

Developing a New Test Culture: The Art of Possible

Evrim ÜSTÜNLÜOĞLU¹

ABSTRACT

Preparing valid, reliable and fair tests has been concern of all educators, trainers, and test developers for many years, specifically at English medium universities, where accurate determination of language proficiency level is important. This study describes a process regarding how end of module exams have been developed in the light of the Common European Framework (CEF) in an institutional setting which has implemented a new system in its Preparatory Programme with rather limited resources. The study describes the design, development and quality procedure of the end of module tests together with the success rates of the exams in detail. The success rates of the exams indicated in this study is worthwhile following the steps in the CEF manual when developing a new test, as well as considering the bands given in the CEF, and the needs of the institution. It is expected that sharing the process of developing a high stake exam will be valuable for other Preparatory Programs in Turkey, especially when the CEF is controversial in Europe since it involves applying one particular framework to a multitude of contexts.

Key Words: CEF, Test Development, Preparatory Programme

Yeni Bir Sınav Kültürü Geliştirmek: Olasılık Sanatı

ÖZET

Geçerli, güvenilir ve adil sınav hazırlamak tüm eğitimcilerin ve sınav yazarlarının yıllardır odak noktası olmuştur. Bu durum, özellikle de yabancı dilde öğrenci yeterlik düzeyini doğru bir şekilde belirlemek durumunda olan İngilizce öğretim yapan üniversiteler için geçerlidir. Bu çalışma, Hazırlık Programında yeni bir yaklaşım (Modüler Sistem) uygulayan bir eğitim kurumunda Avrupa Ortak Çerçeve Programları ışığında öğrenci yeterliklerini belirlemeye yönelik sınavların kısıtlı kaynaklarla nasıl hazırlandığına yönelik bir süreci anlatmaktadır. Çalışma, sınavların tasarlanma, geliştirilme ve kalite süreçlerinin yanı sıra sınav başarı oranlarını da detaylı olarak kapsamaktadır. Çalışmada belirtilen sınav başarı ortalamaları, kurumun ihtiyaçları da göz önüne alınmak koşuluyla Avrupa Ortak Çerçeve Programlarında belirtilen bantları takip ederek sınav geliştirmenin denemeye değer bir çaba olduğunu göstermektedir. Avrupa Ortak Çerçeve Programlarının farklı bağlamlarda ve ortamlarda kullanılmasının sıklıkla sorgulandığı ve tartışıldığı günümüzde bu çalışmada anlatılan güvenilir sınav geliştirme sürecinin özellikle Türkiye’de benzer programı takip eden hazırlık programları için önemli olacağı düşünülmektedir.

Anahtar Sözcükler: Avrupa Ortak Çerçeve Programı, Sınav Geliştirme, Hazırlık Programı

¹ Assist. Prof. İzmir Ekonomi University, Turkey School of Foreign Languages - evrim.ustunluoglu@ieu.edu.tr

INTRODUCTION

Preparing valid, reliable and fair tests has been a concern of all educators, trainers, and test developers for many years. This specifically applies to English medium universities, which use English as a tool of instruction in their settings where English is spoken as a foreign language. These universities have to use scores to determine their students' language proficiency level, and thus they need clear descriptors and a clear policy for the language skills of students. The Common European Framework (CEF) is used as a reference by some of those institutions because it describes six proficiency levels and also presents a basis for the mutual recognition of language qualifications and educational mobility (Council of Europe, 2003).

The framework provides a common basis for the explicit description of objectives, content and methods, and thus, enhances the transparency of courses, syllabus and qualifications. The key words in this process are comprehensiveness (describing the objectives), transparency (explicitness and availability of information), and coherence (harmony among the components). Coherence includes the identification of needs, determination of objectives, definition of content, selection/creation of material, establishment of teaching/learning programmes together with the methods and finally testing, assessment and evaluation (Council of Europe, 2003, 4). Testing, assessment and evaluation are crucial final steps of developing a new program. In particular, achievement tests, proficiency tests and end-of-module tests should be developed carefully as they determine students' proficiency. However, those tests and their level of difficulty based on proficiency level are likely to vary according to the needs of the schools, to the needs of the students, or to other circumstances, such as the faculty of the student, which may become controversial later.

Although the CEF provides a common basis for language syllabi, examinations, and textbooks, the feasibility of taking one particular framework and applying to a multitude of contexts is still in question (Urkun, 2008). Little (2005) points out that there is no definition of how many descriptors should define a level or how many communicative tasks one must be able to perform in order to achieve a level. He also indicates that there is a risk that teachers or students may claim that a certain level has been reached on the basis of performance in just one or two of the descriptors. Weir (2005) shares the similar concerns, indicating the possibility of making false assumptions of equivalence, when the tests constructed for different purposes are located on the same scale point. He also adds that this may lead to performance descriptors that may be inconsistent or not transparent enough.

