



Determination of cutting score in performance tasks

Performans görevlerinde kesme puanının belirlenmesi

Sevda Çetin¹
Ayşegül Altun²

Abstract

Performance tasks are important in terms of meaningful learning since they enable students to use the information they have learned in their daily lives. Performance task is also the most commonly used method that our teachers use apart from written exams. While evaluating performance tasks, it is necessary to determine whether the tasks that students have prepared are adequate or not. In order to determine the measure of this decision, standard-setting studies have been carried out. In this study, passing score of a prepared performance task was determined through one of the standard-setting methods, Angoff method. 26 mathematics teachers who are working in various high schools in different cities participated in the study as experts. Data collection tool is a five-question performance project that is called 'As the Crow Flies' and includes coordinate plane and Pythagoras theorem. In the process of data collection, each expert was shown the questions in the performance task, and asked about the possibility of giving correct answers for the students who were on the border of failing. The consistency of points among experts were determined through Kendall's W coefficient of concordance ($\chi^2 = 137.872$, $sd=9$, $p<0.00$). According to the points acquired from experts, it was concluded that although there was not a high level of consistency among experts, there

Özet

Performans görevleri, öğrencilerin kavradıkları bilgileri günlük hayatla bağdaştırmaları ve anlamlı öğrenmeler açısından önemlidir. Ayrıca öğretmenlerin yazılı sınavlar dışında en çok kullandıkları yöntemlerden biridir. Performans görevlerini değerlendirirken öğrencilerin ortaya koydukları performans görevlerinde yeterli olup olmadıkları hakkında karar verilmesi gerekmektedir. Bu kararın hangi ölçüte göre verileceğini belirlemek için standart belirleme çalışmaları yapılmaktadır. Bu çalışmada, hazırlanan bir performans görevine ilişkin geçme puanı, standart belirleme yöntemlerinden biri olan Angoff yöntemi ile belirlenmiştir. Çalışmaya farklı illerdeki liselerde çalışan 26 matematik öğretmeni uzman olarak katılmıştır. Veri toplama aracı 'Kuş Uçuşu' olarak adlandırılan, koordinat düzlemi ve Pisagor teoremini içeren beş soruluk bir performans görevi kullanılmıştır. Veri toplama sürecinde her bir uzmana performans görevindeki sorular gösterilerek, bu soruları geçme-kalma sınırındaki bir öğrencinin doğru cevaplama olasılığı sorulmuştur. Puanlamada uzmanlar arası uyumun olup olmadığı Kendall'in W uyum katsayısı ile belirlenmiştir ($\chi^2 = 137.872$, $sd=9$, $p<0.00$). Elde edilen sonuca göre, uzmanlar arasında orta düzeyde anlamlı bir uyum olduğu sonucuna varılmıştır. Yapılan hesaplamalar ile nihai kesme puanı 45.10 olarak bulunmuştur.

¹ Dr. Öğr. Üyesi, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, tsevda@hacettepe.edu.tr

² Dr. Öğr. Üyesi, Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, aysegul.altun@omu.edu.tr

was a moderate level of consistency. As a result of the calculations, final cutoff score was found to be 45.10.

Anahtar Kelimeler: Performans görevi; standart belirleme; kesme puanı; Angoff yöntemi

Keywords: Performance task; standard setting; cutoff score; Angoff method.

[\(Extended English summary is at the end of this document\)](#)

1. Giriş

Standart belirleme, belirli bir performans düzeyine ulaşmak için gerekli olan bilgi ve beceri düzeyinin tanımlanması ve o performans düzeyini karşılayan bir puanın belirlenmesi sürecidir. Performans kategorilerini iki ya da daha fazla alana ayırarak bireyleri sınıflandıran bu puanlar, kesme puanı olarak adlandırılmaktadır. Kane (2001)'e göre performans standardı bireylerin performans düzeylerinin kavramsal tanımı iken; kesme puanları, kişileri performans düzeylerine göre ayıran performans standartlarının test puanları cetvelindeki rakamsal (işevuruk) bir tanımdır. Cizek (2001) bu ayrıma çok katılmamaktadır. Ona göre performans standardı; erişim düzeyi, standart, geçme puanı, kesme puanı ile eş anlamlı olarak kullanılabilir.

Geçme puanlarının belirlenmesinde birçok yaklaşım geliştirilmiş ve araştırılmıştır. Bu çalışmalarda görülmüştür ki;

- İster çoktan seçmeli olsun isterse performans dayalı sınavlar olsun, “en iyi” yöntem ya da “en doğru” geçme puanı bulunmamaktadır.
- Farklı yöntemlerle farklı sonuçlara ulaşılmaktadır.
- Yöntem seçimi, sınavın amacına uygun olarak yapılmalıdır.
- Geçme puanı belirleme çalışmaları, değerlendirmenin amacına ve çalışma alanına aşina uzmanlarla yapılmalıdır.
- Seçilen standart belirleme yöntemi, değerlendirme amaçlarıyla ilgili, kolay açıklanabilir ve uygulanabilir, o performans alanında çalışmış uzman yargılarına ve araştırmalarına dayanan bir yöntem olmalıdır (McKinley ve Norcini, 2014).

Standartlar bağıl (norm dayanaklı) ve mutlak (ölçüt dayanaklı) olarak gruplandırılabilir (Livingstone ve Zieky, 1982). Bireyler birbirlerine göre değerlendirildiğinde, örneğin; geçme puanı öğrenci yüzdesine göre belirlendiğinde standartlar “bağıl” olmaktadır. Bu tür standartlar genellikle seçilecek kişi sayısının az olduğu eğitim programlarında tercih edilmektedir. Bu standart türünde, testi alan bireylerin puanı ne kadar artarsa, standart da o kadar yükselecektir. Fakat mezuniyet, lisans, sertifika vb. gibi durumlarda bağıl standartlar kullanışlı değildir; çünkü testi alan grubun yetenek düzeyi ve değerlendirmenin içeriği zamanla değişebilir. Önceden belirlenmiş geçme puanları kullanıldığında ise (sınıf içi değerlendirmeler) öğrencilerin bilgi ve beceri düzeyleri ile testin gücü dikkate alınmamaktadır. Bireylerin yeteneği ve testin gücü geçme puanının uygunluğunu etkileyecek faktörlerdir.

Bağıl standartların aksine “mutlak standartlar”, gruba (testi alanların performansına) bağlı değildir ve ölçüt temelli standartlar olarak da bilinirler. Ölçüt temelli standartlarda, testi alan her bir birey için testi alan diğer bireylerin puanları önemli değildir, çünkü onların puanları standardı etkilememiştir. Mutlak standart belirleme yöntemleri test merkezli ya da öğrenci merkezli olarak ayrılmaktadır. Test merkezli modellerde uzmanlar testin maddelerine odaklanırlar, buna karşın öğrenci merkezli modellerde uzmanlar öğrenci performansına odaklanmaktadır (Livingstone ve Zieky, 1982).

