



## Impact of Standard Setting Methodologies over Passing Scores\*

Sevda ÇETİN\*\* Selahattin GELBAL\*\*\*

**ABSTRACT.** In this study impact of standard setting methodologies over passing scores is studied and the changes on students' pass/fail scores using different setting methodologies are researched. The data of the research in the first phase is collected from test results which are developed by researcher using 25 items per Turkish Language and Mathematics exams and applied to 129, 7th grade, Elementary School students. The data for second phase are collected by calculating cut points standing on different expert judgments per different methodology. It is seen by this study that the decision on standard setting methodologies for the students are related to students group, the structure of test methods and instructors.

**Key Words:** Standard setting methodologies, cut point, Angoff Methods, Nedelsky Methods, 1-0 (Yes/No) Methods

### SUMMARY

**Purpose and Significance:** In this study impact of standard setting methodologies over passing scores is studied and the changes on students' pass/fail scores using different setting methodologies are researched.

---

\* This study composed of part of the master thesis called "Impacts Of Standard Setting Methodologies Over Passing Scores".

\*\* Research Assist. Hacettepe University, Faculty of Education, Department of Educational Sciences, tsevda@hacettepe.edu.tr

\*\*\* Assist. Prof. Hacettepe University, Faculty of Education, Department of Educational Sciences, gelbal@hacettepe.edu.tr

Standards are defined to be able to make decision about students performance levels. There are several standard setting methodologies in literature but schools and teachers do not have enough information about these methodologies, their application and impacts over passing scores. This study will supply important information to institutions, schools and teachers.

**Methods:** The data of the research in the first phase is collected from test results which are developed by researcher using 25 items per Turkish Language and Mathematics exams and applied to 129, 7th grade, Elementary School students. The data for second phase are collected by calculating cut points standing on different expert judgments per different methodology.

At the analyzing phase of the collected data, differences on ratio of pass scores, defined by different standard setting methodologies, to students who have the grades over pass scores and differences about pass scores based on sort of lessons such as applied sciences or social sciences are tested by using dependent ratio test, also significance level is tested by Z- test. Agreement between methodologies is tested using tetra-choric correlation and Cohen's Kappa Methods.

**Results:** By this research, the differentiation on ratio levels between Angoff and Yes/No methods is seen about 0,05 and 0,01 for the other methods when the student pass scores are compared in twos determined by different setting methodologies for Turkish Language Lesson. By the side of Mathematics Lesson, the differentiation on ratio levels is seen about 0,01 for all the methodologies compared in twos.

A significant difference is also seen about a level of 0.01 by the comparison between the contents (applied or social sciences) of the lessons and the students have grades over pass scores.

By the side of agreement between methodologies; while the results for Turkish Lesson by the aspect of Angoff-Yes/No methodologies are positive with a high level significancy and a positive high level agreement range is seen between Angoff-Nedelsky methodologies and Nedelsky-Yes/No methodologies, the results for Mathematics lesson using Angoff and Yes/No methodologies are positive acceptable agreement with an significant level, and a positive low level agreement is seen between Angoff -Nedelsky methodologies and Nedelsky and Yes/No methodologies.

**Discussion and Conclusions:** In this study, it is seen that the decision on standards setting methodologies for the students are related to students group, the structure of test methods and instructors.



## Farklı Standart Belirleme Yöntemlerinin Geçme Puanları Üzerine Etkisi\*

Sevda ÇETİN\*\*

Selahattin GELBAL\*\*\*

**ÖZ.** Bu araştırmada standart belirleme yöntemlerinin geçme puanları üzerine etkisi üzerinde çalışılmış, öğrencilerin geçti/kaldı durumlarının farklı standart belirleme yöntemlerine göre değişiklik gösterip göstermediği incelenmiştir. Araştırmanın kapsamında kullanılan öğrenci verileri, öğrencilerin biri sözel diğeri sayısal bir ders olmak üzere iki farklı dersten almış oldukları test puanlarıdır. Bu puanları elde etmek amacı ile 25'er maddeden oluşan Türkçe ve Matematik başarı testleri kullanılmıştır. Araştırmanın ikinci aşamasında gerekli olan veriler ise her bir yöntem için farklı hesaplanan ve farklı uzman görüşlerine dayanan puanlardır. Araştırma ile öğrenciler hakkında verilecek kararlar için kullanılacak Angoff Yöntemi, Nedelsky Yöntemi ve 1-0 (Yes/No) standart belirleme yöntemlerinden hangisinin seçileceği öğrenci grubuna, derse, kullanılacak testin yapısına ve öğretmenlere bağlı olarak değişebileceği görülmüştür.

**Anahtar Sözcükler:** Standart belirleme yöntemleri, kesme puan, Angoff Yöntemi, Nedelsky Yöntemi, 1-0 (Yes/No) Yöntemi

\* Bu makale "Farklı Standart Belirleme Yöntemlerinin Geçme Puanları Üzerine Etkisi" adlı Yüksek Lisans tezinin bir kısmından oluşmaktadır.

\*\* Araş. Gör. Hacettepe Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, tsevda@hacettepe.edu.tr

\*\*\* Doç. Dr. Hacettepe Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, gelbal@hacettepe.edu.tr

## GİRİŞ

Eğitim, öğretim alanında öğrenciler, öğrenme eksikliklerinin, ilgi ve yeteneklerinin saptanması, öğretim etkililiğinin ve programlarının değerlendirilmesi için birçok sınava tabi tutulmaktadır. ÖSS (Öğrenci Seçme Sınavı), SBS (Seviye Belirleme Sınavı), KPDS (Kamu Personel Dil Sınavı), LES (Lisansüstü Eğitim Sınavı) sınavları ülkemizde öğrenci seçiminde kullanılan güncel sınavlardır. ÖSS, öğrencilerin üniversitelerce seçilmesinde tek etkin sınav rolünderken, SBS ilköğretim kurumlarındaki öğrencilerin daha nitelikli ortaöğretim eğitimi alabilmek için katıldıkları bir sınavdır. Uluslar arası düzeyde yapılan PIRLS (Progress in International Reading Literacy Study), PISA (Program for International Student Assessment) ve TIMSS (Trends in International Mathematics and Science Study) sınavları ise ülkeler arası eğitim düzeyinin karşılaştırılmasında etkindirler.