On the other hand, North (2007) states that it is not the aim of the CEF to tell practitioners what to do in terms of laying down the objectives. Instead, the aim is to discuss the way languages are taught and assessed to ensure higher quality because the CEF facilitates comparisons between different systems of qualifications (Council of Europe, 2001). In this way, specifically, test users can interpret their own exam results and make better sense of their existing level of proficiency in a particular language. However, Alderson (2006) states that in order for claims to linkage with the CEF to be valid, the process of designing tests should be well-structured. According to him, institutions which intend to make use of the CEF in their local context should follow a very realistic approach recommended in the manual. The manual outlines the alignment process as a set of stages: Specification (coverage of the examination), Standardization (common understanding of the meaning of CEFR levels), Empirical Validation (collection and analysis of data), and Familiarization (detailed knowledge of the CEFR) (Council of Europe, 2003, 6-7). Briefly, the

manual and the reference should be applied appropriately; there should be evidence of the quality of the process followed to link tests and exams to the CEF; and the process should be monitored. Urkun (2008) and Jones (2002) note that institutions which are in the process of developing new tests need to ensure accurate interpretation and application of the CEF in order to achieve the desired results, otherwise, different countries' interpretations could be culturally determined and these may result in inconsistencies.

While the discussions are continuing on this issue, some institutions, specifically the schools of foreign languages which are in the process of developing their own tests use the manual as a reference. This study describes how exams have been developed in the light of the CEF by following the steps in the manual in a university setting, where a new system in the Preparatory Programme was implemented with relatively limited resources. The study first gives information about purpose and the setting, and later the development of exams regarding the efforts to reflect an international level of descriptors.

Purpose of the Study

The School of Foreign Languages in the university has been undergoing a significant change, which has led to the creation of clear descriptors and scores to determine students' language proficiency level. With the introduction of the modular system at the Preparatory Programme, it was felt an urgent need to review and revise the testing procedure to produce more reliable tests than the previous system, in which results were considered unreliable. This study aims to describe the process of developing Gateway exams –end of module exams- using the CEF as a resource tool, and using the success rates of the exams as the end product, and also report on the experiences gained and lessons learnt. This purpose is also in parallel with the aims of the CEF, “to facilitate the mutual recognition of qualifications gained in different learning contexts and aid European mobility” (Council of Europe, 2001, 1). By doing so, it is firmly believed that schools of foreign languages which have similar systems and therefore similar problems will benefit from this study. It is also expected that the process explained in the study will contribute to the standardization of proficiency exams, which has been under discussion in Turkey for many years, particularly for transfer students changing university who need to prove language proficiency.

Setting

The study was conducted in a school of foreign languages within an English medium university in Turkey. Students who enroll in the university but who do not meet the required proficiency level of English attend the Preparatory Programme run by the School of Foreign Languages (SFL) and the school prepares the students for their Faculties through an intensive English preparatory year. Over the previous seven years of its existence, the school followed a semester-based system. However, the feedback from faculty members and students themselves about inadequacies in language proficiency led the administration to search for new approaches. Having worked on a new “Modular system” for one and a half years, the administration decided to adopt it for the 2009-2010 academic years and develop Gateway exams, focused on the CEF, B2 level, but also considering the needs of the setting. The Modular system is based on assessing students' performances over short modules (7 weeks) with specific objectives. The assessment becomes more significant in this new system because Gateway exams determine whether students will move on the next module or not. The Gateway exams, from Beginner to Upper Intermediate level, produced and administered by the Testing Unit, are given at the end of each 7 week-module.

Process

First, a project manager from the school was assigned to monitor the process, as it was considered that managing innovation needs a plan. An external consultant from England, the project manager, the school administration, the Testing Unit (TU), the Curriculum and Material Development Unit (CMDU), five Preparatory Programme coordinators and a focus group formed by a group of teachers worked cooperatively throughout the project to set up the modular system for the Preparatory Programme. The Programme Head, the Measurement and Evaluation Unit, the testers together with the CMDU members specifically worked on developing Gateway exams. These were planned to be high-stake institutional examinations designed for students who require evidence of competency in English at an Upper-Intermediate proficiency level (B2+) in reading, writing, listening, and speaking as well as in vocabulary and grammar resources. The process was carried out as follows:

