczny – mają tę samą wartość punktową.

Linguaskill General

Zdecydowanie najmłodszym egzaminem certyfikującym z języka angielskiego, a jednocześnie jednym z najbardziej zaawansowanych technicznie jest *Linguaskill* (Cambridge Assessment English 2017; Cambridge Assessment English 2019c; Ludlow 2020) – egzamin Cambridge English Assessment dostępny w dwóch wersjach: *General* oraz *Business*. Wariant biznesowy zostanie omówiony w dalszych podrozdziałach, ale należy tutaj wspomnieć, że obydwa testy są skonstruowane w identyczny sposób i różnią się jedynie tematyką tekstów oraz charakterem zadań do wykonania.

Linguaskill jest adaptywnym testem wielopoziomowym sprawdzającym biegłość kandydatów na poziomach od A1 do C według skali ESOKJ w formie całkowicie komputerowej. Struktura egzaminu oparta jest na czterech podstawowych sprawnościach językowych: *Reading & Listening*, *Writing* i *Speaking*. Z informacji na stronie Cambridge English Assessment⁶⁹ wynika, że w planach jest jeszcze arkusz leksykalno-gramatyczny: *Use of English*. Kandydat nie musi zdawać wszystkich części egzaminu – sam decyduje, na podstawie swoich indywidualnych potrzeb, do których z trzech arkuszy zamierza przystąpić.

Pytania w teście adaptywnym (*adaptive test*) stają się łatwiejsze lub trudniejsze w zależności od wcześniejszej odpowiedzi zdającego. Tym sposobem każdy zdający otrzymuje unikalną, niepowtarzalną wersję testu z zadaniami dostosowanymi do swojego poziomu zaawansowania. Liczba zadań wytypowanych do rozwiązania dla konkretnego kandydata jest uzależniona od przebiegu i jakości wykonania kolejnych zadań. Taką spersonalizowaną część testu stanowi *Reading & Listening*.

To, co czyni *Linguaskill* jednym z najbardziej zaawansowanych certyfikujących egzaminów językowych, jest wykorzystanie do sprawdzania wypowiedzi pisemnej technologii opartej na sztucznej inteligencji (*artificial intelligence* – AI), dzięki której odpowiedzi zdającego są automatycznie sprawdzane i oceniane przez komputer (*Automarker*). W tej części egzaminu rolą zdającego jest napisanie odpowiedzi na email i zawarcie w niej wymaganych, podanych w instrukcji do zadania informacji. Drugim, dłuższym rodzajem tekstu, który kandydat ma stworzyć podczas egzaminu, może być np. recenzja, artykuł albo wpis na stronie internetowej. Również w tym zadaniu, oprócz jakości użytego języka, zdający jest rozliczany m.in. z przekazu określonych informacji podanych w formie danych wejściowych do zadania.

Egzamin ustny również przeprowadzany jest z użyciem komputera. Kandydat nagrywa swoje odpowiedzi na pytania odtwarzane za pomocą komputera lub zapisane na ekranie. W ramach niepunktowanej „rozgrzewki” odpowiada na kilka pytań dotyczących swojej osoby, po czym proszony jest o głośne odczytanie ośmiu zdań. W tym zadaniu oceniana jest wymowa i intonacja. Kolejne zadanie wymagające wypowiedzi o zwiększonej długości to ok. jednodominutowa wypowiedź na określony temat bazująca na materiałach graficznych. Na przygotowanie tej wypowiedzi, która może mieć formę np. nagrania wiadomości głosowej,

⁶⁹ Zob. https://www.youtube.com/results?search_query=linguaskill [dostęp: 10.07.2020].

przewiduje się ok. 40 sekund. Ostatnie zadanie sprawdza umiejętność wyrażania własnych opinii i oparte jest na pięciu pytaniach dotyczących jednego wspólnego tematu. Z każdym kolejnym zadaniem obserwujemy rosnący stopień trudności.

Poprzez zastosowanie w ocenie zadań zamkniętych sprawdzanych komputerowo, indywidualnego doboru zadań pod kątem umiejętności konkretnego zdającego oraz wspomnianej technologii *Automarkera*, egzamin dostępny jest na życzenie, a wyniki uzyskiwane są w ciągu 48 godzin. Dzięki nowoczesnej technologii nadzorującej do egzaminu można również przystępować poza ośrodkami egzaminacyjnymi, np. w firmach, lub nawet w domu kandydata.

Wyniki, podobnie jak w pozostałych egzaminach Cambridge English, przedstawiane są w formie szczegółowej punktacji w ramach poszczególnych części egzaminu (sprawności), do których kandydat przystępował, oraz wyniku średniego. Zarówno ogólny wynik, jak i wyniki szczegółowe przedstawione są na skali CEFR oraz *Cambridge English Scale*.

Porównanie egzaminów certyfikujących z języka ogólnego

Na podstawie analizy w/w certyfikatów informacje dotyczące omawianych testów z języka ogólnego zebrano w *Formularzu analizy egzaminów GE* (Załącznik 1) oraz w *Zestawieniu analizowanych egzaminów GE i ESP* (Załącznik 3). W porównywanych egzaminach GE daje się zauważyć wiele cech wspólnych, z których najważniejszą jest, że wszystkie opisywane tutaj egzaminy odnoszą biegłość zdającego do skal ESOKJ. Ponadto egzaminy z ogólnego języka angielskiego skonstruowane są w podobny sposób: dla wszystkich analizowanych przykładów punktem wyjścia dla zadań egzaminacyjnych są cztery główne sprawności językowe. Wszystkie analizowane egzaminy sprawdzają umiejętności w ramach czytania, pisania, słuchania i mówienia. Częstym elementem są również zadania badające znajomość leksyki ogólnej oraz struktur językowych. Trzy z sześciu egzaminów posiadają aktualnie odrębną część gramatyczno-leksykalną⁷⁰, a jeden z nich (*Linguaskill*) planuje wprowadzenie takiego elementu w przeszłości.

⁷⁰ Podobnie wygląda sytuacja dot. certyfikacji języka włoskiego, gdzie wszystkie cztery zatwierdzone przez włoskie Ministerstwo Spraw Zagranicznych certyfikaty na omawianym tu poziomie B2 zbudowane są z czterech części opartych na głównych sprawnościach językowych, a trzy spośród nich zawierają dodatkowo moduł sprawdzający kompetencje w zakresie gramatyki i słownictwa (Kaliska 2018: 401–406).

Żaden z analizowanych arkuszy nie integruje sprawności językowych w zadaniach egzaminacyjnych w sposób autentyczny. Jedyne zadanie napotkane w omawianych tutaj egzaminach, które łączy w sobie dwie sprawności, to dyktando w egzaminie *PTE General*. Nie jest to jednak zadanie spełniające w istotny sposób kryterium autentyczności, gdyż zapisywanie dłuższego tekstu ze słuchu jest raczej rzadką czynnością w codziennym życiu, chyba że mamy do czynienia z zawodem asystenta, sekretarki bądź stenotypisty.

Cztery z sześciu egzaminów wymagają od zdającego przystąpienia do wszystkich części egzaminu. Dwa najmłodsze testy – *Aptis* oraz *Linguaskill* – nie zakładają takiego obostrzenia. W przypadku tych egzaminów to kandydat decyduje, w ramach których sprawności chce udokumentować swoją biegłość językową.

O autentyczności zadań egzaminacyjnych można mówić w przypadku niektórych zadań w częściach sprawdzających sprawność pisanie. Zadania takie jak wypełnianie formularzy (*Aptis*), wpisy na czacie (*Aptis*) oraz pisanie korespondencji o różnym stopniu formalności (*Aptis, First, IELTS, PTE, Linguaskill, telc*), to zadania, z którymi zdający mają dużą szansę spotkać się w codziennym życiu – zarówno prywatnym, jak i zawodowym. Również niektóre elementy egzaminu ustnego wykazują się dość dużym stopniem autentyczności zadań egzaminacyjnych. Udział w dyskusji (*telc, PTE*), wspólne planowanie (*telc*) czy nagrywanie wiadomości głosowej (*Linguaskill*) są czynnościami często wykonywanymi w życiu codziennym. Wśród zadań na egzaminie ustnym występują także takie o stosunkowo niskim stopniu autentyczności, jak na przykład bardzo popularny w egzaminach językowych opis obrazka (*Aptis, First, PTE*).

Wiele zadań egzaminacyjnych w ramach testów sprawności czytania i słuchania oraz zadania gramatyczno-leksykalne skonstruowanych jest z jednostek zamkniętych, jednak tylko dwa z opisanych tu egzaminów ograniczają się wyłącznie do tego rodzaju zadań (*telc, Aptis*). Argumentem za zastosowaniem wyłącznie tej formy ewaluacji są niewątpliwie względy praktyczne.

Egzaminy z języka ogólnego winny sprostać ogólnym potrzebom językowym wszystkich użytkowników danego języka obcego, a potrzeby te, wobec niespecyficzności kandydatów przystępujących do tego rodzaju testów, są bliżej nieokreślone i niekoniecznie takie same dla wszystkich użytkowników języka ogólnego. W związku z tym nie do końca zaspaka-

jają potrzeby osób uczących się języka i przystępujących do danego egzaminu. Potwierdza to Komorowska (2019: 159), która stwierdza, że:

[...] egzaminy całościowe, szczególnie zaś egzaminy doniosłe, a także liczne egzaminy międzynarodowe nie uwzględniają potrzeb uczących się w odniesieniu do poszczególnych sprawności i kompetencji, a udział w nich wymaga poddania się przekrojowej ocenie w zakresie wszystkich podstawowych sprawności na wybranym poziomie. Owa wielosprawnościowość systemów egzaminacyjnych i związana z nią certyfikacja powoduje, iż nie sposób mówić w tej sytuacji o jakimkolwiek ukierunkowaniu na osobę uczącego się.

3. Certyfikacja języków specjalistycznych

W poniższym podrozdziale dokonane zostanie zestawienie kilku znacznie różniących się między sobą egzaminów z języków specjalistycznych. Ich analiza wykazała, że istnieją egzaminy ESP, w których badane sprawności i umiejętności oraz rodzaje zadań egzaminacyjnych nie różnią się znacząco od tych zawartych w egzaminach z języka ogólnego. Odróżnia je głównie poszerzenie sprawdzanego słownictwa o zakres tematyczny związany z danym zawodem czy branżą. Z drugiej strony kontinuum, o którym mówi Douglas (2000: 1), znajdują się egzaminy, w których nastawienie specjalistyczne, a w zasadzie specyficzny ich cel, są bardzo mocno widoczne.

Do powstania różnic w stopniu sprofilowania zawodowego pomiędzy poszczególnymi egzaminami specjalistycznymi mogły się przyczynić między innymi skale opisu biegłości, które stanowią bazę dla ustalenia konkretnych umiejętności oczekiwanych od zdających test na określonym poziomie. Część egzaminów opiera się na skalach ESOKJ, niektóre natomiast stosują się do skal opracowanych dla konkretnego zawodu lub egzaminu, co wykazano w pierwszej części niniejszego rozdziału.

Innym czynnikiem, który ma wpływ na kształt egzaminu, są osoby, które go tworzyły. Daje się zauważyć, w których przypadkach o formie egzaminu decydowali wyłącznie glottodydaktycy, a kiedy głos w tej sprawie zabierali również specjaliści z danej branży zawodowej.

Na podstawie wybranych egzaminów ESP wskazane zostaną różnice w podejściu do oceny umiejętności językowych i kompetencji w zakre-

sie języka specjalistycznego. Zaprezentowane zostaną również różne koncepcje oceniania umiejętności językowych studentów w uczelniach wyższych w Polsce oraz formuła egzaminu LSP opracowana w ramach międzynarodowego projektu GULT, kierowana do europejskich uczelni wyższych.

Międzynarodowe egzaminy certyfikujące z języków specjalistycznych

W obecnych czasach, gdy wykorzystanie znajomości języków obcych do celów specjalistycznych i zawodowych jest coraz bardziej istotne i pożądane, pojawiają się na rynku edukacyjnym kolejne egzaminy certyfikujące biegłość językową w zakresie różnych odmian języków do celów specjalistycznych. Na przełomie lat 2014 i 2015 przeprowadzone zostało badanie, którego celem było zebranie dostępnych w tym czasie w Polsce egzaminów certyfikujących z języka angielskiego oraz dokonanie analizy i porównania najbardziej charakterystycznych z nich. Wyniki przedstawiono w artykule *Egzaminy certyfikujące z języków obcych ogólnych i akademickich oraz języków obcych dla celów zawodowych – porównanie sprawdzanych umiejętności, sprawności i kryteriów oceny* (Sendur 2015). W momencie przeprowadzania badania na polskim rynku edukacyjnym dostępnych było – oprócz pięciu różnych testów międzynarodowych z ogólnego języka angielskiego – siedem egzaminów z języka angielskiego akademickiego oraz ok. 30–40 egzaminów specjalistycznych (Sendur 2015: 239–242). Liczba egzaminów z języków zawodowych jest o tyle trudna do precyzyjnego określenia, że zależy od podejścia do ich klasyfikacji. Wynika to z faktu, iż w przypadku niektórych testów mamy do czynienia z kilkoma wersjami dedykowanymi różnym zawodom w ramach jednej szerszej branży. Sytuacja taka dotyczy na przykład egzaminów z języka medycznego sTANDEM, który był dostępny w oddzielnych wersjach dla lekarzy, pielęgniarek i dla farmaceutów, oraz z języka technicznego MONDIALE, oferowanego w wariantach dla szeregu różnych branż technicznych.

Jak wspomniano wcześniej, stopień sprofilowania zawodowego dostępnych egzaminów z języka angielskiego dla celów zawodowych bywa różny: analizując ich formę, treści i rodzaje zadań testowych, można wnioskować, czy dany egzamin jest dziełem glottodydaktyków czuwających nad dostosowaniem arkuszy egzaminacyjnych do tradycji oceniania ogólnej biegłości językowej, czy też w tworzeniu testu brali aktywny udział fachowcy z branży, której dane narzędzie dotyczy.

W dalszej części tego rozdziału zaprezentowano przykłady egzaminów mieszczących się w rozmaitych punktach osi specjalistyczności. Analizie poddano szereg testów skierowanych do różnych profesji. Ich dobór jest całkowicie autorski: przeanalizowano egzaminy, które wydawały się godne zauważenia w kontekście badanego tematu. Starano się wybrać przykłady reprezentatywne dla różnych dziedzin zawodowych. Wzięto pod uwagę także takie kryteria, jak szczególne cechy danego egzaminu oraz jego popularność.

Niektóre z testów kierowane są do całej branży, a nie do konkretnego zawodu (np. branża turystyczna, wojsko, szeroko rozumiany biznes); w innych przypadkach w ramach jednego egzaminu mamy do czynienia z wersjami dla różnych specjalizacji bądź różnych zawodów (np. lekarze, pielęgniarki i farmaceuci; piloci i kontrolerzy ruchu lotniczego). Rozróżnienie pomiędzy tymi dwoma rodzajami testów LSP opisuje ALTE (2018: 3), dzieląc je na kategorie o odpowiednio: szerokim i wąskim zakresie (*broad LSP tests / narrow LSP tests*).

W celu porównywalności wymagań wzięto pod uwagę egzaminy na zbliżonym poziomie zaawansowania; zdecydowano się na wybór najpopularniejszego poziomu certyfikacji, czyli B1–B2 (ew. B2+) wg ESOKJ. Są też wśród badanych przykładów testy wielopoziomowe (MOTET, *Linguaskill*). Niektóre z certyfikatów omówionych w dalszej części rozdziału zostały w międzyczasie wycofane z rynku edukacyjnego, zostały jednak uwzględnione w analizie ze względu na swoje specyficzne cechy.

Opisane egzaminy to: dwa różne testy z technicznego języka angielskiego (MOTET oraz *telc English B2 Technical*), egzamin dla wojska (STANAG 6001), z języka angielskiego dla księgowych (EfA), dla branży turystycznej (EfT), test mierzący umiejętności w ramach języka biznesu (*Linguaskill Business*), jak również egzamin dla pilotów i kontrolerów ruchu lotniczego (ELPAC). Branża medyczna reprezentowana jest przez międzynarodowy certyfikat stworzony w latach 80. XX wieku na zlecenie rządu Australii (OET) oraz przez egzamin opracowany przez międzynarodowy zespół pod auspicjami i z finansowaniem Komisji Europejskiej (STANDEM). Uwzględniono także serię egzaminów skierowanych do wielu różnych zawodów pod szyldem *Vocational English Certificate* (VEC).

W celu oceny poziomu stopnia specjalistyczności zastosowano opracowany do innego badania egzaminów LSP (Sendur 2020) i zmodyfikowany do celów niniejszej publikacji *Formularz analizy egzaminów LSP*

(Załącznik 2). Arkusz ten pozwala na dokonanie oceny każdego egzaminu w ramach czterech kryteriów:

- dobór części egzaminu – w skali 1–5 oceniony jest tutaj stopień uzasadnienia poszczególnych elementów arkusza egzaminacyjnego względem potrzeb zawodowych odbiorców danego testu;
- stopień autentyczności zadań egzaminacyjnych oceniany jest w skali 1–5, gdzie wartość 1 przyporządkowana jest typowym zadaniom testowym, a ocena 5 – całkowicie autentycznym zadaniom zawodowym;
- wykorzystanie wiedzy specjalistycznej w celu wykonania zadania egzaminacyjnego oceniane jest w skali 0–4, gdzie 0 oznacza całkowity brak wykorzystania takiej wiedzy, a 4 to stopień, w którym wiedza fachowa wydaje się być niezbędna w celu wykonania zadania;
- kryteria oceny wykonania zadania oceniane są niżej tam, gdzie położony został większy nacisk na jakość języka, wyżej zaś w tych przypadkach, gdzie poziom wykonania zadania zawodnego odgrywa większą rolę.

Formularz uwzględnia również opis schematu egzaminu z uwzględnieniem zadań sprawdzających w sposób bezpośredni znajomość struktur językowych oraz leksyki ogólnej i specjalistycznej, a także zadań integrujących sprawności językowe, które wydają się istotne w egzaminach LSP.

Za pomocą w/w narzędzia, przeprowadzono analizę i ocenę egzaminów pod kątem stopnia ich specyficzności i dokonano podziału na trzy kategorie: egzaminy o niższym, średnim i wysokim stopniu sprofilowania zawodowego. Należy jednak podkreślić, że mimo próby zobiektywizowania oceny egzaminów i oparcia jej na określonych kryteriach, pozostaje ona do pewnego stopnia subiektywną oceną Autorki.

Egzaminy ESP o niższym stopniu sprofilowania zawodowego

Wśród analizowanych egzaminów specjalistycznych testami o najniższym stopniu sprofilowania zawodowego, a więc testami zbliżonymi do arkuszy używanych w ocenie języków ogólnych, wydają się egzaminy z serii *Vocational English Certificate* (VEC) skierowanych do całej gamy różnych zawodów, test z zakresu księgowości – *English for Accounting* (Efa) oraz z technicznego języka angielskiego – MOTET.

• *Vocational English Certificate* (VEC)

W międzynarodowej ofercie *International Vocational English* (INTERVENG)⁷¹ znajdują się obecnie egzaminy dla 65 różnych branż i zawodów⁷², które pozwalają uzyskać międzynarodowy certyfikat językowy *Vocational English Certificate* (VEC). Dokument poświadczający biegłość językową można zdobyć w tak odmiennych zawodach jak strażak (*Firefighter*), kierowca taksówki (*Taxi Driver*), nauczyciel edukacji przedszkolnej (*Kindergarten Teacher*) czy pracownik telefonicznego centrum obsługi klienta (*Call Centre*) oraz w szerszych branżach zawodowych, jak rolnictwo (*Agriculture*), usługi hotelowe i żywieniowe (*Hotels & Catering*), rybołówstwo i przemysł rybny (*Fishing & Seafood Industry*) oraz bardzo ogólnie sport (*Sports*). Co ciekawe, w opisie egzaminu, jak również na samym certyfikacie, nie ma informacji dotyczącej poziomu języka. Z informacji ustnej uzyskanej od osoby zajmującej się egzaminem, jak również z własnej obserwacji i oceny wyniku, że poziom egzaminu należy określić jako B1, natomiast formalnie nie odnosi się on do żadnej skali.

