XIII Konferencja Diagnostyki Edukacyjnej
Uczenie się i egzamin w oczach uczniów.
Łomża, 5 - 7.10.2007

Monika Stawicka
Uniwersytet Jagielloński

Standaryzacja jako element konieczny na drodze do egzaminów certyfikowanych.

Przykład Jagiellońskiego Centrum Językowego UJ

Celem artykułu jest po pierwsze, ogólne przedstawienie procesu standaryzacji jakiemu zamierza się poddać Jagiellońskie Centrum Językowe UJ w dążeniu do oferowania studentom kończącym lektoraty z języków obcych egzaminów certyfikowanych, oraz po drugie, bardziej szczegółowe omówienie wstępnej części pierwszego etapu (Faza 1) tego procesu, którym jest próba uzyskania możliwie obiektywnej oceny prezentacji ustnej rozumianej tu jako test 'na zaliczenie', a zamykającej kurs języka obcego w jego części specjalistycznej zależnej od kierunku studiów. Artykuł omawia prace już wykonane w procesie standaryzacji prezentacji i wskazuje na etapy, które są do pokonania.

Tło

Celem reformy zainicjowanej w Jagiellońskim Centrum Językowym Uniwersytetu Jagiellońskiego (JCJ) w 2006 roku jest zagwarantowanie jego studentom kończącym tu edukację językową możliwości otrzymania certyfikatów językowych, których wartość informacyjna wychodzić będzie poza grupę studencką czy Centrum i rozszerzy się do granic zjednoczonej Europy. To ambitne zadanie może być zrealizowane tylko przez administrowanie egzaminów certyfikowanych, tj takich, które będą narzędziami wiarygodnymi, rzetelnymi i praktycznymi a więc, docelowo, będą mogły spełniać kryterium zewnętrznej walidacji, co oznacza możliwość interpretowania ich poziomu wg skali CEFR)[1].

[1] **CEFR** - *Common European Framework of Reference Learning, teaching, assessment* jest dokumentem Rady Europy; zaakceptowany ostatecznie w dzisiejszej wersji w 2001 jest on jedynym w Europie oficjalnym narzędziem referencyjnym niezbędnym w procesie odnoszenia *(relating)* poziomów nauczania i ocenia do skal w nim zawartych. Razem z *Manual Preliminary Pilot Version z 2003 roku i Reference Supplement* z 2004 tworzy on źródło teoretycznych i praktycznych informacji, z którymi dobrze się jest zapoznać zajmując się ocenianiem kompetencji językowych w jakimkolwiek aspekcie.

Droga do tego celu będzie niewątpliwie wyzwaniem ponieważ wymagać ona może refleksji nad obecnym modelem nauczania i testowania a w konsekwencji, także ponownego sformułowania celów i standardów nauczania, dostosowania do nich programów nauczania i zaprojektowania nowych egzaminów końcowych. Wszelkie tego typu działania można by ująć słowem s*tandaryzacja*, oznaczającym proces, którego celem jest uzyskanie możliwie zbliżonego sposobu rozumienia standardów nauczania, a przede wszystkim standardów oceniania oraz podobnego posługiwania się testem w celu zminimalizowania działania czynników mających wpływ na wynik.

Elementem stwarzającym dodatkowy problem na drodze JCJ do bardziej obiektywnego oceniania jest element L/ESP[2] między innymi. ze względu na następujące czynniki:

- jedenaście istniejących uniwersyteckich wydziałów wraz z dziesiątkami kierunków - każdy ze specyficznym żargonem fachowym,
- brak adekwatnych programów nauczania i odpowiednich sposobów/skal oceniania L/ESP,
- dużą liczbę nauczycieli - (112 w tym 51 nauczycieli języka angielskiego) - każdy posługujący się nieco odmienną kulturą nauczania i oceniania.

W sytuacji tak znacznej liczby czynników trudnych lub nie poddających się standaryzacji obiektywne ocenianie ESP i włączenie go do egzaminu końcowego wydaje się przedsięwzięciem z gruntu skazanym na niepowodzenie. Stąd powstał pomysł aby ocena elementu ESP (komponent ustny i komponent pisemny) stała się oceną *dopuszczającą* studenta do egzaminu końcowego, tj testem na zaliczenie. Egzamin końcowy natomiast w wersji podstawowej czy standardowej ma docelowo odpowiadać poziomowi B2 wg skali CEFR i ma być egzaminem oceniającym ogólne umiejętności językowe naszych studentów – GE (*General English)*. Nie oznacza to jednak pozostawienie testu zaliczeniowego „samemu sobie".