- A needs analysis was conducted with the faculty members and freshman students to identify the main needs of students and the expectations of faculty members, which later led the curriculum designers to develop the current curriculum accordingly. The need for a renewed curriculum established the need for a new test.
- The team (Programme Head, members of the CMDU and the TU, the Measurement and Evaluation Unit, and the coordinators) involved in the process of developing exams first familiarized themselves with the CEF. Having considered the needs of Faculty, the team first decided on the outcome of Preparatory Programme as B2/CEF. From this point, they worked backwards through each level, discussing the objectives of Beginner (A), Elementary (B), Pre-Intermediate (C), and Intermediate (D) levels. The identified needs/objectives were converted into linguistic requirements – “can do” statements of the CEFR-, specifying not only the knowledge and skills, but also the ability level that the learner is likely to need. This stage was based on knowledge, judgments and expertise of the team regarding their experiences.
- Before writing the test items, clear guidelines (writer guidelines, routine test production process, standard operating procedures for exam production, and grading manuals) were prepared. Test writers were trained on guidelines, checklists, and given examples of valid, reliable, appropriate tasks.
- Test specification was the next step, during which purpose and testing focus was decided upon and documented. Specifications described the content of the tests, the length of each part of the test, item choices, and thus, helped the test developers to choose appropriate task types and test items. Specifications were produced, linking students’ needs to the test requirements stated in the CEFR. Finally, the Gateways were developed for each level, covering listening and note-taking, writing, reading, and speaking skills. The team’s experience suggested that a strong focus on basic competency in grammar was necessary for Beginner, Elementary and Pre-Intermediate groups, but this should be replaced by an increasing emphasis on skills in Intermediate and Upper-Intermediate (E) levels (see Table 1.). The format of the questions consisted of multiple choice, matching, gapped text/cloze, selected response, short answers to open questions, and extended answers (text/monologue). Speaking section of the Gateway exams covered the tasks such as asking for opinions, giving clear and detailed descriptions on a wide range of subjects, synthesizing and reporting information from a number of sources, and taking an active part in discussion. Test tasks were discussed before use by internal group discussions and also supervised by the external consultant.

Table 1. Percentages by skills and levels

Level	Listening	Use of English	Reading	Writing	Speaking
A	20 %	20 %	20 %	20 %	20 %
B	20 %	15 %	25 %	20 %	20 %
C	20 %	10 %	25 %	25 %	20 %
D	25 %	-	25 %	25 %	25 %
E	25 %	-	25 %	25 %	25 %

- After the Gateway exams for each level were developed, they were piloted in order to remove or amend the controversial items and to evaluate the difficulty level of items. After the placement exam, the pilot group was placed in appropriate levels to take the Gateway exam under simulated examination conditions. Useful information was gathered through feedback from the pilot group and decisions were made regarding difficulty level, timing and tasks. The items which did not work and the tasks which were not understood were removed.

- Regarding operational phase, several documents such as test tasks, sample test papers, sample video of the oral exam, sample answer papers, and grading schemes were made available for students and teachers.

- The marking process was the next step. For administrative purposes, reading and listening parts were answered on forms for optical marking. One holistic score for each task was used. For writing and speaking parts of the tests, students created their own answers. Regarding writing and speaking assessment, two core groups of assessors were formed and standardization sessions were arranged before marking. Both groups were trained by two external experts. The writing section was marked by double raters, two markers who sought agreement in case of discrepancy. Accuracy of marking was promoted by training of markers, moderating sessions to standardize judgments, and using standardized examples of test tasks. The speaking part was evaluated by two teachers- one interlocutor and one assessor. Accuracy of marking was promoted by regular checks by coordinators, training of markers, moderating sessions to standardize judgments, using standardized examples of test tasks, and calibrating to the CEF. The criteria used for writing and speaking skills came from the CEFR –“can do” statements– at each level and they were presented as a band scale which describes performance in terms of 5 levels ranging from unsatisfactory to excellent. In the second module, inter-rater reliability coefficient was conducted for only Beginner, Elementary and Intermediate levels writing section due to the limited resources and the values were 0.79 for Beginner level, 0.70 for Elementary level, and 0.80 for Intermediate level.

- Monitoring phase covered obtaining regular feedback from students and teachers and analysis of students’ performance on the Gateway exams. This phase helped revision when there was a need.

- Finally, global and section based scores were reported to the students electronically.