Egzamin dla wszystkich wymienionych tutaj branż, ale także pozostałych w ofercie, wygląda identycznie i jest zbudowany z takich samych elementów niezależnie od grupy docelowej. Każdy arkusz skonstruowany jest w ten sam sposób i składa się z części *Listening*, *Knowledge of Linguistic Means*, czyli z testu sprawdzającego znajomość leksyki specjalistycznej, *Knowledge of Language Functions*, *Reading* oraz *Writing Awareness*. Wszystkie elementy egzaminu składają się z zadań zamkniętych (WW oraz prawda/fałsz/brak informacji). Również część egzaminu, której celem jest sprawdzenie umiejętności w zakresie sprawności pisania (*Writing Awareness*), zawiera wyłącznie zadania zamknięte, w których zdający muszą dopasować wyjęte z tekstu zdania lub fragmenty zdań do luk powstałych po ich usunięciu. Zadanie to sprawdza zatem umiejętności w ramach czytania tekstu, a nie jego tworzenia.

Testy INTERVENG stanowią jedyną w swoim rodzaju, bardzo bogatą ofertę egzaminów ze specjalistycznych wariantów języka angielskiego. Nie są one jednak oparte o analizę potrzeb poszczególnych zawodów oraz branż zawodowych, do których są kierowane. Skonstruowane są z identycznych elementów, niezależnie od odmiennych potrzeb poszczegól-

⁷¹ Zob. <https://www.interveng.com/i-want-to-find-exam-information> [dostęp: 14.10.2020].

⁷² Pełna lista dostępnych egzaminów: <https://www.interveng.com/i-want-to-find-a-course> [dostęp: 14.10.2020].

nych profesji, a różni je jedynie tematyka zadań egzaminacyjnych oraz sprawdzana podczas egzaminu leksyka związana z odnośną branżą lub zawodem. Takie podejście do struktury testów egzaminacyjnych, jak również wykorzystanie w nich wyłącznie zadań zamkniętych, świadczy o wysokim stopniu praktyczności testu przy jednoczesnym niskim stopniu trafności i rzetelności.

• **Egzamin z języka angielskiego z zakresu księgowości – *Award in English for Accounting***

Kolejnym sprawdzianem ESP dość bliskim testom GE jest LCCI IQ⁷³ – *Język angielski dla księgowych – English for Accounting* (EfA). Wycofany z oferty Pearson English Language Assessment pod koniec 2017 roku, jest jednak uwzględniony w niniejszym zestawieniu ze względu na interesujące podejście do testowania języka specjalistycznego.

Egzamin odbiega od testów GE przede wszystkim dlatego, że nie jest skonstruowany wokół poszczególnych sprawności językowych. Sprawdza głównie znajomość słownictwa specjalistycznego, umiejętność rozpoznawania poprawnych struktur gramatycznych, a jeśli chodzi o sprawności językowe, to w niewielkim zakresie bada rozumienie czytania, bazując na bardzo krótkich tekstach składających się z 1–3 rozbudowanych zdań. Zarówno teksty, jak i pojedyncze zdania w części leksykalnej i gramatycznej są tematycznie związane z księgowością.

W przypadku tego egzaminu trudno jednak mówić o zadaniach zawodowych. W teście nie ma żadnego zadania sprawdzającego umiejętność wykonywania zadań zbliżonych do zawodowych. Sama forma zadań nie spełnia kryterium autentyczności, choć niewątpliwie zaspokojone zostało kryterium praktyczności, gdyż arkusz na poziomie B2 składa się całkowicie z zadań zamkniętych wyboru wielokrotnego. Wątpliwości budzą także zadania sprawdzające umiejętności rozpoznawania błędnych form gramatycznych (przykład 11) oraz błędnego zapisu ortograficznego (przykład 12). Należałoby się zatem zastanowić, czy sama znajomość słownictwa specjalistycznego i sprawność w zakresie czytania wyrwanych z szerszego kontekstu zdań związanych tematycznie z rachunkowością, jak również umiejętność identyfikowania błędów językowych są wy-

⁷³ *The London Chamber of Commerce and Industry – International Qualifications*. Egzaminy i certyfikaty z języka angielskiego branżowego przeprowadzane i wydawane obecnie przez firmę Pearson, a firmowane przez Londyńską Izbę Przemysłowo-Handlową.

starczające, by sprostać zadaniom zawodowym w języku obcym stojącym przed posiadaczem uzyskanego certyfikatu.

Przykład 11. Zadanie sprawdzające umiejętność rozpoznania błędnych form gramatycznych z próbnego testu *Award in English for Accounting Level 3*⁷⁴

Each of the following sentences contains FOUR underlined words. Identify which ONE of these words is incorrectly formed.

11. The external (A) auditor may make observings (B) with respect to the efficiency (C) of the financial procedures (D).

Przykład 12. Zadanie sprawdzające umiejętność rozpoznania błędnie zapisanych wyrazów z próbnego testu *Award in English for Accounting Level 3*⁷⁵

How many words are incorrectly spelled in the following texts.

23. The accounts of Schneeman PLC for the last financail year continue to show an upward trend. The year end balance sheet shows that the reorganisation undertaken by the new management has been effective. Although the share price has remained stable the board has decided to increase the dividend by €1 to €11 per share. While market conditions are unpredictable the company remains a safe investmant, although investors should keep a close eye on cache flow for the next quarter.

- A. 4
- B. 5
- C. 6
- D. 7

• Egzamin z technicznego języka angielskiego MONDIALE (MOTET)

Egzaminy ESP, bliskie w swoim charakterze testom z języka ogólnego, to również wachlarz egzaminów MONDIALE⁷⁶, które umożliwiają kandydatom potwierdzanie znajomości technicznego języka angielskiego.

⁷⁴ Źródło: <https://qualifications.pearson.com/content/dam/pdf/LCCI/English-for-Accounting/2010/Specification-and-sample-assessments/Level-3-Award-in-English-for-Accounting-Sample-Paper.pdf> [dostęp: 2.02.2020].

⁷⁵ Źródło: <https://qualifications.pearson.com/content/dam/pdf/LCCI/English-for-Accounting/2010/Specification-and-sample-assessments/Level-3-Award-in-English-for-Accounting-Sample-Paper.pdf> [dostęp 2.02.2020].

⁷⁶ Opisane bardziej szczegółowo w: Sendur 2015.

*MONDIALE Technical English Test (MOTET)*⁷⁷ jest testem wielopoziomowym, sprawdzającym umiejętności językowe na poziomach od B1 do C1. Wyniki przedstawione są w formie ogólnego poziomu biegłości osiągniętego przez kandydata, jak również w formie profilu językowego opisującego umiejętności zdającego w ramach poszczególnych sprawności w odniesieniu do poziomów biegłości ESOKJ. Istotne z punktu widzenia porównania różnych egzaminów LSP jest tutaj właśnie odniesienie do poziomów Rady Europy. Zgodnie z informacją podaną w informatorze dotyczącym egzaminów (*Mondiale Testing* b.r.: 4), „MONDIALE stosuje się do podejścia, treści i opisów poziomów *Europejskiego systemu opisu kształcenia językowego*”⁷⁸.

Podobnie jak większość egzaminów GE, MOTET składa się z części pisemnej i ustnej, mierząc umiejętności w zakresie czterech sprawności oraz znajomość gramatyki i leksyki z szeroko rozumianego języka technicznego. Oddzielna część sprawdzająca kompetencje gramatyczne obecna w egzaminach MOTET jest bardziej charakterystyczna dla egzaminów GE i często w innych egzaminach ESP całkowicie nieobecna (*cf.* STANAG 6001, EfT, OET, sTANDEM, ELPAC). Co natomiast odróżnia opisywany egzamin od egzaminów z języka ogólnego, to zawartość treści o tematyce technicznej oraz fakt, iż kandydat nie musi zdawać testów z wszystkich sprawności językowych, ale może podejść tylko do wybranych przez siebie modułów. Warto jednak zwrócić uwagę, iż część gramatyczno-leksykalna jest elementem obowiązkowym dla wszystkich kandydatów.

Zadania gramatyczno-leksykalne oparte są na zdaniach o tematyce technicznej. Celem zachowania pozorów autentyczności każde pojedyncze zadanie zawiera krótki opis sytuacji, w której podane zdanie miałyby się pojawić. Chociaż opisana sytuacja jest umiejscowiona w kontekście zawodowym, to jednak zapoznanie się z tą informacją wejściową nie jest wcale konieczne do poprawnego wykonania zadania (*cf.* przykład 13). Również w części ustnej trudno mówić o podejściu zadaniowym – rolą zdającego jest tu prezentacja tematu i dyskusja z egzaminatorem, a nie wykonanie zadania w symulowanej sytuacji zawodowej.

⁷⁷ *Mondiale Testing* (online), <http://www.mondiale-testing.de/site/content/view/90/72/lang,en/> [dostęp: 23.02.2015].

⁷⁸ Tłumaczenie własne – A.S.

Przykład 13. Zadanie gramatyczne z próbnego testu MOTET za: *MONDIALE Technical English Mock Test*

<p>Situation</p> <p>You have to repair a refrigeration unit which might take several hours. You call your partner to inform her that she may need to take over your next appointment.</p>
<p>Task</p> <p>Click on the correct answer.</p>
<p>Question 19</p> <p>I might finish on time but ... I can't make the next appointment, I will call you.</p> <p><input type="radio"/> whether</p> <p><input type="radio"/> if</p> <p><input type="radio"/> when</p>

Podsumowując, można stwierdzić, że mało specjalistyczny charakter egzaminów MOTET daje się zauważyć zarówno w braku zadań charakterystycznych dla sytuacji zawodowych, jak i w mnogości zadań sprawdzających poprawność językową.

b) Egzamin o średnim stopniu sprofilowania zawodowego

W zestawieniu egzaminów pośrodku skali specjalistyczności uplasowały się: egzamin z języka angielskiego wg STANAG 6001, europejski test dla branży medycznej STANDEM oraz egzamin z technicznego języka angielskiego – *telc English B2 Technical*.

• **Egzamin z języka angielskiego wg STANAG 6001**

Egzaminy językowe wg STANAG 6001 przeznaczone są przede wszystkim dla żołnierzy zawodowych i pracowników wojska. Porozumienie standaryzacyjne STANAG 6001 nie określa formy egzaminu sprawdzającego poziom znajomości danego języka, opisuje jedynie efekt końcowy, a więc poziom biegłości językowej, jaki powinna posiadać osoba legitymująca się danym certyfikatem. Sam egzamin oparty jest na dokumentach wewnętrznych państwa, w którym jest przeprowadzany, zatem w każdym z krajów członkowskich wygląda nieco inaczej. W Polsce

wojskowe egzaminy resortowe z języka angielskiego oparte są na *Ramowym programie nauczania języka angielskiego w Siłach Zbrojnych RP* (Ministerstwo Obrony Narodowej. Departament Nauki i Szkolnictwa Wojskowego 2009).

Zgodnie z opisem egzaminu dostępnym w *Modelu egzaminu z języka angielskiego poziom 3* (Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej 2013: 6), sprawdzane są umiejętności wykorzystania języka obcego w sytuacjach zawodowych i towarzyskich w środowisku wojskowym (np. w sztabie o składzie wielonarodowym, podczas operacji wsparcia pokoju, wielonarodowych ćwiczeń wojskowych lub podczas szkoleń zagranicznych) w ramach czterech sprawności językowych: czytania, słuchania, pisania i mówienia.

Jak zauważono wcześniej, skala stopni znajomości języków STANAG 6001 nie odnosi się bezpośrednio do zagadnień militarnych i tylko w minimalnym stopniu odwołuje się do umiejętności związanych ze specjalistycznym użyciem języka obcego, dając priorytet umiejętnościom ogólnym. Egzamin z języka angielskiego na interesującym nas poziomie (poziom 3: B2/B2+⁷⁹) mierzy jednak zarówno ogólne umiejętności językowe, jak i kompetencje w zakresie języka specjalistycznego.

Elementy sprawdzające sprawność czytania składają się z zadań wyboru wielokrotnego (WW). Rodzaje zadań egzaminacyjnych w części badającej rozumienie ze słuchu to WW oraz sporządzanie notatek. Analiza *Modelu* (Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej 2013) oraz przykładowych egzaminów⁸⁰ pokazuje, że zakres tematyczny w obydwu tych częściach egzaminu dotyczy zarówno zagadnień wojskowych, jak i specyficznie rozumianych zagadnień ogólnych, takich jak polityka, ekonomia, nauka, technika, kultura itp.

W części badającej umiejętność pisania rolą zdającego jest stworzenie korespondencji formalnej. Wymagane rodzaje tekstu to list, notatka służbowa oraz sprawozdanie. Zgodnie z opisem tej części egzaminu

⁷⁹ Egzaminy wg STANAG 6001 nie odnoszą swoich poziomów do skali ESOKJ, tylko do swojej własnej *Tabeli stopni znajomości języków (Porozumienie Standaryzacyjne NATO nr 6001)*; podane odniesienia do poziomów ESOKJ oparte są m.in. na porównaniu poziomów CEFR i STANAG 6001 opracowanych przez wydawnictwo Macmillan, <http://www.campaignmilitaryenglish.com/Course/teacher.htm> [dostęp: 19.04.2020].

⁸⁰ Arkusz L3 054 i L3 057 (Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej b.r.-a i b).

w *Modelu egzaminu z języka angielskiego* (Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej 2013: 12–13) od zdających oczekuje się umiejętności tworzenia wypowiedzi pisemnej na tematy wojskowe i ogólne. Analizując przykładowe arkusze egzaminacyjne, obserwujemy, że chociaż niektóre z opisanych sytuacji należą do kręgu tematyki ogólnej (np. reklamacja konsumenta), to jednak szczegóły z nią związane wchodzą już w tematykę zawodową (reklamacja dotyczy zakupu koszulek dla batalionu wojska). Można zatem stwierdzić, że zadania egzaminacyjne są bardzo bliskie autentycznym zadaniom zawodowym w tej branży.

Jednym z zadań w części egzaminu badającej sprawność mówienia jest wojskowy *briefing* – zdający wygłasza kilkuminutowy monolog z wykorzystaniem podanych informacji o tematyce wojskowej, po którym następuje rozmowa pomiędzy zdającymi na temat przedstawionych danych. Druga część egzaminu dotyczy sfer aktywności innych niż zawodowe. W tej części zdający prowadzą rozmowę na temat podany w bilecie egzaminacyjnym, a tematyka dotyczy takich zagadnień, jak zdrowie, zakupy, życie w mieście, czas wolny, sport itp.

Egzamin STANAG 6001 skierowany jest do ściśle określonej grupy docelowej, o konkretnych potrzebach w zakresie komunikacji w języku obcym, jednak ujęte w nim zadania egzaminacyjne oraz sprawdzane za jego pomocą umiejętności nie ograniczają się do zawodowych kompetencji językowych i stosunkowo duża część testu sprawdza umiejętności w zakresie języka ogólnego.

• **Egzamin z medycznego języka specjalistycznego dla lekarzy, pielęgniarek i farmaceutów – sTANDEM⁸¹**

*Standardised Language Certificate for Medical Purposes (sTANDEM)*⁸² to egzamin dla branży medycznej. Powstał w wyniku współpracy przedstawicieli europejskich uczelni medycznych jako odpowiedź na brak odpowiedniego narzędzia do mierzenia biegłości w zakresie języka medycznego na rynku europejskim.

Egzamin, w trzech różnych wersjach, przewidziany jest dla lekarzy, pielęgniarek i farmaceutów. Arkusz egzaminacyjny składa się z testu pisemnego sprawdzającego umiejętności w zakresie sprawności czytania

⁸¹ Egzamin pozostaje aktualnie zawieszony (cf. *Zakończenie*).

⁸² Opisany bardziej szczegółowo w: Sendur 2015, The sTANDEM Consortium 2013 oraz www.standem.eu [dostęp: 27.02.2015].

i pisania oraz z części badającej sprawność słuchania i mówienia. Zdający nie musi jednak przystępować do obydwu części egzaminu: może dokonać wyboru jednej z nich i otrzymać certyfikat dotyczący tylko wybranych sprawności lub też zdać obydwa egzaminy i uzyskać dwa niezależne dokumenty poświadczające biegłość językową w ramach różnych sprawności. Takie podejście umożliwia certyfikację w przypadku osób, u których pewne sprawności językowe są bardziej rozwinięte lub bardziej pożądane dla potrzeb wykonywanego zawodu (Charpy i Carnet 2014: 4). Za pomocą egzaminu sprawdzane są trzy poziomy biegłości językowej: B1, B2 i C1. Egzamin sTANDEM odnosi swoje testy i certyfikaty do poziomów biegłości Rady Europy (2003), opierając się na wyselekcjonowanych do tego celu deskryptorach ESOKJ zebranych w dokumencie *Descriptors Recommended for sTANDEM*⁸³.

Zarówno test sprawdzający umiejętność czytania, jak i arkusz badający sprawność rozumienia ze słuchu oparte są na tekstach o tematyce medycznej. Test sprawności pisania ma na celu sprawdzenie umiejętności konstruowania listów formalnych oraz pisania analiz badań naukowych. Rolą zdającego jest napisanie listu według podanych wskazówek, odnoszącego się do zaprezentowanych w poleceniu aspektów, oraz przeanalizowanie i podsumowanie danych przedstawionych w formie graficznej (np. wykresu). Sprawdzane są tutaj zatem – obok znajomości języka medycznego – umiejętności akademickie, co plasuje omawiany egzamin w kategorii języka medycznego akademickiego (*English for Academic Medical Purposes*).

Egzamin ustny bada, w jakim stopniu zdający potrafi komunikować się w języku obcym w środowisku akademickim i zawodowym. Pośród zadań w tej części znajduje się rozmowa na tematy dotyczące pracy lub badań naukowych zdającego oraz wypowiedź na temat informacji zawartych w materiale stymulującym (w postaci np. wykresu o tematyce medycznej). W arkuszu znajduje się również zadanie sprawdzające podwójną kompetencję. Dane wejściowe do tego zadania to szczegółowy scenariusz, na podstawie którego ma zostać odegrana scenka, w której zdającemu i egzaminatorowi przydzielono odpowiednio zaprojektowane role. W bilecie egzaminacyjnym sprecyzowane jest miejsce rozmowy, opisana jest sytuacja i problem, z jakim zgłasza się hipotetyczny pacjent. Zamieszczony jest tam także krótki tekst (np. ulotka informacyjna dot.

⁸³ Zob. http://www.standem.eu/wp-content/uploads/2012/07/sTANDEMdescriptors_03.08.2012.pdf [dostęp: 16.11.2015].

leku, broszura informacyjna dot. choroby, epikryza), z którego informacje powinny być wykorzystane w rozmowie. Poprawne wykonanie zadania wymaga zapoznania się z tekstem, zrozumienia informacji w nim zawartych i wykazania się zarówno umiejętnościami językowymi, jak i kompetencjami związanymi z reprezentowanym zawodem.

W konstrukcji egzaminu sTANDEM daje się zauważyć znacząca współpraca glottodydaktyków zajmujących się nauczaniem języka specjalistycznego ze specjalistami z zakresu medycyny. Potwierdza to zarówno forma zadań (nastawionych na autentyczne działanie zawodowe), jak również bezpośrednio sami autorzy opisujący pracę nad egzaminami (Charpy i Carnet 2014; Read i Knoch 2009). Efektem tej synergii jest egzamin językowy wpisujący się dość mocno w specyficzne potrzeby zawodów, dla których jest przeznaczony.

