

GENELLENEBİLİRLİK KURAMINDA ÇOK YÜZEYLİ DESENLERİN İNCELENMESİ*

Zarife TAŞTAN**

N.Bilge UZUN***

ÖZ

Araştırmanın temel amacı; Genellebilirlik kuramına göre geometri performans görevlerini puanlamada öğrencilerin birden fazla puanlayıcı tarafından puanlanmasıyla oluşturulan desenlerden elde edilen G çalışmaları sonuçlarını incelemektir. Araştırmanın çalışma grubunu, 2014- 2015 öğretim yılında Mersin ili merkez liselerinde öğrenim gören 128 12. sınıf öğrencisi oluşturmaktadır. Ayrıca araştırmada öğrencilerin geometri performans görevlerinin puanlanmasında sekiz puanlayıcı görev almıştır. Araştırmada puanlayıcıların öğrencileri sekiz görev doğrultusunda puanlamasıyla oluşturulan (b:o)xgxp , (b:o)x(g:p) ve (b:(p:o))xg desenleri (b: öğrenci, g: görev, p: puanlayıcı, o: okul) için ayrı ayrı G çalışması yapılmıştır. Elde edilen sonuçlara göre, aynı veri seti üzerinden kurgulanan desenlerde varyans bileşenlerine ilişkin büyüklükleri, varyans değişimlerinin büyüklük açısından sıralamadaki yerleri farklılıklar göstermektedir. Bunun yanı sıra söz konusu desenlerdeki artık varyans değeri de değişmektedir. (b:o)xpxg deseni için kestirilen artık varyans bileşeni, (b:o)x(g:p) ve (b:(p:o))xg desenlerine göre oldukça küçük bulunmuştur. Bu bağlamda G kuramında duruma ilişkin kurgulanan desenin doğruluğunun yapılacak yorumların güvenilirliğini ve geçerliğini etkileyeceği sonucuna ulaşılmıştır. Her bir desen için elde edilen G ve Phi katsayıları dikkate alındığında; (b:(p:o))xg deseninin varyans bileşenlerini açıklama oranı daha az gibi görünse de G ve Phi katsayılarının diğer iki desene göre daha yüksek çıktığı sonucuna ulaşılmıştır.

Anahtar Kelimeler: Klasik test kuramı, genellebilirlik kuramı, çok yüzeysel desen, performans ölçme, puanlama anahtarı

INVESTIGATION OF MULTI-FACET DESIGNS IN GENERALIZABILITY

ABSTRACT

The main purpose of the research is; according to the Generalizability Theory, it is to examine the results of G studies obtained from the designs created by scoring students by more than one rater in scoring geometry performance tasks. The study group of the research consists of 128 12th grade students studying in Mersin central high schools in the 2014-2015 academic year. In addition, eight raters took part in scoring the geometry performance tasks of the students in the study. In the study, a separate G study was conducted for (b:o)xgxp, (b:o)x(g:p) and (b:(p:o))xg designs (b: student, g: task, p: rater, o: school), which were created by scoring students according to eight tasks. According to the results, the sizes of the variance components and the place of variance changes in the order differ in terms of the size of the variance components in the designs built on the same data set. In addition, the residual variance value in these designs is also changing. The residual variance component estimated for (b:o) xpxg design was found to be quite small compared to (b:o)x(g:p) and (b:(p:o))xg designs. In this context, it has been concluded that the accuracy of the design set up in the G theory regarding the situation will affect the reliability and validity of the interpretations to be made. Considering the G and Phi coefficients obtained for each design; Although the (b:(p:o))xg design seems to have less explanation of variance components, it was concluded that the G and Phi coefficients were higher than the other two designs

Keywords: Generalizability theory, classical test theory, multi-facet design, performance measurement, holistic rubric.

GİRİŞ

Teknolojinin hızla geliştiği günümüzde bilimsel bilgi de hızla gelişmektedir. Bu gelişim eğitimde de gelişmelere ve değişmelere sebep olmuştur. Daha önceleri öğrenci başarısını sadesonuca odaklanarak

* Bu çalışma 2017 yılında tamamlanan aynı isimli Yüksek Lisans tezinden türetilmiştir.

** Dr., Mersin Üniversitesi, Spor Bilimleri Fakültesi, Beden Eğitimi ve Spor Bölümü, Mersin. zariphe02@gmail.com, ORCID: <https://orcid.org/0000-0002-7229-2706>

*** Doç. Dr., Mersin Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Mersin. n.bilgeuzun@gmail.com, ORCID: <https://orcid.org/0000-0003-2293-4536>

değerlendiren, öğrencinin bilgisini klasik ölçme araçları (kısa yanıtı, uzun yanıtı, çoktan seçmeli sınavlar vb.) ile sınırlı bir zaman diliminde ölçmeye çalışan ölçme ve değerlendirme yaklaşımları bu değişimlerle beraber öğrencinin değerlendirilmesi konusunda farklı bir bakış açısı kazanmıştır. Artık öğrencilerden beklenen; bilgiyi ezberleyen değil, kazandığı bilgiyi farklı yaşam durumlarına uygulayan, problem çözebilen, eleştirel düşünebilen, verileri analiz edebilen ve etkili sunabilen, sözlü ve yazılı iletişim kurabilen, kendini değerlendirebilen bireyler olmalarıdır. Bu nedenle uygulanan sonuç odaklı klasik değerlendirme yaklaşımları yerine, sözü edilen üst düzey bilişsel özelliklerin ölçülmesinde öğrenme sürecinin de değerlendirilmeye katılması önemli görülmektedir. Bu bağlamda eğitimde performans değerlendirme daha sık yer almaya başlamıştır.

Performans, “eğitimdeki öğrenme çıktılarını kazanmadaki başarıma” olarak tanımlanır ve bilişsel, duyuşsal, psiko-motor davranışlarla üst düzey zihinsel beceriler olmak üzere bütün öğrenme ürünlerini içerir (Turgut ve Baykul, 2010). Performans değerlendirme, öğrencilerin bilgiyi yeniden yapılandırmasını sağladığından, öğrencilerin güçlü ve zayıf yanlarını, başarılarını ve eksikliklerini, öğretim yöntemlerinin yeterliliğini anlamada yardımcı olur. Bu amaçla öğrencileri değerlendirmek için klasik ölçme araçlarının (kısa yanıtı, uzun yanıtı, çoktan seçmeli sınavlar vb.) yanı sıra performans değerlendirme araçları kullanılabilir. Performans değerlendirme araçları ile yapılan işlemlerin güvenilir olması değerlendirmede verilecek kararların doğruluğu açısından oldukça önemlidir. Performans değerlendirmede güvenilir puanlama sağlamak objektif testler kadar kolay olmamakla birlikte gözleme dayalı bir değerlendirme olduğundan birçok hatayı da içinde barındırmaktadır. Performans değerlendirmede öğrencinin performansının puanlayıcılar tarafından değerlendirilmesi söz konusu olduğundan bu çalışmalarda, puanlayıcı güvenilirliği ön plana çıkmaktadır. Puanlayıcılar performansın ölçülmesinde güvenilirliği etkileyen önemli bir hata kaynağı olmakla birlikte ölçümlere karışan diğer değişkenlik kaynakları ve puanlayıcıların birey, görev gibi birçok değişkenlik kaynaklarıyla etkileşimleri de güvenilirlik için önemlidir (Güler, 2008). Bu nedenle güvenilirlik belirlenirken birçok değişkenlik kaynağından gelen hatalar ve bu değişkenlik kaynaklarının birbirleriyle etkileşimleri sonucu çıkabilecek hatalar da dikkate alınmalıdır. Bu bağlamda; davranış ve sosyal bilimlerde (psikoloji, eğitim) birçok ölçme sorununu belirleyebilmek için geniş bir kavramsal sisteme ve güçlü istatistiklere sahip olan Genellebilirlik (G) Kuramı; Cronbach, Gleser, Nanda ve Rajaratnam tarafından (1972) önerilmiştir (Brennan, 1992). G kuramı, sistematik olan ve olmayan hata kaynaklarını ayrı ayrı ele alır ve bunların birbirleriyle etkileşimlerini de belirleyebilir (Brennan, 1992). G kuramına dayalı yapılan çalışmalarda; puanlayıcılardan, diğer değişkenlik kaynaklarından (zaman, test formu, madde, görev, vb.) ve bu değişkenlik kaynakları arası etkileşimden ortaya çıkan hata varyanslarının bir arada göz önüne alındığı kapsamlı tek bir güvenilirlik katsayısı hesaplanmaktadır (Atılğan, 2004). Diğer ifadeyle G kuramı, araştırmacıya puanların potansiyel değişkenlik kaynaklarını tanımlamalarına ve her bir kaynağın katkısının ilişkisini analiz etmelerine olanak sağlamaktadır (Taylor, 2013).

