

**TEST UZDOLNIEŃ
DO NAUKI JĘZYKÓW OBCYCH
(TUNJO)**

Opis, analiza, statystyki

Autor adaptacji: dr Jacek Rysiewicz

Instytut Filologii Angielskiej
Uniwersytet im. A. Mickiewicza
Poznań
rjacek@ifa.amu.edu.pl

Współpraca: studenci seminarium magisterskiego
2006/2008

Poznań, 2011

CHARAKTERYSTYKA TESTU

Nazwa narzędzia:	Test Uzdolnień do Nauki Języków Obcych
Przeznaczenie:	- diagnoza profilu uzdolnień językowych ucznia - prognozowanie powodzenia w uczeniu się języka obcego w warunkach formalnej jak i nieformalnej nauki - badania
Rodzaj testu:	test uzdolnień poznawczych; test mocy z ograniczeniami czasowymi na zadanie test typu 'papier i ołówek'
Populacja docelowa:	od około 17 lat wzwyż
Czas trwania:	średnio 55 min. (od 53 do 60 min)
Przeprowadzanie:	grupowe lub indywidualnie
Dostępność:	ograniczona i tylko dla celów badawczych (po uprzednim kontakcie z autorem)
Ilość zadań w teście:	5
Materiały testowe:	zeszyt z zadaniami arkusz odpowiedzi płyta CD z zadaniem nr I krótki opis interpretacji wyników

WSTĘP

Test Uzdolnień do Nauki Języków Obcych (TUNJO) jest baterią testową składającą się z pięciu testów mierzących predyspozycje (uzdolnienia) do nauczania się języków obcych. Bateria TUNJO jest w dużym stopniu adaptacją na grunt polski testu MLAT (*Modern Language Aptitude Test*) skonstruowanego przez Carrolla i Sapona w 1959 roku w oparciu o model uzdolnień językowych tychże autorów (patrz dalej).

Jest kilka powodów, dla których podjęto się adaptacji testu MLAT. Pierwszym z nich jest brak rodzimego narzędzia, które by w sposób rzetelny i teoretycznie trafny umożliwiło pomiar zdolności do nauczania się języków obcych. Drugim powodem, jest fakt, że pojawiające się w Polsce w ciągu ostatnich lat doraźnie konstruowane testy uzdolnień językowych, w rzeczywistości są testami umiejętności (kompetencji) językowych języka polskiego i, jako takie, nie nadają się do celów diagnostycznych a tym bardziej badawczych. Trzecim powodem jest chęć dostarczenia polskim badaczom narzędzia, które przez fakt, że jest oparte na sprawdzonym modelu uzdolnień językowych, umożliwi im nawiązanie dyskursu naukowego z badaczami z innych krajów poprzez względną porównywalność konstruktów obu narzędzi. Czwarty powód jest typowo praktyczny i łączy się z docelowym zastosowaniem testu jako narzędzia o funkcji prognostycznej i diagnostycznej w praktyce dydaktycznej języków obcych.

Model uzdolnień językowych Carrolla i Sapona

Klasyczny, czteroczynnikowy model uzdolnień językowych Carrolla i Sapona został wyłoniony w rezultacie długotrwałych, bo trwających pięć lat, badań empirycznych. W trakcie tych badań Carroll i Sapon poddali analizie statystycznej ogromną ilość różnych zadań pod kątem ich korelacji ze zmienną kryterialną, tj., wynikami w nauce języków obcych. Wyniki przez nich uzyskane zostały poddane ilościowej analizie korelacyjnej w celu wyłonienia tych zdolności, które okazały się być najbardziej predyktywne w stosunku do wyników testów językowych. W etapie następnym, po zastosowaniu analizy czynnikowej w

celu wydobycia czynników wspólnych, obecnych w kilku różnych testach baterii, badacze poddali wyłonione czynniki interpretacji jakościowej w oparciu o ujawniony empirycznie model wzajemnych zależności pomiędzy czynnikami. Ostatnim krokiem badania było nazwanie wyodrębnionych w ten sposób czynników i określenie ich względem zdolności w nich się przejawiających (Skehan, 1989; Stansfield, Reed, 2004). Opisana procedura doprowadziła autorów do zaproponowania modelu predyspozycji językowych, na który składały się następujące zdolności:

1. kojarzenia formy dźwiękowej z graficzną, tzw. kodowanie fonetyczne (*phonemic coding ability*)
2. rozpoznawania gramatycznej funkcji wyrazów/fraz w zdaniu (*grammatical sensitivity*)
3. formułowania uogólnień w postaci reguł językowych w oparciu o nieznaną (fikcyjną) materiał językowy, tzw., wnioskowanie indukcyjne (*inductive language learning*)
4. wzrokowego zapamiętywania znaczenia słów, przechowywania i odtwarzania zapamiętanych znaczeń (*associative memory*).

Tak wypracowany model uzdolnień językowych posłużył następnie autorom do skonstruowania baterii testowej, która była odwzorowaniem poszczególnych komponentów modelu uzdolnień. Autorom nie do końca udało się takie zoperacjonalizowanie swojego modelu by, z jednej strony, każde zadanie było odwzorowaniem tylko jednej z czterech zdolności, i z drugiej, by zadania nie duplikowały się nawzajem pozostając słabo lub nisko ze sobą skorelowane.

Opis testu MLAT

Bateria MLAT składa się z pięciu testów. Pierwsza część testu (MLAT 1) to „*Number learning*” (Uczenie się liczb). W teście, tym, składającym się z 43 pozycji, zadanie polega na nauczaniu się systemu liczbowego fikcyjnego języka w zakresie jedności (liczby: 0, 1, 2, 3 i 4), dziesiątek (liczby: 10, 20, 30 i 40) oraz setek (liczby: 100, 200, 300 i 400), zaprezentowanych odsłuchowo, przećwiczenia tego systemu w fazie ćwiczeniowej i następnie, w fazie testowej, rozpoznaniu liczb utworzonych w tym systemie a zaprezentowanych odsłuchowo. Test ten mierzy pamięć mechaniczną i wnioskowanie indukcyjne oraz, pośrednio, gotowość słuchową.

Druga część testu (MLAT 2) to „*Phonetic script*” (Zapis fonetyczny) i mierzy zdolność łączenia formy dźwiękowej wyrazu z jego zapisem graficznym (literowym), przechowania tego połączenia w pamięci fonetycznej na czas trwania zadania i odtworzenia go w fazie testowej. W fazie prezentacji zadanie polega na wysłuchaniu trzydziestu serii, czterech jednosylabowych ‘słów’, po pięć w każdej grupie, przy jednoczesnym śledzeniu ich zapisu dokonanego przy użyciu mieszaniny znanych i obco wyglądających ‘liter’. Dźwięki i towarzyszące im ‘litery’ są wprowadzane progresywnie w opozycjach par minimalnych bądź prawie-minimalnych (*sit-set*). W fazie testowej, zadaniem badanych jest odsłuchowe rozpoznanie jednego z czterech wyrazów z każdej serii i zaznaczenie odpowiedniego zapisu literowego na karcie odpowiedzi.

Trzecia część testu (MLAT 3) „*Spelling Clues*” (Co to za słowo?) polega na odgadnięciu jakie słowo języka angielskiego kryje się pod zniekształconym zapisem ortograficznym. Zadaniem zdającego jest wskazanie słowa o bliskoznacznym znaczeniu do słowa ukrytego, tak jak w przykładzie poniżej:

srtñ

(A. *foolish*; B. *honest*; C. *complaint*; D. *sure*; E. *fatal*)

gdzie ‘srtñ’ jest fonetycznym zapisem wyrazu ‘*certain*’ (ang.) [pewien] będącego przymiotnikiem, którego bliskoznacznikiem jest wyraz ‘*sure*’ (ang.). Ta część testu mierzy

kodowanie fonetyczne i, po części, jest testem znajomości słownictwa języka ojczystego. Test zawiera 50 pozycji testowych.

