
Czy testy mierzą to, co mierzyć powinny? O trafności testów maturalnych sprawdzających rozumienie ze słuchu

Do tests measure what they are supposed to? On validity of
listening comprehension tests

Emilia Podpora-Polit
Zespół Szkół Ponadgimnazjalnych nr 2
w Kielcach

Abstrakt

Trafność jest jednym z najważniejszych kryteriów poprawności testu. Jak stwierdza Bolesław Niemierko, trafność gwarantuje, że „zmierzone wszystko to i tylko to, co należało zmierzyć” (1999: 178), a więc test nie sprawdza ani za dużo, ani za mało. Celem artykułu jest przybliżenie pojęcia trafności testu językowego zarówno z teoretycznej, jak i praktycznej perspektywy. Artykuł ma za zadanie odpowiedzieć na następujące pytania: Na czym polega trafność testu? Jak można rozpoznać, czy dany test sprawdzający rozumienie ze słuchu jest trafny? Jak wygląda realizacja kryterium trafności w praktyce, na przykładzie testowania umiejętności rozumienia ze słuchu na egzaminie maturalnym z języka niemieckiego?

Słowa kluczowe: trafność, test, rozumienie ze słuchu.

Abstract

Validity is considered the most important criterion for the quality of a language test. As Bolesław Niemierko (1999: 179) states, validity guarantees that “everything and only what was supposed to be measured

was measured". The aim of this article is to present the validity notion from the theoretical as well as from the practical point of view. The paper should answer the following questions: What does test validity mean in practice? How can it be recognized that a listening comprehension test is valid? Do school leaving exams in German fulfill the validity criteria?

Key words: validity, test, listening comprehension.

Wstęp

Trafność treści i trafność konstrukcyjna to podstawowe kryteria poprawności testu językowego. Test trafny „mierzy tylko to i wszystko to, co należało zmierzyć” (Niemierko, 1999: 178). W przypadku testu maturalnego gwarantem jego poprawności powinna być standaryzacja, czyli wcześniejsze poddanie go procesowi „prób, ulepszeń i normowania” (Niemierko, 1999: 55). Czy testy maturalne z języka niemieckiego spełniają to kryterium? Analizie pod tym kątem poddane zostaną zadania sprawdzające rozumienie ze słuchu na poziomie podstawowym i rozszerzonym, które pojawiły się na egzaminie maturalnym w sesji majowej w ciągu ostatnich pięciu lat, tj. w latach 2009-2014.

Trafność testu

Wyróżnia się dwa podstawowe rodzaje trafności, tj. trafność wewnętrzną i trafność zewnętrzną. Pierwszy wymieniony typ: trafność wewnętrzną, która, zdaniem Komorowskiej, jest „najistotniejszą i wymagającą największej troski cechą testu lub sprawdzianu” (2004: 25), dzieli się na trafność treści, trafność konstrukcyjną oraz trafność fasadową. Ponieważ trafność fasadowa odnosi się do sposobu postrzegania danego testu przez zdających, ustalenie, czy dany test maturalny spełnia to kryterium, wymagałoby zbadania nastawienia abiturientów. Również określenie trafności zewnętrznej testu, w obrębie której wyróżniamy trafność diagnostyczną i prognostyczną, nie jest możliwe bez przeprowadzenia dodatkowego pomiaru. Trafność zewnętrzna odnosi się bowiem „do stopnia, w jakim test mierzy tę samą rzecz co inny, sprawdzony test”

(Przewodnik... 2004: 34). Ten rodzaj trafności ustala się przez porównanie wyników uzyskanych w tym samym czasie (trafność diagnostyczna) lub w niedalekiej przyszłości (trafność prognostyczna) na innym teście sprawdzającym te same umiejętności (Komorowska, 2004: 24-25). Z tego względu niniejszy wywód zostanie ograniczony do dwóch aspektów tego obszernego zagadnienia: trafności treści i trafności konstrukcyjnej.