Ölçme araçlarından elde edilen geçme puanlarına dayalı olarak kişiler hakkında seçme, yerleştirme, tanı, rehberlik gibi çeşitli kurumsal kararlar verilmektedir. Testler, diğer yöntemlere göre daha kolay uygulanabilmesi, objektif olması, geçerli ve güvenilir gözlemler yapmaya olanak

vermesi sebebiyle tercih edilmektedir (Cronbach,1990; Pawar, 2004). Fakat ülkemizde ilköğretim programlarının yapılandırıcı yaklaşıma dayanılarak yenilenmesiyle birlikte süreci ölçmeye ve değerlendirmeye yönelik yaklaşımları da hayatımıza girmiştir. Bu yaklaşımlar içerisinde öğretmenlerin en çok kullandığı yöntemlerinden biri performans değerlendirmedir (Acar ve Anıl, 2009; Yiğit ve Kırımlı, 2014; Altun ve Gelbal, 2014). Performans görevi ise ilköğretimde 2014 yılına kadar zorunlu olarak uygulanırsa da ortaöğretimde uygulanmamıştır. Ortaöğretimde performans görevinin zorunlu olarak uygulanması ise 2013-2014 eğitim-öğretim yılı ile başlamıştır.

Performans görevinin kullanımıyla ilgili iki önemli görüş bulunmaktadır. Bunlardan biri performans görevinin daha iyi bir ölçme sağladığı ikincisi ise öğrenme ve öğretmeyi geliştirdiğidir (Shephard, Flexer, Heibert, Marion, Myfield ve Weston,1995). Performans görevi, öğrencilerin üst düzey zihinsel becerilerini kullanmasına olanak sağlayan; öğrencilerin problem çözme ve işbirliği yapma becerileri kazandıran ve kendilerini geliştirebilme fırsatı veren nitelikte olmalıdır (Kumandaş ve Kutlu, 2010). Bu özelliklere ek olarak performans görevinin ürün, gözlenebilir performans, üst düzey düşünme süreci, sosyal beceriler ve grup çalışması, disiplinler arası geçiş ve bilgi alış-verişini sağlanması gerekir (Berberoğlu, 2007).

Türkiye’de öğretmenlerin en çok kullandığı tamamlayıcı ölçme ve değerlendirme yöntemlerinden biri performans değerlendirmedir (Acar ve Anıl, 2009; Yiğit ve Kırımlı, 2014; Altun ve Gelbal, 2014). Bu nedenle performans göreviyle ilgili yapılan çok sayıda çalışma vardır. Yapılan bu çalışmalar incelendiğinde çalışmaların hepsinin ilköğretim düzeyinde olduğu görülmektedir. Yerli alan yazında öğrencilerin performans görevine karşı tutumu (Kumandaş ve Kutlu, 2010), kullandıkları kaynaklar, karşılaştıkları problemler (Akdağ ve Çoklar, 2009), performans göreviyle ilgili görüşleri (Çetin, Çakan, 2010); öğretmenlerin proje ve performans görevinin uygulamasında karşılaştıkları sorunlar (Arı, 2010; Acar ve Anıl, 2012), performans görevine ilişkin görüşleri (Ersoy ve Anagüm, 2009; Çiftçi, 2010; Bal, 2012; Bozkur, Aslanargun, Akın ve Kılıç, 2014), performans görevine yönelik velilerin tutumları (Tüysüz, Karakuyu ve Tatar, 2010) gibi performans görevini farklı açılardan ele alan çok sayıda çalışma mevcuttur. Ancak performans göreviyle ilgili standart belirleme çalışmasına yerli alanyazında rastlanmamıştır.

Yurt dışında yapılan performans görevlerinde standart belirleme çalışmalarının daha çok tıp eğitimi alanında olduğu görülmektedir. Bu çalışmalardan biri Kaufman ve arkadaşları tarafından yapılmıştır. Araştırmalarında objektif yapılandırılmış bir klinik testi için (OSCE-Objective Structured Clinical Examination) dört standart belirleme yöntemini karşılaştırmışlardır. Dalhousie Üniversitesi Tıp Fakültesi’ndeki son sınıfta okuyan 84 tıp öğrencisine 12 istasyondan oluşan bir sınav uygulamış, kesme puanı oluşturmak için verilere Angoff, sınır grup (borderline), göreceli (relative) ve bütünsel (holistic) olmak üzere dört standart belirleme yöntemi kullanılmıştır. Çalışma sonucunda kesme puanı belirleme işlemlerinin tutarsız sonuçlar verdiği görülmüştür. Angoff ve sınır grup yöntemleri benzer sonuçlar verirken, göreceli ve bütüncül yöntemler çok farklı sonuçlar vermiştir (Kaufman, Mann, Muijtens ve Vleuten, 2000). Başka bir çalışmada ise 119 üçüncü sınıf diş hekimliği öğrencisinin performansını değerlendirmek için 14 istasyonlu objektif yapılandırılmış bir diş hekimliği klinik sınav kullanılmıştır. Kesme puanını belirlemek üzere Angoff, değiştirilmiş (modified) Angoff ve sınır grup regresyon modeli olmak üzere üç standart belirleme yöntemi kullanılmış ve çalışma sonucunda sınır grup regresyon yönteminin her iki Angoff yöntemine göre sınavı geçme oranları ve güvenilirlik açısından bakıldığında daha kabul edilebilir sonuçlar verdiği görülmüştür (Schoonheim-Klein, Muijtens, Habets, Manogue, Van Der Vleuten ve Van der Velden, 2009). Diğer bir çalışmada ise Cusimano ve Rothman (2003) Hofstee yöntemini Angoff ve Ebel yöntemleri ile karşılaştırmak amacıyla 96 uzman tarafından yapılandırılmış objektif klinik sınav için bu üç yöntemle göre kararlar almış ve çalışma sonucunda Hofstee yönteminin Angoff ve Ebel yöntemlerine göre daha gerçekçi kesme puanı belirlediği sonucuna ulaşmışlardır.