Sözö edilen sınavlardan KPDS ve ÜDS benzeri sınavların değerlendirilmesinde mutlak değerlendirme kullanılırken; ÖSS, OKS ve LES benzeri sınavların değerlendirilmesinde bağıl değerlendirme yöntemi kullanılır. Mutlak değerlendirmede öğrencinin mutlak başarı düzeyi kestirilir ve buna göre not verilir. Başarıyı oluşturan davranışların mutlak sıfır noktası ve standart birimleri bulunmamaktadır (Turgut, 1992). Bağıl değerlendirmede ise öğrenci başarısı başka bir ölçüt ile kıyaslanarak not verilir. Bu ölçüt *standart puan*, *kesme puanı* ya da *geçme puanı* olarak adlandırılmaktadır.

Standart puan; puan cetveli üzerinde değişik yeterlilik düzeylerini yansıtan öğrencileri birden fazla sınıflandırmaya tabi tutan bir puanlamadır. Örneğin; İleri Düzey (Advanced) denildiğinde “Yüksek” performans ya da “Pekiyi”, Yeterli (Proficient) düzeyden bahsedildiğinde söz konusu konu ile ilgili “yeterli” bilgiye sahip ya da “İyi”, Temel (Basic) düzeyde sadece temel bilgi ve becerilere sahip ya da “Orta”, Temel Düzey Yakın (Approaching Basic) temel bilgi ve becerilere kısmen sahip ya da “Geçer”, Yetersiz (Unsatisfactory) düzeyde ise temel bilgi ve becerilere “sahip değil” ya da “Zayıf” anlaşılmaktadır. Ülkemiz eğitim kurumlarında kullanılan sınıflandırmalar ise ilköğretim kurumlarında 5-Pekiyi, 4-İyi, 3-Orta, 2-Geçer, 1-Zayıf şeklinde iken yükseköğretim kurumlarında; A-Pekiyi, B-İyi, C- Orta, D- Geçer, F-Zayıf şeklindedir.

Yukarıda örneklenen sınıflandırmalar standart puanların belirlenme sürecinden sonra ortaya konulmaktadır. Bu süreç, *standart belirleme* olarak değerlendirilmektedir. Ancak, standart belirleme yalnızca bu süreçten ibaret değildir. Standart belirleyicilerin işi kategoriler arasında bir sınır aramak ve

sınır oluşturmak ile birlikte aynı zamanda uzman kanılarını ülkesel politik kararlar ile uyumlu hale getirerek puan cetveline yansıtmaktır.

Standart belirlemeye duyulan ihtiyaç, test puanlarının ne amaçla kullanılacağına bağlıdır. Test maddelerine bağlı olarak ölçüldüğünde, kişinin performansı hakkında tanımlayıcı bilgiler elde edilmiş olunur. Test puanları genellikle öğrenciler arasında karşılaştırma yapabilmek ve yapılan bu karşılaştırmalar sonucunda bir karara varabilmek için kullanılmaktadır. “A kişisi B konusu ile ilgili bir testte maddelerin %85’ini doğru cevaplamıştır.” gibi bir tanımlayıcı bilgi değerlendirilerek, kararlar bu değerlendirmeye göre verilmelidir. Örneğin B konusunda bu kadar yüksek performans göstermiş bir kişi takip eden konuda da aynı başarıyı gösterebilecek midir? Kişinin performans düzeyi daha fazla çalışmayı gerektiriyor mudur? Kişinin performans düzeyi öğretmenler ve müfredatın hedeflediği amaçlar için yeterli midir? İşte tüm bu soruları ve daha başka birçok soruyu cevaplayabilmek için standartlar belirlenmelidir. Standart belirleme kişilerle ilgili bilgilerin test ile toplanması, testin diğer bilgi kaynakları açısından iyi bilinmesi ve daha kaliteli olması ile ilgilidir. Standart belirleyicilere göre testlerin uygun kullanımı ile kişiler ve programlar hakkında daha geniş bilgilere ulaşılarak eğitim ve iş alanında daha açık ve kullanışlı bir gereç sağlanmış olur.

Standart belirleme sürecinde dikkate alınması gereken ilk ve en önemli noktalardan biri ise standardı belirlemede kullanılacak yöntemin seçimidir. Bu süreçte birçok faktörün bir arada düşünülmesi gerekmektedir. Berk (1986) bu süreçle ilgili olarak aşağıdaki ölçütleri belirlemiştir:

- Yöntem sınıflamaya uygun olarak iki boyutlu ya da çok boyutlu bilgi verebilmelidir.
- Yöntem öğrenci performanslarını belirlemede hassas olmalıdır. Yani kesme puanları gerçekçi olmalı ve öğrencilerin öğrenmiş oldukları konu ile ilgili olarak sadece test sonuçlarını değil performans gereklerini de yansıtabilmelidir.
- Seçilen yöntem testin güvenilirliğini sağlayabilmek üzere, öğrencilerin öğrenme şansı olduğu konuları içermelidir.
- Yöntem istatistiksel açıdan üzerinde çalışılabilir olmalıdır.
- Testin güvenilirliği ve ölçme hataları kesme puanları belirlenirken dikkate alınmalıdır.

