

Anna Seretny¹, Ewa Lipińska²

TRAFNOŚĆ I RZETELNOŚĆ TESTU JAKO MIARY JEGO POPRAWNOŚCI, CZYLI O BŁĘDACH W KONSTRUOWANIU JĘZYKOWYCH TESTÓW OSIĄGNIĘĆ

Słowa kluczowe: test osiągnięć, funkcja pomiaru dydaktycznego, rzetelność, trafność, czynniki wpływające na trafność/rzetelność testu, błędy konstrukcyjne

Streszczenie. Narzędziem pomiaru najczęściej stosowanym w dydaktyce języka jest test osiągnięć. Za jego pomocą określa się stopień opanowania przez uczących się określonego materiału i/lub wielkość przyrostu ich umiejętności językowych. Błędy popełniane na etapie konstruowania testu osiągnięć w zakresie doboru treści podlegających sprawdzeniu, zadań testowych oraz zastosowanych kryteriów poprawy mogą fałszować rzeczywisty obraz umiejętności zdających, niesłusznie je przeceniając lub dostatecznie ich nie dowartościowując. W niniejszym tekście analizie poddaje się czynniki wpływające na trafność oraz rzetelność testu osiągnięć, które są podstawowymi wyznacznikami jego poprawności.

WSTĘP

Test językowy najczęściej definiuje się jako procedurę dokonywania ewaluacji określonych aspektów umiejętności językowych, która ma umożliwić uczącym się

wykazanie się wiedzą lub sprawnościami w sposób na tyle zwięzły, aby można było ocenić wyniki i z poziomu wykonania zadań wyciągnąć wnioski o ogólnym poziomie działania językowego, jakiego możemy oczekiwać od nich bądź obecnie, bądź też w przyszłości.

(Ingram 1983, s. 225)

¹ aseretny@poczta.onet.pl; Centrum Języka i Kultury Polskiej w Świecie, Wydział Polonistyki, Uniwersytet Jagielloński, ul. Grodzka 64, 31-044 Kraków.

² ewa.lipinska@uj.edu.pl; Centrum Języka i Kultury Polskiej w Świecie, Wydział Polonistyki, Uniwersytet Jagielloński, ul. Grodzka 64, 31-044 Kraków.

Narzędziem pomiaru wiedzy i umiejętności uczących się najczęściej stosowanym w dydaktyce języków obcych, w tym także polszczyzny, są testy osiągnięć, nazywane inaczej sprawdzianami (testami) nauczycielskimi. W przeciwieństwie do testów biegłości są one całkowicie zależne od programu kształcenia, a co za tym idzie – oparte na podręczniku (zob. Seretny, Lipińska 2011). Warunkuje to ich dwojaki cel: pierwszym jest określenie, którą część materiału, będącego przedmiotem nauczania w czasie bezpośrednio poprzedzającym pomiar, a ujętego w założonym programie, uczący się już opanowali, drugim – wskazanie stopnia, w jakim treści te zostały przez nich przyswojone.

Dla wielu uczących opracowanie poprawnego testu stanowi niemały problem (zob. Wysocka 2009). Sięgają więc chętnie po gotowe sprawdziany (czasem towarzyszące podręcznikom). Nie jest to jednak rozwiązanie właściwe, gdyż po pierwsze obejmują one wyłącznie materiał zawarty w książce, pomijając inne, omawiane w czasie lekcji treści dydaktyczne, po drugie – udostępnione uczącym się (wraz z odpowiedziami) zostają przekazane kolejnym rocznikom (Dziuba 2000, s. 15–16). Kłóci się to z istotą kontroli wyników, dlatego każdy nauczyciel powinien samodzielnie opracowywać testy osiągnięć, dbając o *trafność*, *rzetelność* i *praktyczność* dokonywanego pomiaru³.

1. TRAFNOŚĆ I RZETELNOŚĆ NARZĘDZIA POMIARU

Trafność jest to **stopień dokładności**, z jaką określone narzędzie (tu: test) mierzy to, co miało mierzyć (Niemierko 1975, s. 169). Najczęściej dzieli się ją na wewnętrzną oraz zewnętrzną. W ramach trafności wewnętrznej bada się, m.in., zgodność:

- treści testu z treściami nauczania – trafność treści (ang. *content validity*),
- sposobów pojmowania biegłości językowej ze sposobami jej sprawdzania – trafność teoretyczna (ang. *construct validity*).

W obrębie trafności zewnętrznej natomiast podejmuje się próby ustalenia, czy i w jakim stopniu test właściwie ocenia umiejętności językowe uczących się w porównaniu z innymi pomiarami, równoczesnymi (trafność zbieżna, diagnostyczna; ang. *concurrent validity*) lub przesuniętymi w czasie (trafność prognozy; ang. *predictive validity*).

Rzetelność definiuje się jako jednorodność, stałość wyników. Chodzi zatem o **konsekwencję**, z jaką test mierzy to, co miał mierzyć (Niemierko 1975, s. 220). W ustalaniu jej poziomu pomocne są różne metody, np.: retestowa, połówkowa, wersji równoległych, przy czym im mniej rozbieżności pojawia się między ko-

³ Trafność i rzetelność, jako kwestie merytoryczne, omawiamy szeroko poniżej; o kwestiach praktycznych wspominamy na końcu tekstu.

lejno uzyskiwanymi wynikami, tym większa wiarygodność stosowanego testu. Uczący się, dwukrotnie rozwiązując test (ten sam lub jego wersję równoległą), powinien więc uzyskać podobny wynik (w przypadku, gdy składał się zadań zamkniętych)⁴ lub zostać oceniony w podobny sposób (gdy jego zadaniem było udzielenie odpowiedzi na zadania otwarte). Rzetelność dotyczy bowiem zarówno stałości wyników testu, jak i stałości oceny⁵. Trzeba pamiętać, że dbałość o rzetelność egzaminu, to także przejaw troski o jego trafność, ponieważ obie te cechy są od siebie zależne.

2. TEST OSIĄGNIĘĆ – JEGO SPECYFIKA I ZNACZENIE W PROCESIE KSZTAŁCENIA JĘZYKOWEGO

Testy osiągnięć stanowią integralną część toku nauczania/uczenia się. Nie należą do grupy egzaminów wysokiej stawki⁶ (ang. *high-stakes exams*), którymi są wszelkiego rodzaju testy biegłości językowej⁷. Są zawsze ukierunkowane na przeszłość, czyli na okres bezpośrednio poprzedzający kontrolę, przy czym wykorzystuje się je zarówno w celu ewaluacji *kształtującej* (bieżącej, odnoszącej się do ostatniego tygodnia, poprzedniej lekcji), jak i *sumującej* (na zakończenie dłuższego okresu nauki, np. semestru, roku).

Jedną ze specyficznych cech testu osiągnięć jest to, że przygotowuje się go zawsze dla konkretnej grupy uczniów pracujących przez ustalony czas z określoną intensywnością pod kierunkiem nauczyciela, który realizuje konkretny program, korzystając przy tym z materiałów dobranych do potrzeb, dostosowanych do możliwości uczących się. Stąd w klasach/grupach na pozór bardzo podobnych, składających się z uczniów o tym samym języku pierwszym, jednolitych pod względem wieku i stopnia zaawansowania, pomiar może wyglądać nieco inaczej,

⁴ W praktyce, zamiast ponownego testowania uczniów z tego samego materiału w identyczny sposób, stosuje się rozmaite techniki szacowania rzetelności (zob. Niemierko 1999, s. 196–201).

⁵ Podstawowe zagadnienia związane z trafnością i rzetelnością testu w przystępny sposób omawia Komorowska (2002, s. 22–29), bardziej szczegółowo ujmują je Niemierko (1975, 1999), Messic (1989), Weir (2005), Kane (2006), Fulcher, Davidson (2007). Tam też znaleźć można informacje dotyczące sposobów badania trafności i rzetelności pomiaru. Pierwsze wymagają zastosowania analiz jakościowych, drugie – opierają się głównie na procedurach ilościowych (najprostsze z nich to określanie stopnia trudności/łatwości poszczególnych zadań za pomocą **wskaźników trudności/łatwości** i wskazywanie ich **mocy różnicującej** przy wykorzystaniu, np. metody „jednych trzecich” (zob. Pukas-Palimąka, Lipińska 1997; Gałyga 2004).