Son yıllarda alanyazında otuzdan fazla standart belirleme yöntemi betimlenmesine karşın (Crocker ve Algina, 1986), test merkezli standart belirleme yöntemlerinden sıklıkla kullanılan

yöntemler, Angoff, Nedelsky, Ebel, Yes/No, Düzeltilmiş Angoff, İşaretleme (Bookmark) Yöntemi, Mapmark Yöntemi iken, en sık kullanılan öğrenci merkezli standart belirleme yöntemleri ise Sınır Grup Yöntemi ve Karşıt Gruplar Yöntemleridir. Bu çalışmada, hazırlanan bir performans görevine ilişkin geçme puanı, standart belirleme yöntemlerinden Angoff yöntemi ile belirlenmiştir. 1971 yılında William H. Angoff tarafından önerilen yöntemde, madde bir bütün olarak değerlendirilir; uzman her bir soru için geçti-kaldı sınırındaki bir öğrencinin cevaplayabilme olasılığını tahmin eder. Bu olasılık 0-100 arasında olacaktır. Sorular kolaylaştıkça rakam daha da büyüyecektir. Eğer soru çoktan seçmeliyse bu olasılık en az şans başarısıyla soruyu doğru cevaplama olasılığı olan “1/seçenek sayısı” oranı olacaktır. Aşağıdaki Tablo 1’de örnek bir durum için 10 maddelik bir testin bu yöntemle göre kesme puanı hesaplaması verilmiştir.

Tablo 1. Angoff Yöntemi ile kesme puanı hesaplama örneği

Sorular	Doğru cevaplama olasılığı
1	0.95
2	0.80
3	0.90
4	0.60
5	0.75
6	0.40
7	0.50
8	0.25
9	0.25
10	0.40
Toplam:	5.80

Yukarıda bir uzmana ait, her bir soru için geçti-kaldı sınırındaki bir öğrencinin soruları cevaplayabilme olasılıkları verilmiştir. Verilen toplam değer ise o uzmana ait bireysel kesme puanı değeridir. Her uzman için tek tek kesme puanları belirlendikten sonra tüm uzmanların kesme puanlarının ortalaması alınarak nihai kesme puanı hesaplanacaktır. Bu tür bir hesaplamanın dezavantajı uzmanlardan birinin çok yüksek ya da çok düşük puan vermesinin hesaplama sonucunu büyük ölçüde etkileyebilmesidir. Bu dezavantajına rağmen, benzer diğer yöntemlere göre sürecin uzmanlara açıklanması, verilerin toplanması ve analizi bakımından daha basit ve diğer yöntemlere göre daha sık kullanılan bir yöntemdir (Mills ve Melican, 1988). Bu nedenle çalışmada bu yöntemin kullanılması tercih edilmiştir.

Araştırmanın önemi

Performans görevleri öğrencilerin kavradıkları bilgileri günlük hayatla bağdaştırmaları ve anlamlı öğrenmeler açısından önemlidir. Ayrıca öğretmenlerin yazılı sınavlar dışında en çok kullandıkları yöntemlerden biridir (Altun ve Gelbal, 2014). Performans görevlerini değerlendirirken öğrencilerin ortaya koydukları performans görevinde yeterli olup olmadıkları hakkında karar verilmesi gerekmektedir. Bu kararın hangi ölçüte göre verileceğini belirlemek için standart belirleme çalışmaları yapılmaktadır. Standart belirleme çalışmaları, öğrencilerin ilgili konuda yeterli bir performansa sahip olup olmadıklarını belirleyebilmek için geçme puanını hesaplamak üzere yapılır. Alanyazında birçok standart belirleme çalışması bulunmaktadır ve bunların pek çoğu da çoktan seçmeli testler üzerinden geçme puanı hesaplama çalışmalarıdır (Chang,1996; Çetin ve Gelbal, 2010, 2013; Gündeğer ve Doğan, 2014; Hurtz ve Hertz,1999; Impara ve Plake,1997; Taşdelen, 2009; Taşdemir, 2013; Tülübaş, 2009).

Ancak performans değerlendirme ile ilgili standart belirleme çalışmalarına yurt içinde yapılan araştırmalarda karşılaşılmamış, yurt dışında yapılan sınırlı çalışmanın da daha çok tıp eğitimi alanında olduğu görülmüştür (Cusimano, Rothman ve Keystone, 1998; Margolis, De Champlain ve Klass,

1998; McKinley ve Norcini, 2014; Kaufman ve ark., 2000; Schoonheim-Klein ve ark., 2009). Performans değerlendirmeler yanıtların çok boyutlu olması (görevi başarıyla tamamlamak için birden çok beceri gerekmesi), karmaşık ve çok yönlü olması, puanlama rubrikleri arasında karşılıklı bağımlılıklar olması (bir bölümünü eksik olan görevi tamamlayamamış olmak) ve görev düzeyindeki puanların genellenebilirliğinin düşük olması (bir görev grubunda iyi performans göstermek, başka birinde yüksek performans anlamına gelmemesi) gibi özelliklere sahiptir. Bu özellikler, standart belirleme yöntemleri için özel sorunlar yaratır (Hambleton, Jaeger, Plake ve Mills, 2000). Örneğin, popüler standart belirleme yöntemlerinin birçoğu (Livingston ve Zieky, 1982), çok yönlü olarak puanlanan performans değerlendirmeler için geçerli değildir. Yapılması gereken ise, eski standart belirleme yöntemlerini uyarlamak ya da performans değerlendirmenin mevcut özelliklerini karşılamak için yeni yöntemler geliştirmektir (Hambleton ve ark., 2000). Bu nedenle, bu araştırmada verilerin toplanması ve analizi bakımından nispeten daha kolay olan ve daha sık kullanılan bir yöntem olan Angoff yöntemi, örnek bir performans görevine uygulanarak kesme puanı belirlenmiştir. Kullanılan performans görevinin niteliği ve bu göreve uyarlanan Angoff yönteminin kullanımı açısından bu araştırmanın alanyazına önemli katkılar sağlayacağı düşünülmektedir.

2. Yöntem

Çalışma grubu

Standart belirleme çalışmasına farklı illerdeki devlet okullarındaki meslek liselerinde çalışan, hizmet yılları beş yılın üzerinde olan 26 matematik öğretmeni uzman olarak katılmıştır. Çalışma grubu 11 erkek (%42.3) ve 15 kadın (%57.7) öğretmenden oluşmaktadır.

