

## CHAPTER 22: TESTING AND EVALUATION

### A. TESTING AND ASSESSMENT

Öğretmenler ya da eğitimciler **testing, evaluation ve assessment** konusunda oldukça çok zaman geçirmektedirler. Ölçme ve değerlendirme; bazen öğrencilerin bir kursa ya da kuruma yerleşmek için yeterli olup olmadığını ölçer, bazen ilerlemenin ne aşamada olduğunu değerlendirir, bazen de öğrenciler kendi yeterliliklerini görmek için kullanılırlar. Bu değerlendirmeler bazen formal ve public bazen de informal ve derslerde gün gün yapılacak şekilde yapılır.

**Summative ve formative assessment** arasında ayırım yapmam mümkündür. Summative değerlendirmeler öğrencilerin değerlendirildiği sene sonu testleri veya birçok öğrencinin girdiği testlerdir. Formative değerlendirme ise öğretmenin öğrencinin performansını artırmak amacıyla kurs boyunca verdiği bir çeşit geri dönütlerdir (feedback). Bu değerlendirme, öğrenciye yanlışı gösterdiğimiz ve öğrencinin yanlışı bulmasına yardım ettiğimiz aşamalarda her zaman yapılabilir ve öğretmen başarı ve gelişim testlerinin sonucunda da bu değerlendirmeye yer verebilir. Bu tarz değerlendirme sonucunda öğretmenin dersin konusunu yada müfredatın ağırlık noktasını değiştirmesi önerisi ortaya çıkabilir. Başka bir deyişle formative assessment'da öğrenci kadar öğretmen de değişmek ve gelişmek zorunda olur.

#### A.1. Different Types of Testing

Ölçmeyi dört bölüme ayıran, dört ana ölçme amacı vardır. Bu bölümler:

- **Placement tests:** Yeni bir öğrenciyi doğru bir sınıfa yerleştirmek placement test ile sağlanır. Bu testler genellikle öğrencinin takip edeceği materyal ve müfredat temellidir ve seviyeye karar verildikten sonra öğrencilerin **productive (speaking-writing)– receptive(listening-reading)** becerileri değerlendirilir. Bazı okullar yerleştirme sürecinin bir parçası olarak öğrencilerden kendilerini değerlendirmelerini isterler ve ortaya çıkan sonucu en son yerleştirme kararına eklerler.
- **Diagnostic Tests:** Placement testler öğrencinin İngilizcesinin ne kadar iyi olduğunu göstermek için tasarlanırken Diagnostic testler öğrencinin öğrenme zorluklarını, bilgi boşluklarını ve beceri eksikliklerini açığa çıkarmak için kullanılabilir. Bu nedenle problemin ne olduğunu biliyorsak bunun hakkında bir şeyler yapabiliriz.
- **Progress or Achievement Tests:** bu testler takip edilen müfredat kapsamında öğrencilerin dil becerilerini ve beceri gelişimini test etmek için oluşturulur.

Progress testler her bir iki haftada bir öğrencinin ne kadar yapabildiğini görmek için öğretmen tarafından yazılır. Bu yolla formative assesment programının bir parçasını oluştururlar.

Achievement testler sadece öğrencilerin içeriğe alışık oldukları durumda işe yarar. Bu mesela bir reading testinde daha önce gördükleri bir verilmesi anlamına gelmez. Bunun anlamı testlerde benzer metinler ve benzer alıştıırma tiplerini görmeleridir.

Eğer öğrenciler tamamen yeni materyallerle karşılaşılırsa bu test genel dil yeterliliğini ölçmesine rağmen kurstaki öğrenmelerini ölçmez.

Achievement testler dönem sonunda başarısızlığı değil ilerlemeyi yansıtmalıdır. Achievement testlerin amacı zayıflığı ortaya çıkarmak için değil öğrenmeyi destekleme amacıyla uygulanmalıdır. Ayrıca bu testler beklediğimizden daha kötü sonuçlara sahip öğrencilerle alakalı gelecek öğretim programlarında değişikliğe karar vermeye yardımcı olur.