RESULTS

The overall success rates of levels and modules are presented in this section together with the means of item difficulty and reliability. As shown at Table 2, the mean item difficulty of Beginner (A) level is between 0.60 and 0.69, Elementary (B) level 0.53 and 0.61, Pre-intermediate (C) level 0.52 and 0.61, Intermediate (D) level 0.53 and 0.65, and finally,

Upper-Intermediate (E) level 0.57 and 0.69. Reliability coefficient of the exams range from 72% to 92%, which shows the internal consistency of the test results obtained by the test items. As indicated in the statistical studies, reliability coefficients theoretically range in value from zero (no reliability) to 1.00 (perfect reliability) and the score .70 and above good for a classroom test, in the range of most (Erkuş, 2003). The overall success rates of the modules 1, 2, 3 and 4 are 57%, 55%, 52%, and 73% respectively.

Table 2. Means of item difficulty and reliability coefficient in modules

Level	Module 1		Module 2		Module 3		Module 4	
	Mean item difficulty	KR 20	Mean item difficulty	KR 20	Mean item difficulty	KR 20	Mean item difficulty	KR 20
A	0.60	0.90	0.63	0.80	0.69	0.77	0.67	0.81
B	0.59	0.85	0.53	0.82	0.61	0.82	0.56	0.81
C	0.55	0.85	0.58	0.81	0.52	0.75	0.61	0.72
D	0.58	0.92	0.55	0.85	0.53	0.81	0.65	0.77
E			0.64	0.75	0.57	0.72	0.69	0.81