• **Egzamin z technicznego języka angielskiego – *telc English B2 Technical***

Egzaminem o dość wysokim stopniu sprofilowania zawodowego, choć – w porównaniu z innymi – zakwalifikowanym w niniejszym zestawieniu do egzaminów mieszczących się raczej pośrodku skali specjalistyczności, jest *telc English B2 Technical* (Telc Language Tests 2016). Certyfikat kierowany jest do fachowców w takich dziedzinach jak chemia, elektronika, energetyka i środowisko, technologie informacyjno-komunikacyjne oraz branża mechaniczna. Część pisemna składa się z testu sprawdzającego rozumienie tekstów czytanych, znajomości gramatyki i leksyki, rozumienia ze słuchu oraz z arkusza sprawdzającego umiejętność tworzenia wypowiedzi pisemnej w postaci tekstu użytkowego, jak na przykład faks. Wszystkie teksty, na których oparte są zadania egzaminacyjne, są tekstami o tematyce technicznej. Do wykonania zadań egzaminacyjnych wiedza specjalistyczna absolutnie nie jest niezbędna, jednak osoby bez podstawowej znajomości zagadnień technicznych, mimo odpowiedniego poziomu biegłości w zakresie języka ogólnego, mogą mieć problem z ich wykonaniem – będą prawdopodobnie wymagać dłuższego czasu na wykonanie zadań egzaminacyjnych niż zdający, którzy posiadają odpowiednią wiedzę fachową.

Arkusz egzaminacyjny zawiera test sprawdzający umiejętności w zakresie struktur i leksyki, co, jak wskazano wcześniej, jest cechą wyróżniającą egzaminy w zakresie języka ogólnego. Jest on jednak skonstruowany wyłącznie z zadań zamkniętych (WW), zatem wymaga rozpoznania,

a nie stworzenia prawidłowej odpowiedzi, oparty jest na tekstach specjalistycznych i stanowi zaledwie 10% całego arkusza pisemnego.

Sprawdzian ustny składa się z zadań badających umiejętności w zakresie dyskursu towarzyskiego oraz języka ogólnego, reakcji językowych w sytuacjach biznesowych i zawodowych (technicznych) oraz prezentacji materiału (przedmiotu, rysunku, procesu itp.) z wybranej przez zdającego dziedziny technicznej.

Telc English B2 Technical jest mocno osadzony w tematyce specjalistycznej, wymaga od kandydatów orientacji w tematyce ogólnotechnicznej, a zadania egzaminacyjne, sprawdzające umiejętność tworzenia tekstu pisanego oraz mówienia, charakteryzują się wysoką autentycznością.

• *Linguaskill Business*

Linguaskill (Cambridge Assessment English 2017; Cambridge Assessment English 2019c; Ludlow 2020) jest nowym egzaminem Cambridge English Assessment dostępnym w dwóch wersjach: *General* oraz *Business*. W tym miejscu interesuje nas głównie wersja specjalistyczna egzaminu, czyli wariant biznesowy. Egzamin ten jest najmłodszym i jednocześnie najnowocześniejszym testem omawianym w niniejszym opracowaniu. *Linguaskill* to test wielopoziomowy, sprawdzający biegłość kandydatów na poziomach od A1 do C. Występuje wyłącznie w formie komputerowej, jest testem adaptacyjnym i jako jedyny spośród omawianych tutaj przykładów sprawdza i ocenia umiejętności w ramach sprawności produktywnych (obecnie: *Writing* i częściowo *Speaking*) z wykorzystaniem sztucznej inteligencji (*artificial intelligence* – AI). Wyniki są dostępne w ciągu 48 godzin.

Egzamin w formie, jaka obowiązuje w momencie pisania tej monografii (2020), składa się z trzech części opartych na głównych sprawnościach językowych: *Reading & Listening*, *Writing* i *Speaking*. Jak dowiadujemy się z informacji na stronie instytucji opracowującej egzamin, w planach jest jeszcze arkusz leksykalno-gramatyczny: *Use of English*, nie ma jednak precyzyjnej informacji, czy element ten ma być obecny wyłącznie w wersji GE, czy również w egzaminie biznesowym oraz czy będzie modulem obowiązkowym, czy opcjonalnym⁸⁴. Na ten moment kandydat sam decyduje, do których arkuszy egzaminu zamierza przystąpić, w zależności od swoich potrzeb lub wymagań swojego pracodawcy.

⁸⁴ Zob. https://www.youtube.com/results?search_query=linguaskill [dostęp: 10.07.2020].

Część *Reading & Listening* jest testem adaptacyjnym. Pytania w tego rodzaju teście stają się łatwiejsze lub trudniejsze w zależności od wcześniejszej odpowiedzi zdającego. W rezultacie tego spersonalizowanego podejścia każdy kandydat otrzymuje unikalną wersję testu z zadaniami dostosowanymi do swojego poziomu zaawansowania.

Moduł *Writing* wykorzystuje innowacyjną technologię opartą na sztucznej inteligencji. Zdający zapisują swój tekst za pomocą klawiatury komputerowej, a ich odpowiedzi są automatycznie sprawdzane i oceniane przez komputer za pomocą technologii *Automarkera*. Zgodnie z informacją zamieszczoną na stronach Cambridge English Assessment trwają bardzo zaawansowane prace nad tym, by również egzamin ustny był całkowicie przeprowadzany i oceniany z wykorzystaniem AI.

Linguaskill Business zastąpił wygaszony z końcem roku 2019 egzamin BULATS. Tak jak i jego poprzednik, jest on skierowany do osób pracujących w środowisku biznesowym i korporacyjnym. Jak wspomniano wcześniej, wersja biznesowa nie różni się w swojej strukturze ani też rodzajem zadań testowych od wersji testującej znajomość języka ogólnego. Odróżnia ją jedynie tematyka tekstów, na których oparte są zadania, obejmująca takie zagadnienia, jak kupno i sprzedaż produktów lub usług, podróże służbowe, zasoby ludzkie i biuro.

Moduł *Reading & Listening* składa się głównie z zadań zamkniętych WW oraz z jednego zadania z lukami. *Writing* w obydwu wersjach – ogólnej i biznesowej – wymaga od zdających stworzenia dwóch tekstów użytkowych – krótkiego maila oraz drugiego, nieco dłuższego tekstu (180 słów). W przypadku egzaminu specjalistycznego obydwa zadania egzaminacyjne są jednocześnie dość autentycznymi zadaniami zawodowymi polegającymi na konieczności stworzenia tekstu dotyczącego określonej sytuacji zawodowej i skierowanego do konkretnego odbiorcy. Również w egzaminie ustnym pytania egzaminacyjne odnoszą się do ogólnych oraz szczegółowych sytuacji zawodowych w środowisku biznesowym. W tej części kandydat nagrywa swoje odpowiedzi na wyświetlone na ekranie komputera lub odtwarzane w formie dźwiękowej pytania. Zadania w części ustnej przyjmują formę pytań i odpowiedzi, wypowiedzi ustnej opartej na przedstawionym na ekranie materiale graficznym (np. podsumowanie danych na podstawie wykresu) oraz testu poprawności wymowy i intonacji, polegającym na odczytaniu i nagraniu przez kandydata kilku zdań pojawiających się kolejno na ekranie.

Mimo iż znaczna część egzaminu zbudowana jest z zadań wyboru wielokrotnego, co powoduje, że elementy te nie charakteryzują się wysokim stopniem autentyczności, to jednak występują w nim zadania egzaminacyjne bliskie autentycznym zadaniom zawodowym. W zestawieniu z pozostałymi analizowanymi tutaj egzaminami LSP test ten został umiejscowiony stosunkowo wysoko na osi specjalistyczności.

Wyniki, tak jak to ma miejsce w innych egzaminach Cambridge English, przedstawiane są w formie szczegółowej punktacji uzyskanej za poszczególne sprawności oraz wyniku średniego. Wynik ogólny z całego egzaminu oraz wyniki szczegółowe przedstawione zostały na skali CEFR oraz *Cambridge English Scale*.

c) Egzamin o wysokim stopniu sprofilowania zawodowego

Jako przykłady testów o wysokim stopniu specjalistyczności wybrano trzy egzaminy. Egzamin z języka medycznego OET i egzamin z języka angielskiego dla branży turystycznej EfT cechują się dużą autentycznością zadań oraz znaczącym stopniem interakcji pomiędzy znajomością języka i wiedzą specjalistyczną. Egzamin dla branży lotniczej ELPAC jest natomiast przykładem testu o zdecydowanie najwyższym stopniu sprofilowania zawodowego.

• Egzamin z języka angielskiego dla branży medycznej – *Occupational English Test*

Occupational English Test (OET) jest egzaminem z języka angielskiego dla branży medycznej stworzonym w latach 80. XX wieku z myślą o rynku Australii, Nowej Zelandii i Singapuru, jest jednak również wykorzystywany jako potwierdzenie kompetencji językowych osób z branży medycznej w Wielkiej Brytanii, a od 2020 roku również w Stanach Zjednoczonych. Służy jako test przesiewowy dla osób ubiegających się o uznanie kwalifikacji medycznych uzyskanych w krajach nieanglojęzycznych, sprawdzając kompetencje językowe specjalistów z dziedziny stomatologii, medycyny, dietetyki, pielęgniarstwa, farmacji, fizjoterapii, radiologii, logopedii, terapii zajęciowej, optometrii, podiatrii i weterynarii.

OET, podobnie jak egzaminy GE, oparty jest na czterech sprawnościach. Sprawdziany umiejętności czytania i słuchania zbudowane są z tekstów o tematyce medycznej i pozostają takie same dla wszystkich specjalności zawodowych. Testy sprawności mówienia i pisania dostosowane są do potrzeb językowych konkretnej specjalizacji w branży medycznej.

Arkusze sprawdzające umiejętności rozumienia ze słuchu oraz czytania składają się z zadań otwartych i zamkniętych, w tym WW, na dobieranie i krótkiej odpowiedzi (np. uzupełnianie zdań, tekstu lub danych w tabelkach), sprawności te badane są więc w sposób tradycyjny, podobnie jak robi to większość egzaminów GE. Zadania – oparte na tekstach o tematyce medycznej – są takie same dla wszystkich specjalności zawodowych objętych egzaminem.

Testy badające sprawność mówienia i pisania dostosowane są do potrzeb językowych konkretnej specjalizacji w branży medycznej. W części sprawdzającej umiejętność pisania zdający pisze list służbowy na podstawie materiału stymulującego w postaci notatek na temat pacjenta, którego list ma dotyczyć. Z reguły jest to list-skierowanie (*letter of referral*), pisany zgodnie z normami i zasadami stosowanymi w krajach, których egzamin dotyczy. Zadanie to należy zatem uznać za bliskie zadaniom zawodowym w służbie zdrowia w krajach anglojęzycznych.

Część ustna składa się z zadań na odgrywanie ról. Scenariusze oparte są na typowych interakcjach pomiędzy pracownikiem służby zdrowia i pacjentem, gdzie zdający odgrywa rolę adekwatną do wykonywanego przez siebie zawodu.

Egzamin przeprowadzany jest wyłącznie na jednym poziomie, przy czym instytucje uznające uprawnienia określają minimalny wynik wymagany do przeprowadzenia procedury uznawania kwalifikacji. Każda część egzaminu oceniana jest oddzielnie (za pomocą ocen od A do E), a większość wspomnianych instytucji wymaga oceny minimum B za każdą z badanych sprawności.

OET opiera się na tematyce medycznej i zadaniach zawodowych poszczególnych specjalności medycznych, nie sprawdza jednak specjalistycznej wiedzy zawodowej jako takiej. Na takie rozwiązanie nie zezwalają przepisy australijskie, w przeciwieństwie do przepisów w Wielkiej Brytanii (McNamara 1996: 41) i Stanach Zjednoczonych, gdzie połączenie tych dwóch rodzajów kompetencji było możliwe. W systemie australijskim wiedzę i umiejętności w zakresie medycyny są weryfikowane za pomocą dwóch kolejnych etapów procedury potwierdzania kwalifikacji, a mianowicie wspomnianego wcześniej testu wiedzy medycznej oraz praktycznego egzaminu z kompetencji klinicznych (OET 2014b).

Jak wykazują badania (OET 2014b), egzamin jest bardzo dobrze postrzegany zarówno przez zdających, jak i osoby z branży medycznej. Jest on pozytywnie oceniany w zakresie tematyki, użytego języka, zadań eg-

zaminacyjnych i umiejętności językowych niezbędnych do wykonania zadań. Przez ekspertów – zarówno glottodydaktyków, jak i fachowców z branży medycznej – uważany jest za test o bardzo wysokim stopniu specyficzności (OET 2014a: 3–8; Vidaković 2015).

Od lat egzamin ten stanowi przedmiot różnego rodzaju badań naukowych, na podstawie których jest udoskonalany. W ostatnich latach przeprowadzono serię badań dotyczących autentyczności zadań egzaminacyjnych (Elder 2016) oraz na temat autentyczności kryteriów ich oceny (Elder i McNamara 2016; Pill 2016).

• Egzamin z języka angielskiego w branży turystycznej – *English for Tourism*⁸⁵

Egzamin **LCCI IQ: English for Tourism – Level 2** (poziom B1⁸⁶) (LCCI International Qualifications 2008; 2010a; 2010b) skierowany jest do osób pracujących w turystyce lub przygotowujących się do pracy w tej branży. Podczas egzaminu oceniana jest umiejętność efektywnego komunikowania się. Zgodnie z opisem testu (Pearson b.r.: 15) na profil osoby, która uzyskała potwierdzenie kwalifikacji na poziomie B1, składają się umiejętności w zakresie języka ogólnego na poziomie B1 oraz umiejętność rozumienia i stosowania języka specjalistycznego, a także terminologii używanej w branży turystycznej. Test składa się z części ustnej i pisemnej. W obydwu częściach sprawdzane są sprawności zarówno receptywne, jak i produktywne. Egzamin ustny bazuje na materiale pisanym. W jednym zadaniu jest to lista wskazówek i poleceń, według których kandydat ma skonstruować swoją wypowiedź. Drugie zadanie opiera się na szczegółowym opisie sytuacji, na podstawie którego zdający wraz z egzaminującym odgrywają scenkę imitującą autentyczną sytuację zawodową. W celu poprawnego wykonania zadania egzaminacyjnego zdający musi najpierw wykazać się zrozumieniem tekstu pisanego. Sprawność ta nie jest w tej części egzaminu oceniana bezpośrednio, jednak poziom zrozumienia przedstawionych informacji i odpowiednie wykorzystanie ich w wypowiedzi ustnej bądź podczas odgrywania ról ma wpływ na ogólną ocenę tej części egzaminu.

⁸⁵ Egzamin EfT wygaszono pod koniec 2017 roku, jednak został uwzględniony w niniejszej publikacji ze względu na bardzo ciekawe podejście do testowania ESP (cf. *Zakończenie*).

⁸⁶ Egzamin był dostępny na dwóch poziomach: *Level 1* (A2) i *Level 2* (B1).

Bardziej złożona jest część pisemna egzaminu, której podstawowym celem pozostaje ocena umiejętności tworzenia wypowiedzi pisemnej w języku angielskim w branży turystycznej. Wszystkie zadania w tej części egzaminu realizowane są przez zdającego w postaci mniej lub bardziej rozbudowanej formy pisemnej, jednak znowu mamy tu do czynienia z integracją dwóch sprawności. Ocenie podlegają zarówno umiejętności w zakresie pisania, jak i czytania. Zadania oparte są bowiem na dość obszernym materiale tekstowym, z którym zdający musi się najpierw dokładnie zapoznać, aby wyselekcjonować z niego informacje potrzebne do wykonania kolejnych zadań. W tym celu kandydat winien wykazać się umiejętnością rozumienia tekstów pisanych uwzględniających język specjalistyczny. Tworząc swoje własne teksty na podstawie i z wykorzystaniem informacji z przedstawionego materiału, zdający musi również „zaprezentować znajomość języka branży turystycznej, specjalistycznego słownictwa, wyrażeń idiomatycznych, terminologii i skrótów często używanych w branży podróźniczej i turystycznej”⁸⁷ (LCCI International Qualifications 2008: 12).

English for Tourism jest testem, który w wyraźny sposób bada podwójną kompetencję. Zadania egzaminacyjne w dużym stopniu przypominają autentyczne sytuacje zawodowe, w których zdający może się znaleźć podczas wykonywania swojej pracy. Warto zwrócić uwagę, że w egzaminie całkowicie nieobecne są zadania zamknięte. Oceniane są nie tylko umiejętności językowe kandydata, ale także umiejętność poprawnego tworzenia różnego rodzaju tekstów użytkowych, jak formalne listy, notatki, zapytania, potwierdzenia, odpowiedzi na reklamacje, raporty, plany podróży, materiały reklamowe, z uwzględnieniem zasad pisania tego typu tekstów oraz wymagań stylistycznych. Warta zauważenia jest również umiejętność opisana w sylabusie do egzaminu pod numerem 6.1:

Zdający muszą umieć:

czytać teksty, listy, faktury, fragmenty autentycznych opublikowanych instrukcji dotyczących podróżowania i turystyki, broszur reklamowych pod kątem ustalenia kosztów podróży, kosztów usług, biletów, zakwaterowania, liczby pasażerów, długości podróży itp. i zaprezentować te informacje jasno, precyzyjnie i logicznie, adekwatnie do sytuacji. (LCCI International Qualifications 2008: 9)

⁸⁷ Tłumaczenie własne – A.S.

W tym celu dozwolone jest posiadanie podczas egzaminu prostego kalkulatora, co jest niewątpliwie nietypowe w egzaminach z języków obcych, ale właśnie potwierdza fakt oceny podwójnej kompetencji.

• Egzamin z języka specjalistycznego dla pilotów i kontrolerów ruchu lotniczego – ELPAC

Spośród analizowanych dla celów niniejszej pracy egzaminów najbardziej specjalistycznym i zawierającym zadania o najwyższym stopniu autentyczności jest bez wątpienia ELPAC⁸⁸, adresowany do pilotów oraz kontrolerów lotów. Egzamin całkowicie pomija sprawności, które nie są wykorzystywane w pracy przedstawicieli tych dwóch zawodów. Nie ma w nim części sprawdzającej rozumienie czytania i pisanie, gdyż umiejętności te nie są konieczne w pracy w odnośnych profesjach, absolutnie niezbędne są natomiast umiejętności związane z produkcją ustną, i to w bardzo specjalistycznym kontekście, oraz z rozumieniem języka mówionego, od których, w przypadku zawodu pilota i kontrolera lotów, uzależnione jest bezpieczeństwo, a nawet życie ludzkie.

Realia w części ustnej egzaminu ELPAC są maksymalnie zbliżone do autentycznych. Rozmowa pomiędzy egzaminatorem a zdającym – odgrywającymi rolę pilota i kontrolera – odbywa się w warunkach imitujących rozmowę radiotelefoniczną, w której to rozmówcy, nie widząc się nawzajem, polegają jedynie na usłyszanych komunikatach. Zdający i egzaminujący są podczas egzaminu oddzieleni ścianką uniemożliwiającą kontakt wzrokowy.