Model procesu standaryzacji egzaminów

Schemat 1[3] przedstawia ogólny model procesu, jaki JCJ planuje przejść w swoim dążeniu do administrowania własnych certyfikowanych egzaminów. Istotne jest, że u podstaw procesu standaryzacji znajduje się test

---

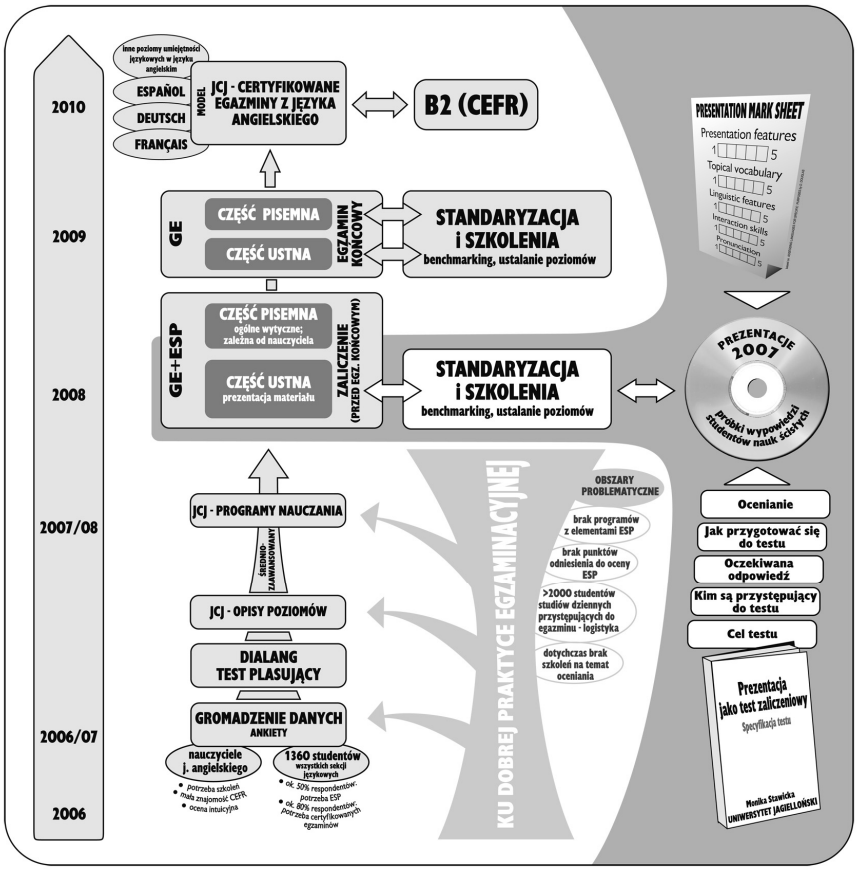
[2] **L/ESP** Language/English for Specific Purposes w naszym (nauczycieli języków obcych w JCJ) rozumieniu jest tą częścią programu nauczania na lektoratach, która dotyczy danego kierunku studiów tworząc tym samym dla specyficzny dla niego kontekst językowy wyrażony głównie przez specyficzne słownictwo. Więcej informacji na temat oceniania LSP znajduje się np. w książce: *Assessing Language for Specific Purposes,*. Douglas, D., 2000.

[3] Schemat 1 jest uproszczoną wersją plakatu - prezentacji, który autorka niniejszego artykułu przedstawiła na IV Międzynarodowej Konferencji EALTA (European Association for Language Testing and Assessment) w czerwcu 2007 w Barcelonie. www.ealta.eu.org

diagnostyczny DIALANG[4], który będąc zwalidowanym narzędziem diagnostycznym pozwala na w miarę rzetelne (samo)określenie wyjściowych umiejętności językowych studentów JCJ. DIALANG został zaprojektowany w odniesieniu do skal kompetencji językowych zawartych w CEFR, dlatego też wydaje się on dawać pewną podstawę do uzyskania spójności paroletniego procesu standaryzacyjnego. Schemat uwzględnia także oś czasu, na której rok 2010 jest zaznaczony jako rok pierwszej edycji takich egzaminów. Wydaje się to być to wersja bardzo optymistyczna, która z pewnością podlegać będzie różnym modyfikacjom merytorycznym i czasowym choćby ze względu na trudny proces pozyskiwania funduszy.