Trafność treści

Trafność treści należy interpretować jako „zgodność treści testu z treściami nauczania” (Komorowska, 2004: 23). W przypadku testów maturalnych, trafność treści oznacza, że test ten nie wykracza poza wymagania zawarte w standardach egzaminacyjnych (Rozporządzenie MENiS 2003) oraz że uwzględnia przynajmniej te istotne.

Poważne uchybienia w tym zakresie wykazuje test sprawdzający rozumienie ze słuchu z 2011 roku. Jego wadą jest jednolitość tematyczna tekstów audialnych. Aż cztery z sześciu prezentowanych tekstów (po trzy teksty na poziomie podstawowym i rozszerzonym) poruszają temat pracy (wybór i aspekty wykonywania danego zawodu, osiągnięcia zawodowe itp.). W standardach egzaminacyjnych praca stanowi jeden z aż 15 zakresów tematycznych (Rozporządzenie MENiS 2003). Test sprawdza zatem umiejętność słuchania w bardzo ograniczonym zakresie tematycznym.

Kryterium trafności narusza również wąski zakres badanych umiejętności w zakresie słuchania. Problem ten ilustruje poniższa tabela, w której przedstawiono wszystkie umiejętności testowane na egzaminie maturalnym w latach 2010-2013¹.

1 W tabeli nie ujęto sprawozdania z 2009 roku, ponieważ nie zawiera ono szczegółowego zestawienia testowanych umiejętności. Natomiast sprawozdanie z 2014 roku w momencie powstawania tego artykułu nie jest jeszcze dostępne.

Standardy egzaminacyjne z języka obcego nowożytnego w zakresie podstawowym i rozszerzonym — rozumienie ze słuchu	Test 2010	Test 2011	Test 2012	Test 2013
	Zadania 1-3 reprezentują poziom podstawowy, zadania 4-6 — poziom rozszerzony			
Określanie głównej myśli tekstu	2.1-2.5	2.1-2.5	3.5	2.1-2.5 3.5 5.1-5.5
Określanie głównych myśli poszczególnych części tekstu	5.1-5.5	5.1-5.5	2.1-2.5 5.1-5.5	
Stwierdzanie, czy tekst zawiera określone informacje	1.1-1.5 4.1-4.5	1.1-1.5 4.1-4.5	1.1-1.5 4.1-4.5	1.1-1.5 4.1-4.5
Selekcjonowanie informacji	3.1-3.5 6.1-6.5	3.1-3.5 6.1-6.5	3.1-3.4 6.1-6.5	3.1-3.4 6.1-6.5
Określanie intencji autora lub nadawcy tekstu	_____	_____	_____	_____
Rozróżnienie formalnego i nieformalnego stylu tekstu	_____	_____	_____	_____
Określanie kontekstu sytuacyjnego (miejsca, czasu warunków, uczestników)	_____	_____	_____	_____

Tabela 1.

Repertuar umiejętności w zakresie słuchania sprawdzanych na egzaminie maturalnym z języka niemieckiego w latach 2010-2013 (oprac. na podstawie sprawozdań o wynikach dostępnych na stronie <http://www.cke.edu.pl/index.php/egzamin-maturalny-left/dla-absolwentow-lat-2005-2014/23-egzamin-maturalny/48-informacje-o-wynikach>)

Jak wynika z tabeli, w zakresie rozumienia ze słuchu w latach 2010-2012 testowane były tylko cztery, a w roku 2013 tylko trzy z siedmiu wymienionych w standardach umiejętności. Testy zazwyczaj sprawdzały umiejętność stwierdzania, czy tekst zawiera określone informacje; okre-

ślania głównej myśli tekstu; selekcjonowania informacji; określania głównych myśli poszczególnych części tekstu. Pozostałe umiejętności: określanie intencji autora lub nadawcy tekstu; określanie kontekstu sytuacyjnego (miejsca, czasu, warunków, uczestników); rozróżnianie formalnego i nieformalnego stylu tekstu nie stanowiły przedmiotu pomiaru.