Veri toplama aracı

Bu çalışmada kullanılan görev NCTM (National Council of Teachers of Mathematics-Ulusal Matematik Öğretmenleri Konseyi) tarafından uzaklık formülünü öğretmek amacıyla tasarlanmıştır. Bu görev öğrencileri uzaklık formülüyle tanıştırmak için uygulanabileceği gibi konuyu tekrar etmek ve öğrencilerin bu formülü ezberlemeden, günlük hayatın içinden bir örnek üzerinde uygulayabilmelerine olanak sağlamaktadır. Bu görevin, değerlendirme amacıyla da kullanılacağı belirtilmiştir (NCTM, 2011). Çalışmada bu görev, performans görevi olarak kullanılmıştır. Kuş uçuşu (As the crow file) olarak adlandırılan bu görev öncelikle çalışmacılardan biri tarafından Türkçe'ye çevrilmiş ve uzman görüşü alınarak görevin uygunluğu sağlanmıştır. Amaç ve görevin genel tanımı NCTM (2011)'deki görev tanımından özetlenmiştir. Daha sonra uzmanlarla birlikte başarı düzeylerinin tanımlanması, soruların puanlanması ve her basamağa kaç puan verileceği belirlenmiştir. Türkçeye çevrilen görev ve yapılan bu düzeylerin tanımlanması ve puanlamaya ilgili olarak bir ölçme-değerlendirme ve dört matematik eğitimcisi olan beş öğretim üyesinden uzman görüşü alınmıştır. Uzmanlardan alınan geribildirimlerle gerekli düzeltmeler yapılarak veri toplama aracına son hali verilmiştir.

Veri toplama aracı amaç, görevin genel tanımı, yeterlik düzeyleri ve performans 'Kuş Uçuşu' olarak adlandırılan bölümlerden oluşmaktadır. Koordinat düzlemi ve Pisagor Teoremini içeren performans görevi beş sorudan oluşmaktadır. Bu beş sorunun aşamalı puanlanması ve her aşamaya kaç puan verileceği de ayrıca belirtilmiştir.

Uzaklık formülüyle ilgili hazırlanan performans görevi bu çalışmada veri toplama aracı olarak kullanılmıştır. Aşağıda bu aracın amacı, görevin genel tanımı ve öğrencilere verilen aktivite yaprağını da içeren bilgiler verilmiştir. Görevin genel tanımı ise matematiksel akıl yürütme, matematiksel düşünme bölümlerinden oluşmaktadır.

Görevin Amacı: Uzaklık formülü genellikle öğrencilere ezberlenecek bir kural olarak sunulur. Bu görev öğrencilerin bu formülü ezberlemeden günlük hayattan bir örnek üzerinden, formülü anlamlandırmalarına yardımcı olmak için tasarlanmıştır. Koordinat sistemini ve Pisagor Teoremini bilen öğrencilere konuyu anlatmak veya tekrar etmek için uygundur. Bu görev ile

öğrencilerin akıl yürütme, anlaşılır olmaya odaklanma ve matematiksel düşünmeye odaklanma becerileri de test edilecektir. Bu beceriler aşağıdaki gibi tanımlanmaktadır (NCTM, 2011).

1) Akıl Yürütme ve Anlaşılır Olmaya Odaklanma

Akıl yürütme alışkanlıkları: Lise matematik programında akıl yürütme ve anlaşılır olma; bir problemi analiz etme, ilişkileri arama; daha önceki öğrendiklerini, bilgilerini uygulama; saklı olan, açıkça verilmeyen yapıları arama, farklı gösterimler arasındaki bağlantıları arama ve gösterme, çözüm üzerinde derinlemesine düşünme, çözümü genelleme olarak tanımlanmıştır.

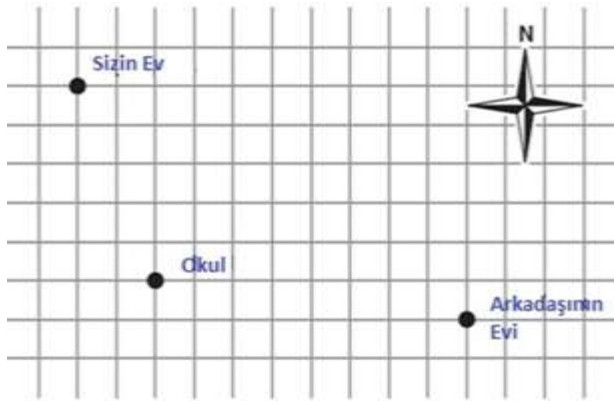
Süreç standartları: Okul matematiğinin prensip ve standartları, problem çözme, problem çözme ile yeni matematiksel bilgiyi inşa etme; bağlantı kurma, matematiksel fikirlerin karşılıklı olarak bağlantılı olduğunu anlama ve mantıklı bir bütün oluşturmak için birbirine dayandırma olarak; simgeleme ise fiziksel sosyal ve matematiksel olguları modelleme ve yorumlama için simge kullanma olarak; matematiksel pratik için standartlar ise soyut ve nicel düşünme, matematiksel modelleme, yapıyı açıklama ve anlamlandırma olarak tanımlanmıştır.

2) Matematiksel Düşünmeye Odaklanma

Lise matematik programında akıl yürütme ve anlaşılır olma; geometriyle akıl yürütme, geometriyi yapılandırma ve değerlendirme; matematik için ortak ve çekirdek standartlar ise cebirsel olarak basit geometrik teoremleri ispatlamak için koordinatları kullanmak ve çokgenlerin çevrelerini hesaplamak için örneğin uzaklık formülünü kullanarak üçgenlerin ve dikdörtgenlerin alanlarını hesaplamaktır.

Görevin genel tanıtımı (Göreve genel bakış): Caddeleri eşit aralıklı kareler şeklinde bölümlenmiş bir şehirdeki iki mesafe arasındaki uzunluğu hesaplayın. Daha sonra bir koordinat sistemi tanımlayın ve koordinat sistemini kullanarak uzaklığı nasıl hesaplayabileceğinizi düşünün.

Kuş Uçuşu Öğrenci Aktivite Yaprağı: Yaşadığınız şehrin, birbirine dik caddeler ve kare şeklinde bloklardan oluştuğunu varsayalım. Aşağıdaki harita; okulunuzu, okulun iki blok batısında beş blok kuzeyinde olan evinizi ve okulun sekiz blok kuzeyinde ve bir blok güneyinde olan en yakın arkadaşınızın evini göstermektedir.