Część czwarta testu („*Words in sentences*” – Słowa w zdaniach) polega na analizie struktury gramatycznej zdań bez odwoływania się do świadomej wiedzy gramatycznej czy do terminologii metajęzykowej. Zadanie polega na wskazaniu, który z pięciu podkreślonych wyrazów/fraz w zdaniu pełni taką samą funkcję jak oznaczony wyraz w zdaniu modelowym, tak jak w przykładzie poniżej,

Jill fell down AND Jack came tumbling after.

[Jill spadła a Jack pokoziołkował za nią]

Now, you may wait out there, or you may come back on Friday if you wish.

A

B

C

D

E

[Więc tak. Możesz sobie tam poczekać, albo przyjsć ponownie w piątek, jeśli chcesz.]
gdzie prawidłową odpowiedzią jest C (funkcja łącząca dwa zdania proste – spójnik). Test ten jest operacjonalizacją postulowanego przez Carrola konstruktów ‘wyczucie gramatyczne’ (*grammatical sensitivity*).

Ostatnia, piąta część testu (MLAT 5) to „*Paired associates*” (Pary znaczeniowe) i jest miarą pamięci rozumianej jako mechaniczne, bez-kontekstowe zapamiętywanie znaczeń i odpowiadających im kombinacji liter (wyrazów) w nieznanym języku. Na etapie prezentacji, zadaniem zdającego jest nauczenie się jak największej ilości znaczeń 24 wyrazów z języka przypominającego kurdyjski. Następnie, w etapie ćwiczeniowym, pisemne przećwiczenie tych wyrazów polegające na wpisaniu ich znaczeń w arkuszu ćwiczeń na liście o kolejności innej niż ta z etapu prezentacji, oraz, w etapie testowym, wybranie jednego z pięciu znaczeń dla każdego z 24 słów języka odlegle przypominającego kurdyjski. W zadaniu tym werstraktory i dystraktory pochodzą wyłącznie z puli 24 słów tegoż języka.

W kontekście wcześniejszych uwag pod adresem prób skonstruowania testu uzdolnień językowych w Polsce ważne jest to, że MLAT nie mierzy nabytej wiedzy/umiejętności językowych języka ojczystego (nie jest testem kompetencji ani umiejętności), ale uzdolnienia do nauki języka obcego. W rozumieniu opozycji terminologicznej: ‘zdolności’ a ‘umiejętności’, żaden z pięciu testów baterii MLAT nie odwołuje się bezpośrednio do już nabytej wiedzy językowej badanego. Tylko zadanie MLAT 3 („*Spelling clues*” – Co to za słowo?) odwołuje się pośrednio do znajomości rodzimego słownictwa, a poziom wykonania zadania MLAT 4 („*Words in sentences*” – Słowa w zdaniach), chociaż opartego na materiale języka ojczystego, nie jest uzależniony od znajomości tegoż języka lecz od umiejętności rozpoznania funkcji wyrazów w zdaniu.

Pomimo ukazania się innych propozycji testów zdolności językowych od czasu powstania MLATu, takich jak PLAB (*Pimsleur Language Aptitude Battery*, Pimsleur, 1966) czy CANAL-F (*Cognitive Ability for Novelty in Language Acquisition – Foreign*, Grigorenko, Sternberg, Ehrman, 2000), model uzdolnień językowych Carrola i Sapona w dalszym ciągu zachowuje wysoką trafność (nie tylko diagnostyczną) i to pomimo radykalnie odmiennych paradygmatów naukowych kształtujących ówczesną a obecną dydaktykę języków obcych.

Zagadnienia adaptacji testów.

Proces dostosowywania narzędzia pomiaru (test, kwestionariusz, ankieta) skonstruowanego w jednym kraju do warunków kraju docelowego nazywa się kulturową adaptacją narzędzia pomiaru. W zależności od tego jak odległe od siebie są kultura wyjściowa i docelowa, oraz w zależności od tego czy pomiar dotyczy wymiarów kognitywnych (inteligencja, umiejętności, zdolności) czy afektywnych (przekonania, postawy, motywacje, lęki itp.) adaptacja narzędzia może być mniej lub bardziej czasochłonna, wymagająca i absorbująca. Decydując się na adaptację testu MLAT raczej niż na skonstruowanie nowego testu opartego na modelu

Carrolla i Sapona, autor kierował się kilkoma przesłankami, które bardzo trafnie oddają Hornowska i Paluchowski (2004:190) w cytacie poniżej:

"Warto poddawać adaptacji kulturowej obce testy, ale tylko te, których nie warto dublować ich polskimi odpowiednikami, gdyż prościej będzie je właśnie zaadaptować, albo te, które zajmują uznaną już w psychologii światowej pozycję, takie jak WAIS-R, (Wechsler Adult Intelligence Scale-Revised) MMPI (Minnesota Multiphasic Personality Inventory), SRPM (Standard Raven Progressive Matrices) czy testy projekcyjne: Rorschacha i Test Apercepcji Tematycznej (TAT) – żeby odwołać się do najbardziej spektakularnych przykładów"

Poza względami natury ekonomicznej oraz tymi odnoszącymi się do ugruntowanej pozycji narzędzia w praktyce badawczej są również inne powody, dla których bardziej wskazane jest dokonanie adaptacji narzędzia niż tworzenie go od podstaw. Jednym z ważniejszych jest brak lokalnych specjalistów z zakresu mierzonej cechy (wymiaru) i, co za tym idzie, niewystarczający poziom rozpoznania teoretycznego wymiaru badanego oraz brak badań podstawowych w tym zakresie. Przy braku specjalistów zajmujących się uzdolnieniami językowymi w Polsce, którzy by byli w stanie skonstruować teoretycznie trafne i rzetelne narzędzie do pomiaru tychże, ten obszar różnic indywidualnych, przez swoją względną atrakcyjność i pozytywną percepcję społeczną, stał się domeną działań nie zawsze kompetentnych i teoretycznie przygotowanych twórców testów uzdolnień językowych, które na ogół tylko z nazwy są testami tego wymiaru.

Wymienione wyżej powody skłoniły autora niniejszego testu do dokonania adaptacji kulturowej baterii MLAT. Do adaptacji narzędzia pomiaru na grunt rodzimy można podejść na dwa sposoby w zależności od specyfiki konstruktów zoperacjonalizowanych w zadaniach baterii. Brzeziński (2002) wyróżnia dwa rodzaje adaptacji: parafrazę i transkrypcję. Parafraza polega na opracowaniu nowego testu po dokonaniu dokładnej analizy założeń leżących u podstaw adaptowanego testu, tj., po zidentyfikowaniu konstruktów teoretycznych, którego test jest operacjonalizacją. Innymi słowy parafraza oddając konstrukt oryginału ale poprzez inny materiał bodźca testowego, może się na pierwszy rzut oka bardzo różnić od testu wyjściowego. Transkrypcja z kolei, jest w miarę wiernym oddaniem oryginalnego testu i zachowuje dużą wierność w tłumaczeniu zawartości jak i w zakresie formy tj., grafiki, materiału bodźcowego, instrukcji i układu testu. Parafrazę stosuje się w przypadku testów 'obciążonych kulturowo' (*culture-laden tests*), a transkrypcję w przypadku testów 'nieobciążonych kulturowo' (*culture-free tests*).