⁶ Od wyników niektórych egzaminów osiągnięć jednak sporo może zależeć, na przykład stopień na świadectwie umożliwiający lub nie promocję do następnej klasy, dyplom szkoły językowej z oceną, która decyduje o przyjęciu do pracy itp.

⁷ To np. testy plasujące (zob. Lipińska, Seretny 2010) czy testy biegłości. W przeciwieństwie do testów osiągnięć są sformalizowane i wystandaryzowane, ukierunkowane na terażniejszość.

przy czym im większe zróżnicowanie uczących się, tym wyraźniejsze mogą być te rozbieżności. W testowaniu osiągnięć sprawą niezwykle istotną jest bowiem waga zagadnień podlegających kontroli, wyrażona w punktach za nie przyznawanych. Ich liczba powinna proporcjonalnie odpowiadać czasowi, jaki przeznaczono na realizację określonych kwestii. Innymi słowy, zadania sprawdzające stopień opanowania treści/umiejętności, na których koncentrowało się nauczanie, powinny być wyżej punktowane niż te ćwiczone w mniejszym zakresie. Owo ukierunkowanie na wiadomego odbiorcę powoduje, że testy osiągnięć są mało uniwersalne, zazwyczaj wręcz jednorazowe o formule często umykającej standaryzacji. Mogą one bowiem mieć bardzo zróżnicowaną objętość, czyli być mniej lub bardziej obszerne (kartkówka vs. egzamin semestralny), sprawdzać stopień opanowania aspektów biegłości językowej selektywnie lub kompleksowo, tzn. koncentrować się na badaniu określonej umiejętności czy sprawności (test z rozumienia ze słuchu, quiz słownikowy) albo kilku z nich (test sprawdzający stopień opanowania sprawności i podsystemów języka); można w nich również stosować otwarte i/lub zamknięte zadania, które zazwyczaj składają się z mniejszych części nazywanych *jednostkami testu*.

Te [testy], które są oparte na zadaniach zamkniętych, czyli niewymagających od uczącego się sformułowania odpowiedzi (produkcji językowej), są bardziej obiektywne niż te, oparte na zadaniach otwartych, w których uczący się muszą samodzielnie udzielić odpowiedzi.

(*Programy nauczania...*, 2011, s. 36)

Testowanie osiągnięć pełni cztery istotne funkcje⁸, mające rudymentarne znaczenie zarówno dla uczących, jak i dla uczniów (zob. tabela 1).

Tabela 1. Funkcje testów osiągnięć

	DYDAKTYCZNA	STERUJĄCO-METODYCZNA	DIAGNOSTYCZNA	WYCHOWAWCZA
NAUCZYCIELE	dowiadują się, w jakim stopniu materiał językowy został opanowany.	mogą dokonać retrospektywnej oceny przebiegu procesu dydaktycznego, wprowadzać zmiany.	mogą określić stopień przyswojenia materiału zarówno w odniesieniu do całej grupy, jak też poszczególnych jej członków.	
UCZNIOWIE				przekonują się, że trzeba pracować systematycznie, porządkować i powtarzać materiał.

źródło: opracowanie własne

⁸ Zob. Komorowska 2002, Seretny, Lipińska 2011.

Zakres oddziaływania testów osiągnięć na proces kształcenia i jego uczestników (ang. *backwash/washback effect*) jest więc bardzo szeroki. Nauczycielom pozwalają uzyskać informacje na temat faktycznego stanu wiedzy i umiejętności językowych uczących się, niezbędne do wiarygodnej i bezstronnej ewaluacji ich osiągnięć. Umożliwiają im również ocenę efektywności stosowanych rozwiązań dydaktycznych, wskutek czego mogą, w razie potrzeby, zoptymalizować swoje działania⁹. Po przeanalizowaniu wyników test „może posłużyć nauczycielowi do takiego zmodyfikowania procesu nauczania, aby uczniowie odnosili możliwie największe korzyści” (Dziuba 2000, s. 16). Poza tym należy pamiętać, że proces ten jest zazwyczaj adresowany do uczniów „średnich”, testy więc pomagają nauczycielowi zauważyć tych, którzy za programem nie nadążają albo jednostki ponadprzeciętne. Uczniom zaś mają pokazać, jak dużo się już nauczyli (a nie udowodnić im, że tej, czy innej rzeczy jeszcze nie wiedzą, nie umieją¹⁰), akcentując mocne strony i pozytywne aspekty uczenia się, unaoczniając im jednocześnie braki i słabsze punkty.

W przypadku uzyskania oceny pozytywnej, uczeń ma poczucie odniesionego sukcesu, co znacząco wpływa na podniesienie jego motywacji do dalszej nauki. Natomiast błędy poprawione w teście pozwalają uczniowi lepiej je sobie uświadomić i unikać ich w przyszłości, co przyspiesza proces uczenia się.

(Dziuba 2000, s. 16)

Testy skłaniają także uczniów do powtarzania przerobionego materiału, co umożliwia im porządkowanie wiedzy, unaocznia zdobyte lub nieopanowane jeszcze umiejętności. Należy więc przekonywać ich o walorach tego typu ewaluacji, szukając sposobów łagodzenia jej negatywnych stron (stres, napięcie, ewentualnie rozczarowanie), przez, na przykład, stwarzanie miłej atmosfery, mobilizowanie do wysiłku, lecz przede wszystkim przez uprzednie informowanie o przebiegu egzaminu, jego zawartości i sposobie oceniania. Nic nie powinno ich zaskoczyć!

Zarówno uczący się, jak i sami nauczyciele nie zawsze w sposób właściwy rozumieją jednak, jak istotne znaczenie w procesie kształcenia ma obiektywne sprawdzanie poziomu osiągnięć. Uczniowie nie dostrzegają w nim odzwierciedlenia swoich umiejętności, lecz stopień, jaki otrzymują/w jakim spełniają określone wymagania. Nie zastanawiają się więc nad tym, czego jeszcze nie umieją. Tymczasem wskazywanie problemów, poszukiwanie źródeł trudności oraz sposobów ich przezwyciężania przyczynia się znacząco do kształtowania u nich postawy autonomicznej (zob. Wilczyńska 1999). Uczący z kolei zbyt często traktują testy osiągnięć wyłącznie jako narzędzie pozwalające na ewaluację postępów swoich

⁹ Jeśli jakieś zagadnienie jest w podręczniku słabo lub w nieodpowiednim dla danej grupy wymiarze opracowane (np. jest zbyt mało ćwiczeń), a egzamin pokazuje, że zdający sobie z nim nie poradzili – uczący musi wrócić do tego zagadnienia, posiłkując się innymi pomocami (np. napisać autorskie ćwiczenia, sięgnąć do innych podręczników itp.)

¹⁰ Zob. definicję testu we wprowadzeniu.

podopiecznych, „podstawę wystawienia oceny”, nie widząc w nich źródła istotnych informacji zwrotnych (ang. *feedback*)¹¹, dzięki którym monitoruje się proces nauczania, zwiększając przy tym jego efektywność. Analiza wyników sprawdzianu może więc i powinna stać się okazją do autorefleksji dla obu stron procesu kształcenia. Dlatego też należy omawiać test i jego wyniki, co stanowi końcowy, choć nie zawsze dający się zastosować¹², etap kontroli wyników. Przynosi to korzyści nie do przecenienia obydwu stronom.

3. TRAFNOŚĆ I RZETELNOŚĆ TESTU – JAK JE ZAPEWNIĆ?

Test osiągnięć, naszym zdaniem, powinna cechować wysoka trafność wewnętrzna oraz rzetelność stałościowa¹³. Musi być także sumiennie punktowany (ocenianny). Warto zatem wiedzieć:

- jakie czynniki wpływają na poprawność zastosowanych w nim rozwiązań, skutkujących wyższym lub niższym poziomem trafności;
- jak należy starać się o wiarygodność narzędzia pomiaru i w jaki sposób ją podnosić.