Do tej pory końcowa ocena kompetencji językowych studenta, zarówno ogólnych jak i języka fachowego, była w gestii nauczyciela, który sam przygotowywał test, sam go też sprawdzał i wystawiał ocenę. Obecnie dążąc do zaprojektowania egzaminów certyfikowanych niezbędne wydaje się być przeprowadzenie standaryzacji oceniania, tj. najpierw wprowadzenia zasad dobrej praktyki egzaminacyjnej, a następnie ich utrzymanie. Ocena elementu ESP – stanowiącego dużą trudność - jest, jak wspomniałam wyżej, oceną na zaliczenie, na które składa się zarówno komponent pisemny i ustny. Według założeń reformy komponent/test pisemny pozostaje niejako dalej zależny od nauczyciela, który otrzymuje wyłącznie wytyczne (*guideline*s) dotyczące specyfikacji. Komponent ustny natomiast, którego głównym elementem jest prezentacja materiału na temat związany z kierunkiem studiowanym przez studenta jest pierwszym testem podlegającym standaryzacji. Innymi słowy ustna prezentacja materiału i rozmowa, która po niej następuje zostają potraktowana jako *test ustny*, którego standaryzacja ma na celu przetworzenie dotychczasowego arbitralnego sposobu oceniania na sposób bardziej rzetelny, mniej subiektywny, a więc prowadzący do uzyskania bardziej porównywalnych wyników. Schemat 1 koncentruje się na procesie standaryzacji prezentacji w jego pierwszej fazie i tego dotyczy dalsza część artykułu.

[4] DIALANG jest testem diagnostycznym (często używanym jako test plasujący) opierającym się na dokumencie CEFR i służącym do samooceny poziomu umiejętności w czytaniu, pisaniu, słuchaniu, gramatyce i słownictwie. Po wykonaniu testu/testów otrzymuje się feedback i radę. Więcej informacji o systemie DIALANG znajduje się w książce: Diagnosing Foreign Language Proficiency J.C. Anderson, 2005.



SCHEMAT 1. Model procesu standaryzacji egzaminów końcowych do poziomu B2 z wyszczególnieniem standaryzacji testu ustnego „zaliczeniowego" w postaci prezentacji materiału[5], Stawicka, M. 2007

[5] Na schemacie znajdują się następujące skróty angielskie:**GE** *(General English)* tj angielski ogólny bez elementów języka specjalistycznego; **ESP** (*English for Specfic Purposes)* język angielski z elementami języka specjalistycznego zależnego od kierunku studiów, np. żargon typowy dla informatyków, chemików czy bibliotekoznawców; **CEFR-**patrz przypis 1-*(Common European Framework of Reference)*czyli Europejski system opisu kształcenia językowego (ESOKJ)-dokument referencyjny Rady Europy zawierający min opisy poziomów umiejętności jezykowych wg skali A, B, C. Poziom A0 jest najniżej na skali i wskazuje na brak biegłości językowych, podczas gdy osiągnięcie poziomu C2 oznacza perfekcyjne opanowania języka przez uczącego się. Poziom **B2,** w naszym kontekście edukacyjnym zwany średnio-zaawansowanym, oznacza wg skali ogólnej CEFR niezależnego użytkownika języka obcego, który „…rozumie znaczenie poszczególnych wątków przekazu zawartego w złożonych tekstach na tematy konkretne i abstrakcyjne, łącznie ze zrozumieniem dyskusji na tematy techniczne z zakresu jej specjalności. Potrafi

Standaryzacja testu ustnego w formie prezentacji

Faza 1.

Przez Fazę 1. rozumie się tutaj całość prac od przygotowania specyfikacji testu do momentu pozyskania nagrań włącznie. Są one niezbędne do przeprowadzenia dalszych etapów standaryzacji każdego testu ustnego.