- **Proficiency Tests:** proficiency testler öğrencinin ilerlemesini ölçmekten daha çok öğrencinin bilgi ve yeteneğinin genel bir çerçevesini sunar. Genellikle yabancı üniversiteye girmeye çalışan öğrencileri, bir işe girmeye çalışanları ya da bir sertifika almak isteyenleri ölçer. Genel (public) sınavlarının çoğu proficiency test türüdür. Bu testler önemli backwash effect (Bölüm D yi inceleyebilirsiniz) e sahiptir.
- **Portfolio Assessments:** achievement ve proficiency testler belirli zaman diliminde öğrencinin yeteneğini ölçmek ile alakalıdır. Bu testlerde öğrencinin ne kadar bildiğini tek seferde göstermeye çalışır.(one shot).Birçok eğitimciye göre bunun gibi testler (sudden death testing) öğrencilerin belirli durumlarda ne kadar iyi sonuçlar gösterip göstermeyeceklerini ölçmoe konusunda yetersiz kalır. Sonuç olarak birçok eğitim kurumu öğrencilerine bir dönemlik veya sömestrilik zaman dilimi içerisinde çalışmalarının bir porfoliosunu oluşturmalarına izin verir ve öğrenci bu periyottaki en iyi 3-4 çalışmasına göre değerlendirilir.

Bu tarz potfolio değerlendirmelerin önemli yararları vardır. Öğrencilerin çabalarına kanıt sağlarlar. Öğrencilerin daha otonom(kendi kendine iş başarabilen) olmalarına yardım eder, ayrıca kendi öğrenmelerini izlemelerine yardım eder. Portfolio teslim edilmeden önce öğrenciler yaptıklarını kontrol etme seçeneğine sahiptir ve bunun sonucunda bu değerlendirme oldukça çok positive washback effect içerir.

Fakat porfolio assesment tamamen dezavantajsız değildir. İlk olarak uzun zaman gerektirir ve öğretmen portfolyodaki seçenekleri nasıl seçeceği ve bunlara nasıl not vereceği konusunda eğitime ihtiyaç duyar. Bazı öğrenciler portfolyolarını en iyi şekilde yapmak için dönem sonuna bırakabilirler, bunların hepsinin ötesinde öğrenci sınıftan ayrı olarak çalıştığı için portfolyolarının kendi çalışması olduğu veya birinden yardım alıp almadıkları net olarak anlaşılabilir.

## A.2. Characteristic of Good Tests

Herhangi bir testin etkililiğine karar vermek için, bazı kriterleri göz önünde bulundurmak gerekir:

- **Validity (geçerlilik):** Bir test etmesi gerekeni test ediyorsa, bu test geçerli bir testtir. Spesifik biyoloji ve tarih bilgisi gerektiren makale sorularıyla yazma becerisini test etmek geçerli değildir. Diğer ölçümlerle aynı sonucu veren test geçerlidir. Eğer bir dinleme testinde heceleme ve gramer için yazılı kısa cevapları puanlanıyorsa bu dinleme için geçerli bir

test değildir çünkü yanlış şey punlanmış olur. Geçerliliğin bir türü face validity( görünüş geçerliliği) bir testin geçerliliği bakıldığında belli olmalıdır. Sadece üç tane çoktan seçmeli sorusu olan test o testin görünüş geçerliliği konusunda öğrenciyi ikna edemez.

- **Reliability (güvenilirlik):** iyi bir test sürekli aynı sonuçları vermelidir. Örneğin, iki gün içerisinde iki kez aynı testi yapan aynı öğrenci grubu her bir durumda aynı sonuca ulaşmalıdır. Eğer öğrenciler benzer başka bir test alırlarsa sonuçlar aynı olmalıdır. Görünüşte benzer olan iki grup testi yaparsa puanlama derecesi aynı olmalıdır.

Pratikte testin güvenilirliği yönlendirmeleri anlaşılır kılarak, sorunun tek bir cevabı oluşturularak ve test koşullarının sürekliliği sağlanarak oluşturulur.

Güvenilirlik(reliability) testi puanlayan kişiye de bağlıdır. Eğer testi puanlayan kişi sonuçlara kendi puanlamasını aşırı derecede katıyorsa bu test güvenilirliğe sahip olmaz.

## **B.TYPES OF TEST ITEM**

Testin ya da sınavın amacı ne olursa olsun ölçmedeki başarıda ya da başarısızlıkta en önemli faktör testin içerdiği soru çeşitleriyle belirlenir.

### **B.1. Direct and Indirect Test Items**

Eğer test içeriği adayların iletişim becerilerini sergilemelerini istiyorsa bu içerik **direct test item**, öğrencinin bilgi ve yeteneğini onların receptive ve productive becerilerinin altında ne yaptığını tespit ederek ölçüyorsa bu test içeriği de **indirect test item** olur. Yani, direct test soruları olabildiğince günlük yaşamda kullanılan dille benzerlik gösterirken indirect items daha kontrollü sorular yoluyla öğrencinin dil bilgisinin ölçmeye çalışır, çoktan seçmeli ve gramma transformation soruları gibi.