Błędy popełniane przez nauczycieli (np. nieprawidłowy dobór zadań, ich niewłaściwe ukierunkowanie, nieodpowiedni stopień trudności) na etapie planowania i konstruowania testu, mającego na celu badanie osiągnięć uczących się, mogą fałszować rzeczywisty obraz ich umiejętności, niesłusznie je przeceniając lub dostatecznie ich nie dowartościowując.

Formuła testu osiągnięć ze względu na możliwość wykorzystywania go w celu dokonania zarówno ewaluacji kształtującej, jak i sumującej, jest, jak już wspomniano, zróżnicowana. W niniejszym tekście analizie poddane zostaną zagadnienia odnoszące się do *nauczycielskich sprawdzianów sumujących* ze względu na ich kompleksowy charakter, a także większą rangę, jaką się im przypisuje przy ustalaniu końcowej oceny danego ucznia. Przeprowadza się je na zakończenie kolejnych etapów procesu nauczania, a ewaluacji mogą podlegać wówczas wszystkie elementy kształcenia językowego.

Przygotowanie konkretnych zadań sprawdzających umiejętności uczących się zawsze powinno poprzedzać opracowanie planu testu (*specyfikacji*), które wiąże się z koniecznością szczegółowej analizy treści określonych w (z)realizowanym programie oraz wytyczonych w nim celów nauczania. Polega ona na

¹¹ Informacja zwrotna – komentarze osób zaangażowanych w proces testowania (egzaminowanych, egzaminatorów), które stają się podstawą jego ewaluacji (zob. *Glosariusz terminów...* 2004, s. 76).

¹² Na przykład w przypadku testu kończącego kurs.

¹³ Jest ona także nazywana rzetelnością testu jako zespołu zadań.

wyodrębnieniu tych zagadnień i umiejętności, które zostaną poddane ewaluacji. Kluczowe znaczenie ma selekcja, rzutująca nie tylko na sposób testowania (dobór technik i ich ukierunkowanie), ale i na rodzaj oraz tematykę materiałów wyjściowych. Będzie też decydująca przy ustalaniu wagi poszczególnych zadań w strukturze określonego komponentu oraz całego testu. Poza ustaleniami merytorycznymi istotne jest również rozstrzygnięcie wielu kwestii technicznych (*Jaka ma być liczba zadań w poszczególnych komponentach? Jak powinny brzmieć instrukcje? Jak będą przeliczane wyniki na oceny?*) i logistycznych (*Jaki czas potrzebny będzie uczącym się na wypełnienie testu pisemnego? Ile będzie trwał egzamin ustny? Ilu egzaminatorów będzie oceniać test?*), gdyż one, w mniejszym lub większym stopniu, też mogą wpływać tak na trafność, jak i rzetelność pomiaru.

Opracowanie szczegółowego planu testu, jego makro- i mikrostruktury, pozwala zatem uniknąć najczęstszych i najpoważniejszych błędów¹⁴, w czym pomocne mogą być poniższe zestawienia (ang. *check lists*)¹⁵. Część z ujętych tabelarycznie zagadnień ma charakter ogólny i dotyczy testu jako całości (tj. jego makrostruktury) (zob. tabela 2). Większość natomiast odnosi się do rozwiązań szczegółowych, czyli budowy i zawartości poszczególnych komponentów (tj. mikrostruktury testu), sprawdzających stopień opanowania poszczególnych sprawności (RS – rozumienie ze słuchu, RT – rozumienie tekstów, P – pisanie, M – mówienie) i kompetencji lingwistycznych (PG – poprawność gramatyczna, PL – poprawność leksykalna) (zob. tabele 3, 4, 5, 6).

¹⁴ Wyniki, ze względu na czynnik ludzki, zawsze są obciążone pewnymi błędami (zob. Niemierko 1975, 1999). Należy jednak zadbać o to, by usterki testu były jak najmniejsze i nie zniekształcały faktycznego obrazu umiejętności uczących się.

¹⁵ Stanowią one poszerzoną i uzupełnioną wersję spisu zamieszczonego w publikacji Seretny, Lipińska (2011). Listy zagadnień można znaleźć także na stronach internetowych Stowarzyszenia ALTE oraz organizacji EALTA. Dotyczą one jednak testów wysokiej stawki.

TABELA 2. MAKROSTRUKTURA TESTU

ZAGADNIENIE	DZIAŁANIA BŁĘDNE – OBNIŻAJĄCE TRAFNOŚĆ (T) i/ lub RZETELNOŚĆ (R):
<p>1. CZAS PRZEZNACZONY NA WYKONANIE PISEMNEJ CZĘŚCI TESTU</p>	<p>R</p> <ul style="list-style-type: none"> ▪ nieokreślenie lub niedopasowanie czasu przeznaczanego na cały egzamin pisemny do możliwości zdających/wielkości testu. ▪ nieokreślenie lub niedopasowanie czasu przeznaczanego na wykonanie poszczególnych części egzaminu (RT, RS, PG, PL) do typów/liczby zadań^a. ▪ brak przerw/zbyt długie przerwy w czasie trwania egzaminu lub wyznaczenie ich w niestosownym momencie, np. w trakcie rozwiązywania jakiegoś komponentu/zadania. <p>Za dużo czasu może spowodować spadek koncentracji uczących się, jego niedostatek – niewykonanie niektórych zadań lub ich zbyt pospieszną realizację. W przypadku części RS należy przewidzieć dodatkowy czas zarówno przed poszczególnymi zadaniami, jak i między nimi, by uczyć się mogli się zapoznać z ich treścią, uzupełnić/sprawdzić swoje odpowiedzi. Przybliżony czas realizacji zadań można zmierzyć w fazie przedegzaminacyjnej, podczas ćwiczeń.</p>
<p>2. CZAS PRZEZNACZONY NA EGZAMIN USTNY</p>	<p>R</p> <ul style="list-style-type: none"> ▪ zaplanowanie zbyt dużo/mало czasu na egzamin. ▪ nieuwzględnianie czasu umożliwiającego uczyć się przygotowanie się do działań językowych. <p>Planując czas, należy wziąć pod uwagę: poziom zaaansowania zdających (im poziom niższy, tym czas krótszy) ich liczbę oraz formułę egzaminu (uczący się mogą zdawać pojedynczo lub w parach; muszą też mieć czas, by przygotować się do realizacji zadania/zadań. Przybliżony czas realizacji zadań można zmierzyć w fazie przedegzaminacyjnej, podczas ćwiczeń.</p>
<p>3. LICZBA CZĘŚCI TESTU</p>	<p>T</p> <ul style="list-style-type: none"> ▪ pominięcie komponentu sprawdzającego stopień opanowania którejś ze sprawności/kompetencji językowej. <p>Test powinien zawierać tyle części, ile przedmiotów (sprawności/kompetencji lingwistycznych) wchodzi w zakres kształcenia.</p>
<p>4. LICZBA PUNKTÓW ZA CAŁY TEST</p>	<p>T</p> <ul style="list-style-type: none"> ▪ wskazanie przypadkowej liczby punktów powstałej, na przykład, po zliczeniu punktacji za wszystkie zadania. <p>Należy zaplanować ogólną liczbę punktów za cały egzamin (warto rozważyć stosowanie wielokrotności liczby 50, gdyż wówczas łatwiej obliczyć wynik procentowy).</p>
<p>5. LICZBA PUNKTÓW ZA POSZCZEGÓLNE CZĘŚCI TESTU</p>	<p>T</p> <ul style="list-style-type: none"> ▪ przypisanie danemu komponentowi i zbyt dużej lub zbyt małej liczby punktów w stosunku do rzeczywistej wartości danej sprawności w procesie kształcenia językowego^b. <p>Liczba punktów powinna odpowiadać randze, jaką danemu komponentowi i nadano w procesie kształcenia.</p>

^a Planując czas przeznaczony na realizację zadań zawartych w komponentach RT, RS, PG, PL należy pamiętać o zróżnicowanej czasochłonności tworzących je jednostek (np. jednostki typu P/F są mniej czasochłonne niż te, które wymagają doborania wielokrotnego; zaznaczenie właściwej formy zabiera mniej czasu niż jej wpisanie itp.).