Test TUNJO

Przed przystąpieniem do adaptacji testu MLAT autor jej podjął kilka ważnych decyzji, które w konsekwencji swojej określiły niejako warunki brzegowe dla całego procesu adaptacji wpływając na jej ostateczny kształt. Pierwszym warunkiem jaki postawiono przed przyszłą adaptacją był aspekt praktycznego zastosowania testu. Założono mianowicie, że adaptowany test nie powinien być wiele dłuższy niż 45 minut, tak by można go było przeprowadzić w warunkach szkolnych bez większego zakłócania rytmu lekcji. W praktyce, oznaczało to zmniejszenie ilości pozycji testowych w dwóch testach baterii (MLAT 3 i 4) i skrócenie czasu przeznaczonego na test MLAT 2. Dwie następne decyzje były ze sobą powiązane i w pewnym sensie sobie przeciwstawne. Otóż, z jednej strony, w swej adaptacji testu autor postanowił zachować jak najwięcej z oryginału, i to zarówno jeśli chodzi o zawartość treściową jak i wygląd i aspekty formalne testu, a z drugiej strony postanowił odwzorować w teście wszystkie cztery składowe modele uzdolnień Carrolla i Sapona. Decyzje te w praktyce doprowadziły do:

1. zachowania w kształcie prawie, że nie zmienionym dwóch nieobciążonych kulturowo testów, a mianowicie MLAT 2, *Phonetic Script* (w TUNJO – *Alfabet Fonetyczny*), i MLAT 5, *Paired Associates* (w TUNJO – *Nowe Słowa*)
2. zrekonstruowania i odtworzenia na materiale języka polskiego dwóch konstruktów, z testu MLAT 3, *Spelling Clues* (w TUNJO *Ukryte Słowa*) i MLAT 4, *Words in Sentences* (w TUNJO – *Słowa w Zdaniach*)
3. pominięcia w polskiej adaptacji testu MLAT 1, *Number Learning*
4. skonstruowania nowego testu TUNJO 2 – *Sztuczny Język*, będącego operacjonalizacją konstruktu ‘wnioskowanie indukcyjne’ z modelu uzdolnień Carrolla i Sapona, który to konstrukt nie uzyskał jednoznacznej reprezentacji w oryginalnej baterii MLAT.

Przejdziemy teraz do nieco dokładniejszego omówienia powyższych punktów pod kątem wprowadzonych zmian i/lub modyfikacji w zadaniach. W zadaniu TUNJO I *Alfabet Fonetyczny*, (MLAT 2, ‘Phonetic Script’) po przeprowadzeniu badań pilotażowych na grupie 60 badanych w wieku od 15 do 19 lat w różnego typu szkołach postanowiono zwiększyć trudność zadania poprzez:

- po pierwsze, ‘udziwnienie’ zapisu literowego materiału dźwiękowego zadania przez wprowadzenie nowych, bardziej ‘egzotycznych’ symboli literowych dla samogłosek i części spółgłosek. Symbole te, takie jak np., š, ɣ, α, œ czy ʒ wzięto z zapisów fonetycznych różnych języków.
- po drugie, skrócenie czasu ekspozycji badanych na materiał dźwiękowy zarówno w fazie prezentacji (uczenia się) jak i w fazie testowej przez skrócenie pauz pomiędzy prezentowanymi wyrazami, co w końcowym efekcie skróciło całkowity czas zadania z 12 do około 9,5 minuty.

Materiał dźwiękowy, a przede wszystkim organizacja jego prezentacji zdającym, oraz ilość pozycji testowych (30), pozostał nie zmieniony. Decyzja o zwiększeniu trudności zadania została podjęta po analizie trudności poszczególnych pozycji testowych zadania jak i mocy różnicującej całego zadania. Obie analizy pokazały, że oryginalne zadanie MLAT 2 jest za łatwe dla zdających i jako takie nie miało żadnej wartości różnicującej w zakresie badanej cechy (zdolności). Wprowadzone zmiany i ponowna analiza zadania wykazały znaczną poprawę właściwości psychometrycznych zadania.

W przypadku zadania TUNJO V *Nowe Słowa* (MLAT V, ‘Paired Associates’) wprowadzone zmiany były niewielkie i polegały głównie na skróceniu czasu nauki nowych wyrazów (z 3 do 2 minut) oraz czasu przeznaczanego na ćwiczenie. Zmiany te, podobnie jak w przypadku zadania poprzedniego, były wynikiem analizy zadania w badaniu pilotażowym. Inne modyfikacje dotyczyły procedury przeprowadzania zadania w związku z odmiennym od oryginału rozkładem materiału w arkuszu odpowiedzi. Nie dokonano żadnych zmian w materiale testowym, który obejmował 24 pozycje.

O ile omówione wyżej dwa zadania zostały ‘przeniesione’ z oryginału w prawie niezmienionym kształcie, to w dwóch następnych zadaniach musiano dokonać daleko idących adaptacji materiału zadaniowego by w miarę wiernie i poprawnie odwzorować konstrukt oryginału. I tak, w zadaniu TUNJO III, *Ukryte Słowa*, (MLAT 3 ‘Spelling Clues’) starano się oddać charakter konstrukt z oryginału zoperacjonalizowanego jako zdolność polegająca na odgadnięciu znaczenia wyrazu zapisanego z pogwałceniem reguł ortografii języka ojczystego. W języku angielskim, który jest językiem o przewadze samogłosek (20) nad spółgłoskami (16), zapis ortograficzny tych ostatnich musi wykazywać dużą nieprzezroczystość fonetyczną, tzn., trudno jest z zapisu odgadnąć wymowę danej litery, lub grupy liter samogłoskowych. Angielski pod tym względem obfituje w sytuacje gdy jedna litera wymawiana jest na kilka różnych sposobów, i odwrotnie, jeden dźwięk może być zapisany za pomocą kilku liter bądź ich kombinacji. Polski pod tym względem jest bardziej przezroczysty (mniej liter/dźwięków realizowanych na więcej niż jeden sposób; tylko dwa różne sposoby realizacji litery/dźwięku)

i posiada mniej samogłosek a więcej spółgłosek. W efekcie, język polski jest językiem z przewagą spółgłoski i ich zbitek, zarówno na początku wyrazu, w środku jak i na końcu. I to właśnie tę cechę w przeważającej mierze wykorzystano w zadaniu III, *Ukryte Słowa*. Słowa wybrane do zadania w których dokonano ‘deformacji’ zapisu w sposób opisany w instrukcji do zadania, zostały wyłonione w toku pilotażu próbnej wersji testu. Zawierała ona 50 słów (czasowników, rzeczowników i przymiotników) o różnej frekwencji użycia w języku polskim i została przetestowana na 60 uczniach szkół średnich. W wyniku analizy trudności zadania wybrano 27 pozycji testowych w taki sposób aby zmaksymalizować moc dyskryminacyjną zadania. Zmniejszono również ilość opcji w formacie odpowiedzi, zamiast pięciu jak w oryginale, zastosowano cztery.

Zadanie TUNJO IV, *Słowa w Zdaniach* (MLAT 4, ‘Words in Sentences’) zostało poddane podobnemu procesowi adaptacji jak ten opisany powyżej. Postulowaną w oryginalnym zadaniu zdolność rozpoznawania funkcji wyrazów w zdaniu zoperacjonalizowano na materiale języka polskiego dokonując przy tym pewnych zmian. Pierwszą z nich było zmniejszenie ilości pozycji testowych z 50 (MLAT) do 23, a drugą ograniczenie wyboru do czterech (z pięciu w MLAT). Pozycje testowe, które weszły do ostatecznej wersji zadania zostały wyłonione w rezultacie badania pilotażowego wstępną wersją zawierającą 50 pozycji.

Zadanie TUNJO II, *Sztuczny Język* bada zdolność językowego wnioskowania indukcyjnego, czyli dochodzenia do reguł składniowych języka. W tym celu stworzono całkowicie nowe zadanie, które operacjonalizuje badaną zdolność tylko w oparciu o materiał językowy, a nie, jak w MLAT I, ‘Number Learning’, w oparciu o materiał nie werbalny (działania na liczbach). Aby zmaksymalizować dyskryminacyjną siłę zadania, zastosowano dwa różne formaty odpowiedzi dla dwóch części zadania. Część pierwsza wymaga od osoby testowanej rozpoznania problemu i dokonania wyboru poprawnej odpowiedzi z trzech podanych, natomiast w części drugiej wymaga się od niego podania częściowego tłumaczenia zdań z języka polskiego na materiał językowy zadania. Tym samym, część pierwsza bada zdolność biernego rozpoznawania postulowanej zdolności, a część druga umiejętność aktywnego zastosowania rozpoznanych reguł języka.