Specyfikacja testu

Alderson (Alderson, 1995) uważa, że specyfikacja testu (*test specifications*) powinna zawierać m.in. takie elementy jak:

- cel testu,
- opis profilu przystępujących do testu,
- określenie poziomu testu
- definicję konstruktu, testu
- opis formatu testu (w naszym wypadku prezentacja jest pierwszą z dwóch części testu podlegających ocenie; drugą natomiast jest rozmowa między interlokutorem a zdającym na tematy odnoszące się do prezentacji oraz luźno związane z przedmiotem studiowanym przez zdającego),
- opis umiejętności językowych, które mają być testowane i zadań w teście, które te umiejętności testują,
- instrukcje dla zdającego wypowiadane przez egzaminatora,
- kryteria oceny wraz z oszacowaniem każdej części testu,
- opis typowych wypowiedzi na różnych poziomach (w obrębie danego poziomu na jakim test jest przeprowadzany; w naszym wypadku jest to w założeniu poziom B2),
- opis kompetencji językowych egzaminowanego w sytuacji poza-egzaminacyjnej,
- przykładowe wypowiedzi egzaminowanych ilustrujące wykonanie poszczególnych zadań, itp.

Elementy te zostały zawarte w skrypcie *Prezentacja jako test Próba standaryzacji* M. Stawickiej[6], która proponuje także sposób oceny

---

porozumiewać się na tyle płynnie i spontanicznie by prowadzić normalną rozmowę z rodzimym użytkownikiem języka, nie powodując przy tym napięcia u którejkolwiek z stron. Potrafi (…) formułować przejrzyste wypowiedzi ustne i pisemne, a także wyjaśniać swoje stanowisko w sprawach będących przedmiotem dyskusji (…).”,*Europejski system kształcenia językowego: uczenie się, nauczanie, ocenianie,* 2002:33.

[6] Skrypt *Prezentacja na zaliczenie Próba standaryzacj*i Moniki Stawickiej został rozesłany on–line do wszystkich członków zespołu sekcji angielskiej JCJ i poddany dyskusji w kwietniu 2007 roku. Jest wewnętrznym dokumentem Centrum przeznaczonym do użytku

prezentacji wraz z arkuszem oceny, kryteriami oceny i opisem procedury oceniania. Procedura ta opiera się na założeniu, że przeprowadzenie testu i jego ocena będzie się odbywać w obecności dwóch niezależnych egzaminatorów (asesora i interlokutora), tj przeszkolonych nauczycieli uczących na tym samym lub zbliżonym kierunku. Ta część skryptu podaje też wskazówki Selinkera[7], który podpowiada na jakie elementy powinni zwrócić uwagę egzaminujący nauczyciele, aby lepiej zrozumieć użycie języka w danej dyscyplinie naukowej.

Dodatkowo skrypt zawiera wyjaśnienie pojęcia prezentacji z punktu widzenia studenta, nauczyciela oraz pod kątem administracyjnym i językowym z implikacjami metodycznymi i uzasadniającej użycie jej jako formy testu; Jest też tam sekcja z materiałami pomocniczymi dla nauczyciela (i studenta) np. repertuar zwrotów angielskich, które można wykorzystać na na różnych poziomach nauczania począwszy od poziomu średnio zaawansowanego, który jest kojarzony z poziomem B wg skali CEFR. W skrypcie znajduje się rozdział zawierający praktyczne rady i uwagi dotyczące przygotowania się do prezentacji, wyboru materiałów, opracowania pomocy wizualnych, przygotowania planu/handoutu, etc. Na końcu skryptu pomocniczy słowniczek wyjaśnia trudniejsze terminy z dziedziny testowania.

U podstaw niniejszej pracy/skryptu leży przeświadczenie o praktycznych korzyściach informacji tam zawartych, których zrozumienie i wykorzystanie przyczyni się do podobnego operowania testem w formie ustnej prezentacji materiału fachowego przez różnych nauczycieli JCJ, a więc do bardziej obiektywnego indywidualnego oceniania z jednej strony, z drugiej zaś umożliwi studentom nabycie praktycznej umiejętności ustnego prezentowania materiału w języku obcym do wielokrotnego wykorzystania w sytuacji docelowej, poza-testowej.

Gromadzenie danych - nagrania wideo[8]

Dokonanie nagrań wideo, które z jednej strony będą podstawą do analizy jakościowej, z drugiej zaś po selekcji i odpowiednim opracowaniu będą

---

przez nauczycieli sekcji angielskiej. Materiały w nim zawarte będą podstawą do szkoleń standaryzujących na temat uczenia a zwłaszcza oceniania prezentacji w następnym roku akademickim.