^b W Centrum Języka i Kultury Polskiej w Świecie UJ procentowy udział poszczególnych sprawności w teście osiągnąć jest zgodny z liczbą godzin przeznaczonych na nie w harmonogramie. Np. na poziomie B2 wynosi: 35% za część PG i PL, 10% za RS, 15% za RT, po 20% za P oraz M. Zob. Seretny, Lipińska 2011.

<p>6. LICZBA ZADAŃ W KAZDEJ CZĘŚCI TESTU</p>	<p>T/R</p> <ul style="list-style-type: none"> ▪ wystąpienie wyraźnych dysproporcji między liczbą zadań testujących poszczególne części (np. jedno zadanie badające umiejętność słuchania ze zrozumieniem, trzy lub cztery przypisane rozumieniu tekstów czytanych). <p>Należy dopasować liczbę zadań do punktacji komponentu, a nie odwrotnie.</p>
<p>7. ZAGADNIENIA ORAZ UMIEJĘTNOŚCI PODLEGAJĄCE TESTOWANIU</p>	<p>T</p> <ul style="list-style-type: none"> ▪ testowanie umiejętności wykraczających poza program nauczania. ▪ położenie nadmiernego nacisku na niektóre umiejętności/kompetencje, niedoszacowanie innych, co znajduje wyraz w założonej punktacji. <p>Umiejętności poddawane testowaniu powinny być zbieżne z tymi, które zostały przypisane do danego poziomu zaawansowania w <i>Programiach nauczania</i> (2011).</p>
<p>8. POLECENIA</p>	<p>R</p> <ul style="list-style-type: none"> ▪ zamieszczanie poleceń niedoprecyzowanych, niejednoznacznych, zawiłych, zbyt krótkich lub zbyt długich, nowych (innych niż na zajęciach). ▪ niejednolitość stosowanych w nich form ('ty' lub 'pani/pan'). ▪ brak rozwiązywania przykładowego (dotyczy RT, RS, PG, PL). <p>Sformułowania użyte w poleceniach muszą być dla zdających zrozumiałe, nie powinny budzić wątpliwości ani stwarzać potrzeby uzyskania dodatkowych wyjaśnień; uczący się muszą wiedzieć, jakiego rodzaju działań się od nich oczekuje. Zamieszczone rozwiązania przykładowe ma im je uaoacznic.</p>
<p>9. SPORZĄDZENIE KLUCZA (DOTYCZY CZĘŚCI RT, RS, PG, PL)</p>	<p>R</p> <ul style="list-style-type: none"> ▪ nieprzygotowanie klucza. ▪ nierozwiązanie testu. <p>Przygotowanie klucza usprawnia poprawę zadań, daje gwarancję, że wszystkie odpowiedzi zostaną ocenione w ten sam sposób. Podczas rozwiązywania testu (weryfikowania klucza) autor może dostrzec uchybienia i niedoskonałości, zarówno w samym kluczu, jak i w zadaniach.</p>
<p>10. SPOSÓB PRZELICZANIA WYNIKÓW PUNKTOWYCH NA OCENY</p>	<p>R</p> <ul style="list-style-type: none"> ▪ brak ustalonego minimalnego progu dla całego testu (liczby punktów koniecznej do zaliczenia testu na ocenę pozytywną), wskazuje go po otrzymaniu wyników. ▪ brak ustalonego minimalnego progu dla poszczególnych komponentów (liczby punktów koniecznej do zaliczenia komponentu na ocenę pozytywną), wskazanie go otrzymaniu wyników. ▪ brak ustalonych przedziałów punktacji, które będą odpowiadały ocenom (zgodnych z systemem oceniania stosowanym w danej instytucji), dostosowanie ich do wyników egzaminu. <p>Zmiana ustalonego progu minimalnego/ustalonych przedziałów punktacji może mieć miejsce tylko wtedy, gdy z istotnych powodów trzeba dokonać „ważenia” wyników egzaminu. Uczący się powinien jednak wiedzieć przed testem, jak będą punktowane poszczególne części egzaminu i jakie przedziały punktacji będą odpowiadały określonym ocenom.</p>

© Zob. *Przewodnik dla autorów* ... (2004: 161).

TABELA 3. MIKROSTRUKTURA TESTU – POZIOM RECEPCJI – KOMPONENT RS (ROZUMIENIE ZE SŁUCHU)
KOMPONENT RT (ROZUMIENIE TEKSTU)

ZAGADNIENIE	DZIAŁANIA BŁĘDNE – OBNIŻAJĄCE TRAFNOŚĆ (T) i/lub RZE/TELNOŚĆ (R):
1. LICZBA ZADAŃ W KOMPONENCIE	<ul style="list-style-type: none"> ▪ zbyt mała/duża liczba zadań sprawdzających stopień opanowania danej sprawności. <p>Sprawności receptywne należy testować za pomocą kilku zadań, by uczyć się mieli do czynienia z więcej niż jednym kręgiem tematycznym, odmiennymi typami dyskursu, a także różnymi technikami testowania.</p>
2. TYPY JEDNOSTEK WYKORZYSTYWANYCH W POSZCZEGÓLNYCH ZADANIACH	<ul style="list-style-type: none"> ▪ wprowadzenie na teście zadań nieznanymi studentom, nieprzećwiczonych i/lub nieodpowiednich (niespełniających swej roli, trudnych do oceny, czasochłonnych). <p>W teście osiągnięć należy wykorzystywać zadania, których format jest dobrze znany uczącym się, gdyż techniki testowania muszą być zgodne z technikami nauczania.</p>
3. PUNKTACJA ZA POSZCZEGÓLNE ZADANIA I JEDNOSTKI TESTU	<ul style="list-style-type: none"> ▪ brak punktacji za zadanie/zadania. ▪ brak punktacji za poszczególne jednostki testu. <p>Uczący się powinni znać punktację za poszczególne zadania oraz za poszczególne jednostki w ich ramach, zwłaszcza gdy mogą być im przypisane różne wartości (dotyczy to tzw. jednostek wielopunktowych nieocenianych dychotomicznie^d).</p>
4. TYPY TEKSTÓW STANOWIĄCYCH MATERIAŁ WYJŚCIOWY, ICH ŹRÓDŁA	<ul style="list-style-type: none"> ▪ opieranie testów na dowolnych typach tekstów, nieadekwatnych dla danego poziomu, z zakresu tematycznego, który nie jest mu przypisany, pochodzących z ograniczonych/nieodpowiednich źródeł. <p>Teksty stanowiące materiał wyjściowy powinny pochodzić z różnych źródeł i należeć do różnych gatunków. Należy unikać tematów, które mogłyby deprymować uczących się lub obrażać ich uczucia. W przypadku słuchania zaleca się użycia tekstów/nagrań znanych studentom, tj. wykorzystujących znane im głosy, nagrane w podobnym tempie^e, o zbliżonym poziomie szumów (np. kolejne wiadomości z tej samej stacji radiowej).</p>
5. CAŁKOWITA DŁUGOŚĆ TEKSTÓW LUB DŁUGOŚĆ TEKSTÓW POSZCZEGÓLNYCH ZADAŃ (W WYRAZACH DLA RT, W MINUTACH DLA RS) / STOPIEŃ ICH TRUDNOŚCI	<ul style="list-style-type: none"> ▪ wykorzystywanie w testach tekstów nieprzystosowanych długością i stopniem trudności do danego poziomu. <p>Teksty muszą mieć właściwą długość (zob. aneks) i odpowiedni stopień trudności. Uwaga ta dotyczy zarówno warstwy leksykalno-gramatycznej tekstu, jak i jego struktury, rejestru itp. Lektura/słuchanie zbyt długich tekstów może spowodować spadek koncentracji, przekładając się na obniżony poziom realizacji zadań.</p>

^d Jednostki, za które można otrzymać punkty, nawet jeśli rozwiązanie nie jest całkiem poprawne, np. za jednostkę wartą 3 punkty uczyć się dostaje 0, 1, 2 lub 3 punkty w zależności od stopnia zgodności odpowiedzi z kluczem.

