

**Sabina Nowak**

Uniwersytet Jagielloński, Jagiellońskie Centrum Językowe

## Trafność egzaminu z języków obcych w gimnazjum

Niniejszy artykuł przedstawia jedno z wielu możliwych podejść do analizy trafności egzaminu z języków obcych w gimnazjum, które oparte jest głównie na sposobie badania trafności zaproponowanym przez Cyrila Weir<sup>1</sup>. Autorka artykułu zwraca się z prośbą o wyrażenie opinii na temat zaprezentowanej poniżej struktury badań nad trafnością egzaminu z języków obcych w gimnazjum w celu ustalenia mocnych i słabych stron owej koncepcji badań.

Celem głównym badania jest określenie trafności egzaminu z języków obcych w gimnazjum, a więc odpowiedź na pytanie, czy test mierzy to, co ma mierzyć lub co tak naprawdę mierzy. Proces weryfikacji trafności egzaminu gimnazjalnego polega na zbieraniu miarodajnych i przydatnych dowodów związanych z badaniem różnych rodzajów trafności. Podstawowym jego celem jest więc odpowiedź na pytanie, czy test może być stosowany do sprawdzenia umiejętności językowych uczniów i czy spełnia kryteria trafności opartej na teorii, trafności kontekstu, oceniania i treści oraz trafności kryterialnej. Źródła danych walidacyjnych przedstawionych na schemacie (Schemat 1.) wzajemnie się uzupełniają i żaden z przedstawionych rodzajów trafności nie jest nadrzędny w stosunku do innego.

Z uwagi na fakt, iż rozumienie znaczenia pojęcia trafności ciągle się zmienia (Standardy, 2007, s. 27–28), w proponowanym badaniu przyjmuje się analizę różnych rodzajów danych, które mogą świadczyć o trafności egzaminu z języków obcych w gimnazjum. Proces ten polegać będzie na ciągłym zbieraniu danych walidacyjnych w latach 2007–2013, aby dostarczyć mocnych podstaw naukowych do interpretacji wyników testu oraz ich użyteczności.

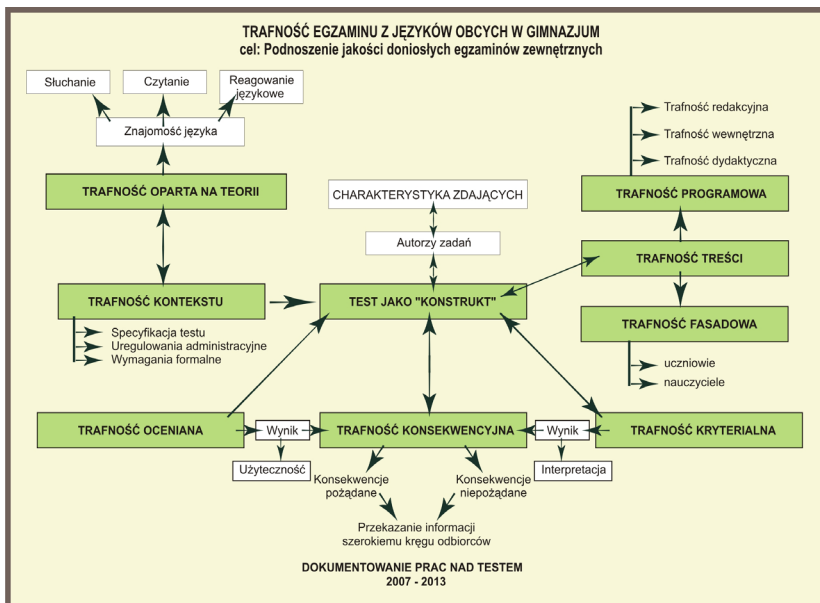
### Test jako „konstrukt”

Punktem wyjścia w procesie walidacji jest próba określenia „konstrukt”, a zatem zdefiniowania umiejętności językowych mających stanowić cel pomiaru, które zgodnie z założeniem ma mierzyć test. Pojęcie „konstrukt” zaproponowane przez autorów wartościowej pozycji *Standardy dla testów stosowanych w psychologii i pedagogice* (2007, s. 300) oznacza „zmienną teoretyczną, o której wnioskuje się na podstawie danych pochodzących z wielu źródeł. Mogą to być: korelacje wyników testowych z innymi zmiennymi, struktura wewnętrzna testu, obserwacja procesu odpowiadania, jak również treść testu”, gdzie „wszystkie wyniki są traktowane jako miara jakiegos

---

<sup>1</sup> Cyril Weir jest profesorem na Uniwersytecie Bedfordshire i dyrektorem Centre for Research in English Language Learning and Assessment (CRELLA).

konstruktu” (tamże, 300). W celu lepszego określenia „konstruktu” proces walidacji polegać będzie na potwierdzaniu trafności na wiele sposobów, przy czym zachowane zostanie tradycyjne nazewnictwo związane z trafnością. Analiza trafności dokonana będzie na podstawie danych empirycznych oraz teorii leżących u podstaw testu.