Genellebilirlik kuramına göre evrende bir ya da birden çok varyans kaynağı bulunmaktadır. Varyans kaynakları olan madde (ya da görev), zaman, puanlayıcı gibi ölçmenin benzer durumlar setine, *değişkenlik kaynakları* denir. Bu değişkenlik kaynakları ölçmenin yüzeylerini (facet) oluşturur (Brennan, 1992). Bu yüzeylerin de *koşulları* (conditions) vardır. *Ölçme objesi*, istenilen kararların alınacağı ölçmenin hedefleridir. Pek çok ölçme durumunda bireyler ölçme objesi durumundadır (Brennan, 1992; Shavelson ve Webb, 1991). G çalışmasında yer alan yüzeylere bağlı koşullardan elde edilen ölçme sonuçlarına *genelleme evreni* denir. Bir başka deyişle, genelleme evreni; araştırmacının genellemek istediği koşulların setidir. Alınabilecek tüm koşullardan elde edilen ölçme sonuçları evrenine ise *kabul edilebilir gözlemler evreni* denir. G kuramında, değişkenlik kaynakları çapraz (crossed) ya da yuvalanmış (nested) şekilde olabilir. Bir değişkenlik kaynağının koşullarının başka bir değişkenlik kaynağının koşullarıyla örtüştüğü duruma *çapraz desen* denilmektedir. Çapraz olarak tasarlanmış bir desende değişkenlik kaynakları arasında “x” işareti kullanılır. Bir değişkenlik kaynağının koşulları diğer değişkenin bazı koşullarıyla örtüşüyorsa değişkenlik kaynakları yuvalanmış olarak tasarlanmalıdır. Yuvalanmış olarak tasarlanmış bir desende değişkenler arasında “:” işareti kullanılır. Benzer şekilde bazı değişkenlik kaynaklarının çaprazlanmış bazılarının ise yuvalanmış olduğu desenlerde düzenlenebilir.

Örneğin; çalışmamızda 2 okul (o) ve okulların her birinden 64 er öğrenci (b) seçilmiştir. Seçilen öğrencilere 8 geometri performans görevi (g) uygulanmıştır. Bu ölçme sürecinde görev alan 8 puanlayıcı (p) da tüm öğrencilerin tüm görevlerini puanlamıştır. Her okuldan 64 öğrenci seçildiği için öğrenciler okulların içine yuvalanmıştır (b:o). Her birey bütün geometri performans görevlerini cevapladığı için ve her performans görevi bütün puanlayıcılar tarafından puanlandığı için okul ve öğrenciler, puanlayıcı ve görev yüzeyleri ile çaprazlanmıştır. Bu desen (b:o)xpxg şeklinde gösterilir. Desenimizde değişkenlik kaynakları (yüzeyler); okul (o), öğrenciler (b), görevler (g) ve puanlayıcılar (p) dir. Değişkenlik kaynaklarından okul yüzeyine ait 2 koşul, öğrenci yüzeyine ait 128 koşul, görev yüzeyine ait 8 koşul ve puanlayıcı yüzeyine ait 8 koşul bulunmaktadır. Desenimizde bireyler, istenilen kararların alınacağı ölçmenin hedefleri olduğundan, ölçme objesidir. Ayrıca, olası tüm sekiz puanlayıcı, sekiz madde ve iki okuldan (çalışma ile aynı sayıda yüzey ve koşullara sahip) oluşan tüm durumlar genelleme evrenini ve benzer durumlar altında olası tüm puanlayıcı, okul ve maddeleri içeren evren de kabul edilebilir gözlemler evrenini oluşturmaktadır.

Cronbach (1972) ında önerdiği gibi tüm olası değişkenlik kaynaklarına ilişkin hatanın kestirimine imkân vermesinden dolayı G çalışmalarında tümüyle çapraz desenler daha çok tercih edilmektedir. Ancak yuvalanmış desenin bazı varyans bileşenlerinin hesaplanmasında serbestlik derecesini artırmasından dolayı yararlı olacağı da ifade edilmektedir (Alkan, 2013).

Bu araştırmanın amacı, G Kuramı'nda performansa dayalı durum belirlemede elde edilen ölçümler için farklı ve çok hata kaynaklı bir ölçme durumu olarak çok yüzeyli farklı desen kurgulamaları ile elde edilen puanların G ve Phi katsayılarının incelenmesidir. Dört değişkenli dengelenmiş veri yapılarında G kuramının nasıl kullanılabileceğini göstermek, araştırma kapsamında kullanılan geometri puanlama anahtarının güvenilirliğini çeşitli çok yüzeyli farklı desen kurgulamalarında incelemek ve birey, okul, puanlayıcı ve görev değişkenlik kaynaklarını ve etkileşimlerini dikkate alarak üç yüzeyli desen uygulamasında Genellenebilirlik katsayıları elde etmek ve yorumlamak amaçlanmaktadır.

Eğitim ve psikolojideki alanyazın incelendiğinde, G Kuramı ile ilgili olan çalışmalarda, G kuramının KTK ile karşılaştırılmasına yönelik çalışmaların ağırlıkta olduğu ve çalışmalarda değişkenlik kaynaklarının çapraz tasarlandığı desenlerin sıklıkla kullanıldığı görülmektedir. G kuramındaki çok yüzeyli çalışmalara yönelik yapılan taramalarda ise; çok yüzeyli çalışmaların çoğunluğunda iki yüzeyli desenler (genellikle maddelerin ve puanlayıcıların yüzey olarak tercih edildiği) ele alınmıştır (Atılğan ve Tezbaşaran 2005, Çakıcı; 2011, Büyükkıdık, 2012; Alkan, 2013). Ülkemizde çok yüzeyli desenler kapsamında ikiden fazla yüzeyin bulunduğu çalışmalar oldukça sınırlıdır (Anıl ve Büyükkıdık, 2012; Büyükkıdık, 2012; Nalbantoğlu ve Başusta, 2015; Aktaş, Başusta ve Alıcı, 2016). Aynı zamanda ikiden fazla yüzeyin bulunduğu ve farklı kurgulanmış desenlerin karşılaştırıldığı herhangi bir çalışmaya ise rastlanmamıştır. Genellenebilirlik çalışmalarında ikiden fazla yüzeyin bulunduğu az sayıda tez çalışması da bulunmaktadır (Büyükkıdık, 2012). Bu çalışmada üç yüzeyli desenlerin incelenmesinin bu anlamda alanyazında önemli bir boşluğu doldurması beklenmektedir. Yurt dışında yapılan G Kuramıyla ilgili çalışmalar incelendiğinde ise, değişkenlerin çapraz tasarlandığı desenlerle G ve K çalışmalarının yapılmasının ağırlık kazandığı, G Kuramının performans değerlendirme çalışmalarında güvenilirlik belirlerken ve ölçeklerdeki madde sayısının belirlenmesinde sıklıkla kullanıldığı görülmektedir.