Jednolitości tematycznej oraz wąskiego zakresu sprawdzanych umiejętności nie można w tym przypadku tłumaczyć limitem czasowym i koniecznością wyboru określonych tekstów. Kluczową rolę odgrywa bowiem dywersyfikacja zadań, czyli zastosowanie tekstów zróżnicowanych pod względem tematycznym oraz skonstruowanie jednostek testowych² badających różne, a nie wciąż te same umiejętności — oczywiście z zakresu podanego w standardach egzaminacyjnych. I tej dywersyfikacji niewątpliwie tutaj zabrakło. Z tego powodu trudno uznać testy maturalne za trafne pod kątem realizacji wymagań programowych.

Trafność konstrukcyjna

Trafność konstrukcyjna, zwana też teoretyczną, „wiąże się z odpowiedzią na pytanie, czy test dokonuje pomiaru konstruktów psychologicznych leżących u podstaw testowanych umiejętności językowych” (*Przewodnik...* 2004: 36). Aby określić trafność konstrukcyjną, należy wziąć pod uwagę konstrukt testowy, który można zdefiniować w różny sposób, m.in. na podstawie programu nauczania danego języka lub teoretycznego modelu kompetencji językowej (Bachman & Palmer, 1996: 67).

Test nie spełnia kryterium trafności konstrukcyjnej m.in. wtedy, gdy nie uwzględnia wszystkich istotnych cech obiektu pomiaru — konstruktowi (*niedoreprezentowanie pojęcia*) oraz mierzy cechy czy umiejętności, które nie konstytuują konstruktowi (*uboczne źródła zmienności jak uboczna trudność lub uboczna łatwość*) (Grotjahn, 2000: 316; Niemierko, 1999: 177).

O niedoreprezentowaniu konstruktowi świadczy m.in. nieuwzględnienie w pomiarze tak istotnej jego części, jaką jest rozumienie spontanicznej mowy obcojęzycznej. W obrębie języka mówionego wyróżnia się bowiem (Dirven 1984: 21-22):

2 Pod pojęciem „jednostka testowa” / „jednostka testu” należy rozumieć „najmniejszy element testu, za który przydzielany jest punkt bądź punkty” (*Przewodnik* 2004: 15).

- spontaniczny język mówiony, np. autentyczne monologi, dialogi, polilogi;
- przygotowany język mówiony, np. wykład;
- niespontaniczny język mówiony, np. teksty wcześniej napisane, wygłoszone z pamięci lub przeczytane (sztuki teatralne, wiadomości telewizyjne).

Teksty audialne, stosowane na egzaminie maturalnym, pochodzą z różnych stron internetowych (m.in. o charakterze podróznym, edukacyjnym, informacyjnym lub młodzieżowym, np. wissen.spiegel.de, www.rp-online.de, www.geo.de). Są to odczytane teksty pisane, które nie wykazują fizycznych i lingwistycznych cech języka mówionego, takich jak np.: przerwy, wahania, wypełniacze ciszy, zróżnicowane tempo mówienia, krótkie, eliptyczne zdania, powtórzenia, ponowne rozpoczęcie zdań czy wyrażenia kolokwialne. Uczeń nie ma do czynienia z tekstami reprezentującymi styl nieformalny, zaczerpniętymi z codziennej komunikacji.

Z kolei uboczną trudność wykazują jednostki testowe, które są zbyt rozbudowane pod względem językowym, przez co mierzą nie tylko stopień opanowania sprawności słuchania, ale także pośrednio sprawdzają umiejętność czytania (tzw. *muddied measurement*, Weir 1990). Należy pamiętać, że odbiór słuchowy jest pod wieloma względami trudniejszy od odbioru wizualnego. Uczeń nie może ani wpływać na tempo mówienia, ani też w zależności od swoich potrzeb decydować o powtórnym wysłuchaniu tekstu. Podczas dwukrotnej prezentacji nagrania zdający musi wychwycić wszystkie konieczne do rozwiązania zadania informacje. Ulotność i krótkotrwałość bodźca słuchowego połączona ze zbyt rozbudowanymi zadaniami w arkuszu egzaminacyjnym może prowadzić do obciążenia pamięci roboczej i w konsekwencji negatywnie wpływać na zrozumienie tekstu (Grotjahn, 2005: 118-119).