Schemat 1. Trafność egzaminu z języków obcych w gimnazjum

## Charakterystyka zdających

Pierwszym ważnym elementem procesu walidacji jest dokładne zebranie informacji na temat **zdających** w szeroko rozumianym kontekście nauczania i uczenia się języków obcych, które zbierane są przede wszystkim za pomocą prowadzonych równolegle badań kontekstowych. Podstawowa charakterystyka populacji zdających obejmuje za O'Sullivanem (2000) charakterystykę *fizyczną/fizjologiczną* (wiek, płeć, dysleksja, lekkie upośledzenia typu niedosłyszenie, niedowidzenie, itd.), *psychologiczną* (osobowość, pamięć, motywację, itd.), oraz *empiryczną* (wykształcenie, stan przygotowania do egzaminu, doświadczenia egzaminacyjne, żywy kontakt z językiem). Informacje tego typu przekazane autorom zadań i konstruktorom arkuszy powinny się przyczynić do lepszego określenia potrzeb uczniów gimnazjum przystępujących do egzaminu w klasie trzeciej. Dodatkowo, dane na temat **autorów zadań**, ich doświadczenia zawodowego i kwalifikacji, przebytych dodatkowych szkoleń w trakcie prac nad tworzeniem testu oraz opisanie procedur dotyczących określania i generowania treści testu będą wartościową informacją potwierdzającą trafność prac nad egzaminem z języków obcych w gimnazjum, ponieważ to głównie na nich spada odpowiedzialność za jakość zadań zamieszczonych w arkuszu egzaminacyjnym.

## Trafność oparta na teorii

*Trafność oparta na teorii* uwarunkowana jest komunikacyjnym modelem testowania<sup>2</sup> znajomości języka, co znaczy, że język którego uczą się zdający powinien ściśle wiązać się z ich obecnymi i przyszłymi potrzebami osadzonymi w sytuacjach jak najbardziej autentycznych (*Przewodnik ...*, 2004, s. 21). Dokonanie takiej analizy potrzeb pozwala na lepsze ukazanie profilu biegłości językowej, jaką zdający mają osiągnąć. Innymi słowy, nie jest możliwe zbadanie testem poziomu umiejętności językowych bez określenia kompetencji językowych zdających. Jednak precyzyjne zdefiniowanie i wyłonienie wszystkich komponentów czy to wiedzy, czy umiejętności, jest sprawą tyle trudną co kontrowersyjną.

Ciekawą propozycję badania *trafności opartej na teorii* przedstawił Cyril Weir<sup>3</sup> w swojej książce *Language Testing and Validation*, który zaproponował oddzielne struktury walidacji testów słuchania, czytania, pisania i mówienia. Według niego każdej z tych umiejętności odpowiadają procesy wewnętrzne zdającego, które związane są z procesami oraz zasobami wykonawczymi, ciągle monitorowanymi przez zdającego pod względem ich efektywności. Tabela 1. przedstawia procesy wewnętrzne dla umiejętności słuchania i czytania.

Tabela 1. Trafność oparta na teorii dla słuchania i czytania wg Weir'a (2005, s. 44–47)

| Procesy wewnętrzne |   |              |  |
|--------------------|---|--------------|--|
| Umiejętności       | Procesy wykonawcze  | MONITROWANIE | Zasoby wykonawcze  |
| Słuchanie          | <ul style="list-style-type: none"> <li>Wyznaczenie celu</li> <li>Identyfikacja wizualna/akustyczna</li> <li>Identyfikacja dźwięków</li> <li>Analiza struktury tekstu</li> </ul> |              | <p><u>Wiedza na temat języka</u></p> <ul style="list-style-type: none"> <li>gramatyczna, rodzaju dyskursu, użyteczna, socjolingwistyczna</li> </ul> <p><u>Wiedza na temat treści</u></p> <ul style="list-style-type: none"> <li>wewnętrzna i zewnętrzna</li> </ul> |
| Czytanie           | <ul style="list-style-type: none"> <li>Wyznaczenie celu</li> <li>Identyfikacja wizualna</li> <li>Analiza struktury tekstu</li> </ul>  |              | <p><u>Wiedza na temat języka</u></p> <ul style="list-style-type: none"> <li>gramatyczna, tekstowa, użyteczna, socjolingwistyczna</li> </ul> <p><u>Wiedza na temat treści</u></p> <ul style="list-style-type: none"> <li>wewnętrzna i zewnętrzna</li> </ul>         |
|                    |   |              |  |