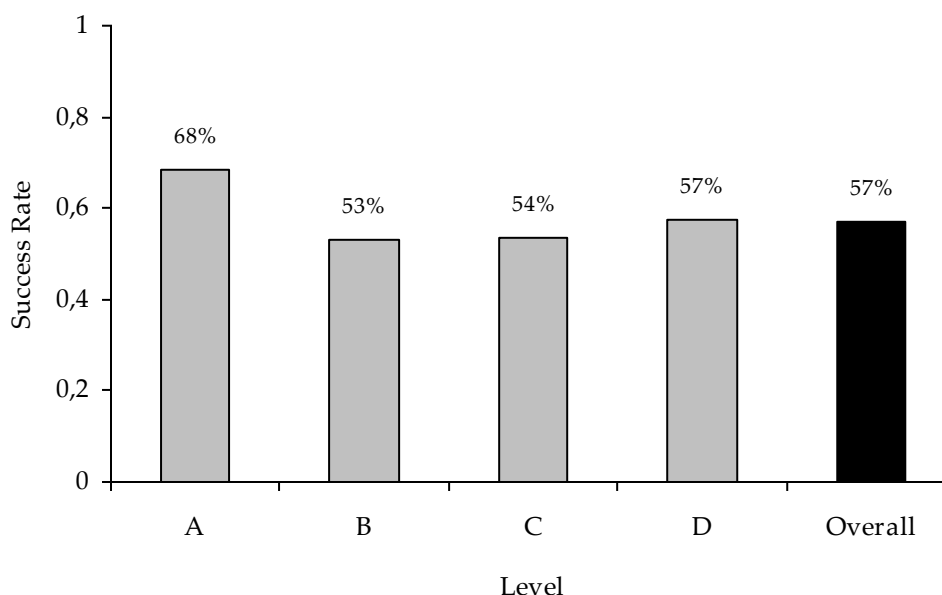


Figure 1. Overall and level-based success rates in module 1

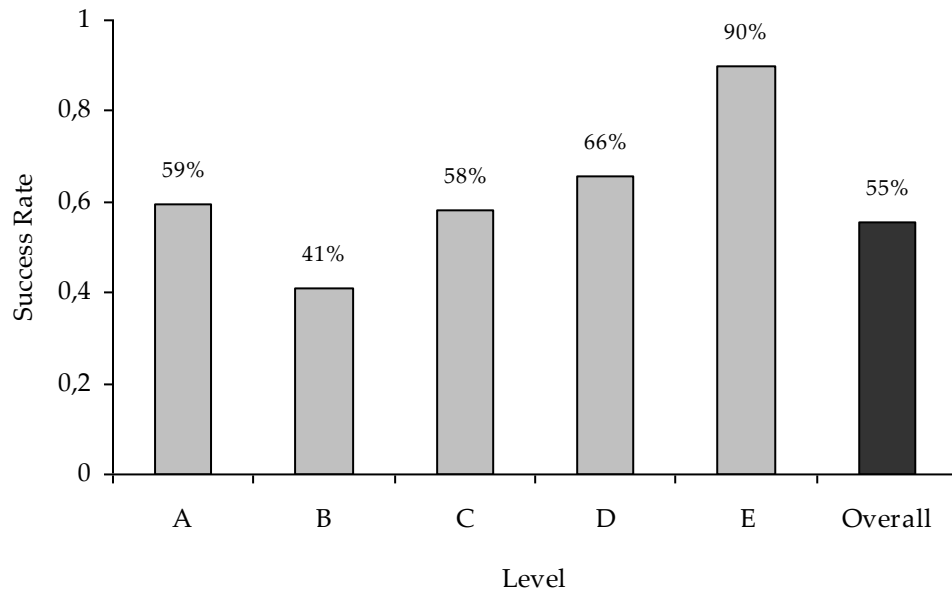


Figure 2. Overall and level-based success rates in module 2

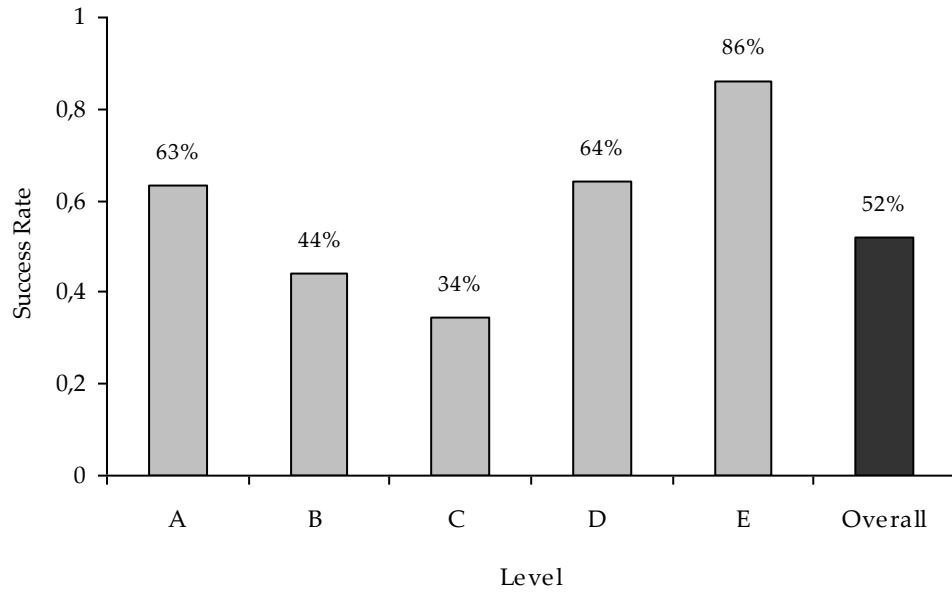


Figure 3. Overall and level-based success rates in module 3

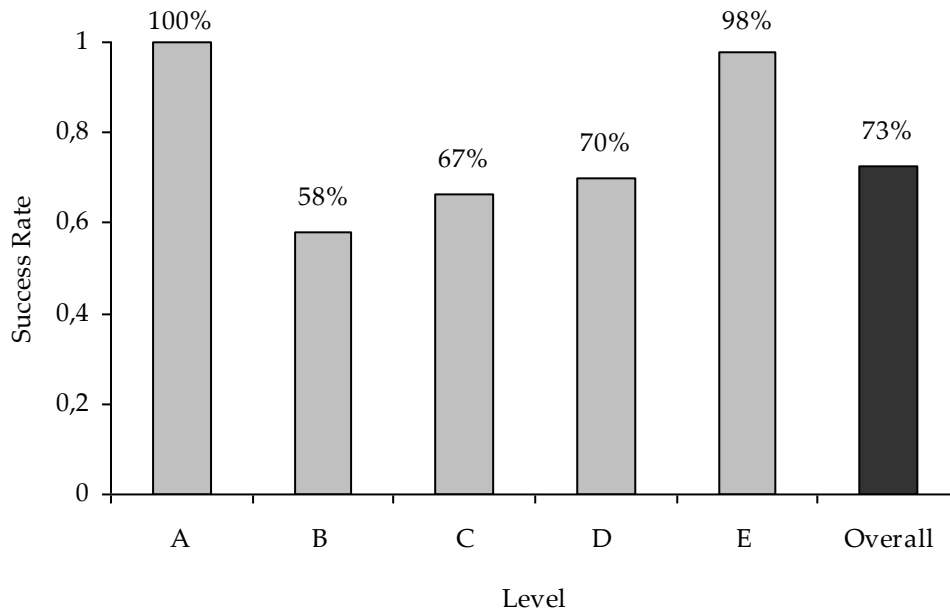


Figure 4. Overall and level-based success rates in module 4

The overall success rates of the first three modules may indicate that the results are below expectations. However, considering that the procedure indicated in the manual was followed without compromise in order to ensure high quality, as well as realistic approach, and taking into account the newness of the modular system for teachers and students, the Gateway exams can be considered to have achieved their aims.