[7] Selinker (Selinker,1997) w swojej pracy pokazuje trudności na jakie napotyka egzaminator oceniający element LSP i zwraca uwagę na konieczność współpracy między nauczycielami języka a fachowcami z danej dziedziny (*informants*) przy konstruowaniu testów z L/ESP.

[8] Dyskusja na temat wyższości nagrań wideo nad nagraniami audio w sytuacji testów ustnych face-to-face oraz sposobów ich przeprowadzania znajduje się w publikacji: *A Qualitative Approach to the Validation of Oral Language Tests*, A. Lazaron, 2002:47-72.

służyć do procesu ustalania poziomów, *benchmarking*[9], przygotowania deskryptorów i szkoleń, są następnym etapem procesu standaryzacji. Takie pozyskiwanie danych w postaci nagrań, które następnie będą dostępne w postaci transkrypcji, niesie ze sobą konieczność podjęcia szeregu decyzji dotyczących sposobu dokonywania nagrań, rodzajów transkrypcji i nauczenia się filozofii transkrypcji. Wszystkie te zadania są trudne logistycznie, technicznie i metodologicznie, ale niezbędne w procesie walidacji. Warto także w tym miejscu wspomnieć o paradoksie obserwatora (*observer's paradox*), który pojawia się tam gdzie gromadzeniu danych towarzyszy jakikolwiek rodzaj obserwacji bez względu na jej cel (Labov, 1972)[10].

W badaniu wzięło udział 30 studentów informatyki UJ, co oznacza, że wszyscy studenci, którzy byli uczestnikami 120 godzinnego rocznego lektoratu z języka angielskiego (w dwóch grupach) zgodzili się na udział w nagraniach wideo. Po przeprowadzeniu nagrania studenci wypełnili kwestionariusz dokonując samooceny i wypowiadając się na temat testu. Materiał badawczy został zebrany w maju i czerwcu 2007 roku i jest obecnie poddawany analizie. Z tego względu fakt zebrania materiału został potraktowane w artykule tylko jako kolejny punkt na liście elementów niezbędnych w procesie standaryzowania tego testu ustnego. Wszelkie procedury pozyskiwania danych przez nagrania wideo i sposoby ich analizy, u podstaw której powinna być rzetelnie wykonana transkrypcja, będą przedstawione wraz z opracowaniem rezultatów w najbliższej przyszłości i stanowić będą **Fazę 2**. Należy zaznaczyć, że zgromadzony materiał będzie mógł służyć następującym celom:

1. badawczym,
2. szkoleniowym,
3. do oszacowań poziomów kompetencji językowych przez sędziów kompetentnych (*raters)*.

---

[9] Proces/ćwiczenie *benchmarking*, czy ściślej mówiąc: *benchmarking local performances* czyli w naszym wypadku szacowanie nagranych na wideo próbek wypowiedzi studentów w odniesieniu do skali CEFR wymaga najpierw przeprowadzenia sesji zaznajamiających z tą skalą przy użyciu wcześniej skalibrowanych próbek wypowiedzi (*familiarisation).* Jeśli wśród sędziów zostanie osiągnięty konsensus co do poziomu poszczególnych próbek, rezultatem *benchmarkin*g, będzie zestaw lokalnie skalibrowanych próbek wypowiedzi w odniesieniu do skali CEFR. Więcej o procesie *benchmarking* można przeczytać np. w dokumencie Rady Europy: *Relating Language Examinations to CEFR. Manual. Preliminary Pilot Version*, 2003: 65-98.

[10] Więcej informacji na temat paradoksu obserwatora, w tym nadmiernego korygowania czy gotowości do korygowania w sytuacji obecności obserwatora interakcji – *hypercorrection,* znajduje się w pracy: Labov, W., *Sociolinguistic Patterns,* 1972.