Również nagrania w części *Listening* sprawiają wrażenie bardzo autentycznych: odtwarzane są wypowiedzi w różnych odmianach języka angielskiego, słyszymy rozmaite akcenty i wymowę, wypowiedzi zakłócające są odgłosami samolotu oraz dźwiękami charakterystycznymi dla wieży kontroli lotów. Dodatkowym elementem, który sprawia, że egzamin jest bliski autentycznej sytuacji zawodowej, jest konieczność udzielenia odpowiedzi po jednokrotnym wysłuchaniu tekstu w zadaniach sprawdzających umiejętność rozumienia ze słuchu (Sendur 2015: 243–244). Ta cecha odróżnia ELPAC od wielu innych egzaminów, w których tekst od-

⁸⁸ Opisany bardziej szczegółowo w: Sendur 2015; Decyzji Nr 46 Prezesa Urzędu Lotnictwa Cywilnego z dnia 28 listopada 2014 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla kontrolerów ruchu lotniczego, praktykantów-kontrolerów ruchu lotniczego i kandydatów na praktykantów-kontrolerów ruchu lotniczego; ELPAC 2013; EUROCONTROL: <https://www.eurocontrol.int/articles/elpac-tests> [dostęp: 20.02.2015].

tworzony jest dwukrotnie. W przypadku zawodu, w którym rozumienie ze słuchu ma tak istotne znaczenie, procedura ta wydaje się w pełni uzasadniona. Kolejnym charakterystycznym elementem, który kontrastuje ELPAC w stosunku do innych egzaminów językowych, jest sposób wystawienia finalnej oceny. W celu zaliczenia egzaminu zdający musi uzyskać pozytywny wynik z wszystkich części, a ocena końcowa to nie średnia ani suma ocen częściowych, ale najniższy osiągnięty wynik (Sendur 2015: 244). Co więcej, egzamin nie musi być oceniany przez lingwistę; jeśli robi to ekspert od języka, wówczas musi on być dodatkowo przeszkolony w zakresie lotnictwa. Oceny może również dokonać ekspert od lotnictwa, który sam posiada kwalifikacje językowe na wysokim poziomie, uprawniające go do odbywania lotów międzynarodowych. Rekomendowanym sposobem ustalenia oceny egzaminu ustnego jest również praca w tandemie obu ekspertów (European Aviation Safety Agency 2011: 28).

ELPAC niewątpliwie znacznie odbiega od testów języka ogólnego. Bardzo wyraźnie widać w nim pełną zgodność testowanych umiejętności z rzeczywistym działaniem zawodowym. Sprawdzana jest tu tzw. podwójna kompetencja (*double compétence*; Carras *et al.* 2007: 53), czyli zarówno umiejętności językowe, jak i zawodowe. W wymaganiach stawianych zdającym oraz w samych zadaniach egzaminacyjnych widać zdecydowaną rolę ekspertów od lotnictwa i ich wpływ na opracowywanie egzaminu.

ELPAC stanowił obowiązkowy egzamin dla pilotów i kontrolerów ruchu lotniczego w Polsce do 2016 roku. Po wyczerpaniu limitów wynikających z umowy z EUROCONTROL⁸⁹, miejsce międzynarodowego egzaminu zajął jego polski odpowiednik: Krajowy System Egzaminów Językowych (KSEJ). Jako że wytyczne dotyczące wymagań językowych przedstawicieli branży lotniczej pozostały bez zmian, egzamin KSEJ skonstruowany jest w bardzo podobny sposób jak ELPAC (Decyzja Nr 20 Prezesa Urzędu Lotnictwa Cywilnego z dn. 4 maja 2015 r.; Decyzja Nr 21 Prezesa Urzędu Lotnictwa Cywilnego z dnia 4 maja 2015 r.; opis egzaminu dostępny na stronie Urzędu Lotnictwa Cywilnego⁹⁰).

⁸⁹ Informacja dostępna na stronie: <https://ulc.gov.pl/pl/publikacje/wiadomosci/4014-egzaminy-z-jezyka-angielskiego-w-systemie-elpac-2> [dostęp: 17.06.2020].

⁹⁰ Zob. <https://ulc.gov.pl/pl/personel-lotniczy/komisja-egzaminacyjna/egzaminy-z-j-angielskiego/3742-krajowy-system-egzaminow-jezykowych-ksej?highlight=WyjrcmFqb3d5Iiwic3lzdGVtIiwij3N5c3RlbSIsInN5c3RlbSdzIiwiZWd6YW1pblx1M-DBmM3ciLCJrcmFqb3d5IHN5c3RlbSIsImtyYWpvd3kgc3lzdGVtIGVnemFtaW5vdYsInN5c3RlbSBIZ3phbWlub3ciXQ> [dostęp: 17.06.2020].

4. Stopień specjalizacji egzaminów przeprowadzanych na uczelniach wyższych

Egzaminy z języków obcych zdawane przez studentów na polskich uczelniach oraz w szkołach wyższych w innych krajach winny sprawdzać nie tylko umiejętności w zakresie języka obcego ogólnego, ale również umiejętności językowe w zakresie studiowanej dziedziny nauki. Uzyskany w wyniku zaliczenia takiego egzaminu dokument powinien poświadczать pewne specyficzne kompetencje. Ze względu na brak ujednoczenia standardów w tej kwestii, a także różnorodność profili uczelni i kierunków studiów, nie można wyciągać jednoznacznych wniosków na temat stopnia specjalizacji tychże egzaminów. Można natomiast próbować przyrzeć się pewnym praktykom oraz wytycznym sformułowanym na potrzeby językowej certyfikacji uczelnianej. Taką próbę podjęto poniżej.

a) Egzaminy językowe przeprowadzane na polskich uczelniach wyższych

Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego w sprawie Krajowych Ram Kwalifikacji dla Szkolnictwa Wyższego (Dz.U. z dnia 24 listopada 2011 r.) opisywało efekty kształcenia, które winna osiągnąć osoba kończąca studia pierwszego i drugiego stopnia. Zgodnie z rozporządzeniem osoba posiadająca zarówno kwalifikacje pierwszego, jak i drugiego stopnia powinna posiadać „umiejętności językowe w zakresie dziedzin nauki i dyscyplin naukowych, właściwe dla studiowanego kierunku studiów”, odpowiednio na poziomie B2 lub B2+ Europejskiego Systemu Opisu Kształcenia Językowego.

Przepis ten wymusił na wyższych uczelniach wprowadzenie do programów nauczania nauk języków specjalistycznych i weryfikację umiejętności studentów w tym zakresie. Na przełomie roku akademickiego 2014/2015 oraz 2015/2016 przeprowadzono badanie (Sendur 2016b), którego celem było sprawdzenie, czy (a jeśli tak – to w jaki sposób) polskie uczelnie wyższe sprawdzają umiejętności językowe swoich studentów w wymaganym zakresie. Analizie poddane zostały 42 uczelnie wyższe, w tym 19 uniwersytetów, 18 uczelni technicznych oraz 5 uczelni ekonomicznych, czyli wszystkie szkoły wyższe wymienionych trzech typów wpisane do rejestru instytucji szkolnictwa wyższego w Zintegrowanym Systemie Informacji o Nauce i Szkolnictwie Wyższym POL-on⁹¹,

⁹¹ Zob. <http://polon.nauka.gov.pl>.

działające w terminie przeprowadzania badania. Analizy dokonano przede wszystkim na podstawie informacji zamieszczonych na stronach internetowych. Przystudiowane zostały udostępnione schematy i opisy egzaminów (tam, gdzie takowe były zamieszczone), regulaminy jednostek zajmujących się nauczaniem języków obcych (w uczelniach, w których takie istnieją i są udostępnione), oraz sylabusy, zarządzenia i uchwały senatu dotyczące nauczania i oceniania w ramach języków obcych. W przypadkach niektórych uczelni dane uzupełnione zostały informacją ustną przekazaną przez pracowników. Szczegółowe wnioski z badania opisano w artykule *Realizacja wymagań dotyczących kształcenia językowego na wyższych uczelniach na podstawie analizy wybranych przykładów* (Sendur 2016b). Do celów niniejszej pracy przytoczono jedynie te dane, które dotyczą sposobu oceny biegłości w ramach języka specjalistycznego.

W wyniku dokonanej analizy można stwierdzić, że większość uczelni dąży do ujednoczenia testowania języków obcych. Jednak zadbanie o to, by nauczać i sprawdzać „umiejętności językowe [studentów] w zakresie dziedzin nauki i dyscyplin naukowych, właściwych dla studiowanego kierunku studiów, zgodne z wymaganiami określonymi dla poziomu B2 Europejskiego Systemu Opisu Kształcenia Językowego”⁹² na egzaminie ogólnouczelnianym przeprowadzanym dla wszystkich kierunków studiów w danej uczelni bądź też na egzaminie wspólnym dla całego wydziału czy kierunku studiów, stwarza spore problemy organizacyjne oraz rodzi trudności związane z samym procesem tworzenia egzaminu. W efekcie, ujednoczenie formuły egzaminów spowodowało, że bez względu na specyfikę danego kierunku umiejętności językowe studenta są w większości uczelni sprawdzane w zakresie języka ogólnego, natomiast język specjalistyczny jest ograniczony głównie do fachowego słownictwa. Takie podejście nie dziwi, gdyż samo słownictwo jest zdecydowanie łatwiej sprawdzić za pomocą testu pisemnego niż szerzej rozumiane kompetencje zawodowe w języku obcym. Potwierdza to badanie przeprowadzone wśród polskich nauczycieli szkół wyższych (Dudzik, Dziecioł-Pędich i Kic-Drgas 2019), które wykazało między innymi, że w zakresie języków specjalistycznych nauczyciele języków obcych na wyższych uczelniach testują głównie znajomość fachowego słownictwa oraz że testowanie LSP jest w ich ocenie

⁹² Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 2 listopada 2011 r. w sprawie Krajowych Ram Kwalifikacji dla Szkolnictwa Wyższego (Dz.U. z dnia 24 listopada 2011 r.).

trudniejsze niż testowanie w ramach języka ogólnego, bez względu na rodzaj nauczanego języka specjalistycznego.

Kompetencje zawodowe w języku obcym powinny być jednak sprawdzane w różnej formie i w różny sposób w zależności od specyfiki kierunku studiów. Jak bowiem zauważają Ligara i Szupelak (2012: 43):

Glottodydaktyka (danego) języka specjalistycznego powinna zatem brać pod uwagę – oprócz terminologii i poziomu zdania, a dalej tekstu specjalistycznego – również gatunki specjalistyczne tekstów oraz interakcyjny dyskurs specjalistyczny.

Zwraca na to uwagę również Alice C. Omaggio (1986), sugerując hybrydowe podejście do konstruowania testu, w którym sprawdzane są z jednej strony słownictwo i gramatyka, z drugiej zaś elementy danego rodzaju dyskursu.

Biorąc jednak pod uwagę fakt, iż nauczaniem języka specjalistycznego, a następnie przygotowaniem testów sprawdzających zajmują się w uczelniach wyłącznie glottodydaktycy, którzy w znacznej większości przypadków nie mają wystarczającej wiedzy merytorycznej dotyczącej kierunku studiów, a więc również potrzeb związanych z konkretnym zawodem czy branżą, nie dziwi, że sami czują się bezpieczniej, mierząc jedynie znajomość specjalistycznego słownictwa studentów.

Badanie pokazało, że większość egzaminów uczelnianych zbudowana jest z takich samych części, jak sprawdziany biegłości z języka ogólnego. Część pisemna z reguły składa się z testu sprawności słuchania, czytania, pisania oraz zadań leksykalno-gramatycznych. Często obowiązkowy jest również egzamin ustny, w trakcie którego badana jest sprawność mówienia. Ocenianie znajomości języka specjalistycznego odbywa się albo w wyraźnie wyróżnionej sekcji egzaminu kończącego lektorat, a dotyczącej słownictwa specjalistycznego, albo też oddzielnie w ramach zaliczenia. W sporadycznych przypadkach badanie sprawności czytania oparte jest na tekstach o specjalistycznej tematyce. Są też uczelnie, w których kompetencje w ramach języka fachowego sprawdzane są podczas egzaminu ustnego. W takim wypadku zadaniem egzaminacyjnym jest z reguły przygotowana wcześniej prezentacja na temat związany z kierunkiem studiów bądź też rozmowa na temat przeczytanego tekstu o tematyce specjalistycznej.

Mimo różnic w szczegółach dotyczących budowy arkuszy egzaminacyjnych można stwierdzić, że większość uczelni podchodzi do swoich

egzaminów w dość standardowy, schematyczny sposób, a testy egzaminacyjne przypominają swoją konstrukcją międzynarodowe egzaminy certyfikujące dla języka ogólnego. W procesie tworzenia schematu egzaminu wyraźnie daje się zauważyć rola nauczyciela-glottodydaktyka, co skutkuje tym, że zdecydowana większość wewnętrznych egzaminów uczelnianych, weryfikująca w swoim założeniu znajomość języka specjalistycznego, nie spełnia zasad, jakim powinny poddać się egzaminy LSP.

Wraz z wprowadzeniem nowych przepisów dotyczących kwalifikacji⁹³ zmianie uległy wymagane od absolwentów studiów pierwszego stopnia efekty uczenia się. Na tym poziomie kwalifikacji (poziom 6) zniknął zapis o konieczności posiadania umiejętności w zakresie języka specjalistycznego. Jest on w dalszym ciągu obecny dla poziomu 7, czyli dla absolwentów studiów magisterskich. Warto byłoby sprawdzić, czy zmiana przepisów pociągnęła za sobą modyfikacje w zakresie egzaminowania umiejętności językowych.

b) Projekt GULT – *Guidelines for University Language Testing*

W latach 2008–2011 pod egidą Europejskiego Centrum Języków Nowożytnych w Grazu (*European Centre for Modern Languages – ECML*), w ramach projektu dotyczącego oceniania w nauczaniu języków obcych w podejściu zadaniowym⁹⁴ (*Task-based Assessment – TBA*), współpracował międzynarodowy zespół, w skład którego wchodził nauczyciele języków obcych i inne osoby zajmujące się testowaniem. Celem prac zespołu było zaprojektowanie formy testowania języków obcych zgodnej z metodologią podejścia zadaniowego, a tym samym „spowodowanie, by testy językowe stały się bardziej relewantne dla zdających oraz umożliwienie testowania stopnia, w jakim egzaminowani są w stanie porozumiewać się prawidłowo w specyficznej sytuacji komunikacyjnej w zakresie swoich studiów lub swojej przyszłej dziedziny zawodowej”⁹⁵ (Fischer *et al.* 2011: 9). Dążeniem zespołu było opracowanie ram egzaminu, który będzie sprawdzał umiejętności studentów w zgodzie z podejściem zadaniowym, a efektem końcowym stała się publikacja *Guidelines for University Language Testing* (GULT; Fischer *et al.* 2011).

⁹³ Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dn. 14 listopada 2018 r. w sprawie charakterystyk drugiego stopnia efektów uczenia się dla kwalifikacji na poziomach 6–8 Polskiej Ramy Kwalifikacji.

⁹⁴ Strona internetowa projektu GULT: <http://gult.ecml.at/Projectdescription/tabid/178/language/en-GB/Default.aspx> [dostęp: 17.06.2020].

⁹⁵ Tłumaczenie własne – A.S.

Opisane w przewodniku zasady testowania w podejściu zadaniowym pokazują kierunek, w jakim według autorów powinno zmierzać nowoczesne testowanie. Przedstawiony w niej projekt egzaminu daje ramy, do których każda instytucja zainteresowana wdrożeniem tego typu oceniania może włożyć własne treści i które może – odpowiednio do swoich potrzeb i zgodnie ze swoimi przekonaniem – modyfikować.

GULT nie jest zatem egzaminem, lecz przewodnikiem pomagającym w jego opracowaniu, stworzonym z myślą o językach specjalistycznych nauczanych na studiach wyższych, chociaż według autorów schemat ten może być zastosowany z równym powodzeniem do egzaminów z języków ogólnych. Jako że przedmiotem projektu są ogólne ramy egzaminu, a nie sam test, nie ogranicza się on także do żadnego konkretnego języka, co stanowi o jego uniwersalności (Fischer *et al.* 2011: 15).

Jak zaznaczono wcześniej, w ocenianiu TBA nacisk położony jest na umiejscowienie aktów komunikacji w specyficznym kontekście i specyficznej sytuacji, ze specyficznym celem, będącym odzwierciedleniem aktualnych i przyszłych komunikacyjnych celów użytkownika języka. Polecenia wykorzystane w testowaniu kompetencji w podejściu zadaniowym powinny być bardzo bliskie zadaniom komunikacyjnym, z którymi zdający mają do czynienia w rzeczywistym świecie lub z którymi najprawdopodobniej zetkną się w przyszłości. Aby uczynić test jak najbardziej autentycznym, zadania postawione przed zdającymi powinny integrować wszystkie sprawności, tak jak to ma miejsce w codziennym życiu, gdzie np. określony tekst czytamy nie tylko po to, by go zrozumieć, ale żeby wykorzystać uzyskane w ten sposób informacje do konkretnego celu. Jeżeli jednak wszystkie sprawności są ze sobą zintegrowane w ramach jednego testu (głównego zadania), a umiejętności zdającego potraktowane holistycznie, to oceniać powinniśmy wyłącznie sprawności produktywne (Fischer *et al.* 2011: 26). Tymczasem w większości językowych egzaminów doniosłych testowane są zazwyczaj cztery sprawności oceniane rozdzielnie. Ponadto mimo iż ESOKJ jest ewidentnym propagatorem nauczania zadaniowego, w ogóle nie odnosi się on na swych łamach do oceniania w takim właśnie podejściu, na co zwracają uwagę autorzy projektu GULT (Fischer *et al.* 2011: 38). Jak zatem odnieść wyniki TBA do poziomów ESOKJ?

Dla celów transparentności, trafności fasadowej, porównywalności i przełożenia na poziomy ESOKJ, jak również na użytek analizy zasadne jest testowanie – obok sprawności produktywnych – również sprawności

receptywnych. Do takich wniosków doszedł zespół GULT, który z tych właśnie powodów do schematu swojego egzaminu wprowadził poszczególne sprawności. Są one testowane indywidualnie, oddzielnie, choć pozostają zarazem ze sobą powiązane za pośrednictwem głównego zadania. Jest to zatem pewien kompromis pomiędzy autentyczną sytuacją językową i tradycją testowania.

Zadania testowe w egzaminie GULT umiejscowione są w specyficznym kontekście i skonstruowane na bazie zadania głównego opartego na działaniu. Składają się z serii pośrednich zadań cząstkowych, które są ze sobą powiązane i które mają prowadzić do wykonania zadania głównego. Zadania cząstkowe, jakie zdający ma wykonać, dążąc do zrealizowania zadania głównego, każdorazowo skupiają się na jednej sprawności. Wykonanie tych czynności pośrednich wymaga koncentracji na całym problemie – tym sposobem sprawności są mocno ze sobą powiązane. Zdający musi zrozumieć pisany lub słuchany tekst, żeby następnie wykonać odpowiednie działania w fazie produkcji. Sprawności są wkomponowane w ogólny kontekst i obudowują zadanie główne (Fischer *et al.* 2011: 25–27).

Struktura egzaminu GULT została opracowana na podstawie francuskich egzaminów uniwersyteckich CLES (*Certificat de Compétences en Langues de l'Enseignement Supérieur*) oraz niemieckich egzaminów Naukowego Komitetu UNiCert. Składa się on z poniższych części:

- 1) prezentacja tematu i zadania głównego,
- 2) test rozumienia tekstu słuchanego i czytanego,
- 3) analiza przypadku – praca w parach lub małych grupach,
- 4) pisanie,
- 5) mówienie (Fischer *et al.* 2011: 28).

W egzaminie GULT część sprawdzająca słuchanie i czytanie otrzymuje nową rolę. Zdający nie zapoznają się z tekstami wyłącznie w celu zrozumienia treści, ale muszą je zrozumieć, wiedząc, że informacje w nich zawarte będą potrzebne do wykonania zadania głównego. Chociaż teksty do słuchania i do czytania odnoszą się do różnych aspektów, to jednak są połączone wspólnym tematem. W części przeznaczony na analizę przypadku zdający pracują w parach lub małych grupach, wspólnie szukając rozwiązania przedstawionego problemu, a następnie w części ustnej razem prezentują rozwiązanie. W testach mierzących sprawności produktywną rolę zdających jest zaprezentowanie konkretnego produktu w określonym kontekście. W tym celu muszą wykorzystać i przetworzyć

informacje analizowane w poprzednich częściach egzaminu. W części sprawdzającej umiejętność pisania zdający samodzielnie tworzą tekst, wykorzystując do tego celu materiały z wcześniejszych części egzaminu oraz swoje notatki. Sprawność mówienia mierzona jest podczas wspólnej grupowej prezentacji, a następnie wynikającej z niej dyskusji (Fischer *et al.* 2011: 28–31).