1)A) Sizin evinizden, arkadaşınızın evine gitmeniz için gideceğiniz yolu çiziniz (5 puan)

B) Sizin evinizden, arkadaşınızın evine gitmeniz için gideceğiniz mesafeyi hesaplayınız. (5 puan)

2)A) Sizin evinizden, arkadaşınızın evine dümdüz bir helikopter ile gitmek istiyorsunuz; bunun için nasıl bir yol izlemeniz gerektiğini çiziniz. (5 puan)

B) Kuş uçuşu olan bu mesafeyi nasıl bulursunuz? (15 puan)

3)A) Okulun orijin olduğu bir koordinat düzlemi çiziniz. (10 puan)

B) Okulun orijin olduğu bir koordinat düzleminde evinizin koordinatları ne olur? (5 puan)

C) Okulun orijin olduğu bir koordinat düzleminde arkadaşınızın evinin koordinatlarını ne olur? (5 puan)

4) Koordinatları, sizin evinizle arkadaşınızın evi arasındaki kuş uçuşu mesafeyi hesaplamak için nasıl kullanırsınız? (20 puan)

5)A) Varsayalım ki amcanızın evi okulun iki blok batısında ve bir blok güneyindedir. Sizde amcanıza uğrayıp oradan arkadaşınıza gitmeye karar verdiğiniz. Sizin evinizden arkadaşınızın evine, amcanızın evine ve sonra kendi evinize olan bu gidiş-dönüş yolculuğunun mesafesini hesaplayıp, çözümünüzü gösteren bir resim çizin. (15 puan)

B) Sadece koordinatları kullanarak gidiş-dönüş mesafesini nasıl bulabileceğinizi gösteriniz. (15 puan)

Verilerin Analizi

Çalışmaya katılan uzmanlara, bu çalışmanın hazırlanan performans görevi için bir geçme puanı belirleme çalışması olduğu belirtilmiş; standart belirleme çalışmasının aşamaları detaylı bir şekilde açıklandıktan sonra her bir uzmana performans görevindeki sorular sunularak, bu soruları geçme-kalma sınırındaki bir öğrencinin doğru cevaplama olasılığı sorulmuştur. Verilen cevaplar dikkate alınarak kesme puanı belirleme sürecinde öncelikle puanlamada uzmanlar arası uyumun olup olmadığı Kendall'ın W uyum katsayısı ile belirlenmiş, daha sonra uzmanların tahminleri doğrultusunda elde edilen verilerinden Angoff yöntemine göre kesme puan hesaplaması yapılmıştır.

3. Bulgular

Uzmanların her bir görev sorusu için verdikleri olasılık tahminleri Tablo 2'de görülmektedir.

Tablo 2. Uzmanların her bir görev sorusu için tahminde buldukları cevaplama olasılıkları

Uzmanlar	Maddeler									
	S 1A	S 1B	S 2A	S 2B	S 3A	S 3B	S 3C	S 4	S 5A	S 5B
1	1.00	0.70	1.00	0.10	0.70	0.05	0.05	0.05	0.02	0.02
2	0.70	0.50	0.80	0.30	0.85	0.70	0.60	0.10	0.60	0.05
3	0.90	0.70	0.80	0.10	0.40	0.20	0.20	0.40	0.30	0.50
4	0.85	0.70	0.70	0.55	0.70	0.60	0.60	0.40	0.60	0.30
5	0.90	0.70	0.80	0.10	0.40	0.20	0.20	0.10	0.02	0.05
6	0.65	0.60	0.55	0.60	0.55	0.50	0.50	0.45	0.30	0.40
7	0.80	0.75	0.70	0.50	0.60	0.60	0.60	0.50	0.40	0.40
8	0.75	0.55	0.90	0.40	0.60	0.35	0.35	0.25	0.15	0.20
9	0.75	0.60	0.90	0.50	0.55	0.45	0.45	0.25	0.20	0.15
10	0.60	0.60	0.20	0.20	0.10	0.10	0.10	0.05	0.05	0.05
11	0.90	0.40	0.90	0.40	0.60	0.60	0.60	0.05	0.70	0.70
12	0.85	0.80	0.90	0.40	0.20	0.10	0.10	0.07	0.02	0.02
13	0.80	0.80	0.80	0.20	0.90	0.60	0.40	0.40	0.30	0.20
14	0.80	0.70	0.70	0.40	0.90	0.70	0.70	0.70	0.40	0.60
15	0.60	0.55	0.65	0.50	0.45	0.40	0.40	0.40	0.40	0.40
16	0.80	0.70	0.80	0.50	0.60	0.50	0.50	0.30	0.50	0.30
17	0.75	0.75	0.90	0.70	0.70	0.50	0.50	0.30	0.20	0.20
18	0.70	0.70	0.50	0.50	0.60	0.60	0.60	0.50	0.65	0.65
19	0.80	0.75	0.90	0.70	0.75	0.70	0.70	0.50	0.40	0.40
20	0.20	0.10	0.40	0.10	0.80	0.60	0.60	0.30	0.40	0.30
21	0.80	0.75	0.60	0.50	0.90	0.85	0.85	0.60	0.40	0.30
22	0.70	0.60	0.60	0.40	0.35	0.30	0.50	0.40	0.20	0.30
23	0.60	0.60	0.70	0.70	0.60	0.50	0.50	0.40	0.60	0.50
24	0.90	0.80	0.80	0.70	0.80	0.70	0.70	0.70	0.60	0.60
25	0.90	0.45	0.70	0.45	0.80	0.70	0.70	0.40	0.40	0.45
26	0.95	0.90	0.90	0.50	0.70	0.60	0.60	0.40	0.30	0.30

Tablo 2 incelendiğinde birinci görev sorusunun A seçeneği için olasılık tahminleri 0.20 ile 1.00 arasında değişirken, aynı görev sorusunun B seçeneği için olasılık tahminleri 0.10 ile 0.90 arasında değiştiği görülmektedir. İkinci görev sorusunun A seçeneği için olasılık tahminleri 0.20 ile 1 arasında değişirken aynı görev sorusunun B seçeneği için olasılık tahminleri 0.10 ile 0.70 arasında değişmektedir. Üçüncü görev sorusunun A seçeneği için olasılık tahminleri 0.10 ile 0.90 arasında değişirken aynı görev sorusunun B seçeneği için olasılık tahminleri 0.05 ile 0.85 arasında, C seçeneği için olasılık tahminleri ise yine 0.05 ile 0.85 arasında değiştiği görülmektedir. Beşinci görev sorusunun A ve B seçenekleri için olasılık tahminlerinin 0.02 ile 0.70 arasında değiştiği görülmektedir. Standartlara göre (AERP, APA & NCME, 1999) uygulanabiliyorsa uzmanlar arası uyum rapor edilmelidir. Tablodaki veriler dikkate alınarak kesme puanı belirleme sürecinde öncelikle puanlamada uzmanlar arası uyumun olup olmadığı Kendall'in W uyum katsayısı ile belirlenmiştir. İki'den fazla puanlayıcının belirli bir davranışı, performansı ya da soruyu puanlaması durumunda kullanılan, parametrik olmayan bir istatistiksel teknik olan Kendall'in uyum katsayısı (Howell, 2012) ile Angoff yöntemiyle elde edilen puanlar arasındaki tutarlık derecesi analiz edilmiş ve sonuç olarak 10 madde için 0.589 olarak bulunmuştur ($\chi^2 = 137.872$, $sd=9$, $p<0,00$). Elde edilen sonuca göre uzmanlar arasında orta düzede anlamlı bir uyum olduğu görülmektedir.