Bu bağlamda; bu çalışmada çok yüzeyli farklı desenlerin (çapraz ve yuvalanmış) kullanılacak olması, ayrıca ölçme objesi (birey), görev ve puanlayıcı yüzeylerinin yanı sıra farklı ve önemli bir değişkenlik kaynağı olduğu düşünülen “okul” yüzeyine de yer verilmesi sebebiyle araştırma sonuçlarının alanyazınına önemli bir katkı sağlayacağı düşünülmektedir.

Problem Cümlesi

G kuramına göre; farklı okullarda bulunan öğrencilerin geometri performansının puanlanmasında kullanılan görevlerin, birden fazla puanlayıcı tarafından puanlanmaları ile oluşturulan farklı desenlerin Genellenebilirlik (G) ve Phi katsayıları sonuçlarının değişimi nasıldır?

Alt problemler

1) Bireyin (b) okul (o) yüzeyinde yuvalandığı, görev (g) ve puanlayıcı (p) değişkenlerinin ise çapraz tasarlandığı (b:o)xgxp deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

2) Bireyin (b) okul (o) yüzeyinde ve görevin (g) puanlayıcı (p) yüzeyinde yuvalandığı ve bu ikisinin çapraz tasarlandığı (b:o)x(g:p) deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

3) Puanlayıcının (p), okul (o) yüzeyinde yuvalandığı, birey yüzeyinin de puanlayıcı (p) yüzeyinde yuvalandığı ve görev (g) yüzeyinin çapraz tasarlandığı (b:(p:o))xg deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

YÖNTEM

Çalışma Grubu

Araştırmanın çalışma grubunu 2014-2015 öğretim yılında Mersin ili farklı iki tür (anadolu lisesi ve fen lisesi) merkez liselerinde okuyan 128 12. sınıf öğrencisi oluşturmaktadır. Öğrenciler gönüllülük esasına dayalı olarak seçilmiştir. Araştırma sorularına yanıt aramak amacıyla lise matematik öğretmeni olarak çalışan ve araştırmaya katılmaya gönüllü 8 matematik öğretmenin katılımıyla puanlayıcı grubu oluşturulmuştur.

Veri Toplama Araçları

Bu çalışmada sekiz geometri sorusundan oluşan açık uçlu sınav ile puanlayıcıların puanlama yapabilmesi için hazırlanan bütünsel dereceli puanlama anahtarı veri toplama aracı olarak kullanılmıştır.

İlk aşamada, 2014-2015 öğretim yılında uygulanmış bir YGS denemesinde yer alan 8 geometri sorusu seçilmiş ve yeni bir soru kağıdı hazırlanmıştır. Hazırlanan soru kağıdı, anlam ve dil bilgisi açısından iki matematik uzmanı tarafından tekrar gözden geçirilmiştir.

Puanlama anahtarının hazırlandığı ikinci aşamada öncelikle soru kağıdında yer alan sorular tarafından ölçülen kazanımlar belirlenmiştir. Öğrencilerin performanslarda göstermeleri istenen, gözlenmek istenen özel davranışlar ve gözlenmek istenmeyen davranışlar tanımlanmıştır. Her soruda örnek cevaptan yetersiz cevaba kadar her bir özelliği dahil eden eksiksiz açıklama yazılmıştır örnek cevaptan yetersiz cevaba kadar süregelen aralıktaki seviyeler ortak özelliklerle tanımlanarak puanlama anahtarı tamamlanmıştır. Hazırlanan dereceli puanlama anahtarı için ayrıca uzman görüşü alınmıştır. Her bir geometri performans görevi 0-4 puan üzerinden ölçülmüştür. Puanlayıcılara, bütünsel dereceli puanlama anahtarının ve puanlama tablolarının nasıl kullanılacağı araştırmacı tarafından açıklanmış, tüm cevap kağıtlarını puanlamaları sağlanmıştır. Puanlama tablolarında, her bir öğrencinin cevap kağıdına verilen, okulunu simgeleyen bir harf ve her bir öğrenci için ayrı ayrı verilmiş bir numaradan oluşan kodlar bulunmaktadır. Diğer bir deyişle, her bir öğrencinin, isminin yer almadığı, bir harf ve bir sayıdan oluşan kod numarasının karşısına her bir madde için verdiği cevaba karşılık gelen puanların yazılacağı tablolar oluşturulmuştur.

Verilerin Analizi

8 geometri performans görevinin 8 puanlayıcı tarafından bütünsel puanlama anahtarı yardımıyla puanlanmasından elde edilen veriler, G Kuramı kapsamında analiz edilmiş ve elde edilen bulgular incelenmiştir. Araştırmanın gerçekleştirilmesinde aşağıdaki adımlar izlenmiştir.

1. 8 geometri performans görevi herbir okulda 64 öğrenci olmak üzere farklı düzeydeki iki okulda toplam 128 öğrenci tarafından cevaplanmıştır.

2. 128 öğrencinin cevapladığı performans görevleri 8 farklı puanlayıcı tarafından puanlanmıştır. Puanlama yapılırken yukarıda açıklamaları verilen bütünsel dereceli puanlama anahtarı kullanılmıştır.

3. Elde edilen veriler desenlere uygun olacak şekilde bilgisayar ortamına aktarılmıştır. G kuramında kullanılmak üzere tamamen dengelenmiş üç desen tasarlanmıştır. G Kuramı'nda analizler her bir desen için ayrı ayrı yapılmıştır.

4. Bu üç farklı desene bağlı olarak yapılan puanlamaların güvenilirliği G ve Phi katsayıları hesaplanarak incelenmiştir.

G kuramı ile desenlere ait varyans bileşenlerinin kestirilmesi, değişkenlerin toplam varyansı açıklama oranlarının hesaplanmasında EduG 6 programı kullanılmıştır.

BULGULAR ve YORUMLAR

Bu bölümde, araştırma sorularının her biri için yapılan analizler sonucu elde edilen bulgulara yer verilmiştir.

Araştırmada ilk alt probleme yanıt aramak için yapılan Genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 1. de verilmiştir.

Tablo 1. (b:o)ıxıxp desenine ait G çalışması bulguları

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	Varyans Yüzdesi
o	730.93372	1	730.93372	0.16445	9.8
b:o	2845.31226	126	22.58184	0.27939	16.6
p	857.05945	7	122.43706	0.10976	6.5
g	2133.90125	7	304.84304	0.26182	15.5
op	46.63757	7	6.66251	0.01151	0.7
og	233.55750	7	33.36536	0.05564	3.3
bp:o	355.03735	882	0.40254	0.02409	1.4
bg:o	3976.02563	882	4.50797	0.53727	31.9
pg	193.79309	49	3.95496	0.02639	1.6
opg	28.30090	49	0.57757	0.00575	0.3
bpg:o	1295.54663	6174	0.20984	0.20984	12.4
Toplam	730.93372	6143			100%

Tablo 1 'de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında b:o ana etkisine ait varyans bileşeni ($\sigma^2(b:o)$) toplam varyansın %16.6'sını açıklamaktadır. Varyans bileşeni $\sigma^2(b:o)$ 'nin büyük olması, farklı okullarda bulunan bireylerin geometri becerileri bakımından farklılık gösterdiği, okullarda bulunan bireylerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı şeklinde yorumlanabilir. Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın %9.8 'ini açıkladığı dikkate alındığında $\sigma^2(b:o)$ 'nin yüksek çıkmasının öğrenci farklılaşmasından kaynaklı olduğu düşünülebilir. Bu durum, farklı türde (anadolu lisesi ve fen lisesi) okullarda bulunan öğrencilerin geometri hazırbuluşluklarının farklı olmasından da kaynaklanabilir. Görev ana etkisine ait varyans bileşeni $\sigma^2(g)$, performans görevlerinin zorluk kolaylık bakımından farklılık gösterip göstermediği hakkında bilgi verir (Shavelson ve Webb, 1991). Görev değişkenine ait varyans bileşeninin toplam varyansın %15.5 olması, performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin değiştiği şeklinde yorumlanabilir. Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın %9.8 ini açıklaması diğer ana etkiler dikkate alındığında nispeten küçük olması sebebiyle, okullarda geometri becerileri bakımından farklılaşmanın az olduğunun bir göstergesidir. Puanlayıcı varyansı, puanlayıcıların diğer puanlayıcılara göre öğrencileri değerlendirmede cömertliklerinin veya katılıklarının değişip değişmediği hakkında bilgi verir (Shavelson ve Webb, 1991). Puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın %6.5 sini açıklaması, tüm ana etkilerden kaynaklı varyans yüzdelerinden nispeten küçük olması sebebiyle, puanlayıcıların öğrenci cevaplarını puanlamada benzer cömertliğe/katılığa sahip oldukları, yani geometri becerilerini puanlama bakımından puanlayıcıdan puanlayıcıya farklılaşmanın fazla olmadığı biçiminde yorumlanabilir.