Z uboczną trudnością mamy do czynienia w zadaniach typu wybór wielokrotny, w których występują rozbudowane przez konstrukcje imiesłowe pytania, oraz w zadaniach na dobieranie, w których również zastosowano zbyt długie pytania. Przykłady zaprezentowano poniżej:

- Zad. 6 z 2011 roku (zadanie typu wybór wielokrotny)

6.2. *Wie ist Herr Schmidt mit seiner **sich abzeichnenden** Arbeitslosigkeit umgegangen? (Jak pan Schmidt radził sobie z rysującym się bezrobociem?)*

6.3. *Wie haben die bekannten von Herrn Schmidt auf seine **zu erwartende** Arbeitslosigkeit reagiert? (Jak znajomi pana Schmidta zareagowali na jego przewidywane bezrobocie?)* (pogrubienia E.P. P.)

- Zad. 5 z 2009 roku (zadanie na dobieranie)

Ist die magische 50 für einen aktiven Menschen wie Sie nur eine Zahl oder doch mehr? (Czy ta magiczna pięćdziesiątka jest dla aktywnego człowieka, takiego jak pan, tylko liczbą czy czymś więcej?)

- Zad. 5 z 2011 roku (zadanie na dobieranie)

Wie erinnern Sie sich an den Tag, an dem die Preisträger bekannt gegeben wurden? (Jak wspomina pani dzień, w którym ogłoszono lauréatów?)

Dodać należy, że zastosowane w zadaniu 6. imiesłowy (wyrazy pogrubione) nie zawierają żadnych istotnych informacji, a pytania pozbawione tych konstrukcji zdają się brzmieć bardziej naturalnie. Prawdopodobnie zostały dodane, aby podnieść poziom trudności tego zadania. Mając na uwadze fakt, że podczas testowania rozumienia ze słuchu stosowanie umiejętności czytania powinno być ograniczone do minimum, taki zabieg nie jest dobrym rozwiązaniem konstrukcyjnym³. Zadania sprawdzające umiejętność słuchania powinna cechować precyzja i oszczędność słowna, zaś w tych pytaniach pojawiają się całkowicie zbędne elementy.

Podobny problem stanowią zbyt długie pytania w zadaniach na dobieranie. W trakcie słuchania zdający musi czytać i porównywać treść siedmiu pytań z zawartością pięciu tekstów audialnych. Im dłuższe pytania, tym więcej tekstu do czytania i tym większe obciążenie pamięci podczas przetwarzania bodźców słuchowych, co może skutkować niepełnym lub błędnym zrozumieniem tekstu audialnego (Paschke, 2001: 162).

3 Z tego względu np. w teście certyfikатовym *Preliminary English Test* (PET) zadania zbudowane są w oparciu o tekst słuchowy i ilustracje.

Innym przejawem naruszenia kryterium trafności jest tzw. uboczna ła-
twość. Pojawia się ona w zadaniach typu wybór wielokrotny, w których py-
tania zadawane przez dziennikarza nie zostały sparafrazowane, ale w nie-
mał niezmienionym brzmieniu zastosowano je w teście, np.:

- Test maturalny z 2010 roku

Uczeń słyszy: *Gefällt dir Wien? (Podoba ci się Wiedeń?)*

Uczeń czyta: 3.3. *Was gefällt Peter an Wien? (Co podoba się Piotrowi w Wiedniu?)*

Uczeń słyszy: *Was sollte man in Wien verändern? (Co powinno zmienić się w Wiedniu?)*