A closer look at the difference between overall success rates indicates that reading, listening and note-taking, writing, and speaking components of Gateways should be reexamined together with the specifications. The difference between the success rates of each level is also noteworthy. The results indicate that Upper-Intermediate students are the most successful compared to Elementary, Pre-Intermediate, and Intermediate levels in the second, third and fourth Modules, while Elementary level students have continuously low rates throughout the year. This finding may also require a careful examination of the cut-off points of placement exam for the following academic year so as to ensure the proper placement of students. A careful examination of the cut-off points of the placement exam may also explain the reason behind low success rate of Pre-Intermediate level with 34% in Module 3 with the mean item difficulty, 0.52. Specifically, those borderline students who were only just able to reach the required standard for Pre-Intermediate level, and subsequently failed the Pre-Intermediate level exam, could also be explained by the cut-off points used in the placement exam. It is clear that there is a striking increase in success rate of the fourth Module, which could be interpreted as either students becoming accustomed to the new system, or by students improving as a result of taking the same formats of exams.

CONCLUSION AND IMPLICATIONS

This study has described the design, development and quality procedure of the Gateway exams at a Preparatory Programme which has implemented a new system with limited human resources. It is clear in this study that the differences in success rates at both level and Module base require a careful review of level objectives, specifications, tasks, and

cut-off points of placement exam for the following year. However, it is important to note here that regardless of exam results, when developing a new test, it is worthwhile following the steps in the manual and the bands in the CEF, as well as taking into account the needs of the institution. It should be highlighted here once more, as indicated by North (2007) and indicated in the CEF (Council of Europe, 2001), that it is not the aim of the CEF to tell practitioners what to do in terms of laying down the objectives and specifications, instead, the aim is to gain an objective approach to language assessment and to discuss about the right way and methodology to ensure higher quality assessment. Despite the fact that the CEFR can be misinterpreted or misused as indicated in previous studies, it is clear that if institutions have a realistic approach to make the best use of the CEFR and maintain a systematic monitoring, they can achieve developing reliable and valid exams if their concern is not only high success rates. Determining specifications, ensuring empirical validation, in other words, developing high stake exams as explained here, is expected to lead to standardization among the universities, particularly for schools of foreign languages in terms of students' language proficiency, which is a controversial issue in Turkey for many years.

The university where this study was conducted is seeking the following ways to make better sense of the existing level of proficiency, to make modifications accordingly, and to maintain the continuity of high stake exams:

- Building up the item bank during the time when items are pretested and calibrated using anchor items to monitor exam difficulty. This will help prepare reliable and valid exams.
- Monitoring, adjusting and keeping the weightings under review during the revision process, based on the statistics, and seeing the programme and test development as an ongoing process.
- Continuously examining objectives and specifications of each level as a natural process of curriculum and exam development.
- Continuing with studies on empirical validity.
- Tracing students' improvement and success rates in each level to develop understanding of cut-off points and make adjustments accordingly.
- Carrying out a similar process for developing proficiency exam, this will ensure that satisfactory standards are met in terms of validity, reliability, impact and practically.

REFERENCES

- Alderson, J. C. (2006). Bridging the gap between theory and practice. *Paper presented at the EALTA Conference, Cracow, Poland.*
- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.
- Council of Europe (2003). *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF).* Manual: Preliminary Pilot Version. DGIV/EDU/LANG 2003, 5. Strasbourg: Language Policy Division.
- Erkuş, A. (2003). *Psikometri üzerine yazılar.* Ankara: Türk Psikologlar Derneği Yayınları.
- Jones, N. (2002). Relating the ALTE framework to the common European framework of references. *Case Studies in applying the Common European Framework.* (Edt: J. C. Alderson). Strasbourg: Council of Europe. pp 167-183.

- Little, D. (2005). The common European framework and the European language portfolio: involving learners and their judgments in the assessment process. *Language Testing*, 22 (3), 321-336.
- North, B. (2007). *The CEFR Common Reference Levels: validated reference points and local strategies, Lecture at the Common European Framework of Reference for Languages (CEFR) and the development of language policies: challenges and responsibilities*. Policy Forum, Strasbourg.
- Urkun, Z. (2008). Re-evaluating the CEFR: an attempt. *TESOL MIWS newsletter*, 21 (2), 13-19.
- Weir, C. J. (2005). Limitations of the Council of Europe's Framework of Reference (CEFR) in developing comparable examinations and tests. *Language Testing*, 22 (3), 282-300.