Birey x görev x okul ortak etkileşiminine ait varyans bileşeni ($\sigma^2(\text{bgo})$), toplam varyansın %31.9'unu açıklamaktadır. Bu değer büyük olması, farklı okullarda bulunan öğrencilerin geometri becerileri bakımından okuldan okula farklılık gösterdiği, birey-görev-okul ortak etkileşiminin farklılaştığı, öğrenciler için geometri performans görevlerinin bir okuldan diğerine aynı zorluk/kolaylıkta olmadığı şeklinde yorumlanabilir. Ana etkiler dikkate alındığında; b:o ana etkisine ait varyans bileşeni $\sigma^2(\text{b:o})$ ve görev ana etkisine ait varyans bileşeni $\sigma^2(\text{g})$ nin diğer ana etkilere göre görece büyük olması, bu etkileşimin büyük olması sonucu ile örtüşmektedir. Bir diğer ifade ile bu etkileşime ilişkin varyans bileşeninin büyük çıkması bu bileşeni oluşturan ana etkilerin varyans bileşenlerinin büyük çıkması bulgusu ile örtüşmektedir.

Tabloda yer alan varyans bileşenlerinden okul-görev ortak etkisine ait kestirilen varyans bileşenin $\sigma^2(\text{og})$ toplam varyansı açıklama yüzdesi %3.3'tür. Bu bulguya göre, soru kağıdında yer alan geometri performans görevlerinin zorluk düzeyi bakımından, farklı okullarda bulunan öğrenciler arasında farklılık yaratmadığı söylenebilir. Ana etkiler dikkate alındığında; söz konusu bu bulgu görev ana etkisine ait varyans bileşeni $\sigma^2(\text{g})$ nin görece büyük olması ile örtüşmemektedir. Bu durum, okul ve görev etkileşim etkisine okul ve görev değişkenlik kaynağı dışında farklı değişkenlik kaynaklarının da etkide bulunduğunu veya farklı seçkisiz hata kaynaklarının (okullarda bulunan öğrencilerin görevlerle daha önceden karşılaşmış olmaları, görevlerin zorluk veya kolaylık bakımından farklı algılanması, okullarda ölçme ortamlarının birbirinden farklı olması vb ...) etkileşime dahil olduğunu göstermektedir.

Puanlayıcı-görev ortak etkisine ait varyans bileşenin $\sigma^2(\text{pg})$ toplam varyansı açıklama yüzdesine bakıldığında, %1.6 'lık bu değer küçük bir değer olduğu, dolayısıyla öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine az miktarda değişiklik gösterdiği, puanlayıcıların bireyleri bir görevden diğerine kararlı puanladıkları şeklinde yorumlanabilir.

Varyans bileşeni $\sigma^2(\text{bp:o})$, toplam varyansın %1.4 'ünü açıklamaktadır. Bu değer küçük olması, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı şeklinde yorumlanabilir.

Okul ve puanlayıcı ortak etkisine ait varyans bileşeninin $\sigma^2(\text{op})$ toplam varyansı açıklama yüzdesi %0.7'dir. Okul ve puanlayıcı etkileşimi her bir okuldaki öğrencileri puanlayanların puanlamalarında meydana gelen değişikliğin göstergesidir. Bu durumda puanlayıcıların bireyleri puanlamasının okuldan okula değişmediğini söylemek mümkündür.

Okul, puanlayıcı ve görev ortak etkisine ait varyans bileşenin $\sigma^2(\text{opg})$ toplam varyansı açıklama yüzdesi 0.3 'tür. Üç değişkenlik kaynağının etkileşiminden elde edilen varyans bileşeninin sıfır (0) olması istenen bir durumdur. Bu oranın sıfıra yakın olması okul, görev, puanlayıcı etkileşiminde, seçkisiz hata kaynaklarının küçük olduğunun bir göstergesidir (Shavelson ve Webb, 1991). Bunun yanı sıra, bu bulgu ölçme objesi olan bireylerin temelde farklılaşması ile de ilgili olabilir. Temel değişkenliğin istendik bir durum olan bireylerden kaynaklı olduğu söylenebilir.

Birey x puanlayıcı x görev x okul $\sigma^2(\text{bpg:o})$, ortak etkisine ait varyans bileşeni artık varyans olarak adlandırılır. Artık varyans bileşeninin %12.4 olması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya seçkisiz hata kaynaklarının yanı sıra başka değişkenlik kaynaklarının da deseni etkilediği şeklinde yorumlanabilir. Bu değişkenlik kaynakları cinsiyet, puanlayıcı nitelikleri, alt beceri alanları (geometri-analitik geometri) olabilir.

Araştırmada ikinci alt probleme yanıt aramak için yapılan Genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 2 'de verilmiştir.

Tablo 2. (b:o)x(g:p) desenine ait G çalışması bulguları

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ort.	Varyans	Varyans Yüzdesi
o	72.25000	1	72.25000	0.13494	6.9
b:o	334.50000	126	2.65476	0.22390	11.4
p	495.47656	3	165.15885	0.52391	26.6
g:p	143.21094	4	35.80273	0.23181	11.8
op	4.10156	3	1.36719	-0.03870	0.0
og:p	24.52344	4	6.13086	0.08528	4.3
bp:o	326.42188	378	0.86355	0.09520	4.8
bg:op	339.26563	504	0.67315	0.67315	34.2
Toplam	1739.75000	1023			100%

Tablo 2 'de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında tüm ana etkiler içerisinde puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın %26.6 'ünü açıkladığı görülmektedir. Bu durum, geometri becerilerini puanlama bakımından puanlayıcıdan puanlayıcıya çok fazla farklılaşma olduğunun bir göstergesidir, dolayısıyla seçilen puanlayıcıların geometri görevleri bakımından bir yaşantı geçirmesinin, uzmanlık ve ilgi alanlarının bu durumu etkilediği düşünülebilir. Etkileşim etkisinde de varyans bileşenleri açıklama yüzdeleri dikkate alındığında en büyük değer $\sigma^2(gp)$ 'ye ait olması da bu durumu desteklemektedir.

Ana etki g:p için kestirilen varyans bileşeni $\sigma^2(g:p)$ toplam varyansın %11.8'ini açıklamaktadır. $\sigma^2(g:p)$ 'nin nispeten büyük olması, geometri görevlerinin puanlanmasının puanlayıcıdan puanlayıcıya farklılık gösterdiği, görev puanlayıcı ortak etkileşiminin farklılaştığı ve öğrencilerin geometri becerilerinin bir puanlayıcıdan diğerine değiştiği şeklinde yorumlanabilir. Bu bulgu, puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ büyüklüğü de göz önüne alındığında; farklı puanlayıcıların farklı görevleri puanlıyor olması sonucu puanlayıcıların puanlamaya konu olan özellikle ilişkili bir durumun varlığını düşündürülebilir. Şöyle ki puanlamayı yapan puanlayıcının puanlama yaptığı göreve ilişkin ilgi düzeyi, temel uzmanlığı, vs. gibi göreve ilişkin geçmiş yaşantıları puanlayıcının görevden göreve puanlamasını değiştiriyor olabilir. Yuvalanmış desenler söz konusu olduğunda varyans bileşenleri birbirinden bağımsız olarak yorumlanamaz (Güler, Kaya Uyanık ve Taşdelen Teker,2012). Görevlerin güçlük düzeyleri arasındaki farklılıklar yorumlanırken, bütün görevlerin bütün puanlayıcılar tarafından puanlanmamasından dolayı her bir puanlayıcının puanladığı görevler üzerinden yorumlanmasının uygun olabileceğini düşündürmektedir.