^e W przypadku RS tempo nagrań musi być adekwatne do poziomu (zob. *Programy nauczania* ..., 2011).

TABELA 4. MIKROSTRUKTURA TESTU – POZIOM PRODUKCJI – KOMPONENT P (PISANIE)

ZAGADNIENIE	DZIAŁANIA BLEDNE – OBNIŻAJĄCE TRAFNOŚĆ (T) i/lub RZETELNOŚĆ (R):	R/T
1. LICZBA ZADAŃ	<ul style="list-style-type: none"> ▪ zbyt mała/duża liczba zadań/zestawów sprawdzających stopień opanowania sprawności. ▪ brak możliwości wyboru zestawu/zadań. 	R/T
2. TYP MATERIAŁÓW WYJŚCIOWYCH	<p>Zadania mogą występować pojedynczo lub być grupowane w zestawy; zdający powinien mieć możliwość wyboru zadania/zestawu</p> <ul style="list-style-type: none"> ▪ niewłaściwy dobór materiałów wyjściowych, mogących utrudnić wykonanie zadania. 	T
3. FORMA I DŁUGOŚĆ ZADAŃ	<p>Materiałem wyjściowym może być bodziec słowny lub wizualny. Zbyt obszerny/szczegółowy może wydłużyć czas wykonania zadania.</p> <ul style="list-style-type: none"> ▪ wymaganie znajomości form pisemnych spoza wykazu zawartego w <i>Programach nauczania...</i> dla danego poziomu i/lub z wyższego poziomu zaawansowania, nieznanymi uczącym się/nieprzećwiczonego na zajęciach. ▪ brak informacji na temat oczekiwanej długości wypowiedzi (mierzonej w wyrazach), jej przeszacowanie lub niedoszacowanie. <p>Docelowo, tj. na zakończenie nauki na danym poziomie uczący się powinni umieć napisać teksty o długości określonej w <i>Programach nauczania ...</i> (zob. aneks).</p>	T
4. TEMATY WYPOWIEDZI PISEMNYCH	<ul style="list-style-type: none"> ▪ wymaganie znajomości tematów spoza katalogu zawartego w <i>Programach nauczania...</i> dla danego poziomu i/lub z wyższego poziomu zaawansowania. <p>Tematy powinny być zróżnicowane i wpisywać się w kręgi omawiane na zajęciach. Nie mogą się powtarzać w ramach zestawów proponowanych do wyboru.</p>	T
5. KRYTERIA OCENY PRAC PISEMNYCH	<ul style="list-style-type: none"> ▪ poprawa i ocena zadań bez uprzednio przygotowanych kryteriów. ▪ stosowanie kryteriów biegłości do oceny testów osiągnięć. <p>Kryteria oceniania powinny zostać przygotowane z uwzględnieniem „wewnętrznej perspektywy kursu”.</p> <p>W ocenie prac pisemnych kontroli poddać należy: kompetencję lingwistyczną, stopień spełnienia wymogów przewidzianych dla danej formy, jej kompozycję i redakcję. Punktacja musi być dopasowana do poziomu i specyfiki grupy.</p>	R
6. LICZBA EGZAMINATORÓW OCENIAJĄCYCH PRACE PISEMNE	<ul style="list-style-type: none"> ▪ poprawianie i ocenianie prac końcowych przez jednego egzaminatora (zwłaszcza na wyższych poziomach zaawansowania). <p>Wypowiedzi pisemne powinny być poprawiane przez dwie osoby (jedną z nich jest uczący w danej grupie).</p>	R

TABELA 5. MIKROSTRUKTURA TESTU – POZIOM PRODUKCJI – KOMPONENT M (MÓWIENIE)

ZAGADNIENIE	DZIAŁANIA BLEDNE – OBNIŻAJĄCE TRAFNOŚĆ (T) i/lub RZETELNOŚĆ (R):	R/T
1. LICZBA ZADAŃ	<ul style="list-style-type: none"> zbyt mała/duża liczba zadań/zestawów sprawdzających stopień opanowania sprawności. brak możliwości wyboru zestawu/zadań. <p>Liczba zadań zebranych w zestaw powinna zawierać się w przedziale 1-3; zdający wybiera jeden zestaw i wykonuje zawarte w nim zadania.</p>	R/T
2. TYP MATERIAŁÓW WYJŚCIOWYCH; ICH ŹRÓDŁA	<ul style="list-style-type: none"> niewłaściwy dobór materiałów wyjściowych, mogących utrudnić wykonanie zadania. niewłaściwa jakość materiałów stymulujących. <p>Materiałem wyjściowym może być bodziec słowny lub wizualny. Zbyt obszerny/szczegółowy może wydłużyć czas wykonania zadania. Materiały ikoniczne muszą być przejrzyste, czytelne i odpowiednio dobrane.</p>	T
3. RODZAJ DZIAŁAŃ JEZYKOWYCH	<ul style="list-style-type: none"> testowanie jednego rodzaju działań (np. tylko umiejętności budowania wypowiedzi monologicznej). <p>Zadania powinny sprawdzać, czy zdający potrafi podejmować różne działania (produkcyjne, interakcyjne i mediacyjne), realizując określone funkcje komunikacyjne^{1/}.</p>	R
4. TEMATY WYPOWIEDZI USTNYCH	<ul style="list-style-type: none"> wymaganie znajomości tematów spoza katalogu zawartego w <i>Programach nauczania</i>... dla danego poziomu i/lub z wyższego poziomu zaawansowania. <p>Tematy powinny być różnicowane i wpisywać się w kręgi omawiane na zajęciach. Nie mogą się powtarzać w ramach zestawów proponowanych do wyboru.</p>	T
5. KRYTERIA OCENY WYPOWIEDZI USTNYCH	<ul style="list-style-type: none"> poprawa i ocena zadań bez uprzednio przygotowanych kryteriów. stosowanie kryteriów biegłości do oceny testów osiągnięć. <p>Kryteria oceniania^{2/} powinny zostać przygotowane z uwzględnieniem „wewnętrznej perspektywy kursu”^{3/}, dostosowane do warunków, w jakich program był realizowany. Punktacja musi być dopasowana do poziomu zaawansowania grupy i jej specyfiki.</p>	R
6. LICZBA EGZAMINATORÓW OCENIAJĄCYCH WYPOWIEDZ USTNĄ	<ul style="list-style-type: none"> ocenie wypowiedzi ustnej przez jednego egzaminatora. <p>Wypowiedzi ustne muszą być oceniane przez co najmniej dwie osoby. Jeden egzaminator nie jest w stanie rzetelnie i obiektywnie ich ocenić, zwłaszcza gdy także pełni funkcję interlokutora.</p>	R

^{1/} Na niższych poziomach powinny dominować działania interakcyjne i produkcyjne, na wyższych produkcyjne i mediacyjne.^{2/} Przykłady siatek oceny wypowiedzi ustnych znaleźć można w pracy Janowskiej (2014).^{3/} Jeśli na danym kursie nacisk był położony na rozwijanie płynności, nie należy surowo oceniać aspektów poprawnościowych i na odwrót.