Faza 3. Ustalanie poziomów i szacowanie próbek wypowiedzi (*standard setting & benchmarking*)

Można założyć, że dokonane na wideo nagranie trzydziestu testów – prezentacji jest pierwszą wersją pilotażową testu. W rezultacie analizy wyników badania pilotażowego, test ten zacznie być poddawany modyfikacjom, prowadzącym do kolejnej, ulepszonej wersji. Ale nawet poprawiona wersja nie daje podstaw do twierdzenia, że test ten, zaprojektowany w założeniu na poziomie B2 wg skali CEFR, tak jak egzamin końcowy, rzeczywiście na tym poziomie się znajduje. Procedurą, która pozwala to stwierdzić jest procedura: *standard setting* - ustalenie poziomu/standardu. W przypadku umiejętności produktywnych (*productive skills*), takich jak pisanie czy mówienie, stosowany jest szczególny rodzaj tej procedury, mianowicie *benchmarking,* gdzie wypowiedzi egzaminowanych, w przypadku testu ustnego nagrane na wideo, będą bezpośrednio[11] oszacowane przez panel sędziów kompetentnych poprzez osiągnięcie przez nich konsensusu co do poziomu każdej wypowiedzi. Poziom kompetencji językowych testowanego oszacowany na podstawie zarejestrowanej na taśmie wypowiedzi i jej transkrypcji będzie osadzany na skali umiejętności językowych CEFR. Uzasadnienie opinii sędziów dotyczących poziomów poszczególnych testów znaleźć będzie można w opracowanych przez nich deskryptorach, które wraz z odpowiadającymi im próbkami nagrań tworzą podstawę materiałów szkoleniowych i standaryzacyjnych dla egzaminatorów. Właściwie przeprowadzony proces *benchmarking* - szacowania próbek wypowiedzi - jest kluczowym elementem zapewniającym standardy czy poziomy referencyjne, do których egzaminatorzy będą się odnosić oceniając wypowiedzi/prezentacje w czasie rzeczywistych egzaminów.

Faza 4. Szkolenia dla egzaminatorów

W celu spełnienia kryteriów rzetelności i trafności testów ustnych, w tym testu, o którym mowa w artykule, zaleca się, aby w procesie standaryzacji egzaminu ustnego znalazły się następujące elementy[12]:

– stosowanie wytycznych dotyczących zachowania egzaminatora (asesora i interlokutora),

[11] Słowo 'bezpośrednio' oznacza tu pominięcie procedury *cut-off scores* – ustalania na skali poziomów krytycznych dzielących jeden poziom od następnego i/lub poprzedniego. Wiecej o tej i innych procedurach ustalania poziomów można przeczytać np. w *Reference Supplement to the Preliminary Pilot Version of the Manual w rozdziale 5* lub w pracy: Cizek,G., & Bunch, M., *What is Standard Setting*? 2007.

[12] Podaję za Csepes, I. & Egyud, G., *Into Europe. The Speaking Handbook,* 2005:24-25.

- stosowanie skali analitycznej, która uprzednio przeszła badanie pilotażowe i została odpowiednio zmodyfikowana,
- szkolenie asesora aby potrafił on właściwie zastosować skalę analityczną i deskryptory,
- wyposażenie interlokutora w precyzyjne wytyczne dotyczące formułowania pytań lub wręcz zestaw pytań czy innych reakcji słownych, spośród których wybiera on te, które umożliwią mu efektywne przeprowadzenie testu ustnego,
- szkolenie interlokutora aby potrafił on zastosować ww. wytyczne i prowadzić egzamin w sposób umożliwiający każdemu egzaminowanemu zaprezentować maksimum swoich kompetencji językowych w sytuacji wysoko stresującej, jaką jest test ustny z języka obcego.