Ana etki b:o için kestirilen varyans bileşeni $\sigma^2(b:o)$, toplam varyansın %11.4'ünü açıklamaktadır. $\sigma^2(b:o)$ 'nin tabloda yer alan diğer varyans bileşenleri dikkate alındığında görece büyük olması, öğrencilerin geometri becerilerinin bir okuldan diğerine değiştiği, okullarda performans görevlerini yanıtlayan öğrencilerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı şeklinde yorumlanabilir.

Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın %6.9 'unu açıkladığı görülmektedir. Okullara ait varyans bileşeninin toplam varyansı açıklama oranının %6.9 olması, okullarda geometri becerileri bakımından farklılaşmanın az olduğunun bir göstergesidir.

Tabloda yer alan varyans bileşenlerinden bazıları negatif değer almıştır. Varyans bileşenlerinin negatif değer alması örneklem hatasından ya da modelin yanlış belirlenmesinden dolayı ortaya çıkabilir (Shavelson ve Webb, 1991). Calkins (1978), Leone ve Nelson (1966) yaptıkları çalışmalarda değişkenlik kaynaklarının düzeyi az sayıda ise (örn. beş) varyans bileşenlerinin negatif çıkabileceğini belirtmişlerdir. Dolayısıyla bu çalışmada bu bulgunun değişkenlik kaynaklarının koşul sayılarının az

olmasından kaynaklandığı düşünülmektedir. Negatif varyans kavramsal olarak mümkün olmadığından negatif değer alan varyans bileşenleri iki farklı yaklaşımla yorumlanabilir. Bu yaklaşımlardan ilki, negatif varyans bileşeni diğer varyans bileşenleri için beklenen kareler ortalamaları denkleminde yer alıyorsa, negatif değer yerine sıfır değeri kullanılarak hesaplama yapılmasıdır (Cronbach, 1972; akt: Shavelson ve Webb, 1991). Diğer yaklaşım ise bu değeri sıfır kabul etmek ancak denkleminde negatif değeri kullanmaktır (Brennan, 1983). Bu araştırmada, ilk yaklaşım dikkate alınmıştır. Okul ve puanlayıcı ortak etkisine ait varyans bileşeninin, $\sigma^2(op)$ negatif olması nedeniyle, toplam varyansı açıklama yüzdesi 0 'dır. Bu durumda puanlayıcı performansının okuldan okula değişmediğini söylemek mümkündür. Puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın %26.6' sını ve okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 6.9 'unu açıkladığı göz önüne alındığında okul-puanlayıcı etkileşim etkisi ile örtüşmediği görülmektedir. Bunun nedeni, puanlayıcıların okullardaki bütün öğrencilerin bütün performans görevlerini puanlamamasından kaynaklanabilir.

Varyans bileşeni $\sigma^2(bp:o)$, toplam varyansın %4.8 'ini açıklamaktadır. Bu değer nispeten küçük olması, bireylerin geometri becerileri bakımından puanlayıcılar tarafından puanlanmasının okuldan okula farklılık göstermediği, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı şeklinde yorumlanabilir.

Artık varyans bileşeni okul $\sigma^2(bg:op)$ toplam varyansın %34.2'sini açıklamaktadır. Artık varyans bileşeninin büyük bulunması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya seçkisiz hata kaynaklarının büyük olabileceğinin göstergesi olabilir. Yanı sıra bu bulgu; desende yer almayan, sonucun genellenebilirliğini etkileyebilecek farklı değişkenlik kaynaklarını da akla getirmektedir. Bu değişkenlik kaynakları; yaşantılar, puanlayıcı nitelikleri, öğrencilerin alanları (sayısal-çeit ağırlık), vb. olabilir.

Araştırmada üçüncü alt probleme yanıt aramak için yapılan Genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 3 'de verilmiştir.

Tablo 3. (b:(p:o))xg desenine ait G çalışması bulguları

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	Varyans Yüzdesi
o	312.84766	1	312.84766	0.58821	30.6
p:o	64.68359	6	10.78060	0.05624	2.9
b:p:o	311.71875	120	2.59766	0.23609	12.3
g	250.95313	7	35.85045	0.25980	13.5
og	18.16797	7	2.59542	0.01410	0.7
pg:o	71.09766	42	1.69280	0.06149	3.2
bg:p:o	595.53125	840	0.70897	0.70897	36.8

Tablo 3 'te verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelere bakıldığında ana etkiler büyükten küçüğe doğru o, g, b:p:o ve p:o varyans bileşenleri şeklinde sıralanmaktadır. Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın %30.6 'sını açıkladığı görülmektedir. Okullara ait varyans bileşeninin çok yüksek olması, okullarda geometri becerileri bakımından farklılaşmanın fazla olduğunun bir göstergesidir. Görev ana etkisine ait varyans bileşeninin $\sigma^2(g)$, toplam varyansın %13.5 'ini açıkladığı görülmektedir. Bu durumda, her öğrencinin cevapladığı geometri görevleri arasında farklılaşmanın fazla olduğu söylenebilir. Elde edilen bu sonuç

performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin değiştiği şeklinde yorumlanabilir.

Varyans bileşeni $\sigma^2(b:p:o)$, toplam varyansın %12.3 ‘ünü açıklayarak dördüncü sırada yer almaktadır. Bu bulgu, farklı okullarda bulunan öğrencilerin geometri becerileri bakımından okuldan okula farklılaştığı, okulların ölçülen özellik bakımından benzeşik olmadığı ve birey-puanlayıcı-okul ortak etkileşiminin farklılaştığı, farklı okullarda bulunan farklı her bir öğrencinin geometri performans görevlerini puanlarken puanlayıcının bir okuldan diğerine tutarlı davranmadığı şeklinde yorumlanabilir. Aynı zamanda bu bulgu; ana etki p:o için kestirilen varyans bileşeni dikkate alındığında varyans bileşeni $\sigma^2(b:p:o)$ yüzdesindeki artışın temel kaynağının bireyler olduğunu düşündürebilir.

Ana etki p:o için kestirilen varyans bileşeni $\sigma^2(p:o)$ toplam varyansın %2.9’ unu açıklamaktadır. Bu varyans bileşeninin küçük olması, puanlayıcıların geometri becerilerini puanlama bakımından okullarda farklılık göstermediği, okulların puanlayıcıların puanladığı özellik bakımından benzeşik olduğu ve puanlayıcı-okul ortak etkileşiminin farklılaşmadığı şeklinde yorumlanabilir.

Varyans bileşeni $\sigma^2(pg:o)$, toplam varyansın %3.2’ sini açıklamaktadır. Bu değerin küçük olması, puanlayıcıların geometri becerilerini puanlama bakımından okuldan okula farklılık göstermediği, okulların ölçülen özellik bakımından benzeşik olduğu ve puanlayıcı-görev-okul ortak etkileşiminin farklılaşmadığı, geometri görevlerinin güçlüklerinin bir görevden diğerine çok fazla değişmediği şeklinde yorumlanabilir.