**Discrete- point testing** ve **integrative testing** arasında da bir ayırım yapılması gerekir. Dahası, discrete- point testing sadece bir tek seferde tek şeyi test eder(öğrenciyi fiilin doğru tense yapısını seçmesini sormak gibi), integrative test soruları ise öğrencilerden verilen bir zamanda çok çeşitli dil becerileri kullanmasını ister. –Örneğin; kompozisyon yazarken ya da iletişimsel konuşma testinde farklı beceriler aynı anda kullanılır.

Öğrencilerin birden fazla farklı test kağıdı aldığı proficiency sınavlarında direct, indirect, discrete- point ve integrative testler bulunur. Test hazırlayanlar bu kombinasyonu öğrencinin yeteneğinin genel bir çerçevesini verdiğini savunurlar. Placement testlerde genellikle var olan dil müfredatına karşı öğrencileri ölçmek için discrete- point testi kullanırlar fakat sonra çerçevenin bütünü oluşturmak için daha fazla direct ve integrative test kullanırlar.

### **B.2. Indirect Test Item Types**

Birçok çeşit indirect test biçimi olmasına rağmen en yaygın kullanılan şunlardır:

- **Multiple - choice questions:** geleneksel çoktan seçmeli kelime testi aşağıdaki gibidir.

The journalist was \_\_\_\_\_ by enemy fire as he tried to send a story by satellite phone.  
a wronged      b wounded      c injured      d damaged

Yıllardır çoktan seçmeli soruların öğrencinin gramer ve kelime bilgisini ölçmek için ideal olduğu düşünülmüştür. Bunun sebebi puanlamasının kolay olmasıdır. Dahası bilgisayarların geliştirilmesi sebebiyle bu testlerin cevap kağıtları makinalar tarafından okunur ve olası puanlama hataları engellenir.

Halbuki çoktan seçmeli sorularda birtakım problemler ortaya çıkar. Öncelikle özellikle çeldiricileri(yanlışları) oluşturma aşamasında bu testlerin hazırlanması zordur. Bu çeldiriciler öğrencinin aklına bunları daha önce görmedikleri düşüncesini getirebilir. İkinci olarak; bu testlerle çocukların çoktan seçmeli soru yeteneklerini geliştirmek mümkünken bu testler İngilizcelerini tam olarak geliştirmeyebilir. Test tekniği gelişmiş öğrenci ile gelişmemiş öğrenci arasındaki puan farkı dil bilgisi ve dil yeteneği gelişmiş olmasından değil tekniğin gelişmiş olmasından kaynaklanabilir.

Çoktan seçmeli sorular hala sıklıkla kullanılır, fakat puanlamadaki pratiklik ve puanlama güvenilirliğine rağmen genel güvenilirlik ve geçerliliği tartışılır.

- **Cloze Procedures:** Cloze procedure bize ideal olan indirect test sorularını sağlar ama integrative test içeriği sunmaz. Cloze Procedure soruları kolayca hazırlanabilir. Eğer içerik hakkında öne sürülenler doğruysa bu prosedürü uygulayan kişinin genel bilgisi etkili bir şekilde ölçülebilir.

Cloze -en sade formunda- her .....inci kelimedede bir boşluk bırakılmasıdır.( Her 7-8 kelimedede bir gibi)

They sat on a bench attached 1 \_\_\_\_\_ a picnic table. Below them they 2 \_\_\_\_\_ see the river gurgling between overgrown 3 \_\_\_\_\_. The sky was diamond blue, with 4 \_\_\_\_\_ white clouds dancing in the freshening 5 \_\_\_\_\_. They could hear the call of 6 \_\_\_\_\_ and the buzzing of countless insects. 8 \_\_\_\_\_ were completely alone. etc.

Cloze test kelimelerin rastgele silinmesinden (gramer, collocation, fix phrases, kelime, okuduğunu anlama vb)dolayı görünüş olarak mükemmel bir test olarak algılanır ve integrative olur. Ancak cloze testing çocukların genel İngilizce bilgisini ölçmeden ziyade silinen belirli kelimelerin puanlanmasına dönüşür. Bazılarının uygulanması diğerlerine göre çok zordur ve bazı durumlarda birden fazla doğru cevap olabilir. Yukarıdaki kısa örnekte bile birinci ve sekizinci boşlukların cevabında şüphe yokken dördüncü boşuk daha az tahmin edilebilir bir boşluktur. Farklı metinler farklı sonuçlar doğurur.