TABELA 6. MIKROSTRUKTURA TESTU – POZIOM RECEPCJI/PRODUKCJI – KOMPONENT PG (POPRAWNOŚĆ GRAMATYCZNA) – KOMPONENT PL (POPRAWNOŚĆ LEKSYKALNA)¹⁶

ZAGADNIENIE	DZIAŁANIA BŁĘDNE – OBNIŻAJĄCE TRAFNOŚĆ (T) i/lub RZETELNOŚĆ (R):
1. LICZBA ZADAŃ W KOMPONENCIE	<p>R/T</p> <ul style="list-style-type: none"> ▪ zbyt mała/długa liczba zadań sprawdzających stopień opanowania danej sprawności. <p>Kompetencje lingwistyczne należy testować za pomocą wielu zróżnicowanych zadań, by uczyć się mieli do czynienia z wieloma kręgami tematycznymi, a także różnymi technikami testowania, gdyż trudność określonych typów zadań nie jest taka sama dla wszystkich zdających.</p>
2. ZAGADNIENIA I/ LUB UMIEJĘTNOŚCI SPRAWDZANE W DANYM ZADANIU Z UWZGLĘDNIENIEM SPOSOBU TESTOWANIA	<p>T</p> <ul style="list-style-type: none"> ▪ niewłaściwa selekcja zagadnień. ▪ brak proporcji między testowaniem zagadnień gramatycznych/leksykalnych na poziomie recepcji i produkcji. <p>Dobierając zagadnienia, należy kierować się ich „przydatnością komunikacyjną”. Na poziomach niższych powinniśmy przeważać zdania sprawdzające produktywną znajomość struktur leksykalno-gramatycznych, na wyższych – zdecydowanie dominować.</p>
3. TYPY JEDNOSTEK WYKORZYSTYWANYCH W POSZCZEGÓLNYCH ZADANIACH	<p>R/T</p> <ul style="list-style-type: none"> ▪ wprowadzenie zadań nieznanych studentom, nieprzećwiczonych i/lub nieodpowiednich (niespełniających swej roli, trudnych do oceny, czasochłonnych). ▪ testowanie gramatyki/słownictwa wyłącznie na pojedynczych zdaniach/izolowanych wyrazach z wykorzystaniem technik automatyzujących typu „dryl”. <p>W teście osiągnięć należy wykorzystywać zadania, których format jest dobrze znany uczącym się, gdyż techniki testowania muszą być zgodne z technikami nauczania, przy czym w jak najszerszym zakresie stosować zadania oparte na tekstach.</p>
4. CAŁKOWITA LICZBA PUNKTÓW ZA DANE ZADANIE WRAZ Z LICZBĄ PUNKTÓW PRZYZNAWANYCH ZA KAŻDĄ JEDNOSTKĘ TESTU	<p>R</p> <ul style="list-style-type: none"> ▪ brak punktacji za zadanie/zadania. ▪ brak punktacji za poszczególne jednostki testu. <p>Uczący się powinni znać wartości za poszczególne zadania oraz za poszczególne jednostki w ich ramach, zwłaszcza, gdy mogą być im przypisane różne wartości (dotyczy to tzw. jednostek wielopunktowych nieocenianych dychotomicznie).</p>

Źródło: opracowanie własne









¹⁶ Błędem jest wybieranie zagadnień mniej istotnych, a trudnych do opanowania, takich na przykład jak wyjątki lub latwych do sprawdzania. /Jednostki, za które można otrzymać punkty, nawet jeśli rozwiązanie nie jest całkiem poprawne, np. za jednostkę wartą 3 punkty, uczący się może dostać 0, 1, 2 lub 3 punkty w zależności od stopnia zgodności odpowiedzi z kluczem.

¹⁶ Nie ma osobnych tabel dla ortografii i fonetyki, gdyż te zazwyczaj testowane są pośrednio, w ramach – odpowiednio – pisania i mówienia.

4. BŁĘDY W KONSTRUKCJI ZADAŃ TESTOWYCH

W niniejszej części zamieszczone zostały przykłady ilustrujące omawiane powyżej zagadnienia. Pochodzą z różnych testów osiągnięć¹⁷, z części sprawdzających znajomość słownictwa i struktur gramatycznych oraz stopień opanowania sprawności pisania, w których stanowiły całe komponenty (słownictwo B1, pisanie B2) lub ich części (słownictwo A2; gramatyka A1). Wykorzystane w nich rozwiązania nie gwarantują odpowiedniej rzetelności pomiaru, cechuje je również stosunkowo niska trafność, co oznacza, że wyniki uzyskane przez uczących się nie dadzą pełnego obrazu ich faktycznych umiejętności językowych.

¹⁷ Zadania sprawdzające znajomość słownictwa i gramatyki zostały wzięte z prac studentek II roku magisterskich studiów uzupełniających *Nauczanie języka polskiego jako obcego* w UJ. Autorskie przykłady zestawów do pisania pochodzą z warsztatów dla zespołu autorów zadań opracowujących testy certyfikatowe z języka polskiego jako obcego oraz z prac słuchaczy podyplomowych studiów *Nauczanie języka polskiego jako obcego* w UJ. Do niektórych przykładów usterki zostały wprowadzone celowo.

<p>POZIOM B1 – SŁOWNICTWO _____ /15 p.</p> <p>I. Proszę podpisać obrazki. _____ /8 p. (8 x 1)</p>  <p>__talerz__</p>        <p>II. Wyszukaj i podkreśl wyraz niepasujący do pozostałych. _____ /6 p. (6 x 1)</p> <p>Przykłady: <i>pomidor, ogórek, jabłko, rzodkiewka, por, sałata</i></p> <ol style="list-style-type: none"> 1. gotować, smażyć, dusić, grillować, jeść 2. widelec, nóż, łyżka, garnek 3. południe, śniadanie, obiad, kolacja, podwieczorek 4. pierogi, naleśniki, ziemniaki, gołąbki, spaghetti 5. kawa, zupa, herbata, woda, sok 6. kieliszek, kubek, suszarka, szklanka 	<p>Te dwa ćwiczenia cechuje niska trafność, gdyż:</p> <ul style="list-style-type: none"> ▪ słownictwo podlegające kontroli sprawdzane jest w sposób nieadekwatny do poziomu zaawansowania – na B1 uczyć się powinien umieć posługiwać się nowo poznawanymi wyrazami w aktach komunikacji, a nie jedynie nazywać przedmioty należące do otaczającej go rzeczywistości^a; ▪ sprawdzane wyrazy zostały niewłaściwie wybrane – powinny być już opanowane na poprzednim poziomie A2 (poza czasownikami <i>smażyć i grillować</i> oraz rzeczownikami <i>patelnia</i> oraz <i>stołek</i>); <p>R z e t e l n o ś ć obniża:</p> <ul style="list-style-type: none"> ▪ zastosowanie nieodpowiedniej punktacji: w obu ćwiczeniach przyznaje się po 1 punkcie za prawidłowe rozwiązanie, nie różniując rodzaju działań (działania receptywne powinny być cenione „niżej” niż produktywne). ▪ brak konsekwencji w formułowaniu poleceń (nieuzasadniona zmiana formy trybu rozkazującego^b); ▪ brak jednego punktu – komponent miał liczyć ich 15.
--	--

^a Podpisywanie obrazków jest techniką bardziej właściwą dla poziomu A, dla wyższych stosuje się ją głównie w przypadku leksyki specjalistycznej.

^b W poleceniach zastosowano różne formy – w pierwszym grzecznością formułę *Proszę podpisać* (tryb oznajmujący w funkcji rozkazującego), w drugim formę trybu rozkazującego – *Wyszukaj*.