Podczas dokonywania oceny, GULT bierze pod uwagę nie tylko poprawność językową wypowiedzi zdającego, ale także poziom wykonania zadania.

5. Stopień sprofilowania zawodowego w egzaminach z języka specjalistycznego

Dzięki powszechnej akceptacji w Europie poziomy biegłości ESOKJ stały się obowiązującym standardem w certyfikacji językowej. Większość testów biegłości opiera się na poziomach i deskryptorach opracowanych przez Radę Europy. Również zdecydowana większość egzaminów ESP odnosi swoje wyniki do poziomów biegłości językowej *Europejskiego systemu opisu* oraz do deskryptorów umiejętności językowych zawartych w tym dokumencie. Dzieje się tak mimo iż wskaźniki skal ESOKJ odwołują się do języka ogólnego, a nie do języków specjalistycznych. W przypadkach, gdy uznano, że skale ESOKJ nie spełniają wymagań danej branży zawodowej, stwarzano niezależną skalę opisu biegłości językowej zawierającą szczegółowe umiejętności wymagane przy wykonywaniu konkretnych zawodów (np. skala ICAO). W innych sytuacjach skale ESOKJ zostały dostosowane do potrzeb egzaminu poprzez selekcjonowanie potrzebnych deskryptorów czy też konkretnych sprawności (np. sTANDEM) bądź też sam egzamin został dostosowany do skal ESOKJ (np. MOTET).

Wart uwagi jest również fakt, że nawet w przypadkach, kiedy na potrzeby certyfikacji biegłości w zakresie języka specjalistycznego zostały stworzone odrębne od ESOKJ skale, bazą dla opisu umiejętności językowych są w tych skalach wskaźniki odnoszące się do biegłości w zakresie języka ogólnego, zaś deskryptory opisujące umiejętności dotyczące języka specjalistycznego stanowią tylko dodatek (ICAO, STANAG 6001).

Ocenianie języków obcych specjalistycznych w swoich założeniach różni się od oceniania w innych obszarach nauczania języka tylko w zakresie pewnych szczegółów, na które wskazuje Douglas (2000: 2) i które zostały omówione bardziej szczegółowo w rozdziale 1. Dwie zasadnicze cechy odróżniające testowanie dla celów ogólnych i celów specjalistycznych to autentyczność zadań oraz interakcja pomiędzy znajomością języka i wiedzą specjalistyczną.

Autentyczność zadań w testach LSP powinna polegać na maksymalnym zbliżeniu zadań egzaminacyjnych do zadań wykonywanych w rzeczywistej sytuacji zawodowej. W takich przypadkach mówimy o sprawdzaniu podwójnej kompetencji zdającego – zarówno znajomości języka, jak i umiejętności zastosowania go w konkretnych zadaniach zawodowych. Zbliżenie zadań egzaminacyjnych do zadań zawodowych jest również powodem całkowitego ograniczenia się w przypadku niektórych egzaminów LSP do zadań badających sprawności priorytetowe, czyli tylko te, które są niezbędne do wykonywania czynności zawodowych w danej branży. Charakterystyczne dla egzaminów LSP, a niewystępujące raczej w egzaminach z języka ogólnego, są także zadania egzaminacyjne uwzględniające interakcję kilku sprawności, co powoduje, że jest ono silnie zbliżone do rzeczywistego zadania zawodowego.

W opisanych egzaminach LSP dają się zauważyć istotne różnice w zakresie stopnia autentyczności zadań, zarówno sytuacyjnej, jak i interakcyjnej. Widzimy przykłady egzaminów, w których zadania testowe są bliskie lub bardzo bliskie rzeczywistej sytuacji zawodowej (np. OET, Eft, ELPAC), oraz egzaminy sprawdzające użycie języka w sytuacji zbliżonej do sytuacji szkolnej (np. VEC, EfA, MOTET, egzaminy uczelniane).

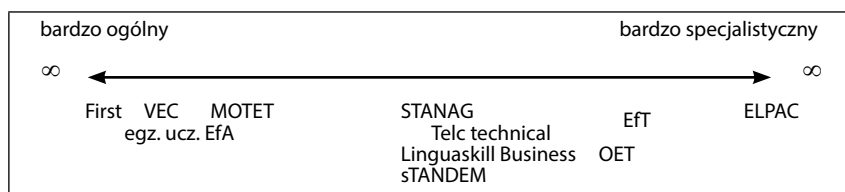
Drugą cechą egzaminów LSP jest interakcja pomiędzy znajomością języka i wiedzą specjalistyczną. W testowaniu LSP, odwrotnie niż w testowaniu języków ogólnych, znajomość zagadnień specjalistycznych jest niezbędna i odgrywa zasadniczą rolę w badaniu umiejętności językowych z danej dziedziny (Douglas 2000: 2). W zakresie powiązania umiejętności językowych i wiedzy specjalistycznej również obserwujemy spore różnice pomiędzy różnymi egzaminami ESP. W porównywanych testach mamy do czynienia z sytuacją, w której brak wiedzy specjalistycznej praktycznie uniemożliwi prawidłowe wykonanie zada-

nia. Takim egzaminem jest ELPAC, w przypadku którego wiedza zawodowa jest absolutnie niezbędna, by poradzić sobie z tym wyzwaniem. Przykładem innego podejścia są VEC i MOTET, gdzie wiedza specjalistyczna jest pewnego rodzaju „otoczką” do zadania i ma zdecydowanie mniejszy wpływ (jeśli w ogóle) na jego realizację.

Znaczenie interakcji pomiędzy znajomością języka i wiedzą specjalistyczną na egzaminach LSP jest istotne nie tylko z punktu widzenia zdającego, ale również z perspektywy egzaminatora. Od stopnia specjalizacji zawodowej egzaminów LSP uzależnione jest to, kto może je oceniać. W przypadku testów z języków ogólnych czy też egzaminów LSP, w których sprawdzana jest wyłącznie znajomość specjalistycznego słownictwa, egzaminator przeszkolony w zakresie kryteriów i sposobu oceny egzaminu będzie w stanie podjąć się tego zadania. Jeśli chodzi natomiast o wysoce specjalistyczne egzaminy jak ELPAC, dokonanie oceny umiejętności zdającego nie jest możliwe bez uprzedniego dogłębnego przygotowania, w tym w zakresie odnośnego zawodu.

Analizując różnice pomiędzy wybranymi egzaminami ESP można dojść do wniosku, że to, w jakim stopniu dany egzamin odzwierciedla autentyczne zadania zawodowe, jest uzależnione od osób, które miały wpływ na jego kształt. Można zaobserwować, jaką rolę przy tworzeniu poszczególnych egzaminów odgrywali glottodydaktycy, a jaką specjaliści z dziedziny zawodowej, której egzamin dotyczy. Im większy wkład ze strony fachowców z zakresu dziedzin specjalistycznych, tym większy wydaje się stopień zadaniowości w testach egzaminacyjnych. Wpływ glottodydaktyków jest widoczny między innymi w obecności na egzaminach zadań sprawdzających poprawność gramatyczną, a nawet ortograficzną, oraz znajomość słownictwa (np. EfA, MOTET, egzaminy uczelniane).

Według Douglasa (2000: 1) każdy test znajduje się w innym punkcie kontinuum specyficzności, od testów sprawdzających język bardzo ogólny, do egzaminów o bardzo specyficznych celach zawodowych. Powyższa obserwacja daje się odnieść do opisanej tutaj certyfikacji ESP. Poniżej zaprezentowano graficznie poziom specyficzności czy też specjalistyczności w omówionych egzaminach, w relacji do przykładowego testu GE, jakim jest *Cambridge B2 First*. Przedstawiony tutaj schemat (Rysunek 3) uzupełnia zestawienie porównawcze stopnia specjalistyczności egzaminów LSP opisane przez Agnieszkę Sendur (2020: 55–60).

Rysunek 3. Stopień specjalistyczności analizowanych egzaminów ESP

6. Główne różnice pomiędzy egzaminami w glottodydaktyce ogólnej oraz specjalistycznej na podstawie analizy przykładowych egzaminów

Zgodnie z wcześniejszymi uwagami ocenianie w zakresie języków ogólnych i języków specjalistycznych rządzi się takimi samymi zasadami: każdy egzamin powinien przestrzegać podstawowych zasad poprawności testu, czyli trafności, rzetelności, ale także praktyczności. Dają się jednak zaobserwować pewne różnice pomiędzy testami z języków ogólnych oraz tymi w glottodydaktyce specjalistycznej.

Po pierwsze, testy ogólne odwołują się do skal biegłości ESOKJ, jako do międzynarodowego, powszechnie rozpoznawalnego i zrozumiałego opisu poziomów biegłości. Egzamin specjalistyczny również odnosi się do tych norm, jednak część z nich opiera się na skalach własnych, stworzonych na potrzeby danego zawodu bądź samego egzaminu.

Po drugie, egzamin z języków ogólnych bada zazwyczaj umiejętności w ramach czterech podstawowych sprawności językowych, traktując je równoważnie. W przypadku egzaminów z języków specjalistycznych sytuacja bywa inna. Egzamin o średnim i wysokim stopniu sprofilowania zawodowego ograniczają się do wybranych sprawności bądź traktują je jako bardziej, a inne jako mniej istotne (sprawności priorytetowe/ograniczone), albo też dają zdającemu możliwość wyboru badanych sprawności.

Po trzecie, egzamin z języków ogólnych często sprawdza znajomość słownictwa oraz struktur gramatycznych w sposób bezpośredni. Egzamin specjalistyczny o średnim i wysokim stopniu specjalistyczności często nie zamieszcza w arkuszu egzaminacyjnym zadań gramatyczno-leksykalnych. W tym wypadku te podsystemy języka podlegają ocenie pośred-

nio, podczas sprawdzania innych kompetencji, na przykład w ramach testu umiejętności pisania czy mówienia.

Po czwarte, zadania w egzaminach ogólnych, z racji swoich niespecyficznych celów, wykazują się mniejszym stopniem autentyczności niż egzaminy z języków specjalistycznych. Zadania egzaminacyjne w testach ogólnych odnoszą się do sytuacji, które niekoniecznie są zbieżne z obecnymi lub przyszłymi potrzebami zdającego. W przypadku egzaminów w glottodydaktyce specjalistycznej okoliczności te są zazwyczaj dużo bliższe sytuacjom zawodowym, w których zdający może się znaleźć w nieodległej przyszłości.

I wreszcie związek pomiędzy wiedzą językową i wiedzą ogólną oraz wiedzą zawodową: w przypadku egzaminów ogólnych wpływ wiedzy ogólnej na rozwiązanie zadań testowych powinien być maksymalnie ograniczony, podczas gdy w przypadku zadań w testach specjalistycznych wiedza ta jest niejednokrotnie niezbędna do prawidłowego wykonania zadania językowego.

W Załączniku 3 zebrane zostały w celach porównawczych szczegółowe informacje dot. struktury wszystkich analizowanych tutaj egzaminów, wraz z wykorzystywanymi w nich rodzajami zadań testowych.

Zakończenie: bilans i perspektywy certyfikacji językowej

Analiza teorii dotyczącej testowania pokazuje, że istota różnicy pomiędzy testowaniem ogólnym a specjalistycznym tkwi w badaniu tzw. podwójnej kompetencji. W działaniu zawodowym kompetencje językowe i zawodowe są ze sobą ściśle powiązane. Dlatego też egzamin, który miałby sprawdzić, czy kandydat jest w stanie skutecznie korzystać z języka obcego w celach zawodowych, powinien być oparty przede wszystkim właśnie na zadaniach naśladujących rzeczywiste działanie w autentycznych sytuacjach zawodowych. Tymczasem, jak wykazuje analiza egzaminów, nie zawsze tak jest. Wręcz przeciwnie, można zaryzykować twierdzenie, że egzaminy silnie ukierunkowane zawodowo są prawdziwą rzadkością.

Rynek edukacyjny, tak jak każdy inny rynek ekonomiczny, rządzi się prawami popytu i podaży, w konsekwencji musi być więc podatny na zmiany. Modyfikacje w ofercie certyfikacji językowej są na porządku dziennym. Egzaminy podlegają regularnym transformacjom wynikającym z nowych doświadczeń, przeprowadzonych badań, pojawiających się nowych możliwości, ale i konieczności. Pewne egzaminy znikają z powodu zmniejszającego się zainteresowania konsumentów, ewolucji potrzeb zawodowych danej branży, jak i zmian zachodzących na szeroko rozumianym rynku pracy. Nie bez znaczenia są oczywiście koszty związane z certyfikacją, które siłą rzeczy są wyższe w przypadku egzaminów kierowanych do wyselekcjonowanych grup zawodowych (jak testy LSP) niż w przypadku certyfikatów z języka ogólnego, które mają dużo szerszego odbiorcę. Certyfikat jest produktem handlowym i jako taki jest uzależniony od praw rynku (Sendur 2020), a praktyczność testu zawsze zdecydowanie przeważa nad jego trafnością (Komorowska 2005b: 48).

Wpływ praw ekonomicznych na losy egzaminów dał się zaobserwować podczas śledzenia zmian w certyfikacji w XX wieku. Nowe egzaminy często funkcjonowały przez pewien okres, wzbogacając ofertę edukacyjną, miały swój znaczący wpływ na rozwój certyfikacji w danym okresie, były udoskonalane, po czym znikają z rynku, ustępując miejsca nowym produktom. Sytuacja taka ma miejsce również obecnie. Z jednej strony, ciągle pojawiają się nowe egzaminy, na przykład w wyniku zmian w przepisach, które wymuszają na pewnej grupie społecznej lub zawodowej uzyskanie określonych poświadczeń (np. ELPAC, KSEJ), bądź też w efekcie zapotrzebowania wynikającego z bieżącej sytuacji politycznej lub ekonomicznej. Przykładem egzaminu, który powstał na potrzeby planowanych wydarzeń gospodarczych, a nie został wprowadzony w życie z przyczyn politycznych, jest test dla policji opracowany przez Brytyjską Izbę Handlową we Włoszech (*British Chamber of Commerce for Italy*), która przygotowała ofertę takiego egzaminu przed wystawą światową Expo 2015 w Mediolanie. Wdrożenie opracowanego już projektu nie doszło do skutku ze względu na zaistniałe w tym czasie zmiany kadrowe we włoskiej policji⁹⁶.

Z drugiej strony, pewne produkty – niejednokrotnie dobrze znane i cieszące się dużym uznaniem w środowisku – znikają z przyczyn ekonomicznych bądź w wyniku konieczności modernizacji. Prawa rynku zdecydowały w ostatnich latach o wycofaniu z rynku kilku egzaminów specjalistycznych. W grudniu 2016 roku przeprowadzono ostatnią edycję egzaminu *Cambridge English Legal* (wcześniej: *International Legal English Certificate* – ILEC). Co ciekawe, organizator (*Cambridge English Language Assessment*) w zamian proponuje zainteresowanym egzamin z akademickiego języka angielskiego – *Cambridge English Advanced*⁹⁷. Dwa kolejne egzaminy ze specjalistycznego języka angielskiego – *English for Accounting* oraz *English for Tourism* (Pearson) zostały wycofane z oferty pod koniec 2017 roku⁹⁸. W miejsce egzaminu dla księgowych Pearson oferuje cały pakiet egzaminów z zakresu finansów i rachunkowości. Są to egzaminy czysto zawodowe, jedynie przeprowadzane

⁹⁶ Informacje z korespondencji własnej z Davidem Gibbonem, osobą odpowiedzialną za certyfikaty Brytyjskiej Izby Handlowej we Włoszech – IPEC (*International Professional English Certificates*) z dn. 27.10.2015.

⁹⁷ Zob. <http://www.cambridgeenglish.org/exams/ilec-discontinued/> [dostęp: 6.07.2020].

⁹⁸ Zob. <http://enews.pearson.com/rsps/m/zaOaTaDE3OfCo54LaMKg6Z-FZVjTsa-D9Y3rStmHRa-k> [dostęp: 6.07.2020].

w języku angielskim, a nie egzaminy językowe: sprawdzają kompetencje zawodowe wymagane do pracy w międzynarodowych korporacjach⁹⁹. Z kolei egzamin dla branży turystycznej przestał być atrakcyjny ekonomicznie dla instytucji egzaminującej w momencie wycofania z polskich przepisów obowiązku poświadczania znajomości języka obcego przez pilotów zagranicznych wycieczek¹⁰⁰. Również w przypadku tej branży organizator proponuje kandydatom udział w sesji dla egzaminów ogólnych. Podczas gdy egzamin dla księgowych nie spełniał norm postawionych przed egzaminami LSP, to w przypadku tego drugiego (EfT) rezygnacja z kontynuacji funkcjonowania testu jest – dla osoby postrzegającej go z perspektywy nie produktu handlowego, ale jakości produktu dydaktycznego – nieodżałowaną stratą.

Z problemami innego typu borykał się egzamin z języka medycznego sTANDEM. Test ten cieszył się dobrymi recenzjami i wysokim zainteresowaniem, zwłaszcza, że potrzeba takiego certyfikatu na gruncie europejskim jest bardzo duża (cf.: Hunter 2010; Smith 2010). Jest on wynikiem współdziałania zespołów z wielu krajów pracujących wspólnie w ramach projektu europejskiego. Kontynuacja tejże współpracy okazała się niemożliwa po zakończeniu projektu, ze względu na brak jednego prawne- go właściciela produktu, co na chwilę obecną spowodowało zawieszenie przeprowadzania egzaminów do odwołania¹⁰¹. Gdyby test dostał szansę na kontynuację oraz dalszy rozwój, mógłby zapewne rozwinąć się w bardzo profesjonalny, wysoko wyspecjalizowany egzamin LSP.

Jeszcze innym powodem zmian w ofercie certyfikacyjnej jest modernizacja egzaminów na skutek pojawiających się nowych możliwości technicznych. Przykładem znanego i uznawanego egzaminu międzynarodowego, który zniknął z rynku, jest test z biznesowego języka angielskiego BULATS¹⁰². Funkcjonujący w branży certyfikacyjnej przez ponad 20 lat (od 1997 do końca 2019 roku) egzamin został zastąpiony przez nowoczesny komputerowy test adaptacyjny oparty na technologii sztucznej inteligencji – *Linguaskill Business*. Nowy egzamin pojawił się również w wersji dla języka ogólnego – *Linguaskill General*.

⁹⁹ Zob. <https://qualifications.pearson.com/en/qualifications/lcci.html> [dostęp: 6.07.2020].

¹⁰⁰ Informacja uzyskana podczas rozmowy telefonicznej z menadżerem ds. rozwiązań egzaminacyjnych w Pearson Poland, p. Dorotą Bruszewską (kwiecień 2017).

¹⁰¹ Informacja uzyskana podczas rozmowy telefonicznej z koordynatorem projektu na terenie Polski, p. Iwoną Misztal (marzec 2017).

¹⁰² Zob. <https://www.cambridgeenglish.org/exams-and-tests/bulats> [dostęp: 6.07.2020].

I tu nasuwa się jeszcze jedna obserwacja dotycząca certyfikacji z języka ogólnego oraz testowania umiejętności z zakresu języków specjalistycznych. Egzaminujący, którym udało się przetrwać całe dekady, a nawet i ponad wiek, to egzaminujący z języka angielskiego ogólnego (lub akademickiego). W czołówce najdłużej funkcjonujących certyfikatów są: powstały w 1939 roku brytyjski *First Certificate* (pierwotnie jako *Lower Certificate in English*, następnie *First Certificate in English*, *Cambridge English First*, a obecnie – *Cambridge B2 First*; Cambridge English Language Assessment), obecny na rynku od 1964 roku amerykański TOEFL (Educational Testing Service) oraz absolutny zwycięzca w tej kategorii – *Cambridge Proficiency in English*, który w 2013 roku świętował swoje setne urodziny!