Bu tarz güvenilirlik problemlerine rağmen cloze test kullanımını tamamen terk edilmeyecek kadar faydalı bir tekniktir, çünkü doğru kelimeyi boşluğa yerleştirmek parçayı anlamayı, o kelimenin bilgisini ve onu nasıl yerleştireceği bilgisine sahip olmayı gerektirir. Silinen kelimeyi parçadan bulunabilmenin sağlandığı rational ya da modified cloze prosedürlerini kullanmak belki daha iyi olabilir. Öğretmen her 8. 10. kelimedede bir kelime silmek yerine parçanın esnekliğine göre 8. ve 10. kelimenin sağında veya solundaki kelimeyi silme opsiyonuna sahiptir.

- **Transformation and Paraphrase:** Yaygın olan test içeriklerinden biridir ve adaylardan cümleyi yüzeysel bir değişiklik ile orijinal anlamını kaybetmeden yazmaları istenir. Mesela aşağıdaki örnek adayın fiil bilgisini ve cümle yapısını test eder.

**I'm sorry that I didn't get her an anniversary present.**  
**I wish \_\_\_\_\_**

Boşluğu başarılı bir şekilde doldurmak için öğrencinin ilk cümleyi anlaması ve gramatik olarak karşılık gelecek yapıyı bilmesi gerekir. Böylece öğrencinin dil bilgisi sistemi hakkında bilgi edinmiş oluruz.

- **Sentence Reordering:** Öğrencilerden kelimeleri uygun cümle yapmak için doğru sıraya koymalarını istemek bize daha çok öğrencilerin kelime dizilişi ve gramer sözcük bilgileri hakkında bilgi verir. Aşağıdaki örnek bunun bir çeşididir.

**Put the words in order to make correct sentences.**  
**called / I / I'm / in / sorry / wasn't / when / you**

Tek bir doğru sıralamayı elde etmek her zaman mümkün olmasa da yeniden sıralama kolay bir şekilde oluşturulur.

Boşluk doldurmanın da içinde bulunduğu birçok indirect teknik vardır, fill in (boşluk doldurma) ( Jan.....to the gym every Tuesday morning), parçanın ya da cümlelerin doğru zamanını seçme (I have arrived / arrived), cümledeki yanlışları bulma ( she noticed about his new jacket) ve kelimenin doğru formunu seçme ( he didn't enjoy being on the (lose).....side). Önerilen bütün sorular çabuk ve etkili puanlanabilir ve öğrencinin bilgisi hakkında bize bir şeyler anlatmayı amaçlar.

### **B.3. Direct Test Item Types**

Direct test içeriğinde geçerlilik ve güvenilirliği sağlamak için testi oluşturanların aşağıdakilere ihtiyacı vardır.

- **Create a "level playing field":** yazılı test durumunda öğretmen ve adaylar aşağıdaki makale sorusu için kesinlikle şikayetçi olacaklardır.

**Why was the discovery of DNA so important for the science of the twentieth century?**

Çünkü bu soru bilimsel bilgisi olan ve 20. yy ın bilimsel tarihi ile ilgili ön tahminleri olan adaylara üstü kapalı da olsa iyilikte bulunur.

Halbuki aşağıdaki konu tüm adaylara aynı şekilde kazanma şansı verir:

Present a written argument or case to an educated non-specialist audience on the following topic:

**Higher mammals, especially monkeys, have rights and should not be used in laboratory experiments.**

You should write at least 250 words.

You should use your own ideas, knowledge and experience to support your arguments with examples and relevant evidence.

Receptive beceri testinin öğrencilerin genel veya özel bilgilerini aşırı derecede talep etmekten kaçınması gerekir. Eğer öğrencinin yazma ve konuşmadaki performansının iyi olması gerektiği durumda receptive yetenek testi uygulanırsa (reading ve listening testi olduğu durumda) bu durumda ölçüyor olduğumuz şeyin receptive beceri olduğuna emin olamayız.

- **Replicate real-life interaction:** İnsanlar gerçek hayatta konuşma yaparken veya yazı yazarken bunları yapmada bir amaç gözetirler. Fakat geleneksel yazı testleri genellikle özel seçilmiş genel makale soruları üzerine temellendirilir ve konuşma testleri sıklıkla adayların bir durum karşısında söyleyecekleri sözlerin sorusunu içerir. Modern olan test hazırlayanlar şuanda gerçek hayat diyalogları içeren görevleri testlere dahil ediyorlar.

Reading ve listening test içerikleri mümkün olduğunca çok gerçek hayatı yansıtmalıdır.

Yani içerik authentic olmasa da mümkün olduğunca gerçekçi olmalıdır. Öğrencinin anladığını test etme yolları olmasına rağmen (eşleştirmeli görevleri kullanmak ya da çoktan seçmeli sorular) bu kriterler yeterli derecede tatmin edici olmayabilir.