<p>POZIOM A2 – SŁOWNICTWO</p> <p>III. Proszę dopisać wyrazy do podanych niżej definicji. _____/2 p. (4 x 0,5)</p> <p>Przykład: <i>związek, stosunek między ludźmi; kontakt</i> _____ <i>relacja</i> _____</p> <ol style="list-style-type: none"> 1. utrzymywać kontakty z innymi osobami _____ 2. miejsce, gdzie rosną kwiaty i pasą się zwierzęta _____ 3. groźne, duże zwierzę, które żyje w górach _____ 4. czuć to samo, co druga osoba _____ 	<p>To ćwiczenie cechuje bardzo niska <i>t r a f n o ś ć</i>, gdyż:</p> <ul style="list-style-type: none"> ▪ słownictwo podlegające kontroli zdecydowanie wykracza poza poziom A2 – jest za trudne (<i>paść się, groźny, utrzymywać, związek, stosunek</i>); ▪ zastosowano nieodpowiednią technikę testowania (tzw. „krzyżówkową”), która nie daje możliwości wiarygodnego sprawdzenia, czy zdający rzeczywiście nie zna danego słowa, czy też nie wpadł na właściwe rozwiązanie; ▪ definicje, a raczej hasła, wykorzystane w ćwiczeniu są nieprecyzyjne; ▪ przykład sugeruje, że wszystkie rozwiązania to rzeczowniki, podczas gdy sformułowania w definicjach wskazują, że mogą to być także czasowniki. <p>Rz e t e l n o ś ć ćwiczenia obniża:</p> <ul style="list-style-type: none"> ▪ zbyt mała liczba tworzących je jednostek.
<p>POZIOM A1 – POPRAWNOŚĆ GRAMATYCZNA</p> <p>IV. Uzupełnij zdania odpowiednimi formami czasownika <i>móc</i>. _____ / 4 p. (4 x 1)</p> <p>Przykład: Ona będzie<i>mogła</i>..... pójść do kina</p> <ol style="list-style-type: none"> 1. Będzie pojechać na wakacje z kolegami. 2. Będziemy podróżować, jak tylko dostaniemy pieniądze. 3. Dziewczyny będą iść popływać, jak tylko poprawi się pogoda. 4. Będę w końcu zarobić na moje wymarzone wakacje. 	<p>Ćwiczenie to ma niską <i>t r a f n o ś ć</i>, gdyż:</p> <ul style="list-style-type: none"> ▪ zastosowano nieodpowiednią technikę testowania, tj. typowe ćwiczenie automatyzujące typu „dryl” sprawdzające odmianę jednego czasownika; ▪ brak informacji dotyczącej rodzaju powoduje, że uczący się mogą wpisać formę znaną sobie, co niekoniecznie oznacza, że znają je obie. <p>Rz e t e l n o ś ć ćwiczenia obniża:</p> <ul style="list-style-type: none"> ▪ zbyt wysoka punktacja za realizację tworzących je jednostek; ▪ zbyt mała ich liczba; ▪ niska autentyczność sytuacyjna struktury zastosowanych jednostek (zdania izolowane^d).

^c Autorkom niniejszego tekstu nie udało się rozwiązać wszystkich jednostek tego ćwiczenia.

^d Tego typu ćwiczenia nadają się na krótkie sprawdziany.

POZIOM B2 – PISANIE Proszę wybrać <u> jeden </u> zestaw i wykonać zawarte w nim zadania ^e .		— /20 p.
Zestaw I		<ul style="list-style-type: none"> ▪ w zadaniu 1. brak określenia, o jakie święta chodzi (należy uściślić: „z okazji Wielkanocy/Bożego Narodzenia”); ▪ jeśli uczestnikami kursu są osoby dorosłe, należy zmienić adresata w zadaniu 1. (np. dla dzieci, siostry, kuzyna).
Zestaw II	<p>1. Proszę napisać życzenia świąteczne dla swoich dziadków/rodziców. (30 wyrazów)</p> <p>2. <i>Komputer – moja miłość</i>. Proszę napisać esej. (270 wyrazów)</p> <p>1. Proszę zaprosić kuzyna z rodziną na 70. urodziny Babci. (40 wyrazów)</p> <p>2. Proszę napisać tekst reklamy swojej ulubionej restauracji (kawiarni). (60 wyrazów)</p> <p>3. <i>Jak spędzały czas dzieci 30 lat temu i dzisiaj?</i> Proszę porównać. (200 wyrazów)</p>	<ul style="list-style-type: none"> ▪ w zdaniu 1. brak określenia miejsca imprezy (należy dodać, np. „które odbędą się w Pani/Pana nowym domu/w restauracji X”); ▪ w zestawie jest za dużo zadań.
Zestaw III	<p>1. Napisz autocharakterystykę. (150 wyrazów)</p> <p>3. Napisz sprawozdanie z wycieczki. (150 wyrazów)</p>	<ul style="list-style-type: none"> ▪ w obu zadaniach zmiana formy z „pan/i” na „ty”; ▪ pomyłka w numeracji (zamiast zadanie 2 jest 3); ▪ brak informacji, o którą wycieczkę chodzi (należy to uściślić, podając np. jej termin, miejsce).
Zestaw IV	<p>1. W krótkim liście proszę podziękować sąsiadce za opiekę nad Pani/Pana mieszkaniem podczas Pani/Pana tygodniowej nieobecności.</p> <p>2. Jakże święto jest dla Pan/i ważne? Dlaczego?</p>	<ul style="list-style-type: none"> ▪ w obu zadaniach nie podano oczekiwanej liczby wyrazów; ▪ w zadaniu 1. brak określenia wieku sąsiadki, który decyduje o wyborze formy oficjalnej lub nie (należy dodać np. „starszej” lub zamienić na „koleżance mieszkającej obok”); ▪ w zadaniu 2. brak wskazania formy wypowiedzi (należy dodać, np.: „Proszę napisać esej. Proszę opisać, jak je Pan/i spędza”); określenie „święto to” w zadaniu 2. jest nieprecyzyjne (należy dodać „święto rodzinne/kościelne/państwowe”); zamiana na l.mn., <i>święta</i>, spowodowałaby częściowe powtórzenie kręgu tematycznego z zadania 1 w zestawie I.

^e W pokazanym teście tematyka jest różnorodna, proponuje się w nim też wiele form. W zeczywistości muszą być one zgodne z treściami nauczania na danym kursie.

^f W tym przypadku doszło by jednak do zdublowania formy – esej pojawia się w zestawie I.

Zestaw V**1. Śladami Stanisława Wyspiańskiego.**

(300 wyrazów)

- zadanie nie ma określonej formy.

Rzeczelnosc pomiaru obniza:

- zamieszczenie w komponencie zbyt wielu zestawów, gdyż piszący tracą czas, zastanawiając się, który wybrać;
- nieuzasadniona, zwiększona liczba zadań (zestaw III);
- brak konsekwencji w formułowaniu poleceń (zestaw IV, zadania 1, 2);
- brak wskazania oczekiwanej długości wypowiedzi (zestaw IV, zadanie 1);
- pomyłki w numeracji zadań (zestaw III).

Na poziomie trafności i rzutuje:

- testowanie niewłaściwych form – pisanie życzeń i zaproszeń (zestaw I, zadanie 1) zazwyczaj sprawdza się na poziomie niższym, choć formy te często wymagają powtórzenia na B2;
- brak określenia formy wypowiedzi (zestaw IV, zadanie 1; zestaw V, zadanie 2), który powoduje, że sprawdzamy niekoniecznie to, zamierzaliśmy, gdyż zdający mogą stosować strategię uników i wybrać znaną sobie/lepiej opanowaną formę;
- brak precyzji sformułowań w instrukcjach (zestaw I, zadanie 1; zestaw II, zadanie 1; zestaw III, zadanie 3; zestaw V, zadanie 1), który powoduje, że pojawi się rozbieżność między tym, co miało być, a tym, co będzie testowane.

5. PODSUMOWANIE

Przygotowanie trafnego i rzetelnego testu osiągnięć sprawdzającego faktyczne umiejętności językowe uczących się wymaga precyzyjnych i przemyślanych działań oraz wiedzy merytorycznej. Przy jego opracowaniu trzeba także pamiętać o kwestiach natury praktycznej, m.in.:

- test musi spełniać określone wymagania techniczne, tj. być przejrzysty, czytelny (kserokopie muszą być dobrej jakości);
- test powinien mieć wyraźnie wydzielone części, wyróżnione polecenia i zaznaczoną punktację;
- miejsca przeznaczone na odpowiedzi uczących się muszą mieć odpowiednią wielkość¹⁸;
- w teście należy ograniczać liczbę zadań, których przygotowanie jest kosztowne (np. wymagają zastosowania kolorowych kserokopii) lub zbyt czasochłonne (np. wymagają klejenia nagrań, programów tv, przekazów Internetowych itp.)¹⁹.