- Yöntem geçerlilik kanıtlarını taşımaları, performans standartları ve bunlara karşılık gelen kesme puanları, başka performans değişkenleri ya da gerçeği yansıtmaya açısından aynı düzeyi karşılayabilmelidir.
- Yöntemin uygulanması kolay olmalı ve tamamlanması çok uzun zaman almamalıdır. Süreçteki her basamak sistematik olmalıdır.
- Yöntemin içerdiği hesaplamalar hesap makinesi ya da bilgisayar destekli bir istatistik yazılımı ile yapılabilecek basitlikte olmalıdır.
- Yöntem ilgili kişiler tarafından kolay anlaşılır ve kolay değerlendirilebilir olmalıdır.

Hambleton ve Eignor (1979) standart belirleme yöntemlerini yargısal, deneysel ve birleştirilmiş yöntemler olmak üzere üç kategoriye ayırmışlardır. Yargısal yöntemde standardı belirleyebilmek için uzmanlardan toplanmış verilere ve uzman görüşlerine ihtiyaç vardır. Deneysel yöntemde öğrencilerden test yolu ile toplanmış veriler ile standartlar belirlenirken, birleştirilmiş yöntemde ise standartlar hem uzmanlardan toplanan veriler, hem de deneysel yolla toplanmış veriler ile belirlenir.

Jeager (1989) ise standart belirleme yöntemlerini “*test merkezli modeller*” ve “*öğrenci merkezli modeller*” olarak ayırmıştır. Bu iki sınıflama arasındaki ayırım uzman kararlarında nelerin göz önüne alındığı ile ilgilidir. Test merkezli modelde uzman kararları testin içeriğine göre verilir. Bu, test maddeleri ayrı ayrı incelenerek ya da testin tamamı bütünsel (holistik) anlamda ele alınarak yapılır. Öğrenci merkezli modellerde ise ele alınan konu ile ilgili olarak beklenen yeterlik düzeylerine göre öğrencilerin yeterlilikleri hakkında kararlar alınır.

Yukarıda sözü edilen modellerden işaretleme yöntemi, test merkezli ve öğrenci merkezli modeller ve bu modeller için standart belirleme yöntemleri aşağıda kısaca açıklanmaktadır:

**1. İşaretleme (Bookmark) Yöntemi:** İşaretleme yöntemi CTB/McGrawHill tarafından 1996’da geliştirilmiştir. Diğer yöntemlere göre çok yeni ve daha az bilinmektedir. Yöntem maddeleri tek tek baz almak yerine testin bütününe esas alır. Test maddeleri IRT (Item Response Theory) kullanılarak zorluk derecelerine göre IRT yetenek ölçeğinde kolaydan zora doğru sıralanır. Uzmanlar bu ölçek üzerinde kesme puanının nerede olması gerektiğini belirler. “İleri Düzey”, “Yeterli” ve “Yetersiz” olmak üzere üç standart belirlendiğinde uzmanlar öncelikle her standart düzeyindeki kişilerin neleri yapabileceklerini tanımlar; daha sonra da bunu karşılayabilecek kesme puanını hesaplarlar. Böylece, maddeleri bir

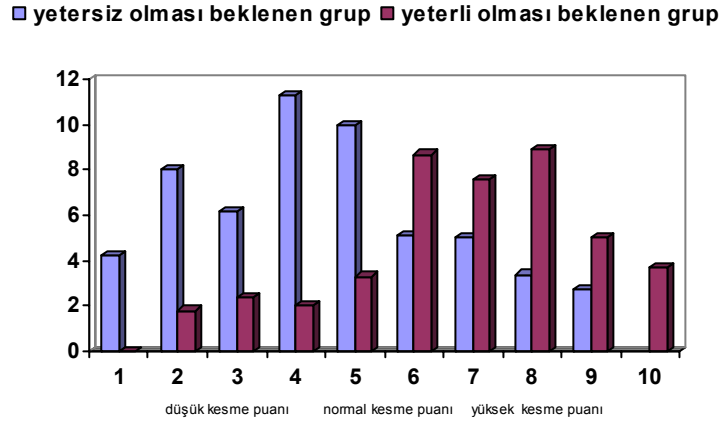
performans düzeyini diğerinden ayırt edebilecek şekilde işaretlemiş olurlar. Burada seçimler yapılırken birçok uzmanın ortak görüşüne başvurulmaktadır. Öncelikle her uzman kendi belirlediği kesme puanını sunmakta, sonra da diğer uzmanların belirledikleri ile uyumsuzluk olup olmadığına bakılmaktadır. Böylece uzmanların performans standardı tanımlamalarında ortak kaniya varmaları sağlanarak işaretleme süreci tekrarlanır ve her düzey için bir kesme puanı belirlenmiş olur. Bu yöntemde delfi (delphi) yöntemi kullanılarak da uzman kararlarının tutarlılığına bakılabilir.

## *2. Öğrenci Merkezli Modeller*

**Karşıt Gruplar Yöntemi:** Karşıt gruplar standart belirleme sürecinde farklı yetenek grupları arasında beklenen performans düzeyi ile var olan performans düzeyi karşılaştırılmakta, kesme puanı da bu karşılaştırmaya göre belirlenmektedir. Örneğin, onuncu sınıf bir matematik testi için, kesme puanı belirleme çalışmasında öncelikle öğrenci grupları özel olarak seçilip belirlenmelidir. İlk gruptaki öğrenciler matematik düzeyi iyi, çeşitli kurslar almış öğrencilerden oluşurken ikinci gruptakiler daha düşük düzeydeki öğrencilerden meydana gelmelidir. İlk grup yeterli olması beklenen grup, ikincisi ise yetersiz olması beklenen bir gruptur. Her iki gruba da test uygulandıktan sonra test puanları grafiğe yerleştirilir. Böylece kişilerin aldıkları puanlarla yeterlilikleri karşılaştırılabilmektedir.