Uzmanların tahminleri doğrultusunda elde edilen Tablo 2 verilerinden her bir uzmana ait kesme puan hesaplaması yapılmıştır. Kesme puan hesaplamaları Angoff yöntemine göre yapılmıştır. Angoff yöntemine göre kesme puanı hesaplamasında her uzmanın olasılık tahminleri toplanarak o uzmana ait kesme puanı elde edilir. Bu çalışmada ele alınan performans görevinin sorularının puanlama ağırlıkları dikkate alınarak uzmanlara ait kesme puanları Tablo 3'deki gibi hesaplanmıştır.

Tablo incelendiğinde kesme puanlarının 14.50 ile 70 puanları arasında değiştiği ve tüm uzmanlara ait kesme puanları kullanarak hesaplanan nihai kesme puanın 45.10 bulunduğu görülmektedir.

Tablo 3: Angoff yöntemi ile uzmanlarca belirlenen kesme puanlar

Uzmanlar	Uzmana ait kesme puan	Uzmanlar	Uzmana ait kesme puan
uzman-1	24.10	uzman-14	62.00
uzman-2	41.25	uzman-15	45.00
uzman-3	39.50	uzman-16	48.00
uzman-4	54.00	uzman-17	46.50
uzman-5	22.55	uzman-18	58.50
uzman-6	48.00	uzman-19	59.25
uzman-7	52.75	uzman-20	35.50
uzman-8	36.75	uzman-21	58.25
uzman-9	39.00	uzman-22	38.50
uzman-10	14.50	uzman-23	55.50
uzman-11	51.00	uzman-24	70.00
uzman-12	23.75	uzman-25	52.75
uzman-13	44.50	uzman-26	51.25
Ortalama nihai kesme puan:			45.10

4. Sonuç ve Öneriler

Bu araştırmada hazırlanan bir performans görevine ilişkin geçme puanı standart belirleme yöntemlerinden biri Angoff yöntemi ile belirlenmiştir. Elde edilen bulgulara göre uzmanlar arasında orta düzeyde anlamlı bir uyum olduğu görülmektedir. Uzman görüşü alınan öğretmenlerin farklı meslek lisesinde görev yapmalarından dolayı kendi öğrenci gruplarını düşünerek olasılık tahminlerinde bulunmaları hakemler arasındaki uyumun çok yüksek çıkmamasının bir nedeni olarak gösterilebilir. Ayrıca uzmanlardan geçme – kalma sınırındaki hipotetik bir grubu ele alıp bu grubun ne kadarının soruya doğru cevap verebileceğini tahmin etmeleri istenmiştir, aslında bu çok önemli bir o kadar da zor bir işlemdir (Çetin, 2011). İşlemin bu zorluğu her uzmanda farklı düzeyde kendini göstermiş olabilir. Bu nedenlerle farklı çalışmalarda lise türü bir değişken olarak alınıp farklı lise türündeki çalışan öğretmenlerin görüşleri arasındaki tutarlılık düzeyleri incelenebilir ve çalışılan lise türüne göre kesme puanının değişip değişmeyeceği araştırılabilir.

Çalışma sonucunda örnek performans çalışmasının geçme puanı 45.10 olarak belirlenmiştir. Milli Eğitim Sınıf Geçme Yönetmeliğine (MEB, 2016) göre geçme puanı 50'dir. Aynı yönetmeliğe göre proje ve performans çalışmaları puanla değerlendirilir ve her dönemde tüm derslerden iki performans puanı verilmelidir. Bu çalışmada kullanılan performans ödevi için geçme puanı olarak bulunan 45.10 mevcut geçme puanından 4.90 puan daha düşüktür. Bu açıdan bakıldığında farklı performans görevleri kullanılarak yapılacak çalışmaların hesaplanan sonuçlara benzer sonuçlar vermesi durumunda, bu bulgular ilgili kurumlarla paylaşılarak geçme puanının benzer bilimsel çalışmaların sonuçları dikkate alınarak yeniden belirlenmesi yönünde bilgi paylaşımı yapılabilir.

Farklı standart belirleme yöntemleri genellikle farklı sonuçlar verir; aynı standart belirleme yöntemini kullanan farklı uygulamalar, karşılaştırılmaz sonuçlar doğuracaktır. Bu nedenle, bir standart belirleme yönteminden elde edilen sonuçlar, aynı yöntemi kullanarak ve standart belirleme yaklaşımlarının bir kombinasyonunu kullanarak eşzamanlı panellerle güçlendirilmelidir. Bazı durumlarda, bir yöntemin performans standardının altında veya üzerinde değer üretilmesine neden olup olmayacağı, büyük olasılıkla önceden tahmin edilebilir. Örneğin, Angoff yönteminde çoktan seçmeli sınavlarda maddeler arası sıfırdan farklı korelasyon nedeniyle, panelistlerin öngördüğünden daha yüksek performans standartları belirlenmesi beklenmektedir (Linn ve Shepard, 1997). Buna ek olarak, benzer şekilde, regresyon etkilerinden dolayı sınır grup (borderline) yönteminin, panelistlerin amaçladıklarına kıyasla, dağılımın ortalamasına daha yakın bir performans standardı ortaya koyması beklenmektedir. Standart belirleme çalışması birden fazla yöntem (örn. Angoff ve sınır grup) kullandığında, sonuçlardan makul bir performans standardının belirlenebileceği kesme puanları belirlenebilir (Hambleton ve ark., 2000). Standart belirlemede aynı anda birden çok yöntem kullanılması özellikle performans görevlerinde, üzerinde çalışılması gereken bir alandır.

Yapılan çalışmalarla görülmüştür ki ister çoktan seçmeli olsun isterse performansa dayalı sınavlar olsun, “en iyi” yöntem ya da “en doğru” geçme puanı bulunmamaktadır. Farklı yöntemlerle farklı sonuçlara ulaşılmaktadır. Kesme puanı belirleme çalışmalarında seçilen standart belirleme yöntemi, değerlendirme amaçlarıyla ilgili, kolay açıklanabilir ve uygulanabilir, o performans alanında çalışmış uzman yargılarına dayalı bir yöntem olmalıdır (Mckinley ve Norcini, 2014). Bu araştırmada test merkezli standart belirleme yöntemlerinden Angoff yöntemi kullanılmıştır. Benzer çalışmalarda diğer test merkezli standart belirleme yöntemleri ve öğrenci merkezli standart belirleme yöntemleri kullanılabilir.