Aşağıdaki direct test içerik çeşitleri yukarıda bahsettiğimiz kriterleri karşılamaya çalışmıştır.

### **SPEAKING**

- An interviewer questions a candidate about themselves.
- Information-gap activities where a candidate has to find out information either from an interlocutor or a fellow candidate. (The role-play on page 359 would not need too much modification to serve as a suitable test item.)
- Decision-making activities, such as showing paired candidates ten photos of people and asking them to put them in order of the best and worst dressed.
- Using pictures for candidates to compare and contrast, whether they can both see them or whether (as in many communication games) they have find similarities and differences without being able to look at each other's material.
- Role-play activities where candidates perform tasks such as introducing themselves or ringing a theatre to book tickets.

### **WRITING**

- Writing compositions and stories.
- 'Transactional letters' where candidates reply to a job advertisement or write a complaint to a hotel based on information given in the exam paper.
- Information leaflets about their school or a place in their town.
- A set of instructions for some common task.
- Newspaper articles about a recent event.

### READING

- Multiple-choice questions to test comprehension of a text.
- Matching written descriptions with pictures of the items or procedure they describe.
- Transferring written information to charts, graphs, maps, etc. (though special care has to be taken not to disadvantage non-mathematically-minded candidates).
- Choosing the best summary of a paragraph or a whole text.
- Matching jumbled headings with paragraphs.
- Inserting sentences provided by the examiner in the correct place in the text.

### LISTENING

- Completing charts with facts and figures from the listening text.
- Identifying which of a number of objects (pictured on the test paper) is being described.
- Identifying who (out of two or three speakers) says what.
- Identifying whether speakers are enthusiastic, encouraging, in disagreement or amused.
- Following directions on a map and identifying the correct house or place.

Güvenilirlikle ilgili olarak listening testlerinde genellikle konuşanın sesine, hızına ve ifadelerine bakmaksızın bütün adayların aynı fırsatlara sahip olduğu teyp veya CD lere başvurulur. Bazen bilgisayarlı TOEFL testlerinde olduğu gibi adaylar kişisel bilgisayarlara bağlı kulaklıkla çalışır. Bir grup öğrencinin aynı kaydı dinlediği ortamlarda ses kaydının açık ve kolayca anlaşılabilir olduğundan emin olmalıyız.

## C. WRITING AND MARKING TESTS

Öğretmenlik kariyerimiz boyunca öğrencilere test hazırlamada ve öğrencilerin tamamladığı testleri puanlamak zorunda kalacağız.

### C.1. Writing Tests

Test hazırlamadan ve bir gruba vermeden önce yapılması gereken bazı şeyler vardır.

- **Asses the test situation:** testi hazırlamaya başlamadan önce testte yer alacak içeriği düşünmeliyiz. Teste ne kadar süre verileceğine, testin ne zaman nerede olacağına, puanlamak için ne kadar zamanın gerekli olacağına karar vermeliyiz.
- **Decide what to test:** testimizin içine neleri ekleyeceğimizi listelemek zorundayız. Yani reading - speaking becerilerden hangisinin teste gireceğine hangisinin dışarıda kalacağına karar vermemiz gerekir. Hangi müfredat içerikleri teste girecek ve ne tür konu ve durumlar öğrencimiz için uygundur bunlar belirlenmelidir. Dönem boyunca öğrencilerimize öğrettiğimiz kelime ve gramer noktalarının listesine sahip olmamız her şeyi test etmek zorunda olduğumuz anlamına gelmez. Eğer biz listeden temsilci örnekler teste koyarsak öğrencinin o sorulardaki başarısı veya başarısızlığı öğretmenlere öğrencilerin ne kadarını öğrendiklerinin bir göstergesi olacaktır.
- **Balance the elements:** eğer direct ve indirect içerik ekleyeceksek kaç tane ekleyeceğimize karar vermeliyiz.

Dengeleme elementleri testin her bir bölümünün ne kadar sürmesini istediğimiz belirlemeyi ve o zaman içinde yer alacak test içeriklerini oluşturma hesaplamalarını içerir. Yeteri kadar zaman ve yer verdiğimiz elementler öğrettiklerimizin önemini yansıtmalıdır.