Horn, Ramas, Blumer, Madaus, (2000) bu yöntemi Şekil 1'deki grafik ile açıklamışlardır.

Şekil 1'deki grafik incelendiğinde yeterli olması beklenen grup öğrencileri yüksek puanlar etrafında yığılma gösterirken, yetersiz olması beklenen grup öğrencileri düşük puanlar etrafında yığılma göstermişlerdir. Bu durumda uzmanlar hangi puanın yeterli ve yetersiz öğrencileri en iyi ayırabileceğini belirlemelidirler. Eğer kesme noktası çok yüksek olursa yeterli olması beklenen birçok öğrenci kesme puanının altında kalır ve yetersiz olarak görülürken; kesme puanının çok düşük belirlendiği bir durumda, tersi söz konusu olur ve yetersiz olması beklenen birçok öğrenci kesme puanı üstünde bir puan alır, yeterli görünür.



Şekil 1: Yeterli olması beklenen grup puanları ile yetersiz olması beklenen grup puanları dağılımı

Bu durumda en iyi kesme puanı, her iki grup için de kesme puanı altında kalan öğrenci sayısını minimize edilebilir. Örneğin şekil 1'deki en düşük kesme puanı 2-3 aralığında iken en yüksek kesme puanı 8-9 aralığında, en iyi kesme puanı ise 5-6 aralığındadır.

**Sınır Grup Yöntemi:** Livingston ve Zieky (1989) tarafından kesme puanı belirleyebilmek için 1981'de geliştirilen bu yöntem çeşitli şekillerde uygulanabilir. İlk aşamada öğrenciler hakkında bilgi sahibi olan uzmanlardan öğrencileri uygulanacak testin içeriğine göre "1 → yeterli", "2 → sınırda", "3 → yetersiz" olmak üzere üç kategoriye ayırmaları istenir. Test, öğrencilere uygulandıktan sonra uzmanlardan geri dönüp daha önce belirlemiş oldukları yeterli ve yetersiz grup öğrencilerinden hangilerinin bu iki sınıflama arasında sınırda olduğunu belirlemeleri istenir. Belirlenen bu öğrenciler "sınır grup" olarak isimlendirilir ve bu gruptaki öğrenci puanlarının ortancası kesme puanı olarak alınır.

### 3. Test Merkezli Modeller

**Ebel Yöntemi:** Ebel (1972) test maddelerini uygunluk ve zorluk olmak üzere iki boyutlu olarak ele almıştır. Uygunluk boyutunda *gerekli, önemli, kabul edilebilir ve tartışılabilir* olmak üzere dört düzey bulunurken; zorluk boyutunda *kolay, orta ve zor* olmak üzere üç boyut bulunmaktadır. Uzmanlardan beklenen ise, her test maddesini bahsedilen düzeylerden oluşan 3x4'lük bir tabloda uygunluk ve zorluk derecelerine göre ilgili hücreye yerleştirmeleri ve her hücre için bir yüzde belirlemeleridir (bu yüzde



minimum yeterlilik düzeyindeki öğrencinin her hücre için doğru cevaplayabileceği madde yüzdesidir). Daha sonra her hücredeki test maddesi toplamı, uzmanlar tarafından belirlenen uygun yüzdelerle çarpılır ve tüm hücrelerin toplam puanının toplam test maddesine bölünmesi ile kesme puanı belirlenmiş olur.

**Nedelsky Yöntemi:** Nedelsky (1954) yönteminde uzmanlar bireysel olarak, minimum yeterlik düzeyindeki öğrencinin (Nedelsky bu öğrencileri D-F yani geçme kalma sınırındaki öğrenciler olarak adlandırmaktadır.) çoktan seçmeli test maddesindeki seçeneklerden eleyebileceği yanlış seçenek sayısını tahmin etmeye çalışırlar.

Puanlamada ise; her madde için elenmeksizin kalan seçenek sayısının toplamı o maddenin puanı olarak alınır. Her öğrencinin Minimum Geçme Puanı (MGP) (Minimum Passing Level) ise o maddede elenmeksizin kalan seçenek sayısıdır ki; bu minimum yeterlik düzeyindeki bir öğrencinin “şans başarısı” puanıdır.

Nedelsky yöntemine göre kesme puanı hesaplaması ise aşağıdaki formüle göre yapılmaktadır.

$$\tau_{cj} = \sum_{i=1}^n (q_i - k_{ij})^{-1}$$

$\tau_{j}$ : j uzmanın belirlediği kesme puanı

$q_i$ : i maddesindeki seçenek sayısı

$k_{ij}$  : j uzmanının minimum yeterlilikteki bir öğrencinin yanlış olarak eleyebileceğini düşündüğü seçenek sayısı

n: testteki soru sayısı

Formül uygulandığında dört seçenekli bir madde için aşağıdaki sonuçlara ulaşılmaktadır.

Tablo 1. *Elenmeden Kalan Seçenek Sayısına Göre Doğru Cevap Olasılığı*

Kalan Seçenek Sayısı	Minimum Geçme Düzeyi
4	.25
3	.33
2	.50
1	1.00

Nedelsky yöntemi minimum yeterlik düzeyindeki öğrencinin istenen testi çözerken eleme sürecini izlediğini savunur. Öğrencinin yanlış seçenekleri artık aralarında seçim yapılamayacak bir sayıya kadar düşüreceği beklenir.