Jaka jest zatem przyszłość certyfikacji językowej? Znajomość języków obcych zajmuje priorytetowe miejsce wśród tak ważnych obecnie kompetencji kluczowych. Duże znaczenie dla instytucji kształcenia oraz na rynku pracy ma dzisiaj możliwość wylegitymowania się certyfikatem językowym. Certyfikat, który może uzyskać akceptację instytucji czy pracodawców jako wiarygodne poświadczenie określonych kwalifikacji, musi spełniać pewne normy i standardy jakości (Sujecka-Zając i Sobolew 2015: 11–12). Aby móc te standardy zapewnić, potrzeba fachowej wiedzy i pracy wielu ekspertów z różnych dziedzin – nie tylko lingwistów i fachowców od testowania, ale również specjalistów od nowych rozwiązań technologicznych oraz w przypadku egzaminów LSP – ekspertów z dziedzin specjalistycznych. Potrzebne są również nakłady na badania, bez których narzędzia pomiaru, jakimi egzaminujący językowe być powinny, nie zawsze będą spełniać podstawowe normy trafności i rzetelności konieczne do prawidłowej diagnostyki i miarodajnej certyfikacji. W efekcie tylko duże instytucje działające na skalę międzynarodową są w stanie zapewnić te standardy i utrzymywać oraz rozwijać egzaminujący na odpowiednim poziomie i na szeroką skalę. Egzaminujący z języków ogólnych mają znacznie więcej odbiorców niż niszowe egzaminujący wysoko specjalistyczne, więc to one są priorytetem dla tych instytucji.

Kierunek, w jakim podąża certyfikacja językowa, to egzaminujący w wersji elektronicznej. Egzaminowanie w wersji komputerowej czy internetowej nie jest nowością: jest ono obecne w certyfikacji językowej od lat. ETS już w 1998 roku wprowadził komputerową wersję testu TOEFL, a w 2005 roku weszła w życie jego internetowa wersja – TOEFL

iBT. Papierowa forma tego egzaminu oferowana jest obecnie wyłącznie w miejscach, gdzie nie ma odpowiedniego dostępu do Internetu.

Kolejnym przełomem w branży egzaminacyjnej jest wprowadzenie do procesu testowania technologii sztucznej inteligencji, co powoduje, że będzie można je przeprowadzać jeszcze szybciej i taniej, na jeszcze bardziej masową skalę (cf. Çetin Köroğlu 2020; Lin, Lin i Tsai 2020; Yannakoudakis, Andersen, Geranpayeh, Briscoe, Nicholls 2018; Zou, Liviero, Hao, Wei 2020). Na przyspieszony rozwój egzaminowania w formie komputerowej ma również wpływ pandemia COVID-19¹⁰³ (Isbell i Kremmel 2020). Konieczność prowadzenia certyfikacji mimo ograniczeń spowodowanych przymusową izolacją ludności skłoniła instytucje, które wcześniej przeprowadzały egzaminy wyłącznie w formie papierowej, do wszczęcia intensywnych prac nad wersją zdalną swoich egzaminów. Przykładem jest *Occupational English Test* (OET), który w maju 2020 roku, po ponad dwudziestu latach funkcjonowania papierowej wersji testu, podjął współpracę z firmą Prometric w celu uruchomienia testu komputerowego *OET@Home*¹⁰⁴. Jak wskazuje nazwa nowego wariantu egzaminu, dzięki odpowiednim zabezpieczeniom, które umożliwiają nowoczesne technologie, do egzaminu będzie można przystąpić przebywając we własnym domu. Niektóre inne egzaminy wprowadziły na czas pandemii szereg modyfikacji umożliwiających przeprowadzanie egzaminu w formie zdalnej. IELTS Consortium oferuje swoim kandydatom wariant egzaminu przeprowadzanego całkowicie w formie zdalnej – *IELTS Indicator*¹⁰⁵, zastrzegając jednak, że wyniki mają charakter ogólnego wskaźnika (stąd nazwa *Indicator*) umiejętności zdającego i nie jest akceptowany przez wszystkie instytucje (Isbell i Kremmel 2020: 6–7). Zgodnie z informacją na stronie instytucji egzaminującej, ma on być dostępny tylko przez okres ograniczeń związanych z pandemią, jednak życie pokaże, czy wprowadzone tymczasowe zmiany nie pozostaną z nami na dłużej. Z kolei ETS, który wprowadził nowy wariant swojego sztandarowego egzaminu – *TOEFL iBT*

¹⁰³ Epidemia zakaźnej choroby COVID-19 wywołanej przez koronawirusa SARS-CoV-2 została uznana w marcu 2020 roku przez Światową Organizację Zdrowia (WHO) za pandemię. Ograniczenia wprowadzone w wielu krajach świata, w celu przeciwdziałania rozprzestrzeniania się choroby, uwzględniały m.in. zamknięcie szkół i uczelni wyższych, jak również ograniczyły czasowo działalność instytucji, w tym tych przeprowadzających egzaminy.

¹⁰⁴ Zob. <https://prod-wp-content.occupationalenglishtest.org/resources/uploads/2020/05/12095339/OET-launches-remote-proctored-test.pdf> [dostęp: 10.07.2020].

¹⁰⁵ Zob. <https://www.ieltsindicator.com/> [dostęp: 22.07.2020].

Special Home Edition – nie ma wątpliwości dotyczących jakości i bezpieczeństwa nowej wersji, zapewniając, że wyniki mogą być akceptowane w takim samym zakresie, jak certyfikat wystawiany na podstawie egzaminu przeprowadzanego w siedzibie instytucji (Isbell i Kremmel 2020: 10).

Wydaje się zatem, że rozszerzenie oferty o egzaminy w wersji elektronicznej, a nawet całkowite wyparcie testów papierowych przez komputerowe, to nieunikniony kierunek zmian nie tylko w certyfikacji językowej.

Swoją idealistyczną wizję egzaminów przyszłości przedstawia Bernard Spolsky (2017: 381–382) w podsumowaniu swojego artykułu dotyczącego historii testowania: „przemysłowe testy staną się bardziej ludzkie i mniej potężne; do celów istotnych decyzji zawodowych stosowane będą wyłącznie trafne testy, które przejdą staranną procedurę walidacji; proste, jednowymiarowe skale i punktacja zostaną zastąpione przez złożony opis profilu każdej testowanej osoby, który pokaże zakres jej kompetencji różnorodnej; testy nie będą wykorzystywane niewłaściwie”.

Spolsky pozostaje jednak realistą, przyznając się do małej wiary w urzeczywistnienie zaprezentowanej wizji. Twierdzi, że informatyzacja testowania pociąga za sobą nowe problemy, jednocześnie nie rozwiązując starych (Spolsky 2017: 375).

Bibliografia

- Alderson, J.C. (2009). Test review: Test of English as a Foreign Language™: Internet-based Test (TOEFL iBT®). *Language Testing* 26(4). 621–631.
- Alderson, J.C., Clapham, C., Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- ALTE (1998). *Multilingual glossary of language testing terms*. Cambridge: Cambridge University Press.
- ALTE (2005). *Materials for the guidance of test item writers*. Online: <https://www.alte.org/resources/Documents/IWG%20July2005.pdf> [dostęp: 9.06.2020].
- ALTE (2009). *Structural competence. Individual component checklist*. Online: http://www.alte.org/attachments/files/structural_comp.pdf [dostęp: 7.12.2018].
- ALTE (2011). *Manual for language test development and examining*. Strasbourg: Language Policy Division. Council of Europe. Online: <https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b> [dostęp: 15.06.2020].
- ALTE (2018). *Guidelines for the development of Language for Specific Purposes tests*. Online: <https://www.alte.org/resources/Documents/6093%20LSP%20Supplement%20-%20WEB.pdf> [dostęp: 28.06.2020].
- Anthony, L. (2018). *Introducing English for specific purposes*. London – New York: Routledge.
- Aubin, S. (2008). De Ce premier numéro. *Synergies Espagne* 1. 19–22.
- Bachman, L.F. (2003). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., Davidson, F., Ryan, K., Choi, I.C. (1995). *An investigation of the comparability of the two tests of English as a foreign language: The Cambridge – TOEFL comparability study*. Cambridge: UCLES/Cambridge University Press.
- Bachman, L.F., Palmer, A.S. (2002). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bartosik, D. (2019). Ocenianie znajomości języka polskiego jako obcego w odmiennie specjalistycznej. *Kwartalnik Polonicum* 31/32. 58–62.
- Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Basingstoke: Palgrave Macmillan.

- Beinstein, J. (1975). Conversations in public places. *Journal of Communication* 25(1). 85–95.
- Belcher, D. (2009). What ESP is and can be: An introduction. W: D. Belcher (red.), *English for Specific Purposes in theory and practice*. Ann Arbor, MI: University of Michigan Press. 1–20.
- Berry, V., O'Sullivan, B. (2016). Language standards for medical practice in the UK: Issues of fairness and quality for all. W: C. Docherty, F. Barker (red.), *Language assessment for multilingualism. Proceedings of ALTE Paris Conference. April 2014*. Cambridge: UCLES/Cambridge University Press. 268–285.
- Boczkowski, A. (2001). Zapewnianie i ocena jakości kształcenia w dziedzinie higieny pracy w Wielkiej Brytanii, USA i Kanadzie. *Medycyna Pracy* 52(2). 145–152. Online: <http://www.imp.lodz.pl/upload/oficyna/artykuly/pdf/full/Bocz11-02m-01.pdf> [dostęp: 7.06.2020].
- Booth, D. (2002). Revisiting the Business English Certificates (BEC) speaking tests. *Cambridge ESOL Research Notes* 8. 4–6. Online: <http://www.cambridgeenglish.org/images/23120-research-notes-08.pdf> [dostęp: 9.06.2020].
- Breszka-Jędrzejewska, K. (2019). Studium przypadku jako narzędzie budowania motywacji uczestników zajęć z języka obcego. *Języki Obce w Szkole* 5. 71–74.
- Brooks, R.L., Hoffman, M. (2014). Government and military assessment. W: J. Kunnan (red.), *The companion to language assessment*. London: John Wiley & Sons. 338–354.
- Cambridge English (2015). *The Cambridge English Scale explained*. Cambridge: Cambridge English Language Assessment. Online: <https://www.cambridgeenglish.org/pl/Images/177867-the-methodology-behind-the-cambridge-english-scale.pdf> [dostęp: 10.07.2020].
- Carr, N.T. (2011). *Designing and analyzing language tests*. Oxford – New York: Oxford University Press.
- Carras, C., Tolas, J., Kohler, P., Szilagyi, E. (2007). *Le Français sur objectifs spécifiques et la classe de langue*. Paris: CLE International.
- Carroll, J.B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. W: H.B. Allen (red.), *Teaching English as a second language*. New York: McGraw Hill. 364–372.
- Çetin Koroğlu, Z. (2020). Assessing language learning in technology-enhanced environments. W: S. Meri Yilan, K. Koruyan (red.), *ICT-based assessment, methods, and programs in tertiary education*. Hershey: IGI Global. 45–64.
- Charpy, J.-P., Carnet, D. (2014). The European sTANDEM project for certification in Medical English: Standards, acceptability and transgression(s). *ILCEA* 19. Online: <http://ilcea.revues.org/2475> [dostęp 15.06.2020].
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge: University of Cambridge Local Examinations Syndicate – Cambridge University Press.
- Cookson, S. (2009). Zagreb and Tenerife. Airline accidents involving linguistic factors. *Australian Review of Applied Linguistics* 32(3). 22.1–22.14. Online:

- <https://benjamins.com/#catalog/journals/aral.32.3.02coo/fulltext> [dostęp: 25.07.2019].
- Coste, D., Courtyllon, J., Ferenczi, V., Martins-Baltar, M., Papo, E. (1976). *Un niveau-seuil*. Paris: Hatier.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Council of Europe Publishing. Online: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989> [dostęp: 15.06.2020].
- Davies, A. (2007). Assessing academic English language proficiency: 40+ years of U.K. language tests. W: J. Fox, M. Wesche, D. Bayliss, L. Cheng, C.E. Turner, C. Doe (red.), *Language testing reconsidered*. Ottawa: University of Ottawa Press. 73–86.
- Davies, A. (2008). *Assessing academic English: Testing English proficiency 1950–89. The IELTS Solution*. Cambridge: Cambridge University Press.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2001). Language for Specific Purposes assessment criteria: Where do they come from?. *Language Testing* 18(2). 171–185.
- Douglas, D. (2013). ESP and assessment. W: B. Paltridge, S. Starfield (red.), *The handbook of English for Specific Purposes*. Chichester: John Wiley & Sons. 368–383.
- Douglas, D. (2014). *Understanding language testing*. London – New York: Routledge.
- Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3). 347–360.
- Dudley-Evans, T., St John, M.J. (1998). *Developments in ESP. A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Dudzik, A., Dzięcioł-Pędich, A., Kic-Drgas, J. (2019). ESP testing in the eyes of Polish tertiary teachers. *Linguodidactica*, 23. 21–34.
- Elder, C. (2016). Exploring the limits of authenticity in LSP testing: The case of a specific-purpose language test for health professionals. *Language Testing*, 33(2). 147–152.
- Elder, C., McNamara, T. (2016). The hunt for “indigenous criteria” in assessing communication in the physiotherapy workplace. *Language Testing*, 33(2). 153–174.
- Elman, B.A. (2000) *A cultural history of civil examinations in late Imperial China*. Berkeley – Los Angeles – London: University of California Press.
- ETS (2010). *TOEFL iBT research insight series*. Vol. 1: *TOEFL iBT™ test framework and test development*. Online: https://www.ets.org/s/toefl/pdf/toefl_ibt_research_insight.pdf [dostęp: 6.06.2020].

- ETS (2011). *TOEFL iBT research insight series*. Vol. 6: *TOEFL® program history*.
Online: https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v6.pdf [dostęp: 6.06.2020].
- Feng, J. (1994). From the Imperial Examination to the National College entrance examination: The dynamics of political centralism in China's educational enterprise. *ASHE Annual Meeting Paper*. Online: <http://files.eric.ed.gov/fulltext/ED375731.pdf> [dostęp: 4.02.2020].
- Fischer, J., Chouissa, C., Dugovičová, S., Virkkunen-Fullenwider, A. (2011). *Guidelines for task-based university language testing*. Graz: European Centre for Modern Languages / Council of Europe Publishing.
- Flowerdew, L. (2013). Needs analysis and curriculum development in ESP. W: B. Paltridge, S. Starfield (red.), *The handbook of English for Specific Purposes*. Chichester: Wiley-Blackwell. 325–346.
- Frendo, E. (2005). *How to teach Business English*. Harlow: Pearson Education Limited.
- Fulcher, G. (1993). *The construction and validation of rating scales for oral tests in English as a foreign language* [PhD thesis]. University of Lancaster. Online: <http://languagetesting.info/articles/store/FulcherPhD.pdf> [dostęp: 7.06.2020].
- Gajewska, E. (2008). Kształcenie nauczycieli języków specjalistycznych. W: A.M. Harbig (red.), *Nauczanie języków obcych w szkole wyższej*. Białystok: WUwB. 75–84.
- Gajewska, E. (2011). Poziomy biegłości a nauczanie języków specjalistycznych. W: S. Piotrowski (red.), *O nauczaniu i uczeniu się języka obcego dla potrzeb zawodowych*. Lublin: Wydawnictwo Werset. 10–21.
- Gajewska, E. (2013). *Courriel vs courrier: la communication écrite en français de la communication professionnelle au temps des nouvelles technologies*. Lublin: Werset.
- Gajewska E., Sendur A.M. (2015). Ocenianie w nauczaniu języków obcych zorientowanym na cele ogólne a ocenianie w nauczaniu zorientowanym na cele zawodowe. *Języki Obce w Szkole*, 3. 49–53.
- Gajewska, E., Sowa, M. (2014). *LSP, FOS, Fachsprache... Dydaktyka języków specjalistycznych*. Lublin: Wydawnictwo Werset.
- Gaszyńska-Magiera, M., Seretny, A. (tłum. i adapt.) (2004). *Przewodnik dla autorów zadań do testów językowych*. Kraków: TAIWPN Universitas.
- Gębal, P. (2016). Od dydaktyki tekstów specjalistycznych do dydaktyzacji potrzeb i działań komunikacyjnych. Kierunki i koncepcje glottodydaktyki specjalistycznej. *Acta Universitatis Lodzianis. Kształcenie Polonistyczne Cudzoziemców*, 23. 19–34.
- Gębal, P.E. (2019). *Dydaktyka języków obcych. Wprowadzenie*. Warszawa: WN PWN.
- Gos, M. (2009). English for Specific Purposes in NATO armed forces. W: S. Szadyko (red.), *Komunikacja specjalistyczna. T. 2: Specyfika języków specjalistycznych*. Warszawa: KJS UW. 122–130.

- Green, A.B. (2012). *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.
- Grucza, F. (2002). Języki specjalistyczne – indykatory i/lub determinanty rozwoju cywilizacyjnego. W: J. Lewandowski (red.), *Języki specjalistyczne 2: Problemy technolingwistyki*. Warszawa: KJS UW. 9–26.
- Grucza, S. (2004). *Od lingwistyki tekstu do lingwistyki tekstu specjalistycznego*. Warszawa: KJS UW.
- Grucza, S. (2013). *Lingwistyka języków specjalistycznych*. Warszawa: Wydawnictwo Naukowe Instytutu Kulturologii i Lingwistyki Antropocentrycznej. Uniwersytet Warszawski.
- Hamp-Lyons, L., Lumley, T. (2001). Assessing Language for Specific Purposes. *Language Testing*, 18(2). 127–132.
- Hawkey, R. (2009). *Examining FCE and CAE: Key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams*. Cambridge: UCLES/Cambridge University Press.
- Heaton, J.B. (1988). *Writing English language tests: a practical guide for teachers*. London: Longman.
- Herzog, M. (b.d.). *An overview of the history of the ILR language proficiency skill level descriptions and scale*. Online: <http://www.govtilr.org/skills/> [dostęp: 6.09.2015].
- Hoffman, L. (1979). Towards a theory of LSP. Elements of a methodology of LSP analysis. *Fachsprache*, 1(2). 12–17.
- Howatt, A.P.R. (1997). *A History of language teaching*. Oxford: Oxford University Press.
- Howatt, A.P.R., Smith, R. (2014). The history of teaching English as a foreign language, from a British and European perspective. *Language and History*, 57(1). 75–95.
- Hughes, A. (1995). *Testing for language teachers*. Cambridge – New York: Cambridge University Press.
- Hunt, T.C. (1976). Assessment for temporary registration. *Proceedings of the Royal Society of Medicine*, 69(11). 812–823. Online: <http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1864725&blobtype=pdf> [dostęp: 20.06.2020].
- Hunter, N. (2010). *1,000 docs let in without language test*. Irishhealth.com. Online: <http://www.irishhealth.com/article.html?id=17983> [dostęp: 10.06.2020].
- Hutchinson, T., Waters, A. (1991). *English for Specific Purposes. A learning-centred approach*. Cambridge: Cambridge University Press.
- Hüttner, J., Mehlmauer-Larcher, B., Reichl, S., Schiftner, B. (2011). *Theory and practice in EFL teacher education: Bridging the gap*. Bristol: Multilingual Matters.
- Hymes, D.H. (1966). Two types of linguistic relativity. W: W. Bright (red.), *Sociolinguistics*. Haga: Mouton. 114–158.
- Hymes, D.H. (1972). *On communicative competence*. W: J.B. Pride, J. Holmes (red.), *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin. 269–293.