Uczeń czyta: 3.5. *Was sollte man in Wien verändern? (Co powinno zmienić się w Wiedniu?)*

- Test maturalny z 2011 roku

Uczeń słyszy: *Wie sieht dann Ihr Alltagsdienst aus? (Jak zatem wygląda pani codzienna służba?)*

Uczeń czyta: 3.3. *Wie sieht Sabines Alltagsdienst aus? (Jak wygląda codzienna służba Sabiny?)*

Uczeń słyszy: *Herr Schmidt, wie haben Sie Ihre Arbeitsstelle verloren? (Panie Schmidt, jak stracił pan swoją pracę?)*

Uczeń czyta: 6.1. *Warum hat Herr Schmidt seine Arbeitsstelle verloren? (Dla-
czego pan Schmidt stracił swoją pracę?)*

- Test maturalny z 2012 roku

Uczeń słyszy: *Wie seid Ihr auf diese Idee gekommen? (Jak wpadliście na ten pomysł?)*

Uczeń czyta: 3.1. *Wo sind die beiden jungen Männer auf ihre Idee gekom-
men? (Gdzie ci obaj młodzi mężczyźni wpadli na swój pomysł?)*

Uczeń słyszy: *Frau Stürmer, sagen Sie uns bitte, wie der Erfolg Ihr Leben
verändert hat. (Pani Stürmer, proszę nam powiedzieć, jak sukces zmienił
pani życie.)*

Uczeń czyta: 6.1 *Wie hat sich Christinas Leben durch ihren Erfolg
verändert?(Jak przez sukces zmieniło się życie Christiny?)*

- Test maturalny z 2014 roku

Uczeń słyszy: *Herr Pietschmann, 2003 haben Sie sich an der Schauspielschule Bochum beworben. Welche Bedingungen mussten Sie erfüllen? (Panie Pietschmann, w 2003 roku ubiegał się pan o miejsce w szkole aktorskiej w Bochum. Jakie warunki musiał pan spełnić?)*

Uczeń czyta: *6.1 Welche Bedingungen musste Andreas erfüllen, um sich bei der Schauspielschule bewerben zu können? (Jakie warunki musiał spełnić Andreas, żeby móc ubiegać się o miejsce w szkole aktorskiej?)*

Uczeń słyszy w nagraniu i jednocześnie widzi w arkuszu egzaminacyjnym to samo lub prawie takie samo pytanie. Takie rozwiązanie konstrukcyjne powoduje zrozumienie pytań pojawiających się w tekście audialnym nie ze słuchu, ale na podstawie zapisu.

Zakończenie

Najczęstszym błędem testów w zakresie trafności jest wąski zakres sprawdzanych umiejętności. Można wręcz odnieść wrażenie, że testy maturalne z języka niemieckiego tworzone są ciągle według tego samego wzorca czy schematu, w którym z pewnych, trudnych do zdefiniowania względów, nie uwzględniono pomiaru umiejętności określania intencji autora lub nadawcy tekstu, określania kontekstu sytuacyjnego (miejsca, czasu, warunków, uczestników) oraz rozróżniania formalnego i nieformalnego stylu tekstu. Ponadto w 2013 roku w testowaniu pominięto również umiejętność określania głównych myśli poszczególnych części tekstu. Tym samym test składający się 30 jednostek testowych (łącznie na poziomie podstawowym i rozszerzonym) sprawdzał tylko trzy umiejętności w zakresie rozumienia ze słuchu.

Istotny problem stanowi także stosowanie tekstów pozbawionych cech naturalnej mowy. Uczeń słyszy teksty odczytane, a nie mówione, „wyczyszczone” z odgłosów w tle, z przerw, powtórzeń, wypełniaczy ciszy itd. Z pewnością takie teksty kompensują wady, w tym także akustyczne niedostatki odbioru słuchowego w sytuacji egzaminacyjnej. W codziennej komunikacji uczeń ma jednakże do czynienia z różnymi tekstami, na

pewno nie tak „czystymi”. Ten fakt również powinien być brany pod uwagę podczas wyboru i nagrywania tekstów w celach testowania.