**Angoff Yöntemi:** Angoff (1971) yöntemi, temel formu ile çok basit bir uygulamadır. Bu yöntemde her uzmandan minimum yeterlilik düzeyindeki bir öğrenci grubunun içinde, öğrencilerden kaç tanesinin (öğrenci sayısı yüzde türünden ifade edilmektedir) testte yer alan maddeleri doğru cevaplayabileceklerini tahmin etmeleri istenir. Bu yüzde, testteki her bir madde için ayrı ayrı olmak üzere, tek bir öğrenci için değil testi alan minimum yeterlilik düzeyindeki öğrencilerin tümü dikkate alınarak verilmelidir. Kesme puanı oluşturulurken her uzmanın bireysel olarak her bir madde için verdiği olasılıklar toplanır ve o uzman için Minimum Geçme Puanı (MGP) belirlenir. MGP'lerin her biri, her uzmanın bireysel kesme puanını ifade eder. Bu kesme puanlarının ortalaması testin nihai kesme puanını verir.

Kesme puanı için ayrıca bir standart hata da hesaplanabilir. Düşük bir standart hata uzmanlar arasındaki uyumu göstereceğinden istenen bir durumdur. Angoff yönteminde uzmanlar sadece doğru cevap tahminleri yaparken Nedelsky yönteminde her seçenek dikkate alınarak bir tahmin yapılmaktadır. Yani başka bir deyişle Angoff yönteminde uzmanlar sadece madde kökü ve doğru cevaba odaklanırken, Nedelsky yönteminde madde kökü ile birlikte tüm seçeneklere odaklanmaktadır.

Angoff Yönteminde uzmanlar çeldiriciler tarafından yanıtılmadığı sürece madde güçlüklerinin daha düşük tahmin edildiği gözlenmiştir (Chang, Lei ;1996). Nedelsky Yönteminde ise uzmanların çeldiricileri değerlendirirken doğru cevabın göz ardı edilmesi olası bir durumdur. Böyle bir durumda uzman doğru cevabın içerdiği ipuçlarını dikkate alamadığından madde güçlüğüne olduğundan büyük tahmin etmektedir. Yöntemler arasında odaklanılan alan farklılıkları sonucunda Nedelsky Yöntemi ile belirlenen kesme puanlarının Angoff Yöntemi ile belirlenen kesme puanlarından daha düşük olduğu saptanmıştır (Chang, Lei ;1996).

Angoff yöntemindeki süreç matematikselleştirildiği zaman aşağıdaki formül elde edilmektedir:

$$\tau_{cj} = \sum_{i=1}^n (p_{ij})$$

Burada  $\tau_{cj}$ : uzmanın belirlediği kesme puanı

$p_{ij}$  : minimum yeterlik düzeyindeki bir öğrencinin j uzmanı tarafından belirlenen i maddesini doğru cevaplayabilme yüzdesi

n : testteki madde sayısı

### **Angoff Çeşitlemeleri**

**1-0 (Yes/No) Yöntemi:** Angoff yönteminin çeşitlemelerinden biri Impara ve Plake (1998) tarafından geliştirilen 1-0 (*Yes/No*) yöntemidir. Bu yöntemde uzmanlardan, öğrenciye uygulanacak olan testteki her bir test maddesini inceleyip, minimum yeterlilik düzeyindeki bir öğrencinin bu maddeleri doğru cevaplayıp cevaplayamayacağını tahmin etmeleri istenir. Öğrencinin maddeyi doğru cevaplayacağı düşünülüyorsa 1 puan, yanlış cevaplayacağı düşünülüyorsa 0 puan verilir. Puan toplamları alınarak o uzman için bir Minimum Geçme Puanı (MGP) belirlenir; tüm uzmanların MGP ortalaması ile de testin kesme puanı belirlenmiş olur.

Yöntemin ilgi çeken yönlerinden biri basitliğidir. Angoff'un tipik çeşitlemelerinde uzmanlardan hipotetik bir grubu ele alıp bu grubun ne kadarının soruya doğru cevap verebileceğini tahmin etmeleri istenir. Aslında bu çok önemli bir o kadar da zor bir işlemdir. Impara ve Plake (1998) 1-0 (*Yes-No*) yöntemi ile işlemin zorluğunu biraz da olsa hafifletmişlerdir. Tahminleri minimum yeterlilik düzeyi sınırındaki tek bir öğrenci üzerinden yapmak büyük bir grubu düşünerek yüzde tahmini yapmaktan çok daha kolaydır. Geleneksel Angoff yöntemi ile yapılan yüzde tahminlerinde uzmanlar verdikleri oranlardan pek fazla emin olamazlarken 1-0 (*Yes-No*) yönteminde iki boyuta düşürülen tahminlerin geçerliliğinin de daha büyük olduğu gözlenmiştir.

Yöntemin Impara ve Plake tarafından sözü edilen kavrama ve uygulama kolaylığı Chinn ve Hertz (2002) tarafından da vurgulanmıştır. Chinn ve Hertz de çalışmalarında uzmanların 1-0 (*Yes/No*) kararlarını geniş aralıktaki yüzde tahminlerine göre çok daha kolay verdiklerini saptamışlardır. Ancak, buldukları kesme puanlarının süreç birkaç kez tekrarlandığında klasik Angoff yöntemindeki yüzde hesaplaması ile bulunanlara göre daha az kararlı olduğunu görmüşlerdir (Chinn; Hertz; 2002).

Bu araştırmada ise uygulama süreçlerinin diğer yöntemlere (özellikle öğrenci merkezli yöntemler) göre kısa olması ve kesme puanlarının hesaplanmasının daha pratik olması nedeni ile test merkezli yöntemlerden Nedelsky, Angoff ve Angoff çeşitlemelerinden biri olan 1-0 (*Yes/No*) yöntemleri ele alınıp karşılaştırılmıştır.