Yeni Bir Sınav Kültürü Geliştirmek: Olasılık Sanatı

Evrım ÜSTÜNLÜOĞLU¹

Giriş

Geçerli, güvenilir ve adil sınavlar hazırlamak test yazarlarının yıllardır odak noktası olmuştur. Bu durum, özellikle İngilizce öğretim yapan ve öğrencilerin yabancı dil yeterlik düzeyini doğru bir şekilde tespit etmek durumunda olan okullar için ayrı bir önem taşımaktadır. Bu nedenle, altı yeterlik düzeyi belirtilen Ortak Avrupa Çerçeve Programları, açık hedefler, içerik ve ölçütler sunduğu için İngilizce eğitim veren bazı okullarda programların yapılandırılması ve sınavların hazırlanması aşamasında bir referans olarak kullanılmaktadır. Ancak, Çerçeve Programlarında belirtilen ortak hedef ve ölçütlere uygun yabancı dil sınav hazırlama ve farklı ortamlarda uygulama konusu pek çok tartışmayı da beraberinde getirmiştir. Little (2005) ve Weir (2005) çalışmalarında konuya eleştirel bir yaklaşım getirerek farklı hedefleri ve amaçları olan sınavların Ortak Avrupa Çerçeve Programlarında belirtilen kriterlerin hazırlanması ve değerlendirilmesinin sıkıntı yaratabileceğini belirtmişlerdir. Çerçeve programların dil öğretimi ve ölçme ve değerlendirme sürecine bir standart getirebileceğini belirten North (2007) yine çerçeve programlar kanalıyla farklı sistemler arasında kıyaslamalar yapılabileceğini ve bu yolla dil öğretimi ve değerlendirmelerinin kaliteyi yakalamada etkin bir kanal olabileceğini vurgulamıştır. Çerçeve programlarının bu amaca hizmet etmesi için Alderson (2006) bu sürecin gerçekçi bir şekilde yapılandırılması gerektiğini belirtmiştir. Yapılandırma aşaması, sınav içeriğinin tespit edilmesi, standardizasyon, verilerin toplanıp analiz edilmesi ve aynı zamanda Ortak Çerçeve programlarının çok iyi bilinmesini kapsamaktadır. Bu aşamada dikkat edilmesi gereken en önemli nokta programların doğru bir şekilde yorumlanması ve kültürel yorumlardan mümkün olduğunca sakınılmasıdır (Jones, 2002). Yabancı dil öğretimi ve yabancı dille öğretimin sıklıkla tartışıldığı Türkiye’de, özellikle yabancı dilde eğitim yapan yüksek öğretim kurumlarında uygulanan sınavların zorluğu, kolaylığı, yeterlik düzeyi gibi konular da sıklıkla bu tartışmalarda yerini almaktadır.

Amaç

Bu çalışmanın amacı, İngilizce öğretim yapan bir üniversitenin modüler sisteme geçen hazırlık programlarında öğrencilerin dil yeterliklerinin daha nitelikli ölçülmesi ve değerlendirilmesi amacıyla Ortak Avrupa Çerçeve Programları temel alınarak güvenilir ve geçerli sınav hazırlama sürecini anlatmak ve bu doğrultuda hazırlanan sınavların başarı oranlarını paylaşmaktır. Bu çalışmanın, sayıları giderek artan yükseköğretim hazırlık programları için örnek teşkil edeceği ve üniversiteler arasında yatay geçiş yapmak isteyen öğrencilerin dil yeterliklerinin sorgulanmasında bu çerçevede hazırlanan sınavların standart oluşturması açısından da önemli olacağı düşünülmektedir. Yabancı Diller Yüksekokulları düzeyinde yapılan toplantılarda verilen yeterlik sınavların zorluk derecesi ve farklı kriter ve değerlendirme sistemlerinin kullanılmakta olduğu sıklıkla dile getirilmekte ve sınavların

¹ Yrd. Doç. Dr. İzmir Ekonomi Üniversitesi Yabancı Diller Yüksekokulu - evrim.ustunluoglu@ieu.edu.tr

standartlaştırılmasına yönelik tartışmalar devam etmektedir. Çalışmanın bu amaca hizmet ederek benzer kurumlar için örnek teşkil edeceği de umulmaktadır.