Reasumując, ostateczna wersja testu TUNJO składa się, tak jak MLAT, z pięciu zadań. Zakres uzdolnień zoperacjonalizowanych w poszczególnych zadaniach baterii MLAT oraz TUNJO podaje zestawienie w Tabeli 1 poniżej:

Tabela 1. Porównanie MLAT i TUNJO

MLAT	Zdolność	TUNJO
MLAT I „Number Learning”	pamięć mechaniczna, wnioskowanie indukcyjne kodowanie fonetyczne	xxxxxxx
MLAT II „Phonetic Script”	kodowanie fonetyczne	TUNJO I „Alfabet Fonetyczny”
MLAT III „Spelling Clues”	kodowanie fonetyczne słownictwo j. ojczystego	TUNJO III „Ukryte Słowa”
MLAT IV „Words in Sentences”	wycucie gramatyczne	TUNJO IV „Słowa w Zdaniach”
MLAT V „Paired Associates”	pamięć mechaniczna	TUNJO V „Nowe Słowa”
zdolność nie ujęta	wnioskowanie indukcyjne	TUNJO II „Sztuczny Język”

Pilotaż TUNJO

Badanie pilotażowa wersją testu zostało już częściowo omówione w sekcji 'Opis testu TUNJO' i zostanie tutaj jeszcze rozszerzone o pewne ważne zagadnienia. Przed każdym właściwym badaniem przy pomocy jakiegokolwiek narzędzia badawczego w postaci testu, kwestionariusza czy też ankiety, trzeba przeprowadzić analizę poszczególnych zadań testu (pozycji testowych). W przypadku gdy narzędziem naszym jest test wymiarów poznawczych badający poziom wykonania lub zdolności do wykonania pewnego zadania, czyli wówczas gdy mamy do czynienia z tzw., skalami wykonaniowymi, przeprowadza się analizę każdej pozycji testowej pod kątem jej współczynnika trudności/łatwości oraz tzw., siły dyskryminacyjnej.

Współczynnik trudności zadania informuje nas o tym jak trudna jest dana pozycja testowa, tzn., ilu badanych nie wykonało jej poprawnie tj., zgodnie z kluczem. Matematycznie jest on stosunkiem osób, które odpowiedziały błędnie do liczby wszystkich badanych. Test w którym wszystkie lub większość pozycji są albo bardzo trudne (nikt lub prawie nikt ich nie wykonał poprawnie), albo bardzo łatwe (wszyscy lub prawie wszyscy wykonali je poprawnie) nie jest dobrym testem z punktu widzenia badacza bo nie jest mu w stanie wiele powiedzieć o badanie cesze w danej grupie badanych. Logicznym jest zatem, aby test był ogólnie, w sposób umiarkowany trudny i zarazem łatwy. Innymi słowy w dobrym teście powinny się znaleźć pozycje testowe oscylujące wokół średniej wartości trudności dla danej populacji z kilkoma trudnymi/bardzo trudnymi i łatwymi/bardzo łatwymi pozycjami testowymi. Zróżnicowana skala trudności poszczególnych pozycji testu stwarza szansę na to, że cecha badana w tym zadaniu ujawni się badaczowi w sposób pozwalający uszeregować badanych na skali od najmniejszego do największego natężenia tej cechy.

Z pojęciem trudności/łatwości pozycji testowej nieodłącznie związany jest współczynnik siły dyskryminacyjnej pozycji testowej. Współczynnik ten informuje nas o tym w jakim stopniu dana pozycja testowa różnicuje osoby badane danym narzędziem. Innymi słowy chodzi o taką sytuację kiedy pozycja testowa (o dodatniej mocy różnicującej) jest częściej rozwiązywana przez tych badanych którzy uzyskali wysoki wynik z całego testu a nie odwrotnie, tj., kiedy ta sama pozycja jest rozwiązywana przez badanych, którzy osiągnęli ogólny niski wynik z całego testu. Pozycja testowa, która ma dodatnią moc dyskryminacyjną różnicuje badanych w zgodzie z innymi pozycjami testu i zwiększa wariancję wyników testowania (Hornowska, http://www.staff.amu.edu.pl/~hornowsk/psychologia//Wyklad_6.pdf, Hornowska 2003).

Jak już powiedziano wyżej badanie pilotażowe przeprowadzono w przeciągu jednego miesiąca (listopad 2006 roku) na łącznej grupie około 300 badanych w wieku od 17 do 18 lat, uczniów szkół średnich. Ze względów organizacyjnych jak i ograniczeń czasowych nie było możliwe przetestowanie wstępnej baterii testu TUNJO na jednej grupie badanych pozyskanych do badania pilotażowego. Ponieważ pilotażowa bateria TUNJO zawierała dokładnie tyle samo pozycji testowych w każdym z pięciu testów ile miał ich test wyjściowy tj., MLAT użycie tej samej grupy do przebadania wszystkich testów wprowadziłoby ogromne zamieszanie w proces dydaktyczny szkoły, której badani byli uczniami zdecydowano się na rozwiązanie mniej zakłócające ten proces. Każdy z pięciu testów baterii został przydzielony dwóm studentom z seminarium magisterskiego prowadzonego przez piszącego te słowa i przetestowany na innej grupie docelowej populacji. Wszystkie grupy badanych liczyły od 50 do 65 osób, wszyscy badani byli uczniami 2 klasy liceum ogólnokształcącego, każdy test przeprowadzony był w innej szkole, proporcje płci były mniej więcej te same i wynosiły od 45 do 65% kobiet i mężczyzn, uczniowie realizowali normalny, nie rozszerzony i nie dwujęzyczny program nauczania języka obcego w szkole, pierwszym językiem obcym nauczonym w szkołach był język angielski a następnie niemiecki, oraz, co najważniejsze, procedura przeprowadzania pilotażu we wszystkich grupach była taka sama.

W wyniku analizy zadań (pozycji testowych) przeprowadzonej po pilotażu dokonano zmian we wszystkich pięciu testach baterii TUNJO częściowo już opisanych w sekcji powyżej. Zmiany te polegały na zmniejszeniu ilości zadań w TUNJO III (US) i IV (SwZ), zwiększeniu tempa ekspozycji na bodziec w TUNJO I (AF), zmniejszeniu ilości opcji w formacie odpowiedzi w TUNJO IV (SwZ), oraz inne. Poza tym, w testach w których dokonano największych zmian adaptacyjnych lub stworzono je od podstaw tj. w SzJ, US i SwZ pozycje testowe uszeregowano od najłatwiejszej do najtrudniejszej co oznaczało, że stopień trudności pozycji testowych wzrastał w miarę jak zdający przechodził do dalszych zadań w teście. Kolejność ta była na bieżąco empirycznie weryfikowana w kolejnych badaniach wersjami testu aby ją zoptymalizować.

Właściwości psychometryczne TUNJO

W tej części podręcznika testowego znajdują się informacje dotyczące: a. statystyk opisowych baterii, b. rzetelności baterii, c. wstępnego badania walidacyjnego baterii i d. innych ważnych informacji psychometrycznych.

Statystyki opisowe

Badania końcową wersją baterii TUNJO rozpoczęły się w marcu 2007 roku. Zamierzonym ich celem było opracowanie norm wiekowych/demograficznych dla docelowej populacji tj., dorosłych od 17 roku wzwyż. W tym celu przebadano testem studentów 1 roku filologii angielskiej, uczniów 2 klasy liceum ogólnokształcącego, słuchaczy szkoły pożarnictwa, żołnierzy zawodowych i służby czynnej oraz uczniów 2 klasy technikum i liceum profilowanego (patrz Tabela 2). W wyniku analizy zadań okazało się jednak, że ten etap badania trzeba poświęcić na dalsze usprawnianie baterii TUNJO. Łącznie, w pierwszym roku prac, przebadano 535 osób trzema różniącymi się wersjami TUNJO. W wyniku tego etapu opracowano ostateczną wersję testu, którą w następnych latach (2008-2010) przetestowano 650 badanych w wieku od 16:10 do 24:05 lat. Przedstawia to tabela 2 poniżej.