Decyzją ekspertów ustanowiono trzy obszary umiejętności uczniów przystępujących do egzaminu z języków obcych w gimnazjum, a mianowicie: odbioru tekstu *słuchanego, czytanego i reagowania językowego*. Nadanie wag owym umiejętnościom (Tabela 2.) oraz wygodna w kalkulacji długość skali wyników (50 punktów) pozwalają na zobiektywizowanie i usprawnianie sposobów sprawdzania osiągnięć zdających.

<sup>2</sup> Do najwybitniejszych twórców modeli komunikacyjnych zaliczyć możemy: Cziko (1982), Canale i Swain (1981,1983), Cummins (1979), Marrow (1979). Jednak jednym z najlepszych modeli komunikacyjnych jest model Palmera i Bachmana z 1994 roku (Źródło: *Przewodnik*, 2004, s. 24–32).

<sup>3</sup> Weir C., *Language Testing and Validation*. Palgrave Macmillan 2005.

**Tabela 2. Założone znaczenie treściowe wyników egzaminu**

| Umiejętność         | Waga | Liczba punktów |
|---------------------|------|----------------|
| Słuchanie           | 20%  | 10             |
| Czytanie            | 40%  | 20             |
| Reagowanie językowe | 40%  | 20             |
| Razem               | 100% | 50 pkt.        |

Badanie trafności opartej na teorii polegać będzie na próbie odpowiedzi na pytanie, czy oczekiwania wobec zdających są precyzyjnie i wyraźnie wyznaczone w odniesieniu do umiejętności słuchania, czytania i reagowania językowego.

### Trafność kontekstu

Do określenia trafności kontekstu wykorzystano socjokognitywną strukturę procesu walidacji Cyrila Weir. Proces ten polega głównie na określeniu *specyfikacji zadań, uregulowaniach administracyjnych i wymaganiach formalnych*. Tabela 3. przedstawia trzy dziedziny, które muszą zostać wzięte pod uwagę od momentu tworzenia testu do czasu jego przeprowadzania.

**Tabela 3. Trafność kontekstu (Weir, 2005, s. 44–47)**

| Trafność kontekstu  |  |  |
|---|--|--|
| I. Specyfikacja zadań   | II. Uregulowania administracyjne   | III. Wymagania formalne  |
| 1. instrukcje<br>2. cel<br>3. format odpowiedzi<br>4. znajomość kryteriów oceniania<br>5. wagowanie<br>6. kolejność zadań<br>7. limit czasu | 1. jednolitość warunków i procedur przeprowadzania egzaminu<br>2. bezpieczeństwo testu<br>3. dostosowanie testu dla uczniów o specjalnych potrzebach | 1. rodzaj dyskursu<br>2. długość tekstu/liczba nagrań<br>3. związek pomiędzy autorem a odbiorcą tekstu<br>4. rodzaj informacji<br>5. wiedza ogólna zdającego (leksykalna, strukturalna, użyteczna)<br>6. akcent, liczba mówiących w nagraniu |

### Trafność oceniania

Trafność oceniania jest według Weir'a terminem obejmującym wszystkie aspekty dotyczące rzetelności testu, które mają wpływ na ogólną jego trafność (Weir, 2005, s. 22). Trafność ta dotyczy wyników testu, które są niezmiennie wobec upływu czasu, konsekwentne jeżeli chodzi o dobór treści oraz wolne od stronniczości (*bias*). Do podstawowych danych na temat rzetelności należy „określenie głównych źródeł błędów, zbiorcze dane statystyczne dotyczące wielkości tych błędów oraz przedstawienie zakresu, w jaki te dane można uogólniać na inne formy testu, inne osoby oceniające i inne procedury czy inne znaczące wymiary” (Standardy, 2007, s. 59).