- International Civil Aviation Organization (2010). *Manual on the implementation of ICAO language proficiency requirements*. Quebec, Canada: ICAO. Online: <https://skybrary.aero/bookshelf/books/2497.pdf> [dostęp: 25.06.2020].
- Isbell, D.R., Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 4(37). 600–619. Online: <https://journals.sagepub.com/doi/full/10.1177/0265532220943483> [dostęp: 24.10.2020].
- Jacoby, S., McNamara, T. (1999). Locating competence. *English for Specific Purposes*, 18(3). 213–241.
- Jang, E.E. (2014). *Focus on assessment*. Oxford: Oxford University Press.
- Janowska, I. (2011). *Podejście zadaniowe do nauczania i uczenia się języków obcych. Na przykładzie języka polskiego jako obcego*. Kraków: TAIWPN Universitas.
- Jaroszewska, A. (2014). O glottodydaktyce słowami glottodydaktyków. *Języki Obce w Szkole*, 4. 52–66.
- Jasso-Aguilar, R. (2005). Sources, methods and triangulation in needs analysis: A critical perspective in a case study of Waikiki hotel maids. W: M.H. Long (red.), *Second language needs analysis*. Cambridge: Cambridge University Press. 127–158.
- Kaliska M. (2016). Rozwój kompetencji interkulturowej a perspektywa działania w nauczaniu języków obcych. W: W. Stec, A. Hau (red.), *W poszukiwaniu tożsamości językowej*, t. 2. Gdańsk: WUG.
- Kaliska, M. (2018). *Model uczenia języków obcych w szkole wyższej na przykładzie języka włoskiego. Założenia teoretyczne, metodologia nauczania i zintegrowany rozwój kompetencji*. Warszawa: WN IKSUW.
- Khan, M., Aljarallah, B.M. (2011). Evaluation of modified essay questions (MEQ) and multiple choice questions (MCQ) as a tool for assessing the cognitive skills of undergraduate medical students. W: *International Journal of Health Sciences*, 5(1). Online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312767/> [dostęp 7.06.2020].
- Komorowska, H. (1982). *Metody badań empirycznych w glottodydaktyce*. Warszawa: PWN.
- Komorowska, H. (1984). *Testy w nauczaniu języków obcych*. Warszawa: WSiP.
- Komorowska, H. (2005a). *Metodyka nauczania języków obcych*. Warszawa: Fraszka Edukacyjna.
- Komorowska, H. (2005b). Testowanie językowych osiągnięć uczniów – trudności i zagrożenia. *Języki Obce w Szkole*, 6. 41–65.
- Komorowska, H. (2007). *Sprawdzanie umiejętności w nauce języka obcego. Kontrola – Ocena – Testowanie*. Warszawa: Fraszka Edukacyjna.
- Komorowska, H. (2016). Zmienne losy innowacji edukacyjnych na przykładzie CEFR. W: D. Gabrys-Barker, R. Kalamarz (red.), *Ocenianie i pomiar biegłości językowej. Wybrane aspekty teoretyczne i praktyczne*. Katowice: WUŚ. 13–24.
- Komorowska, H. (2019). Ocenianie w nauczaniu języków obcych. Fakty, mity i trudności. *Neofilolog*, 52(2). 153–170.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests: A teacher's book*. London: Longman.

- Monaghan, R. (2012). Language and interoperability in NATO: The Bureau for International Language Co-ordination (BILC). *Canadian Military Journal*, 13(1). 23–31.
- Mourlhon-Dalies, F. (2008). *Enseigner une langue à des fins professionnelles*. Paris: Didier.
- Niemierko, B. (1975). *ABC testów osiągnięć szkolnych*. Warszawa: WSiP.
- Niemierko, B. (1997). *Pomiar wyników kształcenia zawodowego*. Warszawa: Biuro Koordynacji Kształcenia Kadr. Fundusz Współpracy.
- Niemierko, B. (1999). *Pomiar wyników kształcenia*. Warszawa: WSiP.
- Niemierko, B. (2002). *Ocenianie szkolne bez tajemnic*. Warszawa: WSiP.
- Niemierko, B. (2006). Znaczenie edukacyjne egzaminu doniosłego. *Teorie i badania. Meritum*, 3. 4–11.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, B. (2006). *Issues in testing Business English: The revision of the Cambridge Business English Certificates*. Cambridge: UCLES/Cambridge University Press.
- O'Sullivan, B. (2011). Language testing. W: J. Simpson (red.), *Routledge handbook of applied linguistics*. Oxford: Routledge. 259–273.
- OET (2009). *An overview of the Occupational English Test*. Melbourne: University of Melbourne. Online: https://www.google.pl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0ahUKEwjR4NHbh-DPAhXINpoKHQf_B9kQFgg4MAM&url=https%3A%2F%2Fwww.ahpra.gov.au%2Fdocuments%2Fdefault.aspx%3Frecord%3DWD10%252F4003%26dbid%3DAP%26chksum%3DaxNx1hDhgzMvaMcq8mAtZA%253D%253D&usq=AFQjCNGp7c49sS-gbTlg-Di7I-cLklhoAA [dostęp: 16.06.2020].
- OET (2014). *The impact and relevance of the Occupational English Test (OET) for the medical and nursing workplace* [plakat przedstawiony na konferencji IAMRA w Londynie w 2014 przez Ivanę Vidakovic]. Cambridge English Language Assessment.
- Okoń, W. (2007). *Nowy słownik pedagogiczny*. Wyd. 10 uzup. i popr. Warszawa: Wydawnictwo Akademickie „Zak”.
- Omaggio, A.C. (1986). *Teaching language in context. Proficiency-oriented instruction*. Boston: Heinle & Heinle.
- Pearson (2015a). *Global Scale of English learning objectives for Academic English*. (b.m.): Pearson Education Limited.
- Pearson (2015b). *Global Scale of English learning objectives for adult learners*. (b.m.): Pearson Education Limited.
- Pearson (2015c). *Global Scale of English learning objectives for professional English*. (b.m.): Pearson Education Limited.
- Pearson (2015d). *Global Scale of English learning objectives for young learners*. (b.m.): Pearson Education Limited.
- Pearson (b.r.). *Internationally recognised English language qualifications*. Online: <http://pearson.com/wp-content/uploads/2014/07/English-Language-Qualifications-web1.pdf> [dostęp: 4.12.2015].

- Pfeiffer, W. (2001). *Nauka języków obcych. Od praktyki do praktyki*. Poznań: Wagros.
- Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing*, 33(2), 175–193.
- Pomykała, W. (red.) (1993). *Encyklopedia pedagogiczna*. Warszawa: Fundacja Innowacja.
- Poszytek, P. (2005a). Historia testowania biegłości językowej a teorie uczenia się języków i metod nauczania. *Języki Obce w Szkole*, 6, 23–29.
- Poszytek, P. (2005b). Cechy testów. *Języki Obce w Szkole*, 6, 29–33.
- Poszytek, P. (2008). *European standards of testing language proficiency and the Polish „Nowa Matura” examinations in English*. Kraków: TAIWPN Universitas.
- Pulliam, R., Ich, V.T. (1968). The Defense Language Proficiency Tests: Background, present programs, and future plans. W: *Proceedings of the 10th Annual Conference of the Military Testing Association*. 69–82. Online: <https://www.yumpu.com/en/document/view/9505268/proceedings-international-military-testing-association/9> [dostęp: 7.07.2020].
- Rada Europy (2003). *Europejski system opisu kształcenia językowego: uczenie się, nauczanie, ocenianie*. Warszawa: WCODN.
- Read, J. (2015). *Assessing English proficiency for university study*. Basingstoke: Palgrave Macmillan.
- Read, J., Knoch, U. (2009). Clearing the air: Applied linguistic perspectives on aviation communication. *Australian Review of Applied Linguistics*, 32(3), 21.1–21.11. Online: <http://arrow.monash.edu.au/hdl/1959.1/787566> [dostęp: 7.07.2020].
- Roach, J. (1971). *Public examinations in England, 1850–1900*. Cambridge: Cambridge University Press.
- Sendur, A.M. (2015). Egzaminacje certyfikujące z języków obcych ogólnych i akademickich oraz języków obcych dla potrzeb zawodowych – porównanie sprawdzanych umiejętności, sprawności i kryteriów oceny. W: M. Srebro, E. Typek, L. Zielińska (red.), *Przyszłość nauczania języków obcych na uczelniach wyższych*. Kraków: Fundacja Uniwersytetu Ekonomicznego w Krakowie. 237–252.
- Sendur, A.M. (2016). Realizacja wymagań dotyczących kształcenia językowego na wyższych uczelniach na podstawie analizy wybranych przykładów. W: D. Gabryś-Barker, R. Kalamarz (red.), *Ocenianie i pomiar biegłości językowej. Wybrane aspekty teoretyczne i praktyczne*. Katowice: WUŚ. 167–179.
- Sendur, A.M. (2017). Language needs analysis as the first step in designing an LSP test for police officers. W: *Pixel. Conference Proceedings. 10th Conference Edition*. Florencia: LibreriaUniversitaria. 284–287.
- Sendur, A.M. (2020). Certyfikacja znajomości języków obcych dla celów specjalistycznych – między teorią a rzeczywistością. *Neofilolog*, 54/1, 47–69.
- Sendur, A.M. (red.) (2016a). *English exam practice. Próbné egzaminacje z języka angielskiego dla studentów Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego*. Kraków: Oficyna Wydawnicza AFM.

- Sinaiko, H.W. (1966). *Foreign language training: An investigation of research and development for Vietnam Institute for Defense Analyses*. Arlington, Va: Research and Engineering Support Division.
- Smith, M., Stansfield C.W. (2016). Testing aptitude for second language learning. W: E. Shohamy, I. Or, S. May (red.), *Language testing and assessment. Encyclopedia of language and education (3rd ed.)*. Cham: Springer. 77–90.
- Smith, R. (2010). NHS confused over language testing of foreign doctors. *The Telegraph*, 4th February. Online: <http://www.telegraph.co.uk/news/health/news/7157882/NHS-confused-over-language-testing-of-foreign-doctors.html> [dostęp: 10.06.2020].
- Sowa, M. (2010). Podejście zadaniowe a kształtowanie kompetencji ucznia w nauce języka obcego dla potrzeb zawodowych. *Roczniki Humanistyczne* 58, 5. 89–105.
- Sowa, M. (2012). Kształtowanie się koncepcji nauczania języka francuskiego dla potrzeb zawodowych. *Lingwistyka Stosowana / Applied Linguistics / Angewandte Linguistik* 6. 155–170.
- Sowa, M. (2015). *Savoir-évaluer*, czyli co i jak oceniać w języku obcym. *Języki Obce w Szkole*, 3. 43–48.
- Sowa, M. (2016). Nomenklatura na usługach glottodydaktyki. Koncepcje nauczania języka francuskiego dla potrzeb zawodowych. *Języki Obce w Szkole*, 3. 4–11.
- Spolsky, B. (1975). *Language testing: Art or science?*. Referat zaprezentowany podczas IV Światowego Kongresu AILA w Stuttgarcie.
- Spolsky, B. (1993). *Testing the English of foreign students in 1930*. Referat zaprezentowany podczas dorocznego Language Testing Research Colloquium w Cambridge.
- Spolsky, B. (1995). *Measured words. The development of objective language testing*. Oxford: Oxford University Press.
- Spolsky, B. (2007). On second thoughts. W: J. Fox, M. Wesche, D. Bayliss, L. Cheng, C.E. Turner (red.), *Language testing reconsidered*. Ottawa: University of Ottawa Press. 9–18.
- Spolsky, B. (2017). History of language testing. W: E. Shohamy et al. (red.), *Language testing and assessment. Encyclopedia of language and education*. 375–384.
- Sujecka-Zajac, J., Sobolew, L. (2015). Ramy teoretyczne i praktyczne ewaluacji biegłości językowej. Przykład egzaminów certyfikacyjnych z języków obcych w UW. W: J. Sujecka-Zajac (red.), *Ewaluacja biegłości językowej. Od pomiaru do sztuki pomiaru*. Warszawa: Rada Koordynacyjna ds. Certyfikacji Biegłości Językowej UW. 11–31.
- Szerszeń, P. (2014). *Platformy (glotto)dydaktyczne. Ich implementacja w uczeniu specjalistycznych języków obcych*. Warszawa: WN IKiLAUW.
- Taylor, L. (2013). *Assessing health professionals* [prezentacja PowerPoint]. Online: https://www.beds.ac.uk/__data/assets/pdf_file/0018/250137/Lynda-Taylor.pdf [dostęp 12.11.2016].

- Taylor, L., Pill, J. (2014). Assessing health professionals. W: J. Kunnan (red.), *The companion to language assessment*. London: John Wiley & Sons. 496–511.
- Trim, J. (2011). *The Origin of the CEFR* [wywiad z Johnem Trimem nagrany w maju 2011 w Cambridge z okazji 10-lecia publikacji ESOKJ]. Online: <https://www.youtube.com/watch?v=mSJNGHDK6lM> [dostęp: 20.06.2020].
- Urbanikowa, J. (2011). Europejska polityka językowa i jej implikacje dla polskich szkół wyższych na przykładzie Uniwersytetu Warszawskiego. W: W. Martyniuk (red.), *Internacjonalizacja studiów wyższych*. Warszawa: FRSE. 93–124.
- van Ek, J.A., Alexander, L.G. (1975). *Systems development in adult language learning: The threshold level in a European unit/credit system for modern language learning by adults*. Strasbourg: Council for Cultural Co-Operation of the Council of Europe.
- Vidaković, I. (2015). *Language ability and content knowledge in an ESP classroom: Two sides of the coin?* [referat wygłoszony na konferencji IATEFL – BEASIG and TEASIG]. Manchester.
- Weir, C.J. (1990). *Communicative language testing*. Hemel Hempstead: Prentice-Hall International.
- Weir, C.J. (1993). *Understanding and developing language tests*. Hemel Hempstead: Prentice-Hall International.
- Weir, C.J. (2002). Innovation and continuity: CPE – past and present. *Cambridge English Research Notes*, 10. 2–6. Online: <http://www.cambridgeenglish.org/images/23124-research-notes-10.pdf> [dostęp: 8.06.2020].
- Weir, C.J. (2003). A survey of the history of the certificate of proficiency in English (CPE) in the twentieth century. W: C.J. Weir, M. Milanovic (red.), *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*. Cambridge: UCLES/Cambridge University Press. 1–56. Online: <http://www.cambridgeenglish.org/images/23131-research-notes-15.pdf> [dostęp: 7.06.2015].
- Weir, C.J. (2005). *Language testing and validation. An evidence-based approach*. New York: Palgrave Macmillan.
- Weir, C.J. (2013a). Measured constructs: A history of Cambridge English language examinations 1913–2012. *Cambridge English Research Notes*, 51. 2–7. Online: www.cambridgeenglish.org/images/130828-research-notes-51-document.pdf [dostęp: 7.06.2020].
- Weir, C.J. (2013b). *Tests, testers and testing: The British Council and language testing 1941–2013* [wykład zaprezentowany w listopadzie 2013 w Pekinie podczas Sympozjum *New Directions English: Language Assessment*]. Online: <https://www.youtube.com/watch?v=Wps9IymFwXA> [dostęp: 30.06.2020].
- Weir, C.J., O’Sullivan, B. (2017). *Assessing English on the global stage. The British Council and English language testing, 1941–2016*. Sheffield: Equinox Publishing.
- Widdowson, H.G. (1984). *Explorations in applied linguistics 2*. Oxford: Oxford University Press.

- Wilbrink, B. (1997). Assessment in historical perspective. *Studies in Educational Evaluation*, 23. 31–48. Online: http://www.english.org/assessment/Manual/_docs/Wilbrink.pdf [dostęp: 4.11.2016].
- Yannakoudakis, H., Andersen, Ø.E., Geranpayeh, A., Briscoe, T., Nicholls, D. (2018). Developing an automated writing placement system for ESL learners. *Applied Measurement in Education*, 31(3). 251–267.
- Zawadowska-Kittel, E. (2011). Egzaminatory doniosłe w oczach badaczy polskich. *The Linguistic Academy Journal of Interdisciplinary Language Studies*, 1. 76–103.
- Zou B., Liviero S., Hao M., Wei C. (2020). Artificial intelligence technology for EAP speaking skills: Student perceptions of opportunities and challenges. W: M. Freiermuth, N. Zarrinabadi (red.), *Technology and the psychology of second language learners and users. New language learning and teaching environments*. Cham: Palgrave Macmillan.

Materiały i dokumenty egzaminacyjne oraz podręczniki

- Acklam, R., Crace, A. (2009). *Total English upper intermediate. Student's book*. Harlow: Pearson.
- British Council (2020). *Aptis. Candidate guide*. Online: www.britishcouncil.org/aptis [dostęp: 7.06.2020].
- Business English Certificate Higher (2015). *Sample paper. UCLES 2015 Cambridge English level 2 certificate in ESOL International (Business English)*. Cambridge: Cambridge English Language Assessment. Online: <http://www.cambridgeenglish.org/exams/business-certificates/business-higher/exam-format/> [dostęp 5.06.2020].
- Cambridge Assessment English (2017). *Computer-adaptive testing of listening and reading skills. A guide for educational institutions and employers*. Online: <https://www.cambridgeenglish.org/Images/421557-computer-adaptive-testing-of-listening-and-reading-skills.pdf> [dostęp: 20.06.2020].
- Cambridge Assessment English (2019a). *B2 First. Handbook for teachers*. Cambridge: Cambridge Assessment English. Online: https://www.cambridgeenglish.org/Images/CER_6168_V1_APR19_Cambridge_English_First_Handbook_WEB_v3.PDF [dostęp: 5.06.2020].
- Cambridge Assessment English (2019b) *Business Higher. Reading. Sample paper 1*. UCLES. Online: <https://www.cambridgeenglish.org/pl/exams-and-tests/qualifications/business/business-higher/exam-format/> [dostęp: 14.05.2020].
- Cambridge Assessment English (2019c). *Linguaskill. Simplify your English language testing. The accurate English test with fast results*. Online: <https://www.cambridgeenglish.org/Images/466596-linguaskill-simplify-your-english-language-testing.pdf> [dostęp: 20.06.2020].
- Cambridge English Language Assessment (2015a). *First Certificate in English. Reading and use of English. Sample test 1. D251/01*. Cambridge: UCLES.

- Online: <http://www.cambridgeenglish.org/exams/first/exam-format/> [dostęp: 20.06.2020].
- Cambridge English Language Assessment (2015b). *Cambridge English First. Information for candidates*. Cambridge: Cambridge English Language Assessment. Online: <http://www.cambridgeenglish.org/pl/images/181310-first-information-for-candidates-.pdf> [dostęp: 7.06.2020].
- Centralna Komisja Egzaminacyjna (2008). *Materiały pomocnicze dla kandydatów na egzaminatorów*. Warszawa: CKE.
- Centralna Komisja Egzaminacyjna (2013). *Egzamin w klasie trzeciej gimnazjum z języka angielskiego poziom poszerzony. Kwiecień 2013*. Online: <https://cke.gov.pl/files/file/Arkusze-2013/GA-R1-132.pdf> [dostęp: 19.09.2020].
- Centralna Komisja Egzaminacyjna (2017) *Informator o egzaminie ósmoklasisty z języka angielskiego od roku szkolnego 2018/2019*. Warszawa: Centralna Komisja Egzaminacyjna. Online: https://cke.gov.pl/images/_EGZAMIN_OSMOKLASISTY/Informatory/Informator_P1_angielski.pdf [dostęp: 14.05.2020].
- Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej (b.r.-a). *Egzamin z języka obcego wg STANAG 6001. Język angielski. Poziom: 3. Sprawność: Reading. L3 054*. Online: <http://www.wsnjo.wp.mil.pl/pl/49.html> [dostęp: 11.11.2016].
- Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej (b.r.-b). *Egzamin z języka obcego wg STANAG 6001. Język angielski. Poziom: 3. Sprawność: Reading. L3 057*. Online: <http://www.wsnjo.wp.mil.pl/pl/49.html> [dostęp: 11.11.2016].
- Centralna Komisja Egzaminacyjna Języków Obcych Ministerstwa Obrony Narodowej (2013). *Model egzaminu z języka angielskiego poziom 3 wg STANAG 6001*. Online: <http://www.wsnjo.wp.mil.pl/pl/49.html> [dostęp: 11.11.2016].
- Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne (2010). *Informator o egzaminie gimnazjalnym przeprowadzanym od roku szkolnego 2011/2012*. Warszawa: CKE. Online: https://www.cke.edu.pl/images/stories/0012_Gimnazjum/100826_Informator%20gimnazjalny.pdf [dostęp: 11.09.2016].
- Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne (2013). *Informator o egzaminie maturalnym z języka angielskiego od roku szkolnego 2014/2015*. Warszawa: CKE. Online: https://www.cke.edu.pl/images/_EGZAMIN_MATURALNY_OD_2015/Informatory/2015/Jezyk-angielski.pdf [dostęp: 11.09.2016].
- Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne (2015a). *Informator o egzaminie eksternistycznym przeprowadzanym od sesji jesiennej 2016 z zakresu szkoły podstawowej. Język angielski*. Warszawa: CKE. Online: https://www.cke.edu.pl/images/_EGZAMIN_EKSTERNISTYCZNY/Szkoła_podstawowa/Informatory/angielski_SP.pdf [dostęp: 7.10.2016].