Tabela 2. Grupy badane testem TUNJO w latach 2008-2010

	1 IFA	1 ILS	1 AE	2 LO	3 LO	Razem
2008/2009	171	89	52	34		346
2009/2010	128					128
2010/2011	148				28	176
Razem	447	89	52	34	28	650

Największą grupę badanych, bo prawie 70%, stanowili studenci pierwszego roku filologii angielskiej, których testowano zawsze w pierwszych dwóch tygodniach od rozpoczęcia roku akademickiego. Następną grupą badanych byli studenci 1-go roku lingwistyki stosowanej (89), akademii ekonomicznej (52), a ostatnią grupę stanowili uczniowie 2 i 3 klasy liceum ogólnokształcącego. W tabelach 3 poniżej podane są podstawowe statystyki opisowe dla baterii TUNJO dla tych grup, a w tabeli 4 to samo lecz z rozbiem na płeć.

Tabela 3. Statystyki opisowe dla ostatecznej wersji TUNJO (kursywą podano średnie w punktach procentowych)

Testy	Średnia						Odch. Stand.					
	1ang	1ekon	1ling	3lo	2lo	razem	1ang	1ekon	1ling	2lo	3lo	razem
AF	25,7	25,5	24,2	24,9	23,9	25,4	2,60	2,40	2,96	2,77	3,38	2,74

	88,6	87,9	83,4	85,9	82,4	87,6						
SzJ	10,1 59,4	9,5 55,9	10,1 59,4	8,8 51,8	11,3 66,5	10,1 59,4	3,38	3,27	3,42	2,91	3,41	3,37
US	22,1 73,7	20,7 69	19,9 66,3	22,1 73,7	22,3 74,3	21,7 72,3	3,70	4,06	4,52	3,60	3,53	3,91
SwZ	14,5 63	12,3 53,5	14,7 63,9	12,8 55,7	13,9 60,4	14,2 61,7	3,25	3,32	3,43	4,09	3,33	3,39
NS	17,0 70,8	15,9 66,3	17,0 70,8	16,3 67,9	17,8 74,2	16,9 70,4	5,09	4,86	4,41	4,15	4,31	4,91
TUNJO	89,4 72,7	83,8 68,1	86,0 69,9	84,9 69	89,1 72,4	88,2 71,7	11,60	9,60	12,64	12,39	13,04	11,83

lang - studenci 1 roku filologii angielskiej N=447; wiek 19:08

lekon - studenci 1 roku akademii ekonomicznej N=52; wiek 19:10

ling - studenci 1 roku lingwistyki stosowanej N=89; wiek 19:08

lo - 2klasa liceum ogólnokształcącego N=34; wiek 17:04

3lo - 3klasa liceum ogólnokształcącego N=28; wiek 18:04

razem wszyscy badani N=650

Tabela 4. Statystyki opisowe dla ostatecznej wersji TUNJO dla czterech grup badanych z podziałem na płeć

Testy	Średnia										Odch. Stand.									
	lang		lekon		ling		lo		razem		lang		lekon		ling		lo		razem	
	K	M	K	M	K	M	K	M	K	M	K	M	K	M	K	M	K	M	K	M
AF	25,6	25,9	24,4	26,2	24,3	23,8	24,7	24,1	25,3	25,5	2,65	2,48	2,62	2,00	2,66	4,16	2,89	3,40	2,72	2,81
SzJ	10,2	9,9	10,1	9,2	10,0	10,6	9,6	10,5	10,1	9,9	3,36	3,45	3,15	3,35	3,27	4,16	3,23	3,55	3,32	3,51
US	22,0	22,4	19,9	21,2	20,1	19,1	22,3	21,9	21,6	21,9	3,62	3,88	3,69	4,25	4,61	4,11	3,50	3,67	3,86	4,03
SwZ	14,8	13,7	13,7	11,4	14,8	13,8	14,1	12,0	14,7	13,1	3,06	3,64	3,06	3,20	3,33	3,84	3,48	3,98	3,14	3,72
NS	17,4	15,6	16,1	15,7	17,4	15,3	17,0	16,9	17,3	15,7	4,89	5,41	6,07	4,03	4,36	4,33	4,45	4,00	4,82	4,94
TUNJO	90,1	87,4	84,1	83,7	86,7	82,4	87,7	85,4	89,1	86,1	11,3	12,4	10,1	9,5	12,1	14,7	12,2	13,9	11,5	12,4

lang - studenci 1 roku filologii angielskiej K=332, M=115

lekon - studenci 1 roku akademii ekonomicznej K=20, M=32

ling - studenci 1 roku lingwistyki stosowanej K=73, M=16

lo - 2 i 3 klasa liceum ogólnokształcącego K=39, M=23

razem wszyscy badani K=464, M=186

Analiza różnic pomiędzy średnimi dla K i M obydwu zestawu danych nie wykazała istnienia różnic w poziomie wykonania poszczególnych zadań baterii jak i w wynikach całościowych, które byłyby statystycznie istotne.

Wskaźniki rzetelności

Z wielu możliwych metod szacowania wskaźników rzetelności narzędzia a omawianych w literaturze przedmiotu (Alderson i inni, 1995, Anastasi i Urbina, 1999, Hornowska 2003) wybrano metodę połówkową jako najbardziej odpowiednią dla testów badających wymiary poznawcze, oraz takich które są w zasadzie bardziej testami mocy niż szybkości. Metoda ta polega na podzieleniu testu na dwie połowy, obliczenia korelacji pomiędzy nimi i następnie obliczenia współczynnika rzetelności wg., formuły Spearmana-Browna. Z wielu różnych sposobów podzielenia testu na dwie połowy wybrano najprostszą tj., 'parzyste - nieparzyste'. Takie 'przepełnienie' testu wydawało się tym bardziej właściwe, że szereguje w pary pozycje testowe z tego samego miejsca w teście. Gdyby 'przepełnienie' testu nastąpiło np.,

według zasady ‘pierwsza połowa testu - druga połowa testu’, to dałoby to, przy uszeregowaniu pozycji testowych w niektórych zadaniach od łatwych do trudnych, nierzetelne oszacowanie współczynnika rzetelności. Wskaźniki rzetelności testu wyrażane są liczbą dziesiętną i zawierają się pomiędzy 0 a 1. W literaturze przyjmuje się na ogół następującą interpretację tych wskaźników:

- < 0,60 - brak akceptowalnej rzetelności
- 0,60-0,70 - niska rzetelność
- 0,70-0,80 - satysfakcjonująca rzetelność
- 0,80-0,90 - wysoka rzetelność
- > 0,90 - bardzo wysoka rzetelność

Według klasycznej teorii testu współczynnik rzetelności testu w wysokości np., 0,75 oznacza, że 56,25% zmienności wyników uzyskanych przy pomocy badanego narzędzia ($0,75^2 \times 100\% = 56,25\%$) można przypisać stałej zmienności (wariancji) wyników u której podstaw leży reakcja badanych na mierzoną cechę ujawniająca się poprzez dane narzędzie. Pozostałe 43,75% zmienności należy przypisać wpływowi błędu pomiaru, który wynika z niedoskonałości narzędzia. W tabeli 5 poniżej podano współczynniki rzetelności obliczone dla poszczególnych grup badanych. Mieszają się one wszystkie na poziomach wysokiej rzetelności.

Tabela 5. Współczynniki rzetelności połówkowej TUNJO

	1ang08	1ang09	1ang10	1ekon	1ling	2lo08	3lo10
N	163	136	148	52	89	34	28
Wiek	19:08	19:09	19:08	19:10	19:08	17:4	18:4
TUNJO	0,86	0,90	0,88	0,82	0,90	0,86	0,91

1ang08 - studenci 1 roku filologii angielskiej badani w roku 2008

2lo08 - 2 klasa LO badana w 2008

3lo10 - 3 klasa LO badana w 2010

Interkorelacje pomiędzy częściami baterii TUNJO

W wielowymiarowych narzędziach pomiaru, tj., w takich które mierzą więcej niż jeden aspekt (wymiar, cechę) szerszego konstruktury analiza interkorelacji pomiędzy poszczególnymi częściami (testami, skalami) baterii (narzędzia) pozwala zorientować się w jakim stopniu części baterii są niezależnymi od siebie aspektami mierzonej cechy. Logika korelacyjna każe spodziewać się, że jeżeli poszczególne testy baterii mierzą odrębne cechy to korelacje pomiędzy nimi będą niskie lub umiarkowane (powiedzmy rzędu do 0,60). Jeżeli byłoby odwrotnie, tzn., gdyby korelacje pomiędzy częściami były wysokie (powiedzmy rzędu > 0,80) to oznaczałoby to, że jeden z tych dwu testów nie przyczynia się do wyjaśnienia zmienności wyników całego testu. Innymi słowy oznaczałoby to, że te dwa testy mierzą podobne do siebie konstrukty (cechy).