- **Weight the scores:** Testimizin ne kadar iyi ayarlasak da öğrencilerimizin başarı ve başarısızlıklarını ölçmemiz testin her bir bölümüne kaç puan verildiğine bağlıdır. Eğer çoktan seçmeli soruların onundan her birine 2 puan verirse transformation sorularının her birine 1 puan verebiliriz. Bu öğrenciler için önceki bölümü sonrakinden daha iyi yapmak gerektiğini anlamına gelir.
- **Make the test work:** Her bir soruyu veya testin tümünü gerçek adaylara sormadan önce meslektaşlarımız veya başka öğrenciler üzerinde denememiz hayati öneme sahiptir. Bir test hazırladığımızda yapacağımız ilk iş onu meslektaşlarımıza göstermektir. Genellikle onlar bizim farkında olmadığımız problemleri tespit ederler veya bizim tahmin edemediğimiz olası cevap ve alternatifleri sunarlar. Meslektaşlarımızın uyarılarına bağlı olarak değişiklikleri yaptıktan sonra testi öğrenciler üzerinde denemek isteriz. Tabi bunu test yapacağımız gerçek öğrenciler üzerinde yapamayız. Fakat hemen hemen aynı seviyede veya bir seviye üstte bir sınıf bulabilirsek kısa zamanda hangi soruların gereksiz problemlere neden olduğunu ve testin ne kadar zaman aldığını tespit edebiliriz. Bu tarz denemeler felaketten kaçınmak ve çeşitli test sonuçlarına çok fazla olası cevaplar verilmesine engel olmak için yapılır.

## C.2. Marking Tests

Yıllar önce Cryil Weir ,TESOL da yaptığı sınavlardan birisinin 8 yazılı kopyasını izlenimci olan mastır öğrencilerine puanlamaları için verdiği puanlar toplamda olması gerekenden 20 puan az ya da çok çıkmıştır. Yazılı kağıtlarını okuyanlardan bazıları diğerlerinden daha fazla puan vermiştir. Bazı el yazısı kağıtlarda puanlama aşırılığa kaçmıştır. Mesela birinde en düşük puan 5 iken diğerinde 20 olmuş. Bunun üzerine testleri puanlarken dikkat edilecek hususlar ortaya çıkarılmıştır.

Aşağıda böyle subjektif durumların önüne geçmek için bazı önlemler verilmiştir.

- **Training:** Eğer puanlayıcılar farklı seviyelerde değişik metinler görüp, hangi puanların verilmesi gerektiği hakkında tartışsalar di sonuçlar yukarıdaki gibi çok farklı olmazdı. Eğer puanlayıcılara videolaştırılmış sözlü testler izlettirilirse doğru bir şekilde konuşulan İngilizce örneklerinin ve performansın ön tanımlarının değerlendirilmesi öğretilir.
- **More Than More Scorer:** Güvenilirlik birden fazla puanlayıcıyla sağlama alınır ve ne kadar çok kişi o testi puanlarsa testin o kadar güvenli puanlanması sağlanır. Sözlü testleri iki gözlemcinin puanlaması tek gözlemcinin puanlamasına göre daha güvenilirdir.



Birçok genel sınav kurulu puanlama için moderatörleri kullanılırlar, bu moderatörlerin işi bireysel test okuyucularının genel test standartlarına uygunluğunu görmek için çalışmalarını kontrol etmektir.

- **Global Assesment Scales:** productive beceri çalışmalarını özel puanlama yöntemleri kullanarak ölçmek mümkündür. Bu yönlendirmeler öğrencilere gerekli puanı alabilmeleri için hangi yeteneklere ihtiyaçları olduklarını söyler. Aşağıdaki örnekte konuşma ölçütü verilmiştir.

Score	Description
0	The candidate is almost unintelligible, uses words wrongly and shows no sign of any grammatical understanding.
1	The candidate is able to transmit only very basic ideas, using individual words rather than phrases or fuller patterns of discourse. Speech is very hesitant and the pronunciation makes intelligibility difficult.
2	The candidate transmits basic ideas in a fairly stilted way. Pronunciation is sometimes problematic and there are examples of grammatical and lexical misuse and gaps which impede communication on occasions.

Score	Description
3	The candidate transmits ideas moderately clearly. Speech is somewhat hesitant and there are frequent lapses in grammar and vocabulary use. Nevertheless, the candidate makes him/herself understood.
4	The candidate speaks fairly fluently, showing an ability to communicate ideas with not too much trouble. There are some problems of grammatical accuracy and some words are inappropriately used.
5	The candidate speaks fluently with few obvious mistakes and a wide variety of lexis and expression. Pronunciation is almost always intelligible, and there is little difficulty in communicating ideas.

Global değerlendirme problemsiz değil fakat öğrencinin konuşması telaffuz açısından zayıf ama gramer olarak doğru olduğu durumda ölçü örnekteki tabloyla eşleşmeyebilir. Ayrıca farklı öğretmenlerin örnekteki tanımların anlamlarıyla aynı fikirde olmaması tehlikesi vardır. Global değerlendirme kendi başına başarmak istediğimiz güvenilirliğin birazını düşürebilir.