Podsumowanie

Trudno jest nie zgodzić się z twierdzeniem, że standaryzacja jest procesem żmudnym i w zasadzie nigdy nieskończonym, bo czyż istnieje stan idealnej stuprocentowej standaryzacji? Ale w naszej sytuacji, to jest w sytuacji pierwszej próby zbliżenia się do uzyskania porównywalnych wyników testu/zaliczenia w części ustnej, proces ten ma spełnić także inne zadania. Przede wszystkim, ma przygotować zespół do podjęcia dużego wyzwania, którym niewątpliwie będzie przygotowanie standaryzowanego egzaminu końcowego sprawdzającego ogólne kompetencje językowe naszych studentów. Można założyć, że próba standaryzowania prezentacji będzie okazją do zapoznania się zespołu z istotą LT *(Language Testing)*, w tym projektowania zadań egzaminacyjnych, procesów standaryzacji zadań i egzaminatorów w celu osiągania porównywalnych wyników, co z kolei, będzie nieść dużą wartość informacyjną dla uczących się, nauczających i zarządzających JCJ. Wartością dodaną procesu będzie niewątpliwie podniesienie kwalifikacji poszczególnych nauczycieli już przez samo uczestniczenie w procesie, nie mówiąc o niezbędnych sesjach szkoleniowych i standaryzacyjnych. Ponadto, rzetelnie przeprowadzona standaryzacja powinna zaowocować licznymi produktami w postaci dokumentów, instrukcji, przykładowych testów, ułatwiając pracę nauczycielom i czyniąc proces dydaktyczny spójny i przejrzysty przede wszystkim dla studentów. Programy nauczania mające swoje odzwierciedlenie w treści i formie egzaminów końcowych, materiały szkoleniowe i standaryzacyjne w postaci nagrań i skryptów z deskryptorami są także przykładami takich właśnie produktów.

Nowoczesne egzaminy ustne, a do takich należy omawiany w artykule test ustny – 'prezentacja na zaliczenie', zwracają uwagę nie tylko na to, co mierzą, ale też w jaki sposób to robią. Standaryzacja jest kluczowym elementem kontroli jakości i dobrej praktyki egzaminacyjnej. Tylko przez wprowadzenie i utrzymanie wspólnie ustanowionych standardów tj norm dotyczących przygotowania i przeprowadzania egzaminów możemy twierdzić, że nasze egzaminy ustne

spełniają kryterium rzetelności, a dany wynik oznacza ten sam poziom kompetencji językowej bez względu na to, który egzaminator dokonywał oceny. Można tego dokonać, spełniając jeszcze jeden warunek, jakim są efektywne szkolenia i treningi nie tylko dla osób projektujących i konstruujących testy, ale przede wszystkim dla egzaminatorów nie tylko po to, aby uniknąć lub zminimalizować pozastandardowe zachowania czy niewłaściwą interpretację skali, ale przede wszystkim, aby wdrażać dobrą praktykę egzaminacyjną, tym samym przybliżając nasz egzamin końcowy do popartego dowodami empirycznymi umiejscowienia go na skali CEFR.

Bibliografia:

1. Alderson, J.C*., Diagnosing Foreign Language Proficiency The Interface between Learning and Assessment*, Bodmin, Cornwall, MPG Books Ltd, 2005.
2. Alderson, J.C., Clapham, C.M. & Wall, D., *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press, 1995.
3. Cizek, G. & M. Bunch, *What is Standard Setting?* w Cizek, Gr. & M. Bunch, Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Test,.London, SAGE Publications, 2007.
4. Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*, Cambridge: Cambridge University Press, 2001:33.
5. Clapham. C., *The development of IELTS: a Study of the effect of Background Knowledge on Reading Comprehension*, Cambridge: Cambridge University Press,1996.
6. Csepes, I. &. Egyud, G., *Into Europe. The Speaking Handbook*,. Alderson ,C.J., (red) Budapest: Teleki Laszlo Foundation, British Council-Hungary, 2005:24-25.
7. Douglas, D., Assessing *Languages for Specific Purposes*, Alderson,J.C & Bachman, L.F. (red) Cambridge: Cambridge University Press, 2000.
8. Kaftandjieva, F., *"Standard setting" .Reference Supplement to North* et al. (2003). Strasburg: Language Policy Division, Council of Europe, 2004.
9. Labov, W., *Sociolinguistic Patterns Philadelphia*, PA: University of Pennsylvania Press, 1972.
10. Lazaraton, A., *A Qualitative Approach to the Validation of Oral Language Testing*, Studies in Language Testing 14, 2002:47 -72.
11. North, B., Figueras, N, Takala, S., Van Avermaet, P., Verhelst, N. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF)*. Manual, Preliminary Pilot Version, Strasburg: Language Policy Division, Council of Europe 2003.
12. Selinker, L., *On the Use of Informants in Discourse Analysis and Language for Specific Purposes* w International Review of Applied Linguistics 17, 1979:189-215
13. Stawicka, M., *Prezentacja jako test. Próba standaryzacji*, nieopublikowany skrypt do użytku wewnętrznego JCJ UJ, Kraków 2007.
14. www. ealta. org