Analiza interkorelacji pomiędzy częściami TUNJO dla poszczególnych grup badanych jak i dla wszystkich badanych (N=650) w tabelach od 7 do 11 poniżej, pokazuje, że wszystkie części testu TUNJO są od siebie względnie niezależne, co oznacza, że każdy z nich mierzy nieco inny aspekt uzdolnień językowych.

Tabela 7. Interkorelacje testów baterii TUNJO: studenci 1 roku anglistyki

N=447	AF	SzJ	US	SwZ	NS
SzJ	0,32				
US	0,25	0,26			

SwZ	0,21	0,33	0,24		
NS	0,19	0,31	0,24	0,23	
TUNJO	0,54	0,67	0,63	0,61	0,72

Tabela 8. Interkorelacje testów baterii TUNJO: studenci 1 roku AE

N=52	AF	SzJ	US	SwZ	NS
SzJ	0,21*				
US	0,14	-0,12			
SwZ	-0,07	0,03	0,07		
NS	0,13	0,21	0,13	0,18	
TUNJO	0,42	0,46	0,51	0,46	0,73

Tabela 9. Interkorelacje testów baterii TUNJO: studenci 1 roku lingwistyki

N=89	AF	SzJ	US	SwZ	NS
SzJ	0,40				
US	0,30	0,15			
SwZ	0,28	0,51	0,26		
NS	0,28	0,34	0,33	0,34	
TUNJO	0,62	0,68	0,60	0,69	0,72

Tabela 10. Interkorelacje testów baterii TUNJO: uczniowie LO

N=62	AF	SzJ	US	SwZ	NS
SzJ	0,42				
US	0,42	0,24			
SwZ	0,41	0,44	0,47		
NS	0,45	0,39	0,28	0,29	
TUNJO	0,74	0,69	0,67	0,73	0,71

Tabela 10a. Interkorelacje testów baterii TUNJO: 2 klasy LO (wiek 17:04)

N=34	AF	SzJ	US	SwZ	NS
SzJ	0,42				
US	0,50	0,17			
SwZ	0,53	0,34	0,46		
NS	0,41	0,28	0,28	0,35	
TUNJO	0,78	0,59	0,69	0,78	0,69

Tabela 10b. Interkorelacje testów baterii TUNJO: 3 klasy LO (wiek 18:05)

N=28	AF	SzJ	US	SwZ	NS
SzJ	0,61				
US	0,37	0,32			
SwZ	0,34	0,55	0,49		
NS	0,57	0,44	0,28	0,16	
TUNJO	0,80	0,79	0,67	0,67	0,71

Tabela 11. Interkorelacje testów baterii TUNJO: wszyscy badani

N=650	AF	SzJ	US	SwZ	NS
-------	----	-----	----	-----	----

SzJ	0,33				
US	0,29	0,21			
SwZ	0,22	0,34	0,25		
NS	0,21	0,31	0,25	0,25	
TUNJO	0,57	0,66	0,63	0,62	0,71

* tłustym drukiem zaznaczono korelacje statystycznie istotne

Trafność TUNJO

Najogólniej rzecz biorąc trafność narzędzia pomiaru to dokładność z jaką test mierzy to, co ma mierzyć. Jest kilka sposobów wykazania trafności testu, w zależności od celów stawianych testowi jak również od dostępnych informacji niezbędnych do przeprowadzenia walidacji narzędzia. W literaturze przedmiotu wyróżnia się następujące obszary, analiza których może nas doprowadzić do wykazania trafności testu:

1. analiza zawartości treściowej testu - „zakres, w jakim treść testu stanowi reprezentatywną próbę dziedziny, która ma być przedmiotem pomiaru”
2. analiza procesu rozwiązywania testu - „analizowanie zadań testowych z punktu widzenia procesów przetwarzania informacji”
3. analiza struktury wewnętrznej testu - „jeżeli można przyjąć w świetle teorii, że mierzona cecha jest cechą homogeniczną, to wysoki współczynnik zgodności wewnętrznej może być traktowany jako dowód trafności teoretycznej”
4. analiza związków (korelacji) testu z innymi testami - na przykład „ocena trafności teoretycznej na podstawie wysokości współczynników korelacji z podobnymi testami (wysokie) i współczynnikami korelacji z testami mierzącymi inne cechy (niskie); zwłaszcza metoda analizy macierzy ‘wielu cech-wielu metod’ Campbella i Fiskego”
5. analiza związków (korelacji) testu ze zmiennymi (testami) zewnętrznymi względem testu a stanowiącymi dla niego kryterium oceny trafności czyli tzw., trafność kryterialna - „jest to stopień zgodności pomiarów otrzymanych za pomocą testu, którego trafność oceniamy oraz pomiarów otrzymanych w wyniku oceny kryterium”

(cytaty z: Hornowska 2003)

W celu określenia trafności omawianego narzędzia zdecydowano się na analizę jego siły predyktywnej, tj., na określenie trafności kryterialnej poprzez obliczenie korelacji pomiędzy wynikami na teście TUNJO a tymi osiągniętymi na testach badających biegłość językową tej samej grupy badanych. Decyzja o użyciu wyników testów językowych raczej niż ocen półrocznych czy semestralnych wystawianych przez nauczycieli języków obcych została podjęta świadomie w obliczu powszechnie znanego faktu, że oceny wystawiane uczniom/studentom na półrocz, semestr czy koniec roku są tylko w niewielkim stopniu odzwierciedleniem rzeczywistej biegłości językowej ocenianego. Oceny, wystawiane na ogół na skali od 1 do 6 (szkoły) czy od 2 do 5 (szkoły wyższe), są często odbiciem pozamerytorycznych aspektów dydaktyki języka obcego i zawierają w sobie silny element uznaniowo-motywacyjny. Jako takie, są niezbyt rzetelną miarą docelowej domeny przewidywanej przez test uzdolnień językowych.

W tabelach 12, 12a, 13 i 13a przedstawiono dane wstępnej analizy trafności baterii TUNJO, którą przeprowadzono w 2008/2009 roku na czterech grupach badanych: studentach pierwszego roku anglistyki, lingwistyki stosowanej i ekonomii, oraz uczniach 2 klasy liceum ogólnokształcącego. Dla ogólnej liczby 346 przebadanych w owym roku testem TUNJO osób (patrz Tabela 2) pozyskano dane kryterialne w postaci wyników testów z języka obcego dla 239 badanych, jak w tabeli 12 poniżej.