## YÖNTEM

### Verilerin Elde Edildiği Grup

Araştırmada, standart belirleme yöntemleri ile bu yöntemlerin geçme puanları üzerindeki etkileri belirlenmeye çalışılmıştır. Çalışma grubu Ankara Mimar Kemal İlköğretim Okulu ile Ankara Namık Kemal İlköğretim Okulu'ndan 125 yedinci sınıf öğrencisinden oluşmaktadır. Ayrıca her yöntem için 4 farklı uzman olmak üzere toplam 24 Türkçe ve Matematik uzmanının görüşlerine başvurulmuştur.

### Verilerin Toplanması ve Veri Toplama Araçları

Araştırmanın kapsamında kullanılan öğrenci verileri, öğrencilerin biri sözel diğeri sayısal bir ders olmak üzere iki farklı dersten almış oldukları test puanlarıdır. Bu puanları elde etmek amacı ile 25'er maddeden oluşan Türkçe ve Matematik başarı testleri kullanılmıştır. Kullanılan bu testler araştırmacı tarafından geliştirilmiştir. Araştırmanın ikinci aşamasında gerekli olan veriler ise her bir yöntem için farklı hesaplanan ve farklı uzman görüşlerine dayanan puanlardır.

Oluşturulan nihai testler öğrencilere uygulandıktan sonra test maddelerinin her birini farklı standart belirleme yöntemine göre değerlendirmek ve her iki test için de farklı üç standart belirleme yöntemine göre kesme puanı belirlemek için uzman görüşüne başvurulmuştur. Buna göre her yöntem için farklı dört alan uzmanı test maddelerini değerlendirmiş ve kesme puanı belirlemiştir.

İlk yöntem olan Angoff Yöntemi için uzmanlardan her bir test maddesini minimum yeterlik düzeyindeki 7. sınıf öğrencilerini göz önüne alarak, bu öğrencilerden yüzde kaçının bu test maddesini doğru cevaplayabileceği konusunda görüş alınmıştır. Buna göre dört farklı Matematik ve Türkçe alan uzmanından bu yöntemle göre soruları değerlendirmeleri istenmiştir. İkinci yöntem olan Nedelsky Yöntemi için de dört alan uzmanından her bir test maddesindeki seçeneklerden minimum yeterlik düzeyindeki öğrencilerin eleyebileceği yanlış seçenek sayısını tahmin etmeleri istenmiştir. Üçüncü yöntem olan 1-0 (Yes/No) Yöntemi için ise dört uzmanından öğrencilere uygulanan testteki her bir test maddesini inceleyip, minimum yeterlik düzeyindeki bir 7. sınıf öğrencisinin bu maddeleri doğru cevaplayıp cevaplayamayacağını tahmin etmeleri, öğrencinin maddeyi doğru cevaplayacağı düşünülüyorsa 1 puan, yanlış cevaplayacağı düşünülüyorsa 0 puan vermeleri istenmiştir. Tüm bu değerlendirmeler için her bir yöntem için random olarak belirlenen farklı

uzmanlar olmak üzere, 12 Matematik dersi uzmanından ve 12 Türkçe dersi uzmanından yardım alınmıştır.

### **Verilerin Analizi**

İlk olarak üç ayrı standart belirleme yöntemine göre geçme puanlarının altında ve üstünde olan öğrencilerin geçme yüzdeleri karşılaştırılmıştır. Farklı standart belirleme yöntemlerine göre geçme puanı üstünde puan alan öğrenci yüzdeleri arasındaki ilişkileri belirlemek amacı ile bağımlı iki oran arasındaki farkın testi yapılmış ve anlamlılığı z testi ile sınanmıştır. Test işleminde önem düzeyi 0,05 alınmıştır.

İkinci olarak derslerin sayısal ve sözel içerikli olması ile geçme puanlarının belirlenmesi arasındaki ilişkileri ortaya koymak amacı iyine bağımlı iki oran arasındaki farkın testi yapılmış ve anlamlılığı z testi ile sınanmıştır. Bu hipotez testinde de önem düzeyi 0,05 alınmıştır.

Son olarak da farklı standart belirleme yöntemleri arasında uyumu belirlemek amacı ile Cohen's Kappa testi yapılmıştır.

### **BULGULAR**

Bu alt probleme cevap aramak için 7. sınıf Matematik ve Türkçe testlerine ait her soruya 4 ayrı uzmanın vermiş oldukları minimum yeterli düzeyleri alınmıştır. Uzmanların vermiş oldukları minimum yeterli düzeyleri Angoff Yöntemi için Tablo 6'da, Yes/No Yöntemi için Tablo 9'de, Nedelsky Yöntemi için ise Tablo 10'da verilmiştir. Bu yöntemlere göre hesaplanan kesme puanına göre elde edilen başarı yüzdeleri z testi ile karşılaştırılmıştır.

Uzmanlar her test maddesini minimum yeterlilikteki 7. sınıf öğrencisinden yüzde kaçının yapabileceğine dair görüşlerini bildirmişlerdir. Uzmanların, 7. sınıf bazındaki tahmini başarı yüzdelerinin ortalama değeri alınarak her bir uzman için bireysel kesme puanı bulunmuştur. Matematik testi için I.uzman KP = 16,30 II.uzman KP =15,30 III.uzman KP =12,95 IV.uzman KP=12,79 ve bunların ortalaması olan 14,34 nihai kesme puanı olarak alınmıştır. Aynı işlem Türkçe testi için de yapılmış ve uzmanların bireysel kesme puanları I.uzman KP = 16,95 II.uzman KP=20,60 III.uzman KP=18,75 IV.uzman KP=17,80 olarak saptanmış ve nihai kesme puanı 18,53 olarak alınmıştır.