Kaynakça

- Acar, M, Anıl, D.(2009). Sınıf Öğretmenlerinin Performans Değerlendirme Sürecindeki Değerlendirme Yöntemlerini Kullanabilme Yeterlikleri. *TÜBAV Bilim*, 2(3) ,354-363.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

- Akdağ, H. ve Çoklar, A. N. (2009). İlköğretim 6. ve 7. Sınıf öğrencilerinin Sosyal Bilgiler Dersi Proje ve Performans Görevlerini Hazırlarken Yararlandıkları Kaynaklar, İnternetin'in yeri ve Karşılaştıkları problemler. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2(2)
- Altun, A. ve Gelbal, S. (2014). Öğretmenlerinin Kullandıkları Ölçme ve Değerlendirme Yöntem veya Araçlarının İkili Karşılaştırma Yöntemiyle Belirlenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 1-11.
- Arı, A. (2010). Öğretmenlere Göre Performans Görevlerinin Uygulanmasında karşılaşılan Sorunlar. *Elektronik Sosyal Bilimler Dergisi*, 9(34), 32-55.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores In R.L. Thorndike (Ed.), *Educational Measurement*, Washington D.C: American Council on Education.
- Bal, A.P. (2012). İlköğretim Öğrencilerinin Matematik Dersi Performans Görevi Hazırlama Sürecine İlişkin Görüşleri ve Yaşadıkları Sorunlar. *Pegem Eğitim ve Öğretim Dergisi* 2(1).
- Berberoğlu, G. (2007). *Sınıf İçi Ölçme ve Değerlendirme Teknikleri*. Morpa, Ankara.
- Bozkurt, S., Aslanargun, E., Akın, S. ve Kılıç, A. (2014). İlköğretimde Uygulanan Performans Ödevlerine İlişkin Sınıf Öğretmenlerinin Görüşleri. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 18, 157-173
- Chang, L. (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, USA.
- Cizek, G.J., (2001). *Setting Performance Standards Concepts, Methods, and Perspectives*. Lawrence Erlbaum Ass., Publishers, Mahwah, NJ
- Cronbach, Lee. J. (1990). *Essentials of Psychological Testing*. Fifth edition, Harper Collins Publishers
- Çetin, M. O. ve Çakan, M. (2010). Fen ve Teknoloji Dersi Başarılarının Farklı Yaklaşımlarla Ölçülmesi ve Bu Yaklaşımlara İlişkin Öğrenci Görüşleri. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 93-99
- Çetin, S. (2011). İşaretleme ve Angoff Standart Belirleme Yöntemlerinin Karşılaştırılması. Yayınlanmamış Doktora tezi, Hacettepe Üniversitesi, Ankara.
- Çetin, S., ve Gelbal, S. (2010). Farklı standart belirleme yöntemlerinin geçme puanları üzerine etkisi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 43 (1), 79-95.
- Çetin, S. ve Gelbal, S. (2013). İşaretleme ve Angoff standart belirleme yöntemlerinin karşılaştırılması. *Kuram ve Uygulamada Eğitim Bilimleri*, 13(4), 2159-2175
- Çiftçi, S. (2010). İlköğretim birinci kademe 4. ve 5. Sınıf öğretmenlerinin performans görevlerine ilişkin görüşleri. *İlköğretim Online*, 9(3), 934-951.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Florida: Holt, Rinehart and Winston, Inc.
- Cusimano, M. D., Rothman, A., & Keystone, J. (1998). Setting Standards for Performance Assessment: Defining Standards of Competent Performance on an OSCE. *Academic Medicine*, 73(10), S112-113.
- Cusimano, M. D. ve Rothman, A. I. (2003). The Effect of Incorporating Normative Data into a Criterion-Referenced Standard Setting in Medical Education. *Academic Medicine*, 78 (10), 88 - 90
- Ebel, R. L. (1972). *Essentials of Educational Measurement*. New Jersey: Prentice Hall.
- Ersoy, A. ve Anagün, Ş.S. (2009). Sınıf Öğretmenlerinin Fen ve Teknoloji Dersi Ödev Sürecine İlişkin Görüşleri. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 3(1), 58-79.
- Gündeğer, C. ve Doğan, N. (2014). Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 53-60.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24(4), 355-366.

- Hambleton, R. K., ve Pitoniak, M. (2006). *Setting performance standards*. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 433-470). Westport, CT: Praeger.
- Howell, D.C. (2012). *Statistical methods for psychology (8th ed.)*. Belmont, CA: Wadsworth Publishing ; 2012.
- Hurtz G. M. ve Hertz N. R. (1999). How Many Raters Should be Used for Establishing Cutoff Scores with the Angoff Method? A Generalizability Theory Study. *Educational And Psychological Measurement*, 59, 885 – 897.
- Impara, J. C. ve Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. *Journal of Educational Measurement*, 35 (1), 69 – 81.
- Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. *Setting performance standards: Concepts, methods, and perspectives*, 53-88. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kaufman, D. M., Mann, K. V., Muijtjens, A. M., & van der Vleuten, C. P. (2000). A Comparison of Standard-setting Procedures for an OSCE in Undergraduate Medical Education. *Academic Medicine*, 75(3), 267-271.
- Kumandaş, H. ve Kutlu, Ö. (2010). İlköğretim 5. Sınıf Öğrencilerinin Performans Görevine ilişkin Tutumlarını Etkileyen Faktörler. *İlköğretim Online*. 9(2), 714-722
- Linn, R. L., ve Shepard, L. A. (1997). Item-by-item standard setting: Misinterpretations of judges' intentions due to less than perfect item intercorrelations. Paper presented at the *Large Scale Assessment Conference*, Colorado Springs CO.
- Livingston, S.A., Zieky J. M., (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*, ETS
- McKinley, D. W., ve Norcini, J. J. (2014). How to Set Standards on Performance-Based Examinations: AMEE Guide No. 85. *Medical teacher*, 36(2), 97-110.
- Margolis, M. J., De Champlain, A. F., ve Klass, D. J. (1998). Setting Standards for Performance Assessment: Setting Examination-level Standards for a Performance-based Assessment of Physicians' Clinical Skills. *Academic Medicine*, 73(10), S114-116.
- MEB (2016). Millî Eğitim Bakanlığı Ortaöğretim Kurumları Yönetmeliği 10 Eylül 2017 tarihinde https://ogm.meb.gov.tr/meb_iys_dosyalar/2016_11/03111224_ooky.pdf adresinden erişildi.
- Mills, C. N., & Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. *Applied Measurement in Education*, 1(3), 261–275.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19
- NCTM (2011). As the crow Flies . 15 Nisan 2016 tarihinde <http://www.nctm.org/Standards-and-Positions/Focus-in-High-School-Mathematics/Reasoning-and-Sense-Making-Task-Library/Crow/> adresinden erişildi.
- Pawar, M. (Ed.). (2004). *Data collecting methods and experiences: a guide for social researchers*. New Delhi: Sterling Publishers Pvt. Ltd.
- Shephard, L. A, Flexer, R.J., Heibert, E.H., Marion, S.F., Myfield, V., ve Weston, T.J. (1995). Effects of Introducing Classroom Performance Assessment on Student Learning. *Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education*, New Orleans, April, 1994.
- Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., Van Der Vleuten, C., ve Van der Velden, U. (2009). Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting methods. *European Journal of Dental Education*, 13(3), 162-171.
- Taşdelen, G. (2009). Nedelsky ve angoff standart belirleme yöntemlerinin genellenebilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma. Yayımlanmamış yüksek lisans tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