Tabela 12. Wskaźniki trafności TUNJO dla poszczególnych grup badanych

	Grupa badana	N	Opis kryterium	r
1.	studenci 1 roku fil. angielskiej	106	test gramatyczno-leksykalny wypracowanie egzamin ustny ocena końcowa	0,33 0,31 0,20 0,35
2.	studenci 1 roku lingwistyki	48	test gramatyczno-leksykalny	0,63
3.	studenci 1 roku ekonomii	52	test gramatyczno-leksykalny test sprawności (czytanie/słuchanie) ocena końcowa	0,11* 0,28 0,18*
4.	uczniowie 2 klasy LO	33	test gramatyczno-leksykalny test sprawności (czytanie i słuchanie) ocena końcowa	0,01* 0,30* 0,11*

* korelacje statystycznie nieistotne

W związku z wynikami z testów językowych (kryterium) użytymi do walidacji TUNJO kilka słów wyjaśnienia. Po pierwsze, ze względów oczywistych różnic pomiędzy badanymi grupami dotyczących poziomu zaawansowania językowego, kontekstu nauki, oraz organizacji procesu dydaktycznego nie było możliwe zastosowanie jednolitego narzędzia do pomiaru zmiennej kryterialnej. Po drugie, a w wyniku pierwszego punktu, otrzymano zróżnicowany zakres treściowy mierzonego kryterium poczynając od wyników tylko dla testu gramatyczno-leksykalnego dla studentów 1 roku lingwistyki, poprzez wyniki dla testu gramatyczno-leksykalnego i testu sprawdzającego sprawności czytania i rozumienia ze słuchu (studenci 1 roku ekonomii i uczniowie 2 klasy LO), a kończąc na teście gramatyczno-leksykalnym, wypracowaniu pisemnym i ocenie sprawności mówienia dla studentów anglistyki. Dla grup z więcej niż jednym kryterium podano również wynik całościowy (końcowy). Po trzecie, zanim obliczono korelacje pomiędzy zmiennymi, wyniki miar kryterialnych w postaci czy to surowych wyników czy też punktów procentowych zamieniono na wyniki standardowe¹ aby wyeliminować efekt różnych skal użytych w poszczególnych testach kryterium. Po czwarte, nie było możliwe uzyskanie od nauczycieli przeprowadzających testy językowe w swoich grupach danych, które by pozwoliły oszacować wskaźniki rzetelności dla wszystkich użytych testów językowych.

Wartości r dla grupy studentów anglistyki można określić za Guilfordem (1964) i Góralskim (1974) jako niskie/słabe (od 0,20 do 0,35), a dla grupy studentów lingwistyki jako przeciętne/umiarkowane (0,63). Wartości współczynnika korelacji 0,20, 0,35 i 0,63 interpretuje się jako odpowiednio 4, 12,3 i 39,7 procent wyjaśnianej wariancji w zmiennej kryterialnej. Nie wydaje się to dużo, szczególnie dla pierwszej wartości ale biorąc pod uwagę fakt, że czynnik zdolności językowych jest tylko jednym z wielu wpływających na proces uczenia się języka obcego wartości te nabierają innego znaczenia.

W związku z korelacjami z tabeli 12 jeszcze jedna rzecz wymaga komentarza i wyjaśnienia. Chodzi mianowicie o fakt, że wszystkie statystycznie istotne wartości współczynnika korelacji (z wyjątkiem jednej) występują w dwu pierwszych grupach. I chociaż dla grupy uczniów LO wartość r dla testu sprawności 'otarły się' o wartości krytyczne dla 32 stopni swobody, to logika wnioskowania indukcyjnego każe jednak ten fakt przypisać raczej przypadkowi niż rzeczywiście ujawniającej się zależności pomiędzy zmiennymi. Widać to wyraźniej w tabeli 12a poniżej gdzie obliczono korelację dla

¹ Wszystkie obliczenia i analizy dla potrzeb tego badania wykonano przy użyciu programu STATISTICA StatSoft, Inc. (2010). wersja 9.1.

połączonych grup studentów ‘językowych’ i ‘nie językowych’. W grupie tych ostatnich jedynie test sprawności językowych (czytanie/słuchanie) koreluje istotnie, acz nisko, z TUNJO ($r = 0,22$; 4,8% wyjaśnianej wariancji). Natomiast w grupie studentów ‘językowych’ korelacja TUNJO z testem gramatyczno-leksykalnym (jedynie wspólne kryterium dla tej grupy) wynosi 0,44 tj., 19,6% wyjaśnianej wariancji w kryterium. Wynik całkiem przyzwoity, bo ze średniego zakresu wartości, które otrzymali autorzy oryginalnego testu MLAT w swoim badaniu walidacyjnym (Carroll i Sapon, 1959).

Tabela 12a. Wskaźniki trafności TUNJO dla grup ‘językowych’ i ‘nie językowych’

	Grupa badana	N	Opis kryterium	r
1+2	studenci 1 roku fil. angielska + lingwistyka	154	test gramatyczno-leksykalny	0,44
3+4	studenci 1 roku ekonomii + 2 klasa LO	85	test gramatyczno-leksykalny test sprawności (czytanie/słuchanie) ocena końcowa	0,00* 0,22 0,12*

* korelacje statystycznie nie istotne

Na pierwszy rzut oka wydaje się, że uderzająca różnica w wielkości r dla grup w 12 i 12a związana jest charakterem grup badanych: ‘językowcy’ - ‘nie językowcy’. Wydawać by się mogło, że wyższe korelacje dla studentów kierunków językowych niż nie językowych spowodowane są właśnie tym, że mamy do czynienia ze studentami o wyższych uzdolnieniach językowych i że to w ich przypadku związek pomiędzy TUNJO a miarami biegłości językowej uwidacznia się silniej i wyraźniej. Otóż jest wręcz przeciwnie. To, że w przypadku studentów językowych mamy do czynienia z wyselekcjonowaną pod względem interesującej nas zdolności grupą badanych powinno raczej zwrócić nam w analizie korelacyjnej niższą wartość współczynnika niż w grupie nie językowej. Tak podpowiada logika leżąca u podstaw analizy korelacji: w grupach homogenicznych pod względem obecności mierzonej cechy rozkład wyników jest ograniczony do wyższych wartości skali i przy w ten sposób ograniczonej wariancji wyników zależności nawet istniejące, często się nie ujawniają lub ujawniają się w postaci niższych korelacji. Natomiast, w przypadku grup o zróżnicowanym zakresie występowania danej cechy (heterogenicznych), tzn., takich gdzie rozkład wyników z danego testu jest zbliżony do normalnego w taki sposób, że wyniki sytuują się na szerszym spektrum skali wariancja jest duża i interesująca nas cecha ma większe ‘szanse by się ujawnić’ w postaci współczynnika korelacji. To, że współczynnik korelacji dla grup językowych (wykazujących się niską wariancją z powodu homogeniczności grupy) jest mimo wszystko dość wysoki kazałoby oczekiwać, że korelacja TUNJO z testem językowym dla grup heterogenicznych (o większym zróżnicowaniu występowania badanej cechy) będzie większa lub co najmniej zbliżona. A tak, jak pokazuje tabela 12 i 12 a, nie jest.

Jak się okazuje wyjaśnienie tej rozbieżności związane jest z rodzajem testów z języka obcego a nie rodzajem grupy. Okazuje się, że testy językowe zdawane przez grupy językowe i nie językowe były w naturze swej różne. Testy dla studentów anglistyki i lingwistyki były testami biegłości językowej (*proficiency tests*) podczas gdy ocena studentów ekonomii i uczniów liceum odbyła się na podstawie testów osiągnięć (*achievement tests*). W tych pierwszych, postulowana zależność pomiędzy testem zdolności a testem językowym jeżeli się unaoczniła to miała na to większe szanse bo testowana domena kryterialna (opanowanie języka) zależy bardziej tylko od zdolności niż innych czynników wpływających na uczenie się. Natomiast w przypadku testów osiągnięć zależność ta, nawet jeżeli w rzeczywistości istnieje (jak mocno wierzymy), nie musiała się ujawnić bo domeną testowaną jest opanowanie przerobionego w określonym a ostatnim czasie materiału językowego. Osiągnięcie wysokiego wyniku na tym kryterium nie musi być wcale uzależnione tylko od zdolności ale również, a

nawet w przeważającej mierze, od czynników takich jak pracowitość, systematyczność, czy wręcz odwrotnie, nagły zryw przede-egzaminacyjny. Innymi słowy, na teście osiągnąć mogło się zdarzyć tak, że ten kto otrzymał niskie wyniki na TUNJO a przygotował się z określonego materiału na egzamin/test zdał go z dobrym rezultatem, a to doprowadziło do tego, że analiza korelacyjna stwierdzi brak korelacji.