1-0 (Yes/No) Yöntemi için matematik ve Türkçe testleri daha öncekilerden farklı 4'er uzmana verilmiş ve minimum yeterlilikteki 7. sınıf öğrencilerinin her bir soruyu doğru cevaplayıp cevaplayamayacağına ilişkin görüşleri alınmıştır

Matematik testi için I.uzman KP=20 II.uzman KP=15 III.uzman KP=20 IV.uzman KP=13 ve bunların ortalaması olan 17 nihai kesme puanı olarak alınmıştır. Aynı işlem Türkçe testi için de yapılmış ve uzmanların bireysel kesme puanları I.uzman KP=19 II.uzman KP=15 III.uzman KP=15 IV.uzman KP=21 olarak saptanmış ve nihai kesme puanı 18 olarak alınmıştır.

Nedelsky yöntemine ait uzman görüşleri Tablo 12’de görülmektedir. Uzmanlar her test maddesinin seçeneklerini incelemiş, minimum yeterlilikteki 7. sınıf öğrencilerinin bu seçenekler içinden hangilerini yanlış cevap olarak eleyebilecekleri konusunda görüşlerini bildirmişlerdir. Elde edilen bu tahmini rakamlarla her bir uzman için bireysel kesme puanı bulunmuştur. Matematik testi için I.uzman KP= 10,46 II.uzman KP=10,28 III.uzman KP=10,14 IV.uzman KP=11,13 ve bunların ortalaması olan 10,52 nihai kesme puanı olarak alınmıştır. Aynı işlem Türkçe testi için de yapılmış ve uzmanların bireysel kesme puanları I.uzman KP=12,82 II.uzman KP=14,33 III.uzman KP=8,76 IV.uzman KP =14,64 olarak saptanmış ve nihai kesme puanı 12,637 olarak alınmıştır.

Her üç yönteme göre kesme puanları belirlendikten sonra öğrenci test puanlarına bakılarak öğrencilerin her yönteme göre geçti/kaldı durumları bulunmuştur.

Tablo 2. Öğrencilerin Yöntemlere Göre Testlerden Geçer Not Alma Sayı ve Oranları

	Türkçe			Matematik		
	Angoff	Nedelsky	Yes/No	Angoff	Nedelsky	Yes/No
<b>Kesme puanları</b>	18,53	12,63	17,5	14,34	10,5	17
<b>Geçen ögr. Sayısı</b>	71	102	76	49	72	28
<b>Geçen ögr. Oranı</b>	0,55	0,79	0,59	0,38	0,56	0,22

Tablo 2’de de görüldüğü gibi Türkçe testinde Angoff yöntemine göre kesme puanı üzerinde puan alan öğrenci oranı 0,55 iken Nedelsky yönteminde bu oranın 0,79’a yükseldiği ve 1-0 yönteminde 0,59 olarak saptandığı gözlenmiştir. Matematik testinde ise Angoff yöntemine göre kesme puanı üzerinde puan alan öğrenci oranı 0,38 iken bu oran Nedelsky yönteminde 0,56’ya çıkmış, 1-0 yönteminde ise 0,22’ye kadar düşmüştür. Derslere ve yöntemlere göre başarı yüzdelerinin değiştiği görülmektedir.

Araştırma verilerinin analizi ile elde edilen sonuçlar, alt problemlerin çözümünde aşağıdaki gibi kullanılmıştır. Tablo 3 incelendiğinde, Türkçe dersi için, yöntemler ikiye ayrılarak karşılaştırıldığında, Angoff ve Yes/No yöntemleri arasındaki yüzdeler arasında 0,05 düzeyinde, diğer yöntemler arasında 0,01 düzeyinde fark olduğu görülmüştür. Matematik dersine bakıldığında yöntemlerin ikiye ayrılarak karşılaştırmalarının tümünde 0,01

düzeyinde farklılaşma görülmektedir. Bu durum farklı yöntemlere göre kesme puanı üzerinde puan alan öğrenci yüzdeleri arasında önemli farklılıklar olduğunu göstermektedir.

Derslerin sayısal ve sözel içerikli olması ile geçme puanlarının belirlenmesi arasındaki ilişkileri belirlemek amacı ile bağımlı iki oran arasındaki farkın testi yapılmıştır. Bu teste ilişkin sonuçlar Tablo 4’de verilmiştir.

Tablo 3. *Türkçe ve Matematik Derslerinde Üç Ayrı Yöntem ile Belirlenen Kesme Puanlarına Göre Başarı Yüzdelerinin Karşılaştırılması*

Dersler	Gruplar	$x_i$	p	z
Türkçe	Angoff	71	0,55	-5,56**
	Nedelsky	102	0,79	
	Angoff	71	0,55	-2,33*
	Yes/No	76	0,59	
	Nedelsky	102	0,79	5,06**
	Yes/No	76	0,59	
Matematik	Angoff	49	0,38	-4,73**
	Nedelsky	72	0,56	
	Angoff	49	0,38	4,54**
	Yes/No	28	0,22	
	Nedelsky	72	0,56	6,51**
	Yes/No	28	0,22	

\*  $p < 0,05$

\*\*  $p < 0,01$

Tablo 4’de de görüldüğü gibi tüm z değerleri anlamlı bulunmuştur. Derslerin sayısal ve sözel içerikli olması ile geçme puanı üzerinde puan alan öğrenci yüzdeleri ikili olarak karşılaştırıldığında bu yüzdeler arasında 0,01 düzeyinde anlamlı bir ilişki olduğu görülmektedir.

Bu durum, derslerin sayısal ya da sözel içerikli olması ile kesme puanı üzerinde puan alan öğrenci yüzdeleri arasında önemli farklılıklar olduğunu gösterir. Verilere bakıldığında Yes/No yöntemi ile belirlenen kesme puanlarının derslerin sözel ya da sayısal içerikli olmasına göre oldukça farklılaştığı görülmüştür. Diğer yöntemlerde sayısal ve sözel derslerin kesme puanları farklılaşma gösterirken bu farklılaşma Yes/No yönteminde görülen farklılaşma kadar büyük ölçüde olmadığı gözlenmiştir.