- Centralna Komisja Egzaminacyjna. Okręgowe Komisje Egzaminacyjne (2015b). *Informator o egzaminie eksternistycznym przeprowadzanym od sesji jesiennej 2016 z zakresu liceum ogólnokształcącego. Język angielski*. Warszawa: CKE. Online: https://www.cke.edu.pl/images/_EGZAMIN_EKSTERNISTYCZ-NY/Liceum/O_egzaminie_od_2016/angielski_LO.pdf [dostęp: 7.10.2016].
- Clare, A., Wilson, J.J. (2006). *Total English intermediate. Student's book*. Harlow: Pearson.
- Clare, A., Wilson, J.J. (2011). *Speakout intermediate. Student's book*. Harlow: Pearson.
- Eales, F., Oakes, S. (2011). *Speakout Upper Intermedite. Student's book*. Harlow: Pearson.
- ELPAC (2013). *English Language Proficiency for Aeronautical Communication. Test Taker Preparation v3.0*. b.m.: ICAO. Online: <http://www.eurocontrol.int/sites/default/files/article/content/documents/atm-training/elpac-test-taker-preparation-v3.pdf> [dostęp: 20.02.2015].
- Eurocontrol (2013). *ELPAC English Language Proficiency for Aeronautical Communication. Test Taker Preparation v3.0*. [b.m.]: [b.w.]. Online: <http://www.eurocontrol.int/sites/default/files/article/content/documents/atm-training/elpac-test-taker-preparation-v3.pdf> [dostęp: 20.02.2015].
- IELTS (2019). *Information for candidates. Introducing IELTS to test takers*. ielts.org. Online: <https://www.ielts.org/-/media/publications/information-for-candidates/ielts-information-for-candidates-english-uk.ashx> [dostęp: 10.06.2020].
- LCCI International Qualifications (2008). *Level 2 Certificate in Written English for Tourism. Syllabus – Effective for examinations to be held from Series 2, 2010*. Online: <http://qualifications.pearson.com/content/dam/pdf/LCCI/Spoken-English-for-Tourism-WEFT/2010/specification-and-sample-assessments/WEFT-Syllabus-Level-2-April-2012.pdf> [dostęp: 4.12.2015].
- LCCI International Qualifications (2010a). *Level 2 Certificate in Written English for Tourism. Syllabus – Effective for examinations to be held from Series 2, 2010*. Online: <http://qualifications.pearson.com/content/dam/pdf/LCCI/Spoken-English-for-Tourism-WEFT/2010/specification-and-sample-assessments/WEFT-Syllabus-Level-2-April-2012.pdf> [dostęp: 4.12.2015].
- LCCI International Qualifications (2010b). *Written English for Tourism. Level 1. Series 4 Examination 2010*. Online: <http://qualifications.pearson.com/en/qualifications/lcci/english-language/written-english-for-tourism-weft.coursematerials.html#filterQuery=Pearson-UK:Category%2FPast-papers-and-model-answers> [dostęp: 4.12.2015].
- Ludlow, K. (2020) *Official quick guide to Linguaskill*. Cambridge: Cambridge University Press.

- Mondiale Testing (b.r.). *Introduction to the MONDIALE Online Technical English Test*. Online: http://www.sjo.pw.edu.pl/EGZAMINY/Dodatkowe/Angielski_Mondiale/mondiale_informator.pdf [dostęp: 23.02.2015].
- OET (2014). *Helping you select international healthcare professionals with the right level of English*. Melbourne: UCLES. Online: <http://www.admissionstestingservice.org/images/206878-oet-stakeholder-guide.pdf> [dostęp: 16.10.2016].
- PTE General (2018) *Guide to Pearson Test of English General. Level 3 (Upper Intermediate)*. Online: <https://qualifications.pearson.com/content/dam/pdf/pearson-test-of-english/pte-general-guides/guide-level-3.pdf> [dostęp: 7.06.2020].
- Sendur, A.M., Warecka, A. (2012). *English for safety, security and law enforcement*. Kraków: Oficyna Wydawnicza AFM.
- Telc Language Tests (2012a). *Telc B1–B2. Handbook. English B1–B2*. Frankfurt am Main: telc. Online: https://www.telc.net/fileadmin/user_upload/Handbooks/handbook_english_b1-b2.pdf [dostęp: 12.06.2020].
- Telc Language Tests (2012b). *Mock examination 1. English B1–B2*. Frankfurt am Main: telc. Online: http://www.telc.net/fileadmin/user_upload/telc_english_b1-b2.pdf [dostęp: 12.06.2020].
- Telc Language Tests (2014). *Telc B1–B2. Tips for test takers. English & English Business. Examination preparation B1–B2*. Frankfurt am Main: telc. Online: https://www.telc.net/fileadmin/user_upload/Tipps_fuer_Teilnehmer/tips_for_test_takers_b1-b2.pdf [dostęp: 12.06.2020].
- Telc Language Tests (2016) *Mock examination 1. English Technical. B2. 2016*. Frankfurt am Main: telc. Online: <https://www.telc.net/en/candidates/language-examinations/tests/detail/telc-english-b2-technical.html#t=1> [dostęp: 31.05.2020].
- Telc Language Tests (2019) *Mock examination 1. English B2. Revised edition 2019*. Frankfurt am Main: telc. Online: <https://www.telc.net/en/candidates/language-examinations/tests/detail/telc-english-b2.html#t=2> [dostęp: 31.05.2020].
- The sTANDEM Consortium (2013). *sTANDEM. A resource manual for candidates and stakeholders*. b.m.: sTANDEM.
- USMLE (2019). *Step 2 Clinical Skills (CS). Content description and general information*. [b.m.]: United States Medical Licensing Examination. Online: <http://www.usmle.org/pdfs/step-2-cs/cs-info-manual.pdf> [dostęp: 12.06.2020].

Akty prawne

Decyzja Nr 20 Prezesa Urzędu Lotnictwa Cywilnego z dn. 4 maja 2015 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla pilotów i kan-dydatów na pilotów w Krajowym Systemie Egzaminów Językowych (KSEJ).

- Online: https://ulc.gov.pl/_download/personel_lotniczy/komisja-egzaminacyjna/Decyzja_Nr_20_Prezesa_ULC_260816.pdf [dostęp: 17.06.2020].
- Decyzja Nr 21 Prezesa Urzędu Lotnictwa Cywilnego z dnia 4 maja 2015 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla kontrolerów ruchu lotniczego, praktykantów – kontrolerów ruchu lotniczego i kandydatów na praktykantów – kontrolerów ruchu lotniczego w Krajowym Systemie Egzaminów Językowych (KSEJ). Online: https://ulc.gov.pl/_download/personel_lotniczy/lkeang/ksej/dec_21_2015.pdf [dostęp: 17.06.2020].
- Decyzja Nr 40 Prezesa Urzędu Lotnictwa Cywilnego z dn. 12 listopada 2014 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla pilotów i kandydatów na pilotów. Online: http://www.ulc.gov.pl/_download/personel_lotniczy/lkeang/Materials/decyzja_nr_40_prezesa_ulc.pdf [dostęp: 20.02.2015].
- Decyzja Nr 46 Prezesa Urzędu Lotnictwa Cywilnego z dnia 28 listopada 2014 r. w sprawie sposobu organizacji egzaminu z języka angielskiego dla kontrolerów ruchu lotniczego, praktykantów – kontrolerów ruchu lotniczego i kandydatów na praktykantów – kontrolerów ruchu lotniczego. Online: http://www.ulc.gov.pl/_download/personel_lotniczy/lkeang/Materials/decyzja_nr46_prezesa_ulc.pdf [dostęp: 20.02.2015].
- Dyrektywa 2006/23/WE Parlamentu Europejskiego i Rady z dnia 5 kwietnia 2006 r. w sprawie wspólnotowej licencji kontrolera ruchu lotniczego, Online: http://eur-lex.europa.eu/legal-content/PL/TXT/?uri=urisrv:OJ.L_.2006.114.01.0022.01.POL [dostęp: 21.04.2020].
- European Aviation Safety Agency (2011). *Acceptable Means of Compliance and Guidance Material to Part-FCL. Annex to ED Decision 2011/016/R*. Online: <https://www.easa.europa.eu/sites/default/files/dfu/AMC%20and%20GM%20to%20Part-FCL.pdf> [dostęp: 21.09.2020].
- Ministerstwo Obrony Narodowej. Departament Nauki i Szkolnictwa Wojskowego (2009). *Ramowy program nauczania języka angielskiego w Siłach Zbrojnych RP. Edycja III. (Poziomy 1–3 wg STANAG 6001). Wprowadzony Decyzją Nr 28 Dyrektora Departamentu Nauki i Szkolnictwa Wojskowego MON z dn. 5 listopada 2009 r.* Online: http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/ramowy_program_nauczania_jezyka_angielskiego_2009.pdf [dostęp: 11.11.2016].
- Ministerstwo Obrony Narodowej. Departament Nauki i Szkolnictwa Wojskowego (2015). *Decyzja Nr 251/MON Ministra Obrony Narodowej z dnia 26 czerwca 2015 r. w sprawie kształcenia i egzaminowania ze znajomości języków obcych w resorcie obrony narodowej*. Online: http://www.dz.urz.mon.gov.pl/zasoby/dziennik/pozycje/tresc-aktow/pdf/2015/06/Poz._193_dec._Nr_251.pdf [dostęp: 11.11.2016].
- NATO Standardization Agency (2009). *Standardization Agreement. STANAG 6001 NTG (edition 3) – Language Proficiency Levels*. Online: http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/Porozumienie_standaryzacyjne_STANAG_6001_edycja_3_angielski.pdf [dostęp: 11.11.2016].

- Organizacja Paktu Północnoatlantyckiego, Wojskowa Agencja Standaryzacji (1976). *STANAG 6001. Porozumienie o standaryzacji (STANAG). Stopnie znajomości języków*. Online: <https://silo.tips/download/porozumienie-o-standaryzacji> [dostęp: 12.12.2020].
- Porozumienie Standaryzacyjne NATO nr 6001 – STANAG 6001. Online: http://wsnjo.wp.mil.pl/plik/file/Do_pobrania_-_Dokumenty_i_Programy/Porozumienie_standaryzacyjne_STANAG_6001_edycja1.pdf [dostęp: 11.11.2016].
- Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dn. 14 listopada 2018 r. w sprawie charakterystyk drugiego stopnia efektów uczenia się dla kwalifikacji na poziomach 6–8 Polskiej Ramy Kwalifikacji (Dz.U. z dn. 28 listopada 2018 r.). Online: <http://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20180002218/O/D20182218.pdf> [dostęp: 12.12.2020].
- Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 2 listopada 2011 r. w sprawie Krajowych Ram Kwalifikacji dla Szkolnictwa Wyższego (Dz.U. z dn. 22 listopada 2011 r.). Online: <http://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20112531520/O/D20111520.pdf> [dostęp: 12.12.2020].

Netografia¹⁰⁶

- British Council: www.britishcouncil.org.
- Cambridge English: www.cambridgeenglish.org.
- Cambridge English Polska: www.cambridgeenglish.org/pl.
- Campaign, English for the Military: www.campaignmilitaryenglish.com.
- Centralna Komisja Egzaminacyjna: www.cke.edu.pl.
- Educational Commission for Foreign Medical Graduates: www.ecfm.org.
- ELT News: www.eltnews.gr.
- EUROCONTROL: www.eurocontrol.int.
- Global Scale of English: <https://www.english.com/gse>.
- Glossary of Educational Reform for Journalists, Parents, and Community Members, The: www.edglossary.org.
- GULT. Guidelines for University Language Testing: gult.ecml.at.
- IELTS: www.ielts.org.
- IELTS Indicator: <https://www.ieltsindicator.com/>.
- IELTS Polska: <https://ieltspolska.pl/pl/skala-ielts>.
- Interagency Language Roundtable: www.govtilr.org.
- International Vocational English: www.interveng.com.
- Mondiale Testing: www.mondiale-testing.de.
- OET. Occupational English Test: www.occupationalenglishtest.org.
- Okręgowa Komisja Egzaminacyjna w Krakowie: www.oke.krakow.pl.
- Pearson: www.pearson.com.
- Pearson PTE Academic: <http://pearsonpte.com/>.
- Pearson Qualifications: <http://qualifications.pearson.com/en/home.html>.

¹⁰⁶ W bibliografii tekstów internetowych (netografia) zamieszczono adresy stron głównych wykorzystanych stron internetowych. Szczegółowe adresy tych stron umieszczone są w przypisach w tekście.

sTANDEM: www.standem.eu.

telc Language Tests: www.telc.net.

Urząd Lotnictwa Cywilnego: www.ulc.gov.pl.

Wojskowe Studium Nauczania Języków Obcych: www.wsnjo.wp.mil.pl.

Zintegrowany System Informacji o Nauce i Szkolnictwie Wyższym POL-on:
<http://polon.nauka.gov.pl/>.

Załączniki

Załącznik 1. Formularz analizy egzaminów GE	216
Załącznik 2. Formularz analizy egzaminów ESP	217
Załącznik 3. Zestawienie analizowanych egzaminów GE i ESP	218

Załącznik 1. Formularz analizy egzaminów GE

OPIS EGZAMINU												
EGZAMIN	FORMA		CZEŚĆ EGZAMINU sprawności językowe oraz podsystemy języka sprawdzone bezpośrednio						Integracja sprawności językowych	Zadania egzamina- cyjne w częściach C, S, G i L		Elementy egzaminu obowiązkowe / opcjonalne
	papie- rowa	kom- pute- rowa	Czytanie (C)	Rozumienie ze słuchu (S)	Pisanie (P)	Mówienie (M)	Znajomość struktur języ- kowych (G)	Znajomość leksyki ogól- nej (L)		Znajomość funkcji języ- kowych (F)	zamknięte	
Aptis General		✓	✓	✓	✓	✓	✓	✓		✓		obowiązkowy: G+L opcjonalne: C, S, P, M
Cambridge B2 First	✓		✓	✓	✓	✓	✓	✓		✓	✓	wszystkie obowiązkowe
IELTS General Training	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	wszystkie obowiązkowe
Linguaskill General		✓	✓	✓	✓	✓	✓	✓		✓	✓	opcjonalne: S+C, P, M
PTE General	✓		✓	✓	✓	✓	✓	✓	S+P (dyktando)	✓	✓	wszystkie obowiązkowe
telc English B2	✓		✓	✓	✓	✓	✓	✓		✓	✓	wszystkie obowiązkowe

Załącznik 3. Zestawienie analizowanych egzaminów GE i ESP

EGZAMIN	GE/ ESP	WERSJA P – papier. K – komp.	SKALA odniesienia poziomów biegłości	DOSTĘPNE POZIOMY	POZIOM analizowanego egzaminu	CZĘŚCI EGZAMINU	TYPY ZADAŃ	% lub liczba punktów	CZĘŚĆ EGZAMINU obowiązkowa / opcjonalna
Aptis General	GE	K	ESOKJ	test wielopoz. A1-C	cały arkusz	S – słuchanie C – czytanie M – mówienie P – pisanie G – gramatyka L – leksyka F – funkcje jęz.	WW – wyboru wielokrotnego D – dobieranie P/F – prawda-falsz P/F/BI – prawda-falsz-brak informacji L – z luką KO – krótkiej odpowiedzi RO – rozszerzonej odpowiedzi	50 50 50 50 50	OB OPC OPC OPC OPC
Cambridge B2 First	GE	P/K	Cambridge English Scale/ ESOKJ	B2	B2	S C+G+L P M	pytanie-odpowiedź, wypowiedź oparta na materiałach stymulujących D, WW, L D, WW, L, KO RO	20% 40% 20% 20%	OB OB OB OB
IELTS General Training	GE	P/K	IELTS band scores/ ESOKJ	test wielopoz. 1-9 (A1-C2)	cały arkusz	S C P M	wywiad, rozmowa na podst. materiału stymulującego, dyskusja WW, D, L WW, D, P/F/BI, L, KO RO	25% 25% 25% 25%	OB OB OB OB
Linguaskill General	GE	K	Cambridge English Scale/ ESOKJ	test wielopoz. A1-C	adaptacyjny test wielopoz. A1-C	S C P M	WW WW, D, L RO wypowiedzi monologowe, głośne czytanie, wypowiedź monologowa na podst. materiału stymulującego	100% 100% 100%	OPC OPC OPC

MOTET	ESP	K	ESOKJ	test wielopoz. B1-C1	cały arkusz	General Techni- cal Module	G+L, spec.	WW	50	100%	OB		
							S	WW				10	OB
							C	WW				10	OB
OET	ESP	P**	brak	jeden poziom, brak odnie- sienia do skali	jeden poziom, brak odnie- sienia do skali	P	RO	100%	100%	OPC			
							M	wypowiedź monologowa, dyskusja			OPC		
							S	WW, D, KO			ocena na skali 0-500	OB	
STANAG 6001	ESP	P	tabela stopni znajomości języków STANAG 6001	poz. 1-5	poziom 3 (B2-B2+)	M	WW, LKO	25%	25%	OB			
							C	WW			25%	OB	
							P	RO			25%	OB	
sTANDEM* ESP	ESP	P	destr. Re komend. dla sTANDEM (na bazie ESOKJ)	B1, B2, C1	B2	M	wypowiedź monologowa, dyskusja/wywiad	25%	25%	OB			
							S	WW, D, L			25	OPC	
							C	WW, D,			25	OPC	
							RO	25			OPC		
							wywiad, wypowiedź monologowa, odgrywanie ról	25			OPC		

telc English B2 Technical	ESP	P	ESOKJ	B2	B2	S	WW, L, PF	40	OB
						C	WW, PF	40	OB
						P	RO	10	OB
						G+L.spec.	WW	10	OB
						M	wypowiedź monologowa, wywiad, reagowanie językowe, prezentacja na podst. materiału stymulującego	25	OB
VEC						L.spec.	WW	36	OB
						F	WW	16	OB
	ESP	P	brak	brak	brak	S	WW	20	OB
						C	WW, P/F/BI	18	OB
						P	D lub WW	10	OB

* Egzamin wycofany z oferty lub czasowo zawieszony. ** Stan na lipiec 2020: OET jest w trakcie przygotowania wersji komputerowej: *OET@home* (<https://www.occupationalenglishtest.org/at-home-remote-proctored-testing/> [dostęp: 10.07.2020].

Praca ta stanowi cenne kompendium wiedzy na temat certyfikacji w zakresie języka angielskiego zarówno tej o charakterze ogólnym, jak i specjalistycznym (LSP), i może okazać się niezwykle przydatna dla glottodydaktyków, nauczycieli, studentów przygotowujących się do wykonywania zawodu nauczyciela języka angielskiego oraz osób zamierzających przystąpić do egzaminu certyfikującego.

Z recenzji dr hab. Marty Kaliskiej

ISBN 978-83-66007-62-8



9 788366 007628

OFICyna
WYDAWNICZA