- **Analitic Profiles:** öğrencinin performansının detaylı bir şekilde analiz edilmesi güvenilir puanlamayı sağlar. Genel değerlendirmenin yerine çeşitli elementler üzerinden puan verilir. Sözlü değerlendirme için öğrencilerin konuşmasını telaffuz, akıcılık, kelimelerin kullanımı, gramer ve zeka gibi farklı yollarla değerlendirebiliriz. Öğrencilerin yeteneklerini düzeltme yapmadan ve onlar için hazırladığımız taskı ne kadar başarılı tamamladıklarını puanlamak, değerlendirmek isteriz.

Criteria	Score (see analytic scales)
Pronunciation	
Fluency	
Use of vocabulary	
Use of grammar	
Intelligibility	
Repair skills	
Task completion	

Şimdi her bir kriter için ayrı bir analitik değerlendirme yapabiliriz.

Score	Description
0	The candidate cannot get words or phrases out at all.
1	The candidate speaks hesitatingly in short, interrupted bursts.
2	The candidate speaks slowly with frequent pauses.
3	The candidate speaks at a comfortable speed with quite a lot of pauses and hesitations.
4	The candidate speaks at a comfortable speed with only an occasional pause or upset.
5	The candidate speaks quickly with few hesitations.

Global ve analitik puanlamanın kombinasyonu bize güvenilir puanlama yapma şansı verir. Fakat kriterlerin çokluğu test puanlamasını uzun ve elverişsiz yapabilir. Test oluşturanlar ve yönetenler güvenilirlik ve uygulanabilirlik durumlarını ölçmek zorundadır.

- **Scoring and interacting during oral tests:** testlerin güvenilirliğini global değerlendirme puanları ve analitik profillerin yanı sıra puanlayıcının rolünün interlocutor'un(the examiner who guides and provokes conversation)rolünden ayırmakta yardım eder. Bu belki uygulama problemlerine sebep olabilir fakat bu puanlayıcıya adayla etkileşim kurmadan gözlemleyip değerlendirme şansı verir. Birçok konuşma testinde genuine interaction sağlaması ve öğrencinin rahatlaması sebebiyle öğrenciler önceden belirli görevler için eşli ve grup olarak test ediliyorlar.Böylece interlocutor ve aday etkileşimi de olmamış oluyor.Fakat yorumculardan bir tanesi öğrencilerin bu yolla eşleştirilmesi adayların gerçek seviyelerinin altında performans sergileme endişesi taşır ve aynı ana dilden öğrenciler eşleştiğinde zekaların eski seviyesinden aşağıya düşebilir.

## D.TEACHING FOR TESTS

Öğretmen ve test oluşturanların zihnini meşgul eden durumlardan birisi de **washback** veya **backwash effect** diye adlandırılan etkidir.Bu öğretmenlerin öğrencilerinin testlerini ve girecekleri sınavları geçebilmelerini istemeleri gerçeğine ve öğretmenlerin öğretiminin test ağırlıklı hale gelmesi anlamına geliyor. Öğrencilerini sınava hazırlamayan öğretmenler çok değişik aktiviteler ve teknikler kullanırken öğrencilerini sınava hazırlayan öğretmenler washback etkisi ile sınav format aktivitelerine bağlı kalırlar.Bu durumda sınavın formatı dersin de formatını belirler.

Washback etkisini tartışırken 2 noktanın dikkate alınması gerekiyor.İlk olarak özellikle direct test içeriklerinin içinde bulunduğu modern testler, testin önceki bazı örneklerinden ziyade asıl olan sınıf aktiviteleri ve metotlarla oluşturulur.Fakat, ikinci olarak öğrencileri belirli bir test formatına hazırlasak bile öğrencilerin hedeflerini gerçekleştirebilmesi için sınav format derslerinin içinde çeşitlilik ve eğlence ortamının hazırlanması gerekir.

Daha öteye gidersek; birçok öğretmen kesin,net amaçları olan ve olabildiğince yüksek motive edilen öğrencilerin bulunduğu sınav öğretme sınıflarını tatmin edici bulurlar ve bu

öğrencilere amaçları net olmayan öğrencilere göre bir şeyler öğretmek daha kolaydır. Bütün sınıfın hedeflediği bir şey olduğunda öğrenciler hedefin olmadığı durumdan daha azimle sebat ederek çalışır. Dahası öğrencilerin sınav becerilerini eğitirken kendi başlarına çalışma, yaptıklarını gözden geçirme, referans araçları kullanma -mesela sözlük, gramer kitapları, internet, günlük, bağımsız öğrenme kayıtları vs. gibi yöntemler kullanılır. Biz öğretmenler otonom öğrenmeye katkıda bulunan yaklaşımları destekleriz.