- Taşdemir, F. (2013). Angoff (1-0), Nedelsky ve sınır değerleri saptama yöntemleri ile bir testin sınıflama doğruluklarının incelenmesi. Yayımlanmamış Doktora Tezi. Ankara Üniversitesi, Ankara
- Tülübaşı, G. (2009). Psikolojik testlerde Angoff ve sınır grup yöntemleri ile kesme puanlarının belirlenmesi. Yayımlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- Tüysüz, C. , Karakuyu, Y. ve Tatar, E. (2010). Fen ve Teknoloji Dersindeki Performans Görevlerine Yönelik Veli Tutumlarının Belirlenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)* 4(1), 108-122.
- Yiğit, F. ve Kırımlı, B. (2014). Türkçe Öğretmenlerinin Alternatif Ölçme ve Değerlendirme Yöntemlerini Uygulama Biçimleri ve Uygulamada Karşılaştıkları Sorunlar. *Turkish Studies-International Periodical For The Languages, Literature and History of Turkish or Turkic* 9(3) 1621-1639.

Extended English Summary

Introduction

Standard-setting is the definition of information and skills level that is necessary to reach a specific performance level, and it is a process of setting a point that corresponds to the performance level. These points, which classify individuals by dividing performance categories into two or more, are called cutoff score. Based on the passing scores acquired from assessment instruments, various institutional decisions about individuals such as selection, placement, identification, and guidance are made. Tests are preferred since they are easily applicable, objective, valid, and reliable to make observations compared to other methods (Cronbach, 1990; Pawar, 2004); however, in recent years, authentic assessment and evaluation methods have been introduced with the renovation of elementary school curricula based on constructivist approach. As one of these approaches, although performance tasks were compulsorily carried out in elementary schools until 2014, they were not carried out in secondary schools. Compulsory application of performance projects in secondary education started in 2013-2014 school year. There are two important perspectives on the use of performance tasks. One of these perspectives is that performance task enables a better evaluation; and the other is that it improves learning and teaching (Shephard, Flexer, Heibert, Marion, Myfield, and Weston, 1995). Performance tasks should enable students to use their high level mental abilities; they should help students to gain abilities in problem solving and cooperation, and provide them an opportunity to improve themselves (Kumandaş, and Kutlu, 2010). In addition to these qualities, a performance task should provide a product, observable performance, high level process of thought, social skills, teamwork, interdisciplinary transition, and information exchange in order to be a performance task (Berberoğlu, 2007).

There are various number of studies which approach performance tasks from different perspectives. However, there is not any standard-setting study about performance tasks in the literature. In this study, passing score of a prepared performance task will be determined through one of the standard-setting methods, Angoff method. Suggested by William H. Angoff in 1971, each item is evaluated as a whole in this method; an expert estimates the possibility of giving correct answers for the students who were on the border of failing.

Compared to other similar methods, Angoff method is more commonly used since it is easier in terms of explaining to experts, and collecting and analysing the data (Mills & Melican, 1988); for this reason, this method was preferred in the study.

The importance of the research

Performance tasks are important in terms of enabling students to use and adapt the information they have learned in their daily lives. Besides, apart from written exams, it is the most common method used by teachers (Altun and Gelbal, 2014). While evaluating performance tasks, it is necessary to determine whether the tasks that students have prepared are adequate or not. In

order to determine the measure of this decision, standard-setting studies have been carried out. Standard-setting studies are the studies which aim to calculate passing scores in order to determine whether students have the necessary performance about the subject topic or not. There are many standard-setting studies in the literature, and most of them are studies that calculate passing scores through multiple choice tests (study will be added); however, there are limited number of studies about performance evaluation and there is not any study in our country. From this perspective, it is assumed that this study will make valuable contributions to the literature.

Method

Study group

26 mathematics teachers who are working in high schools in different cities participated in the standard-setting study as experts.

Data collection tool

In order to introduce and teach the distance formula to the students, NCTM ([National Council of Teachers of Mathematics](#)) designed an assignment that aims to help students to repeat the subject and to use the distance formula on an example in daily life rather than helping them to memorize the formula. It was indicated that this assignment would be used with the aim of evaluation (NCTM, 2011). In the study, this assignment was used as performance task. Called as the crow flies and included coordinate plane and Pythagoras theorem, this five-question assignment was firstly translated into Turkish. Later on, with the cooperation of the experts, identification of success level, grading of the questions, and the point of each stage were specified.

Analysis of the Data

The participant experts were informed about that this is a study which identifies a passing score for the prepared performance task; after explaining the stages of the standard-setting study in details, each expert was shown the questions in the performance task, and asked about the possibility of giving correct answers for the students who were on the border of failing. The consistency of grading among experts were determined through Kendall's W coefficient of

concordance ($\chi^2 = 137,872$, $sd=9$, $p<0,00$). According to the points acquired from experts, it was concluded that although there was not a high level of consistency among experts, there was a moderate level of consistency. As a result of the calculations, final cutoff score was found to be 45.10.

Conclusion and Suggestions

In this study, passing score of a prepared performance task was determined through Angoff method which is one of the standard-setting methods. According to the acquired findings, passing score of the performance task was determined as 45.10. According to National Education Regulations for Passing a Grade Level (MEB, 2016), passing score is 50. According to the same regulation, study tasks and performance tasks are evaluated through points and two performance scores for each lesson should be given in each term. From this perspective, passing score for the performance task used in this study was found as 45.10, and it is 4.90 points less from the present passing score. When considered from this point of view, if similar studies to be carried out with different performance tasks give similar results, these results can be shared with decision makers in order to re-determine the passing score under the light of the findings of similar scientific studies. In this study, Angoff method, which is one of the test-centred standard-setting methods, was used. Other test-centred standard-setting methods and student-centred standard-setting methods can be used in similar studies.