Süreç

Yabancı Diller Yüksekokulu, Hazırlık Programında geliştirilmesi planlanan Modül sonu sınavları bir proje kapsamında yürütülmüştür. Proje; program başkanı, koordinatörler, Müfredat ve Materyal Geliştirme Birimi, Ölçme ve Değerlendirme Birimi, Sınav Hazırlama Birimi ve bir grup öğretmenden oluşan bir ekip tarafından yürütülmüştür. Modüler sistem yaklaşımına dayalı İngilizce Hazırlık Programı öğrenme çıktısı Ortak Avrupa Programları çerçevesinde B2 + olarak belirlenmiştir. Her modül için belirlenen hedefler, hedef davranışlara dönüştürülmüştür. Test maddeleri yazılmadan önce ilkeler açık olarak belirlenmiş ve test yazarları geçerli ve güvenilir sınav yazma konusunda eğitilmişlerdir. Sınavların amaçları ve odak noktaları belirlenerek sınav içerikleri, sınav uzunluğu, madde türleri Ortak Avrupa Çerçeve programlarında belirtilen sınav basamakları da göz önüne alınarak tartışılmıştır. Her seviye için geliştirilen modül sonu sınavları, dinleme ve not alma, yazma, okuma ve konuşma becerilerini içermiş, ilk seviyelerde dil bilgisi de ölçülmüştür. Her bir düzey için geliştirilen sınavların pilot çalışmaları yapılarak sınav süresi ve maddelerde değişiklikler ve düzeltmeler yapılmıştır. Sınavlar uygulanmadan önce sınav örnekleri, konuşma sınavı örneği videosu, cevap anahtarları, değerlendirme ölçütleri web sitesine öğretmen ve öğrencilerin erişilerine açılmıştır. Sözlü ve yazılı bölümler için standardizasyon çalışmaları ile birlikte değerlendirmeler de yapılmıştır.

Sonuç

İstatistik sonuçları madde zorluk ortalamalarının başlangıç düzeyi için 0.60 ve 0.69, az bilenler için 0.53 ve 0.61, orta düzey altı için 0.52 ve 0.61, orta düzey için 0.53 ve 0.65 ve son olarak da orta düzey üstü için 0.57 ve 0.69 arasında olduğunu göstermektedir. Sınavların güvenilirlik katsayıları % 72 ve % 92 arasında bulunmuştur. Bu aralık sınavların iç tutarlığının yüksek olduğunu göstermektedir. İlk üç modülün toplam başarı oranları beklentilerin altında bulunmakla beraber (Modül 1. % 57, Modül 2. % 55, Modül 3. % 52) çerçeve programlarında belirtilen basamakların ödün verilmeden takip edilmesi sınavların amacına hizmet ettiğini göstermektedir. Son Modülde sınav başarı ortalamasının (% 73) yükselmesi sınav sisteminin, öğrenci ve öğretmenler tarafından giderek benimsendiği şeklinde yorumlanabilir.

Modüllerin başarı ortalamaları beklentilerin altında çıkmakla birlikte, kaliteli sınav geliştirmek adına Ortak Çerçeve Programlarında belirtilen sınav geliştirme basamaklarının ödün verilmeden ve gerçekçi bir yaklaşımla takip edilmesi, madde güçlüklerinin normal ve güvenilirlik katsayısının yüksek olması sınav geliştirme sürecini denemeye değer kılmıştır. Ancak düzeyler ve modüller arasındaki başarı farklılıkları ilke, amaç, soru tipleri, bölüm ağırlıklarının tekrar gözden geçirilmesini gerekli kılmaktadır. Güvenirliği yüksek sınavların geliştirilmesi için soru bankası oluşturulması, istatistik çalışmaları sonuçlarına dayalı olarak sınavların bölümleri arasında ağırlıkların gözden geçirilmesi, müfredat ve sınav geliştirmenin doğal bir süreci olarak her bir seviyenin hedef ve amaçlarının düzenli olarak gözden geçirilmesi, yerleştirme sınavı sonucunda seviyelerine yerleşen öğrencilerinin başarılarının izlenmesi ve yerleştirme puanlarının takip edilmesi ve yeterlik sınavı hazırlanması aşamasında benzer basamakların kullanılması çalışmada önerilmektedir. Özellikle üniversiteler arasında yatay ve dikey geçişlerin yapıldığı ve Yabancı Diller Yüksekokulları tarafından verilen yeterlik sınavların sıklıkla sorgulandığı günümüzde bu

alıřmada belirtilen basamakların kullanılması yeterlik sınavlarına standardizasyon getirilmesi aısından nem kazanmaktadır.

Atıf iin / Please cite as:

stünlüođlu, E. (2011). Developing a new test culture: the art of possible [Yeni bir sınav kltürü geliřtirmek: olasılık sanatı]. *Eđitim Bilimleri Arařtırmaları Dergisi - Journal of Educational Sciences Research*, 1 (1), 19–31. <http://ebad-jesr.com/>.