Kolejnym powtarzającym się w testach błędem jest wykorzystywanie w takiej samej lub nieznacznie zmienionej formie zdań i wyrażań z tekstu słuchowego w pytaniach testowych. Taki zabieg stanowi znaczne ułatwienie odbioru tekstu słuchowego (uboczna łatwość), dlatego trudno na podstawie tak skonstruowanych jednostek testowych wnioskować o stopniu opanowania sprawności słuchania ze zrozumieniem.

Najwięcej błędów w zakresie trafności zawartych jest w teście z 2011 roku. Obok powyżej opisanych, występują również inne istotne niedociągnięcia, jak np. jednorodność tematyczna tekstów audialnych i unikanie tematów bardziej abstrakcyjnych, jak np. kultura, państwo i społeczeństwo, oraz zastosowanie zbyt rozbudowanych jednostek testowych (uboczna trudność).

Na koniec warto jeszcze raz wyraźnie podkreślić, że „nadrzędnym celem jakiegokolwiek formy testowania jest takie zbadanie poziomu umiejętności językowych zdającego, by można było określić, jak poradzi on sobie w sytuacjach pozatestowych. Test, który czyni to w sposób satysfakcjonujący, jest trafny” (*Przewodnik...* 2004: 32). W świetle powyższego rodzi się pytanie, czy testy maturalne z języka maturalnego rzeczywiście w sposób satysfakcjonujący mierzą umiejętność rozumienia ze słuchu.

Bibliografia

- Bachman, L. F. & Palmer, A. S. 1996. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Dirven, R. 1984. Was ist Hörverstehen? Synopse vorhandener Theorien und Modelle. In Schumann, A. & Vogel, K. & Voss, B. (eds). 1984. *Hörverstehen. Grundlagen, Modelle, Materialien zur Schulung des Hörverständnisses im Fremdsprachenunterricht der Hochschule*. Tübingen: Narr, 19-40.
- Grotjahn, R. 2000. Testtheorie: Grundzüge und Anwendungen in der Praxis. In Wolff, A. & Tanzer, H. (eds). *Sprache — Kultur — Politik: Beiträge der 27. Jahres-*

- tagung *Deutsch als Fremdsprache 3*. — 5. Juni 1999 an der Universität Regensburg. (*Materialien Deutsch als Fremdsprache Bd. 53*). Regensburg: Fachverband Deutsch als Fremdsprache, 304-341.
- Grotjahn, R. 2005. Testen und Bewerten des Hörverstehens. In Ó Dúill, M. & Zahn, R. & Höppner K.D.C. (eds). *Zusammenarbeiten. Eine Festschrift für Bernd Voss*. Bochum: AKS-Verlag, 115-144.
- Komorowska, H. 2004. *Sprawdzanie umiejętności w nauce języka obcego. Kontrola — Ocena — Testowanie*. Warszawa: Fraszka Edukacyjna.
- Niemierko, B. 1999. *Pomiar wyników kształcenia*. Warszawa: WSiP.
- Paschke, P. 2001. Zum Problem der Authentizität in L2-Hörverstehenstests. *Fremdsprachen lehren und lernen*. 30: 150-166.
- Przewodnik dla autorów zadań do testów językowych*. 2004. Gaszyńska-Magiera, M., Seretny, A. (tłum. i adapt.). Kraków: Universitas.
- Rozporządzenie Ministra Edukacji Narodowej i Sportu z dnia 10 kwietnia 2003 roku zmieniające rozporządzenie w sprawie standardów wymagań będących podstawą przeprowadzania sprawdzianów i egzaminów. Załącznik nr 3. Dz. U. nr 90, poz. 846.
- Weir, C.J. 1990. *Communicative language testing*. London: Prentice Hall.