W tabelach 13 i 13a podane są korelacje dla zmiennych: TUNJO i jego części oraz test językowy. Podano również interkorelacje pomiędzy częściami TUNJO dla grupy 'językowców' i 'nie językowców'.

Tabela 13. Korelacje i interkorelacje dla TUNJO i kryterium dla studentów anglistyki i lingwistyki (N=154)

	AF	SzJ	US	SwZ	NS	TUNJO
SzJ	0,32					
US	0,39	0,27				
SwZ	0,11*	0,36	0,12*			
NS	0,12*	0,34	0,21	0,35		
TUNJO	0,54	0,71	0,64	0,59	0,69	
test gr-leks	0,26	0,34	0,34	0,23	0,22	0,44

Tabela 13a. Korelacje i interkorelacje dla TUNJO i kryterium dla studentów ekonomii i uczniów LO (N=85)

	AF	SzJ	US	SwZ	NS	TUNJO	test	spraw
SzJ	0,29							
US	0,25	-0,04*						
SwZ	0,22	0,16*	0,23					
NS	0,23	0,23	0,18*	0,24				
TUNJO	0,59	0,53	0,55	0,62	0,69			
test gr-leks	-0,04*	0,11*	-0,08*	0,02*	-0,01*	0,00*		
sprawności	0,04*	0,02*	0,22	0,22	0,14*	0,22	-0,25	
L2 total	-0,01*	0,12*	0,04*	0,14*	0,06*	0,12*	0,86	0,29

*korelacje statystycznie nie istotne

Obszary zastosowań baterii TUNJO

Potencjalne obszary zastosowań baterii TUNJO są dość szerokie. Najogólniej rzecz biorąc test TUNJO może być stosowany dla celów diagnostycznych, prognostycznych oraz badawczych, przez nauczycieli, administratorów kursów językowych i naukowców zajmujących się zagadnieniami akwizycji języków obcych. Jednakże, niektóre z tych zastosowań, na obecnym etapie prac nad baterią, są albo niemożliwe albo ograniczone. I tak, ponieważ nie przeprowadzono jeszcze badań normalizacyjnych testu, to znaczy nie zaopatrzone TUNJO w pełne wiekowe jak i demograficzne normy osiągnięć dla poszczególnych kategorii wiekowych i społecznych (rodzaj ukończonej szkoły, 'wieś-małe miasta-duże miasta', i inne) nie jest możliwe pełne wykorzystanie baterii dla celów diagnozy polegającej na porównaniach grupowych osiągnięć badanych w stosunku do badanych z innych grup wiekowych czy społecznych. Podobnie, ponieważ badanie walidacyjne baterii zostało dokonane tylko na ograniczonych próbkach badanych, które, jak zostało omówione wcześniej, nie są reprezentatywne dla całej docelowej populacji dla badania której test jest przeznaczony, decyzje prognostyczne muszą być z konieczności przybliżone i trzeba je

rozpatrywać w indywidualnym raczej niż ogólnym kontekście próby badanej. Natomiast jeżeli chodzi o użycie TUNJO do celów badawczych, które nie skutkują podejmowaniem ważkich decyzji dydaktycznych czy pedagogicznych, to bateria jest w pełni funkcjonalna i może być z powodzeniem zastosowana w badaniach zarówno ilościowych jak i jakościowych, podłużnych jak i przekrojowych, eksploracyjnych jak i confirmacyjnych.

W kontekście powyższych uwag, jednym z powodów dla których bateria TUNJO jest udostępniana badaczom jest rozszerzenie empirycznej bazy danych niezbędnej do dalszej walidacji testu jak i normalizacji wyników nim uzyskanych.

Bibliografia

- Alderson, J. C., Clapham, C., Wall, D. 1995. *Language test construction and evaluation*. Cambridge: CUP
- Anastasi, A., Urbina, S. 1999. *Testy psychologiczne*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Brzeziński, J. 2002. *Metodologia badań psychologicznych*. Warszawa: PWN.
- Brzeziński, J. (red.) 2004. *Metodologia badań psychologicznych. Wybór tekstów*. Warszawa: PWN.
- Carroll, J. B., Sapon, S. M. 2002, [1959]. *Modern Language Aptitude Test*. Bethesda, MD: Second Language Testing.
- Dornyei, Z. 2001. *Teaching and researching motivation*. Harlow: Longman.
- Dornyei, Z. 2003. *Questionnaires in second language research: Construction, administration, and processing*. Mahwah: Lawrence Erlbaum Associates.
- Dornyei, Z. 2005. *The psychology of the language learner*. Mahwah: Lawrence Erlbaum Associates.
- Góralski, A. 1974. "Metody opisu i wnioskowania statystycznego w psychologii". Warszawa: PWN, s. 34.
- Grigorenko, E. L., Sternberg, R. J., Ehrman, M. E. 2000. "A theory-based approach to the measurement of foreign language learning ability: The CANAL-F theory and test". *Modern Language Journal* 84(3). s. 390–405.
- Guilford, J. G. 1964. "Podstawowe metody statystyczne w psychologii i pedagogice". Warszawa: PWN, s. 157.
- Hornowska, E. 2003. *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Hornowska, E., Paluchowski, W. J. 2004. *Kulturowa adaptacja testów psychologicznych*. W: Brzeziński, J. (Red.). *Metodologia badań psychologicznych. Wybór tekstów*. Warszawa: PWN. s. 151-191.
- Pimsleur, P. 1966. [2003]. *The Pimsleur Language Aptitude Battery*. N. Bethesda, Maryland: Second Language Testing, Inc.
- Rysiewicz, J. 2006. "FL aptitude research and testing. A survey of issues and problems." W: Dziubalska-Kołaczyk, K. "IFAtuation. A life in IFA. A festschrift for Professor Jacek Fisiak on the occasion of his 70th birthday." Poznań: Wydawnictwo Naukowe UAM. s. 545-564.
- Rysiewicz, J. 2006. "Uzdolnienia językowe. Nieporozumienia a rzeczywistość badawcza." *Neofilolog* 29. s. 17-25.
- Rysiewicz, J. 2007. "A study on the prediction of success in the learning of English as a SL." W: Arabski, J. "On foreign language acquisition and effective learning." Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Rysiewicz, J. 2007. "Statistical analyses in individual differences research methodology – FL

- aptitude and how to effectively predict FL learning success? W: Stalmaszczyk, P. i Witczak-Piasecka, I. (red). PASE Studies in linguistics. Łódź: Wydawnictwo Uniwersytetu Łódzkiego. s. 239-254.
- Rysiewicz, J. 2008. "Cognitive Profiles of (un)successful EFL learners: A cluster analytical study". *Modern Language Journal* 92/1. s. 87-99.
- Rysiewicz, J. 2008. "Measuring foreign language learning aptitude. Polish adaptation of The Modern Language Aptitude Test by Carroll and Sapon". *Poznań Studies in Contemporary Linguistics* 44(4). s. 569-595.
- Rysiewicz, J. 2009. "Measuring aptitude - Polish adaptation of Modern Language Aptitude Test by Carroll and Sapon". W: Edwardes, M. (red.). *Materiały konferencyjne BAAL Annual Conference 2008*. London: Scitsiugnil Press. s. 93.
- Rysiewicz, J. 2010. „Uzdolnienia do nauki języków obcych, pomiar i implikacje dydaktyczne”. *Języki Obce w Szkole* 2. s. 3-15.
- Rysiewicz, J. 2010. „Krytyczna analiza dwóch narzędzi do pomiaru uzdolnień do nauki języków obcych”. *Studia Psychologiczne* 48, 3. s. 49-60.
- Skehan, P. 1989. *Individual differences in language learning*. London: Arnold.
- Stansfield, C.W., D.J. Reed. 2003. "Adaptation of the Modern Language Aptitude Test and The Pimsleur Language Aptitude Battery for examinees whose first language is not English". (Referat wygłoszony podczas corocznej konferencji East Coast Organization of Language Testers, 20–21 March 2003, Washington, D.C.)
- Strelau, J. (red.) 2004. *Psychologia: podręcznik akademicki. Tom 2*. Gdańsk: GWP.