Gotowy sprawdzian należy poddać analizie pod kątem *autentyczności* i *interakcyjności* wykorzystanych w nim zadań. Przez *autentyczność* rozumie się stopień, w jakim działania podejmowane przez uczących się w celu realizacji zadań egzaminacyjnych, przypominają te, które mogliby wykonywać w rzeczywistości; *interakcyjność* natomiast odnosi się do poziomu zaangażowania intelektualnego i emocjonalnego towarzyszącego zdającym w czasie realizacji zadań – im bliższy naturalnemu, tym lepiej (Bachman, Palmer 1996, s. 23-25). Warto go też skonsultować z kolegami/współpracownikami mogącymi dostrzec błędy, które umknęły autorowi, „wnieść uwagi dotyczące konstrukcji, czy podsunąć lepsze rozwiązania” (Dziuba 2000, s. 22). W czasie przeprowadzania testu wskazane jest notowanie wszelkich spostrzeżeń pomocnych przy układaniu kolejnego (np. czas realizacji określonych zadań; czas, w jakim pierwszy/ostatni uczeń oddał pracę pisemną itp.). Bardzo ważna jest również, jak już wspomniano, analiza uzyskanych wyników. Jeśli bowiem test został rozwiązany przez wszystkich na ocenę bardzo dobrą, musiał składać się z zadań zbyt łatwych. Gdy zaś większość radziła z nim sobie z wyraźnymi problemami – tworzące go zadania cechował za wysoki poziom trudności, przy czym w obu przypadkach rezultaty nie pozwalają właściwie różnicować ocen przyznawanych zdającym²⁰.

¹⁸ W przypadku zadania z jednostkami typu „uzupełnianie luk” wklejanie skopiowanego tekstu z zamalowanymi wyrazami jest przejawem braku profesjonalizmu.

¹⁹ Gdy jednak są one niezbędne, należy rozważyć takie przygotowanie testu/jego fragmentów, by można było go/ich użyć ponownie. Wiąże się to wówczas z koniecznością przygotowania oddzielnych arkuszy odpowiedzi, na których uczniowie będą rozwiązywać zadania.

²⁰ Należy jednak pamiętać, że rozkładu normalnego, tzw. krzywej Gaussa (zob. *Przewodnik dla autorów zadań do testów językowych*, 2004) oczekiwać można jedynie w przypadku, gdy grupa piszących jest większa niż 30 osób. W grupach mniej licznych wyniki bywają bardziej splaszczone, lecz na pewno nie powinny być jednorodne.

Na zakończenie warto raz jeszcze podkreślić, że test osiągnięć jest bardzo ważnym narzędziem dydaktycznym. Jego tworzeniu powinna więc zawsze towarzyszyć refleksja nad celowością stosowanych rozwiązań oraz ich możliwymi konsekwencjami.

BIBLIOGRAFIA

- Bachman L., Palmer A., 1996, *Language Testing in Practice*, Oxford.
- Dziuba D., 2000, *Testy w nauczaniu języków obcych*, „Języki Obce w Szkole”, nr 4, s. 15–22.
- Fulcher G., Davidson F., 2007, *Language Testing and Assessment: An Advanced Resource Book*, London–New York.
- Gałyga D., 2004, *Rzetelność i trafność pomiaru dydaktycznego na wybranych przykładach testów*, w: W. Martyniuk, E. Lipińska, A. Seretny (red.), *Opisywanie, rozwijanie i testowanie znajomości języka polskiego jako obcego*, Kraków, s. 79–87.
- Glosariusz terminów z zakresu testowania biegłości językowej*, 2004, Kraków.
- Ingram E., 1983, *Testy językowe*, w: J. B. P. Allen, S. Pit Corder, A. Davies (red.), *Kurs edynburski językoznawstwa stosowanego*, t. II. Warszawa, s. 225–251.
- Janowska I., 2014, *Jak oceniać ustne działania językowe uczących się? Ewaluacja wypowiedzi ustnej w podejściu ukierunkowanym na działanie*, w: A. Mielczarek, A. Roter-Bourkane, M. Zduniak-Wiktorowicz (red.), *Sukcesy, problemy i wyzwania w nauczaniu języka polskiego jako obcego*, Poznań, s. 227–238.
- Kane M., 2006, *Validation*, w: R. L. Brennan (red.), *Educational Measurement: Fourth Edition*, Westport, CT, s. 17–64.
- Komorowska H., 2002, *Sprawdzanie umiejętności w nauce języka obcego. Kontrola – ocena – testowanie*, Warszawa.
- Lipińska E., Seretny A., 2010, *Testy plasujące jako narzędzie pomiaru dydaktycznego na przykładzie języka polskiego jako obcego – ocena dotychczasowych doświadczeń*, „Neofilolog” nr 34, s. 229–242.
- Messick S., 1989, *Validity*, w: R. L. Linn (red.), *Educational Measurement: Third Edition*, New York, s. 13–103.
- Niemierko B., 1975, *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe*, Warszawa.
- Niemierko B., 1999, *Pomiar wyników kształcenia*, Warszawa.
- Programy nauczania języka polskiego jako obcego. Poziomy A1–C2*, 2011, Kraków.
- Przewodnik dla autorów zadań do testów językowych*, tłum i adaptacja M. Gaszyńska-Magiera i A. Seretny, Universitas, Kraków 2004.
- Pukas-Palimąka D., Lipińska E. (1997), *Ewaluacja testów biegłości do nauczania języka polskiego jako obcego*, „Przegląd Polonijny”, z. 2, s. 115–122
- Seretny A., Lipińska E., 2011, *Kontrola bieżąca (formative evaluation) a kontrola globalna (summative evaluation) – testy osiągnięć w procesie dydaktycznym*, w: U. Pluskota, K. Taczyńska (red.), *Nowe perspektywy w nauczaniu języka polskiego jako obcego II*, Toruń, s. 53–70.
- Weir C. J., 2005, *Language Testing and Validation. An Evidence-Based Approach*, Basingstoke.
- Wilczyńska W., 1999, *Uczyć się czy być nauczonym? O autonomii w przyswajaniu języka obcego*, Warszawa.
- Wysocka M., 2009, *Badania innowacyjne (action research) jako czynnik stymulujący pracę twórczą nauczycieli języków obcych*, w: M. Pawlak, A. Mystkowska-Wiertelak, A. Pietrzykowska (red.), *Nauczyciel języków obcych dziś i jutro*, Poznań-Kalisz, s. 509–517.

ANEKS²¹

1.

Długość tekstów stanowiących podstawę wykonania zadań z RS i RT powinna mieścić się w poniższych zakresach (+/- 10%):

- | | |
|---|---|
| 1. dla poziomu A1: 80 – 100 wyrazów | 4. dla poziomu B2: 250 – 300 wyrazów |
| 2. dla poziomu A2: 120 – 150 wyrazów | 5. dla poziomu C1: 350 – 350 wyrazów |
| 3. dla poziomu B1: 200 – 250 wyrazów | 6. dla poziomu C2: 300 – 400 wyrazów |

2.

Długość wypowiedzi pisemnych uczących się powinna mieścić się w poniższych zakresach (+/- 10%):

- | | |
|---------------------------------------|---------------------------------------|
| 1. dla poziomu A1: 100 wyrazów | 4. dla poziomu B2: 300 wyrazów |
| 2. dla poziomu A2: 150 wyrazów | 5. dla poziomu C1: 400 wyrazów |
| 3. dla poziomu B1: 200 wyrazów | 6. dla poziomu C2: 500 wyrazów |

Anna Seretny, Ewa Lipińska

FACTORS AFFECTING VALIDITY AND RELIABILITY OF ACHIEVEMENT TESTS

Keywords: achievement test, reliability, validity, test planning, mistakes of the planning phase, factors affecting test reliability and validity, factor analysis

Summary. A language achievement test is the most commonly used evaluation tool in the field of language learning/ teaching. Such tests measure students' mastery of a particular instructional domain in order to make decisions regarding their advancement and/or competency. A well-constructed test will give both the teacher and the students an appraisal of their respective achievements. Administering tests will lose its importance if the items do not pose a particular challenge to the students and/or if they do not adequately reflect the given body of content. The text presents a thorough analysis of the factors affecting validity and reliability of achievement tests. Validity is the most important quality of test use, which concerns the extent to which meaningful inferences can be drawn from test scores. No test can achieve its intended purpose if its results are unreliable.

²¹ Dane pochodzą z *Programów nauczania...* (2011): A1 – s. 36; A2 – s. 67, B1 – 101, B2 – s. 136, C1 – s. 173, C2 – s. 209.