İyi sınav hazırlayacak öğretmenler testlerle kendilerini öğrencilerine tanıtmalı ve öğrencilerin ilgi ve endişelerine cevap vermelidir. Hepsinden öte öğretmenlerin iyi sınav hazırlama çizgisinde yürümeleri gerekir ve washback etkisiyle sürüklenmemeleri gerekir. Bu bağlam içinde sınav format sınıflarında yapabileceğimiz şeyler şunlardır:

- **Train For Test Types:** Çocuklara farklı test çeşitleri gösterebiliriz ve onlar bu sayede neyin gerekli olduğu konusunda fikir sahibi olurlar. Testi oluşturanın amacını anlamaları için yardım edebiliriz. Puanlama cetveli göstererek başarıyı neyin getirdiğini farkında olmalarını sağlarız. Öğrenciler test çeşidini tanımladıktan sonra sorulara kaç puan ve neden verebileceğini onlara söyleyebiliriz. Öğrencileri test sorularına etkili bir şekilde yaklaşmaları konusunda eğitebiliriz. Mesela konuşma aktivitesi için bu tür tasklerdeki uygunsuz iletişimin üstesinden gelebilmeleri için öğrencileri uygun etkileşim diliyle hazırlarız. Eğer test ya da sınav çoktan seçmeli soruları içeriyorsa onlara öncelikle ilk çeldiriciyi bulmanın avantajını anlamalarına yardım edebiliriz. Onlar sonra diğer çeldiricilerin benzerlik ve farklılıklarının ne olduğunu bulabilir. Böylece hedeflenen anlam ve gramer alanını tanımlayabilirler. Bu çalışmaların tümündeki görevimiz öğrencilerin karşılaşacakları test soruları ile ilgili esaslı bir araştırma yapmalarını sağlamaktır. Böylece öğrenciler ellerinden gelenin en iyisini yapacaklar ve test de onların İngilizce seviyelerini ortaya çıkaracaktır.
- **Discuss General Exam Skills:** Birçok öğrenci genel test ve sınav becerileriyle alakalı bilgilendirilmekten fayda sağlarlar böylece çalışmaları da boşa gitmez. Onlardan ne beklediğini tam olarak bilmeleri için soruları dikkatlice okumaları gerekir. Öğrencilerin kendilerini ilerletmeye ihtiyaçları vardır, bu yüzden sınavın sadece bir kısmına orantısız zaman ayırmamalıdır. Örneğin writing de taskler için process becerilere başvurmaları gerekir. Öğrencilerin sınava hazırlanırken çalışmalarını organize etmeleri gerekir. Böylece çalışmalarını etkili bir şekilde tekrar gözden geçirebilirler.
- **Do Practice Tests :** Öğrencilerin test ve sınavları pratik yapma şansları olmalıdır böylece deneyim kazandıklarını hissederler. Öğrenciler kursun çeşitli noktalarında pratik testleriyle gerçek sınavın provasını yapabilirler.
- **Have Fun:** Yukarıda bahsedilen durumların yaşanması, öğrencilerin bazı test tiplerini pratik etmeye ihtiyacı olması bunların sıkıcı ve endişelendirici olması anlamına gelmez. Sınavları ve testleri eğlenceli yapma yolları vardı. Eğer bir test tipi öğrencilerden kelimeleri düz cümle haline getirmelerini istiyorsa, öğretmen fiziksel olarak bir araya geldiğinde cümleler oluşturacak resimler vererek sınıfı bu tarz sorular için hazırlar.

Öğrenciler şuna kadar gördükleri örneklere ve üzerinde çalıştıkları dile bağlı kendi test sorularını yazmaya cesaretlendirilebilirler. Ne kadar zor ve ne derece iyi yazıldığını görmeleri için yeni test soruları diğer öğrencilere verilebilir. Bu, öğrencilere soru yazarları ve onların testleri hakkında bilgi verir.

- **Ignore The Test:** Sınav için çalışan öğrenciler eğer sadece test tipine bakarlarsa, sınav tekniğini tartışıp testi alıştırma gibi görürlerse dersler monoton hale gelebilir. Sınava hazırlanma pahasına genel İngilizce gelişiminin tehlikeye atılma tehlikesi vardır. Bir öğrenciyi sınava hazırlarken zaman zaman sınavları göz ardı etmek gerekir. Böylece genel dil durumları üzerinde çalışma şansını elde ederiz ve bu sayede öğrenciler İngilizce dersleri için uygun olan motive edici aktivitelerde yer alabilirler.