Tradycyjnie wyróżnić można cztery kategorie współczynników rzetelności: „(a) współczynniki otrzymane na podstawie niezależnego, dwukrotnego badania formami równoległymi testu; (b) współczynniki otrzymane na podstawie dwukrotnego badania tym samym testem (*test-retest*); (c) współczynniki otrzymane na podstawie jednokrotnego badania testem i oparte na wielkościach korelacji między wynikami dla poszczególnych pozycji testu czy skal” (tamże, s. 60). Czwartą kategorią jest rzetelność egzaminatorów dotycząca zgodności ocen między egzaminatorami (*Inter-rater reliability*) oraz ocena porównywalności/zgodności wyników tego samego egzaminatora przy ocenie tej samej cechy (*Intra-rater reliability*).

### Trafność treści

W niniejszym badaniu przyjęto podział *trafności treści* na *trafność programową* i *fasadową*. Porównanie treści zawartych w teście z programem kształcenia odgrywa szczególną rolę w egzaminach doniosłych. Trafność treści definiuje się jako „stopień zgodności treści testu z treściami nauczania” (*Przewodnik...*, 2004, s. 33). W przypadku egzaminu gimnazjalnego będzie to analiza porównawcza materiału zawartego w teście z materiałem zawartym w podstawie programowej oraz standardach egzaminacyjnych. O stopniu zgodności tych treści decydować będą *sędziowie kompetentni*, zaś samo badanie trafności treści testu będzie poniekąd oceną sposobu postrzegania, rozumienia biegłości zaproponowanej przez twórców testu. Tak więc pytanie, *czy test mierzy to, co ma mierzyć i w jakim zakresie* będzie tu miało olbrzymie znaczenie.

*Trafność programową* podzielono na *trafność dydaktyczną*, *wewnętrzną* i *redakcyjną* (por.: Szyling, 2004, s. 17). Trafność *dydaktyczna* określona zostanie na podstawie zgodności standardów egzaminacyjnych z podstawą programową i koncepcją egzaminu gimnazjalnego oraz stwierdzeniu, czy wybrane umiejętności odgrywają istotne znaczenie w procesie kształcenia. Trafność *wewnętrzna* oparta zostanie na przyporządkowaniu obszarów umiejętności, które określają najistotniejsze kompetencje uczniów gimnazjum, które są przypisane do danych standardów w sposób uzasadniony i konsekwentny, a także zbadanie, czy obszary standardów są rozłączne. Trafność *redakcyjna* sprawdzi poprawne sformułowanie standardów, a więc czy stopień uszczegółowienia pozwala na ich zrozumienie i jednoznaczną interpretację, czy standardy są jednoznaczne pod względem pojęć i terminów oraz czy standardy są poprawnie sformułowane pod względem językowym i stylistycznym.

W wyniku przeprowadzenia badania *trafności fasadowej* otrzymamy informację zwrotną m.in. o tym, jak *uczniowie i nauczyciele* oceniają zadania zamieszczone w arkuszu egzaminacyjnym. Będzie to nie tylko ocena łatwości zadań, ale także ocena zrozumienia poleceń do zadań oraz oszacowanie czasu przeznaczanego na wykonanie zadań określonego typu np. zadań testujących rozumienie tekstów słuchanych, jak również czasu potrzebnego na wykonanie pełnego arkusza egzaminacyjnego. Rola nauczyciela będzie tu bardziej złożona, gdyż oprócz tych samych pytań, które zostały skierowane do uczniów, nauczyciele zostaną poproszeni o udzielenie odpowiedzi na temat typologii zadań oraz adekwatności standardów egzaminacyjnych.

## Trafność kryterialna

Ustalenie trafności kryterialnej będzie się wiązać po pierwsze z procedurą ustalenia korelacji pomiędzy wynikami nowego egzaminu z języków obcych w gimnazjum z zewnętrznym kryterium, jakim będzie ocena szkolna, po drugie z porównaniem wyników dwóch wersji egzaminu gimnazjalnego (za pomocą IRT), po trzecie z porównaniem wyników tego samego egzaminu przeprowadzonego o różnych porach dnia oraz porównaniem wyników testu z przyszłym wynikiem z innego testu (np. egzaminem maturalnym).

## Trafność konsekwencyjna

Zbieranie danych nawiązujących do trafności konsekwencyjnej jest bardzo istotne w przypadku doniosłych egzaminów zewnętrznych. Potwierdzenie dowodów dokonywane będzie tutaj w trzech obszarach:

1. Trafności różnicowej polegającej na określeniu stronniczości testu (*test bias*<sup>4</sup>) w kontekście jego bezstronności (*fairness*), tak aby egzamin gimnazjalny jednakowo sprzyjał zdającym z miasta i ze wsi, niezależnie od ich płci czy wykształcenia i aspiracji rodziców, itd. oraz aby wyniki zdających były jak najmniej zależne od wybranej formy zadań (sprytu testowego, łatwości pisanie, czy w końcu zgadywania) i od egzaminatora, który będzie oceniał wyniki.
2. Szeroko rozumianego wpływu (tzw. *washback/impact*<sup>5</sup>), jaki wywierać może egzamin na wszystkich jego beneficjentów (zwanym również *udziałowcami*, ang. *stakeholders*).
3. Badania efektu, jaki egzamin wywarł na społeczeństwo