Okul ve görev ortak etkisi için kestirilen varyans bileşeninin $\sigma^2(og)$ toplam varyansı açıklama yüzdesi %0.7’ dir. Bu durum, kitapçıkta yer alan 8 geometri performans görevinin zorluk düzeyi bakımından birbirinden farklılık göstermediği, okullar arasında farklılık bulunmadığı biçiminde açıklanabilir. Bu bulgu; okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın %30.6 ve görev ana etkisine ait varyans bileşeninin $\sigma^2(g)$, toplam varyansın % 13.5 ‘ini açıklaması ile örtüşmemektedir. Bu durumun nedeni; okul ve görev değişkenlik kaynakları dışında başka değişkenlik kaynaklarından kaynaklanabilir. Bu değişkenlik kaynakları; okulların ölçme ortamları, bireylerin performans görevlerine yönelik geçmiş yaşantıları olabilir.

Artık varyans bileşeni $\sigma^2(bg:p:o)$ toplam varyansın %36.8 ‘ini açıklamaktadır. Artık varyans bileşeninin büyük çıkması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya seçkisiz hata kaynaklarının büyük olabileceğinin göstergesi olabilir.

(b:o)xpxg , (b:o)x(g:p) ve (b:(p:o))xg desenlerinden elde edilen G ve Phi Katsayıları incelendiğinde 8 puanlayıcıdan her birinin farklı iki okulda bulunan 64 er öğrenciyi 8 görev doğrultusunda puanlamasıyla tasarlanan (b:o)xpxg desenine göre yapılan G çalışmasında G katsayısı 0.84, Phi katsayısı ise 0.78 olarak kestirilmiştir. Araştırmada 8 puanlayıcı içinde belirlenen 4 puanlayıcıdan her birinin farklı okullarda bulunan 64 öğrenciyi 2 ‘şer görev doğrultusunda puanlamasıyla tasarlanan (b:o)x(g:p) desenine göre yapılan G çalışmasında G katsayısı 0.75, Phi katsayısı ise 0.56 olarak kestirilmiştir. Araştırmada farklı iki okulda bulunan 64 öğrenciyi 8 puanlayıcıdan her birinin 16 şarlı olarak (8x16=128) 8 görev doğrultusunda puanlamasına ilişkin (b:(p:o))xg desenine göre yapılan G çalışmasında G katsayısı 0.90, Phi katsayısı ise 0.87 olarak kestirilmiştir. Elde edilen bu katsayılar benzer durumda benzer nitelikteki puanlayıcıların bulunduğu durumlarda öğrencilerin geometri görevlerine ilişkin performanslarının genellenebilir nitelikte olduğunun bir göstergesidir. Bu araştırmada ölçme objesi öğrencilerdir. Her bir desen için bireye ait varyans bileşenleri farklı hesaplandığından, evren puan varyansları da farklıdır.

TARTIŞMA, SONUÇ ve ÖNERİLER

(b:o)xgxp , (b:o)x(g:p) ve (b:(p:o))xg desenlerinden G çalışmasıyla elde edilen varyans ve toplam varyansı açıklama oranları incelendiğinde;

Okul ana etkisine ait varyans bileşeninin toplam varyansı açıklama oranı (b:(p:o))xg deseninde diğer iki desende kestirilen toplam varyansı açıklama oranlarından yüksek çıkmıştır. (b:(p:o))xg deseninde kestirilen okul ana etkisine ait varyans bileşeni diğer iki desende kestirilen varyans bileşenlerine göre

daha yüksek kestirildiğinden, (b:(p:o))xg desenine göre diğer iki desende okullarda ölçülen özellik bakımından farklılaşmanın az olduğu sonucuna ulaşılmıştır. Çalışma grubunu Anadolu lisesi ve Fen lisesi okullarında öğrenim gören öğrenciler oluşturduğundan Lgs sınavı sonrası okullara öğrenci alım puanları ve öğrencilerin hazırbulunuşlukları dikkate alındığında okullarda ölçülen özellik bakımından farklılaşma beklenen bir durumdur.

b:o etkisine ait varyans bileşeninin toplam varyansı açıklama oranı (b:o)xgxp deseninde, (b:o)x(g:p) deseninde kestirilen varyans bileşeninden daha yüksek çıkmıştır. Her iki desenden de elde edilen varyans bileşenleri için toplam varyansı açıklama oranlarına bakıldığında, okullarda performans görevlerini yanıtlayan öğrencilerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değiştiği sonucuna ulaşılmıştır.

Görev ana etkisine ait varyans bileşeni (b:o)xgxp deseninde (b:(p:o))xg deseninden daha yüksek kestirilmiştir. Ancak her iki desende de kestirilen varyans bileşeni oranlarına bakıldığında, her öğrencinin cevapladığı geometri görevleri arasında farklılaşmanın olduğu sonucuna ulaşılmıştır. Elde edilen bu sonuç, performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin öğrenciden öğrenciye değiştiği şeklinde yorumlanmıştır.

Puanlayıcı ana etkisine ait varyans oranlarına bakıldığında (b:o)x(g:p) deseninde (b:o)xgxp desenine göre çok yüksek bir değer kestirilmiştir. Bunun temel nedenlerinden birinin puanlayıcıların farklı görevleri (g:p) puanlaması olabileceği sonucuna ulaşılmıştır. Bir başka ifade ile, (b:o)x(g:p) deseninde puanlayıcıların farklı görevleri puanlamasından dolayı puanlayıcıların görevlerle ilgili farklı yaşantılara sahip olmasından kaynaklanabileceği sonucuna ulaşılmıştır.

op ortak etkisine ait varyans değeri (b:o)xgxp deseninde sıfıra yakın çıkarken (b:o)x(g:p) deseninde ise negatif değer aldığından dolayı sıfır olarak kabul edilmiştir. Varyansların sıfır ve sıfıra yakın çıkması, okul-puanlayıcı ortak etkisinin farklılaşmadığı, puanlayıcıların performanslarının okuldan okula farklılık göstermediği; diğer bir ifadeyle puanlayıcıların farklı okullarda bulunan öğrencilerin geometri performanslarını puanlarken benzer cömertliğe/katılığa sahip olduğu sonucuna ulaşılmıştır.

og ortak etkisine ait varyans bileşeninin toplam varyansı açıklama oranı (b:o)xgxp ve (b:(p:o))xg desenlerinde küçük kestirildiğinden, her iki desende de okul-görev ortak etkisinin farklılaşmadığı, geometri performans görevlerinin farklı okullarda bulunan bireyler tarafından aynı güçlükte algılandığı sonucuna ulaşılmıştır.

bp:o ortak etkisine ait varyans değeri (b:o)x(g:p) deseninde (b:o)xgxp deseninden daha yüksek kestirilmiştir. Bu durumun (b:o)x(g:p) deseninde performans görevlerinin puanlayıcı yüzeyinde yuvalanmasından; diğer bir ifadeyle farklı puanlayıcıların farklı görevleri puanlamasından kaynaklı olarak yüksek çıktığı görülmektedir. Her iki desen için bp:o varyans bileşeninin toplam varyansı açıklama oranlarına bakıldığında bu değer küçük olması, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı sonucuna ulaşılmıştır. Bunun yanı sıra (b:(p:o))xg deseninde b:p:o ortak etkisine ait varyans bileşeninin toplam varyansı açıklama oranı diğer iki desendeki bp:o ortak etkisine ait varyans bileşeninin toplam varyansı açıklama oranından yüksek çıkmıştır. Bu bulgu, farklı okullarda bulunan bireylerin geometri becerileri bakımından okuldan okula farklılaştığını, okulların ölçülen özellik bakımından benzeşik olmadığı ve birey-puanlayıcı-okul ortak etkileşiminin farklılaştığını, puanlayıcıların farklı okullarda bulunan her bir öğrencinin geometri performans görevlerini puanlarken puanlayıcının bir okuldan diğerine tutarlı davranmadığını göstermektedir. Bu bulgular dikkate alındığında, desenlerin yapısının varyans bileşenlerinin oranlarını, dolayısıyla yorumlanmasını farklılaştırdığı görülmüştür. Bir başka ifade ile desenlerde yuvalanmış veya çaprazlanmış yüzeylerin sayısı değiştikçe varyans bileşenlerinin açıkladıkları varyans oranlarının da değiştiği sonucuna ulaşılmıştır.