Tablo 4. Türkçe Ve Matematik Derslerinde Üç Ayrı Yöntem İle Belirlenen Kesme Puanlarının Karşılaştırılması

Yöntemler	$x_i$	p	z
Angoff (T)	71	0,55	
Angoff (M)	49	0,38	Z = 3,47**
Nedelsky (T)	102	0,79	
Nedelsky (M)	72	0,56	Z = 4,70**
Yes/No (T)	76	0,59	
Yes/No (M)	28	0,22	Z = 6,27**

\*\* p < 0,01

Farklı standart belirleme yöntemleri arasında uyumu belirlemek amacı ile Cohen's Kappa testi yapılmıştır. Türkçe dersi için Angoff ve Yes/No yöntemleri arasında pozitif ve yüksek düzeyde ( $Kappa > 0,81$ ), Angoff ve Nedelsky yöntemleri ile Nedelsky ve Yes/No yöntemleri arasında ise pozitif ve orta düzeyde ( $Kappa < 0,60$ ) uyum görülmüş iken; Matematik dersi için Angoff ve Yes/No yöntemleri arasında pozitif ve orta düzeyde ( $Kappa > 0,41$ ), Angoff ve Nedelsky yöntemleri ile Nedelsky ve Yes/No yöntemleri arasında pozitif ve düşük düzeyde ( $Kappa < 0,40$ ) anlamlı bir uyum görülmüştür.

Nedelsky yöntemindeki düşük düzeydeki uyumun nedeni Nedelsky Türkçe alan uzmanları arasındaki tutarsızlık ve Nedelsky Matematik alan uzmanları arasındaki düşük düzeydeki tutarlılık olabilir. Ayrıca diğer bir neden olarak da; Nedelsky Yöntemi puanlamasının, diğer yöntemlere göre daha karmaşık ve güç olması gösterilebilir. Angoff ve Yes/No yöntemleri arasındaki uyumun yüksek düzeyde olması, her iki yöntemin de kolay puanlanabilir olmasının sonucu olabilir.

## SONUÇLAR

Farklı standart belirleme yöntemlerine göre geçme puanı üstünde puan alan öğrenci yüzdeleri farklılaşma göstermektedir. Türkçe dersi için Angoff ve Nedelsky, Nedelsky ve Yes/No yöntemleri arasında ve Angoff ve Yes/No yöntemleri arasında anlamlı bir fark görülmüş iken; Matematik dersi için Angoff ve Nedelsky yöntemleri ile Angoff ve Yes/No yöntemleri arasında ve Angoff ve Nedelsky yöntemleri arasında anlamlı bir fark görülmüştür.

Derslerin sayısal ve sözel içerikli olması ile geçme puanı üzerinde puan alan öğrenci yüzdeleri ikili olarak karşılaştırıldığında, yüzdeler arasında farklılaşma görülmüştür. En yüksek düzeydeki farklılaşma, Türkçe ve



Matematik dersleri için Yes/No yöntemiyle belirlenmiş kesme puanları arasında görülmüştür.

Kesme puanı belirleme yöntemleri arasındaki uyum anlamlı bulunmuştur. Türkçe dersi için Angoff ve Yes/No yöntemleri arasında pozitif ve yüksek düzeyde, Angoff ve Nedelsky yöntemleri ile Nedelsky ve Yes/No yöntemleri arasında ise pozitif ve orta düzeyde uyum görülmüş iken; Matematik dersi için Angoff ve Yes/No yöntemleri arasında pozitif ve orta düzeyde, Angoff ve Nedelsky yöntemleri ile Nedelsky ve Yes/No yöntemleri arasında pozitif ve düşük düzeyde anlamlı bir uyum görülmüştür.

#### **KAYNAKLAR**

- Angoff, W.H. (1971), Scales, Norms and Equivalent Scores, In R. L. Thorndike (Ed.) Educational Measurement, Washington, DC: American Council on Education
- Berk, R.A.; (1982), Criterion Referenced Measurement: The State of The Arts
- Chang, Lei. A.; (1996), Comparison Between The Nedelsky And Angoff Standart Setting Methods, Paper Presented At The Annual Meeting Of The National Council On Measurement In Education, New York
- Cizek, G.; Bunch M.; Koons H.; (2004), Setting Performance Standard: Contemporary Methods, Educational Measurement, Issues And Practice, Proquest Education Journals
- Chinn, R. N., & Hertz, N. R. (2002). Alternative approaches to standard setting for licensing and certification examinations. Applied Measurement in Education, Volume 15
- Ebel, R. L. (1972). Essential of Educational Measurement. Englewood Cliffs, New Jersey:Prentice- Hall, Inc.
- Eignor, D., Hambleton, R.K., (1979), Effects of Test Length and Advancement Score on Several Criterion-Referenced Test Reliability And Validity Indices. Laboratory of Psychometric And Evaluation Research Report No:86, Paper Presented at The Annual Meeting of The National Council on Measurement in Education, San Francisco.
- Horn,C.; Ramos,M; Blumer,I; Madaus,G.; (2000), Cut Scores: Results May Vary, Nbetpp Monograpps, Volume 1, Number 1
- Impara, J.C., Plake, B.S., (1998), Teachers Ability to Estimate Item Difficulty: A Test Of Assumptions in The Angoff Standart Setting Method, Journal of Educational Measurement, Volume35
- Jaeger, R. M. (1989) Certification of student competence. In R.L. Linn (Ed.) Educational Measurement (3rd ed.), Newyork
- Nedelsky, L. (1954) Absolute grading standards for objective tests. Educational and Psychological Measurement, Volume 14
- Turgut, M. Fuat. (1992), Eğitimde Ölçme ve Değerlendirme Metodları, Ankara, Saydam Matbaacılık.