W określeniu *trafności konsekwencyjnej* nie jest wymagane dostarczenie danych empirycznych, lecz możliwe jest przewidywanie „na podstawie przesłanek” (Szyling, 2004, s. 15). W niniejszym badaniu przyjęto analizę *pożądanych i niepożądanych konsekwencji testowych egzaminów* w nawiązaniu do kontekstu polskiego egzaminu z języków obcych w gimnazjum. Analiza *pożądanych i niepożądanych konsekwencji* wynikających z wprowadzenia nowego egzaminu w gimnazjum przedstawiona zostanie według następujących obszarów: wpływ na efektywność pracy w szkołach, motywacja uczniów, modyfikowanie/ulepszenie programów nauczania oraz polityka państwa (Standardy, 2007, s. 42–43).

<sup>4</sup> „Stronniczość może być rozumiana jako niedoreprezentowanie konstruktów bądź obecność elementów niezwiązanych z konstruktem, które w różny sposób wpływają na poziom wykonania testu przez różne grupy osób badanych” (Standardy, 2007, s. 298).

<sup>5</sup> *Washback* w kontekście edukacyjnym definiowany jest jako zmiany w zachowaniu lub wpływ na program i metody nauczania, strategie uczenia się, materiały do nauki, praktyka związana z ocenianiem i instrukcje. Natomiast „wpływ” (*impact*) związany jest z wartościami, motywacją i potencjalnymi konsekwencjami testu w dziedzinach społecznych (Liz Hamp Lyons, Research Notes no 31, Feb. 2008, s. 42).

## Przekazanie informacji na temat egzaminu szerokiemu kręgu odbiorców

Ze względu na różne grupy odbiorców i ich oczekiwania wszelkie informacje przekazywane do wiadomości publicznej powinny spełniać kryteria użyteczności, czytelności, dostępności, różnorodności form, drożności kanałów informacyjnych oraz skuteczności przepływu informacji. Publikacje obejmować powinny całość zagadnień informujących uczniów o wszelkich sprawach związanych z egzaminem.

Informacja na temat egzaminu powinna mieć charakter analityczny (opisowy), który odzwierciedli w adekwatny sposób jakość nabytych kompetencji, umiejętności i wiadomości we wszystkich testowanych obszarach, tj. odbiorze tekstu słuchanego, odbiorze tekstu czytanego oraz reagowaniu językowym. Pozwoli to także wyciągnąć konkretne wnioski na temat jakości procesu nauczania oraz skuteczności osiągania celów i na refleksyjną ewaluację procesu nauczania nad metodami nauczania, doбором celów, treści, programów nauczania i podręczników.

Dostarczenie różnym grupom odbiorców informacji na temat samego testu, jak również materiałów ułatwiających przygotowanie się do egzaminu powinno przyczynić się do efektywniejszego przygotowania zdających do egzaminu, a więc do podniesienia jakości egzaminu zewnętrznego, jakim jest egzamin z języków obcych w gimnazjum.

### Bibliografia:

1. Biuletyn CKE nr 1, 2004, [http://www.oke.krakow.pl/inf/filedata/files/biuletyn\\_1\\_2004.pdf](http://www.oke.krakow.pl/inf/filedata/files/biuletyn_1_2004.pdf)
2. *Listy zagadnień kontrolnych do analizy testów biegłości językowej*, tłum. Gaszyńska-Magiera M., Seretny A., Universitas, Kraków 2004.
3. O'Sullivan B., *Towards a model of Performance in Oral Language Testing*. Niepublikowana rozprawa doktorska. Uniwersytet Reading 2000, (w:) Weir C., *Language Testing and Validation*) 2005.
4. *Przewodnik dla autorów zadań do testów językowych*, tłum. Gaszyńska-Magiera M., Seretny A., Universitas, Kraków 2004.
5. *Standardy dla testów stosowanych w psychologii i pedagogice*, przekł. Hornowska E., Gdańskie Wydawnictwo Psychologiczne 2007.
6. Szyling G., Biuletyn Badawczy CKE 2004 nr 1.
7. Weir C., *Language Testing and Validation*. Palgrave Macmillan 2005.