Okul-puanlayıcı-görev ortak etkileşiminin açıkladığı varyans oranına bakıldığında da her üç desen için farklı sonuçlar elde edilmiştir. Bu durumda desenlerin farklılaşmasıyla varyans bileşenlerinin açıklama oranlarının değiştiğinin bir başka kanıtı olarak ortaya çıkmıştır.

Bütün bulgular dikkate alındığında, aynı veri seti üzerinden kurgulanan desenlerde varyans bileşenlerine ilişkin büyüklükler, varyans değişimlerinin büyüklük açısından sıralamadaki yeri farklılıklar göstermektedir. Bu durum, söz konusu bileşenlere ilişkin yorumlamaları da değiştirmektedir. Örneğin (b:o)xp_xg ve (b:o)x(g:p) desenlerinde okula ilişkin varyans değişiminin diğer bileşenlere göre küçük olması okullardan kaynaklı farklılığın az olduğu biçiminde yorumlanırken; (b:(p:o))xg deseninde okullara ait varyans bileşeninin çok yüksek olması, okullarda geometri becerileri bakımından farklılaşmanın fazla olduğu biçiminde yorumlanmıştır. Bunun yanı sıra söz konusu desenlerdeki artık varyans değerinin de değiştiği görülmektedir. (b:o)xp_xg deseni için kestirilen artık varyans bileşeni, (b:o)x(g:p) ve (b:(p:o))xg desenlerine göre oldukça küçük bulunmuştur. Bu bağlamda G kuramında duruma ilişkin kurgulanan desenin doğruluğunun yapılacak yorumların güvenilirliğini ve geçerliğini de etkileyeceği sonucuna ulaşılmıştır. Desenlerin çaprazlanmış ve/veya yuvalanmış olmasının da ana etkiler ile etkileşim etkilerinin varyans bileşenlerinin büyüklüklerinde farklılıklar yarattığı görülmektedir. Bir başka ifade ile, çaprazlanmış desenlerde ana etkilere ait varyans yüzdeleri çok büyük ya da çok küçük olabilirken yuvalanmış desenlerin açıkladığı varyans yüzdeleri değişebilmektedir. Bu durum Chang ve Hocevar (2000), Nalbantoğlu (2009), Alkan (2013), Nalbantoğlu ve Başusta (2015) ve Doğan ve Anadol (2017), tarafından yapılan çalışmalarla da desteklenmektedir.

(b:o)xp_xg, (b:o)x(g:p) ve (b:(p:o))xg desenlerinden elde edilen G ve Phi Katsayıları sırasıyla; 0.84 - 0.78; 0.75 - 0.56 ; 0.90 - 0.87 olarak kestirilmiştir. Çalışmada yer alan bütün desenlerin yuvalanmış olmasından dolayı bireylere ait varyans bileşenleri diğer değişkenlik kaynaklarından bağımsız olarak elde edilememekte ve yorumlanamamaktadır. Aynı zamanda desenlerin yapısı gereği içerdikleri varyans kaynakları sayısı da farklılık göstermektedir. G ve Phi katsayıları hesaplanırken paydada bulunan varyans bileşen sayısı da farklı olduğundan katsayılar da farklı büyüklüklerde bulunmuştur. Paydada bulunan varyans kaynak sayısı arttıkça katsayılar küçülecek; varyans kaynak sayısı azaldıkça da katsayılar büyüyecektir. Birinci desene ait G ve Phi katsayılarının ikinci desene göre yüksek olmasının sebebi; artık varyans değerinin ikinci desene göre oldukça düşük olması; birinci desenin G ve Phi katsayılarının üçüncü desene göre düşük olması ise varyans bileşenlerinin sayıca fazla olması ile açıklanabilir. Bu çıkarımlar doğrultusunda üçüncü desenin varyans bileşenlerini açıklama oranı daha az gibi görünse de G ve Phi katsayıları diğer iki desene göre daha yüksek çıkmıştır. Bunun nedeni de desenin yapısındaki varyans bileşen sayısının az olması olarak açıklanmıştır. Bir desende ne kadar çok çaprazlanmış yüzey varsa ayrı ayrı elde edilen varyans bileşen sayısının o kadar artacağı görülmektedir. Ele alınan üç farklı desen ve her bir desenden elde edilen varyans bileşenlerinin sayısı farklı olduğundan elde edilen G ve Phi katsayılarında farklı olduğu sonucuna ulaşılmıştır.

Bu çalışmanın bulgularından yola çıkılarak araştırmacılara; farklı desenler ve farklı değişkenlik kaynakları ele alınarak benzer çalışmaların yapılması; benzer desenler ve benzer değişkenlik kaynakları ele alınarak dengelenmemiş desenlerde benzer çalışmaların yürütülmesi önerilebilir.

KAYNAKLAR

- Aktaş, M., Başusta, B. ve Alıcı, D.(2016). Farklı puanlama anahtarlarının güvenilirliğinin genellenebilirlik kuramı ile incelenmesi. *1-3 Eylül 2016 Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi. Akdeniz Üniversitesi Eğitim Fakültesi, Antalya.*
- Alharby, E. R. (2006). *A comparison between two scoring methods, holistic vs. Analytic using two measurement models, the generalizability theory and the many facet rasch measurement within the context of performance assessment.* (Yayınlanmamış doktora tezi). The Pennsylvania State University.
- Alkan, M. (2013). *Pisa 2009 okuma becerileri açık uçlu sorularının puanlanmasında genellenebilirlik kuramındaki farklı desenlerin karşılaştırılması.* (Doktora tezi) Hacettepe Üniversitesi, Ankara.
- Anıl, D. ve Büyükkıdık, S. (2012). Genellenebilirlik kuramında dört facetli karışık desen kullanımı için örnek bir uygulama. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Kış* , 3(2), 291-296.

- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Atılğan, H. ve Tezbaşaran, A. A. (2005). Genellenebilirlik kuramı alternatif karar çalışmaları ile senaryolar ve gerçek durumlar için elde edilen g ve phi katsayılarının tutarlılığının incelenmesi. *Eurasian Journal Of Educational Research*, 18, 236-252.
- Atılğan, H. (Ed.). (2009). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması*. Ankara: Pegema Yayıncılık.
- Brennan, R. L. (1992). *Elements of generalizability theory (rev. ed)*. Iowa City, IA. American College Testing.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlog.
- Brennan, R. L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied Measurement In Education*, 14(2), 191-203.
- Brennan, R. L. (2011). Generalizability theory and classical test theory. Center for advanced studies in measurement and assessment. *Applied Measurement In Education*, 24: 1-21.
- Brualdi, A. (1998). *Implementing performance assessment in the classroom*. Practical Assessment, Research and Evaluation, 6(2). <http://Pareonline.Net/Getvn.Asp?V=6&N=2>.
- Büyükkıdık, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenilirliğin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması*. (Yüksek Lisans Tezi) Hacettepe Üniversitesi, Ankara.
- Büyükoztürk, Ş. (2007). Performansa dayalı durum belirleme nedir? *İlköğretmen Dergisi*, 8, 28-32.
- Chang, L. and Dennis, H. (2010). Models of generalizability theory in analyzing existing faculty evaluation data. *Applied Measurement In Education*, 13(3), 255-275.
- Chien, Y. (2008). *An investigation of testlet-based item response models with a random facets design in generalizability theory (Doctoral thesis)*. The University Of Iowa.
- Cohen, R. J. and Swerdlik, M. E. (2013). *Psikolojik test ve değerlendirme* (Çev. E. Tavşancıl). Nobel Yayıncılık (7. Basım).
- Crocker, L., Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Javanovich College Publishers, USA.
- Çakıcı, D. (2011). *Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlılığın karşılaştırılması*. (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Doğan, C.D ve Anadol, H.Ö. (2017). Genellenebilirlik kuramında tümüyle çaprazlanmış ve maddelerin puanlayıcılara yuvalandığı desenlerin karşılaştırılması. *Kastamonu Eğitim Dergisi Cilt:25 No: 1*, 361-372.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: TPD Yayınları.
- Fan, X., Sun, S. (2013). Generalizability theory as a unifying framework of measurement reliability in adolescent research. *Journal of Early Adolescence* 2014, Vol 34(1), 38-65
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve rasch modeli üzerine bir araştırma*. (Doktora Tezi). Hacettepe Üniversitesi, Ankara.
- Güler, N, Kaya Uyanık G. ve Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. Pegema Yayıncılık.
- Karasar, N. (1998). *Araştırmalarda rapor hazırlama yöntemi*. Ankara: Pars Matbaacılık
- Kim, Y. (1993). *The use of a 2-parameter graded irt model and generalizability theory in the development and validation of a mathematics performance test*. (Doktora tezi). University of Pittsburgh.
- Moskal, Barbara M., JonA. Leydens (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research and Evaluation*, 7(10). Available Online: <http://Pareonline.Net/Getvn.Asp?V=7&N=10>
- Mushquash, C. and O'Connor, B. P. (2006). Spss and sas programs for generalizability theory analysis. *Behavior Research Methods*. 38 (3), 542-547.
- Nalbantoğlu, F. (2009). *Performans ölçümlerinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

- Nalbantoğlu, F. ve Başusta, B. (2015). Genellenebilirlik kuramıyla dikiş atma ve alma becerileri istasyonu güvenilirliğinin değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6 (1), 107-116.
- Özberk, E.H. (2012). *Genellenebilirlik kuramı karar çalışmalarında kullanılan farklı katsayıların karşılaştırılması*. (Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Shavelson, J. R. and Webb N. M. (1991). *Generalizability theory :a primer*. Newbury Park. CA: Sage Publications.
- Taylor, M. A., Pastor, D. A. (2013). an application of generalizability theory to evaluate the technical quality of an alternate assessment. *Applied Measurement In Education*, 26, 279–297.
- Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi.
- Turgut, M.F., Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Pegem Yayınılık.
- Yin, Y., Shevalson, R. (2004). *Application of generalizability theory to concept-map assessment research*. 26 Mayıs 2016 tarihinde <http://www.Stanford.Edu/Dept/SUSE/SEAL> Adresinden Erişildi.
- Yin, Y. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement In Education*, 21, 273–291.

Extended Abstract

With the reflection of the changes experienced in recent years on education, from the students; They are expected to be individuals who do not memorize knowledge, but apply the acquired knowledge to different life situations, can solve problems, think critically, analyze data and evaluate themselves. For this reason, it is important to include the learning process in the assessment in the measurement of high-level cognitive characteristics, instead of classical assessment approaches. In this context, performance evaluation in education has started to take place more frequently. Performance evaluation is a type of evaluation that aims students to fulfill some complex tasks that require higher level thinking; It has an important place in the field of measurement and evaluation. For this purpose, performance assessment tools can be used in addition to classical assessment tools (short answer, long answer, multiple choice exams, etc.) to evaluate students. Providing reliable scoring in performance evaluation is not as easy as objective tests, but it also contains many errors since it is an observational evaluation. For this reason, when determining reliability, errors from many variability sources and errors that may arise as a result of the interaction of these variability sources should also be taken into account. In this context; Generability (G) Theory, which has a broad conceptual system and strong statistics to determine many measurement problems in behavioral and social sciences (psychology, education); It was proposed by Cronbach, Gleser, Nanda, and Rajaratnam (1972) (Brennan, 1992). In studies based on G theory; A single comprehensive reliability coefficient is calculated in which the error variances arising from the raters, other variability sources (time, test form, item, task, etc.) and the interaction between these variability sources are taken into account (Taylor, 2013; Atılgan, 2004). This research has the characteristics of a descriptive research, since the Generalizability (G) and Phi coefficients of the scores obtained by multi-faceted different design constructions for the measurements obtained in determining the performance based situation in the Generality Theory are examined. The study group of the study consists of 128 12th grade students studying in two different types of central high schools (anatolian high school and science high school) in Mersin in the 2014-2015 academic year. In order to find answers to the research questions, a rater group was formed with the participation of 8 mathematics teachers who were working as high school mathematics teachers and who volunteered to participate in the study. In this study, an open-ended exam consisting of eight geometry questions and a holistic rubric that was prepared for raters to score was used as a data collection tool. The data obtained from the scoring of the 8 geometry performance tasks by the help of the holistic rubric by the 8 raters, the (b:o) xgxp design where the individual (b) is nested on the facet of the school (o), and the task (g) and the rater (p) are designed crossed, the (b:o)x(g:p) design where the individual (b) is nested on the facet of the school (o), the task (g) is nested on the facet of the rater (p) and the (b:o) and the (g:p) are designed crossed and the (b:(p:o))xg design where (p) is nested on the facet of the school (o), the facet of the individual is also nested on the facet of the scorer (p) and the facet of the task (g) is designed crossed. In this study, the variance components and the percentages of the total variance estimation were investigated. Considering all the findings, the sizes of variance components and the place of variance changes in the order differ in terms of the size of the variance components in the designs built on the same data set. This situation also changes the

interpretation of these components. For example, in (b:o)xpxg and (b:o)x(g:p) designs, the fact that the variance change related to the school is small compared to other components is interpreted as the difference due to schools is small; in (b:(p:o))xg designs, the fact that the variance component of the schools is very high was interpreted as the differentiation in terms of geometry skills in schools. In addition, it is seen that the residual variance value in these designs has also changed. The residual variance component estimated for (b:o)xpxg design was found to be quite small compared to (b:o)x(g:p) and (b:(p:o))xg designs. In this context, it has been concluded that the accuracy of the design constructed in the G theory regarding the situation will also affect the reliability and validity of the interpretations to be made. It is seen that the crossed and / or nested designs of the designs also cause differences in the size of the variance components of the main effects and interaction effects. In other words, while the variance percentages of the main effects in crossed designs may be too large or too small, the variance percentages explained by the nested designs may change. This situation is also supported by studies by Chang and Hocevar (2000), Nalbantoğlu (2009), Alkan (2013), Nalbantoğlu and Başusta (2015) and Doğan and Anadol (2017). G and Phi Coefficients obtained from (b:o) xpxg, (b:o)x(g:p) and (b:(p:o))xg designs respectively; 0.84 - 0.78; 0.75 - 0.56; 0.90 - 0.87. Since all of the designs in the study are nested, the variance components of individuals cannot be obtained and interpreted independently of other sources of variability. At the same time, the number of sources of variance they contain are different. When calculating the G and Phi coefficients, the number of variance components in the denominator is different, and the coefficients were found in different sizes. As the number of sources in the denominator increases, the coefficients will be smaller; The coefficients will grow as the number of variance sources decreases. G and Phi coefficients of the first design are higher than the second design; the variance value is much lower than the second design; the lower the G and Phi coefficients of the first design compared to the third design, the greater the number of variance components. According to these inferences, although the ratio of explaining the variance components of the third design seems less, G and Phi coefficients were higher than the other two designs. This is because the variance in the structure of the design is explained by the small number of components. If there is a crossed facet in a design, it will be seen that the number of variance components obtained separately will increase. It was concluded that there are three different designs and variance components obtained from each design and different from G and Phi coefficients. Based on the findings of this study, the researchers; similar studies can be suggested by considering different designs and different sources of variability. In addition, it may be suggested to carry out similar studies in unbalanced designs by considering similar designs and similar variability sources.