

# Vee Diyagramından Elde Edilen Puanların Güvenirliğinin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre İncelenmesi\*

## The Examination of Reliability of Vee Diagrams According to Classical Test Theory and Generalizability Theory

Betül POLAT DEMİR\*\*

### Öz

Araştırmada, vee diyagramından elde edilen puanların güvenirliliğinin incelenmesi amaçlanmıştır. Bu amaçla araştırmada klasik test kuramı ve genellenebilirlik kuramından yararlanılmıştır. Araştırmada betimsel araştırma modeli kullanılmıştır. Çalışma Niğde ilinde öğrenim gören 32, 7. ve 8. sınıf öğrencisi ile yürütülmüştür. Vee diyagramları puanlama anahtarları kullanılarak puanlanmıştır. KTK’da, Cronbach alfa iç tutarlılık güvenilirlik katsayısı; puanlayıcılar arası güvenilirlik için Kendall’in uyum katsayısı (W), G kuramında ise tek değişkenli modelle  $E_p^2$  (G) ve  $\Phi$  (Phi) katsayıları kestirilmiştir. Cronbach alfa güvenilirlik katsayıları birinci puanlayıcı için 0,77, ikinci puanlayıcı için 0,76 ve üçüncü puanlayıcı için 0,76 bulunmuştur. Puanlayıcılar arası uyum ise Kendall’in uyum katsayısı ile belirlenmiş; birinci, ikinci, üçüncü ve dördüncü vee diyagramları için Kendall W katsayısı sırasıyla; 0,959, 0,924, 0,963 ve 0,954 olarak bulunmuştur. Genellenebilirlik kuramı analizinde varyansları ve varyans yüzdelerini hesaplamak amacıyla, tümüyle çaprazlanmış desen uygulanmıştır. G Kuramına göre matematik başarısının ölçülmesiyle vee diyagramından elde edilen puanların Genellenebilirlik (G) katsayısı 0,918 ve Güvenirlik (Phi) katsayısı 0,899 bulunmuştur. G çalışması sonucunda, öğrenci (ö) ana etkisi için kestirilen varyans bileşenin toplam varyansın % 42,4’ünü açıkladığı ve toplam varyans içinde en yüksek değere sahip varyans bileşeni olduğu belirlenmiştir. Puanlayıcı (p) ana etkisi için kestirilen varyans bileşenin ise toplam varyansın %0,1’ini açıkladığı ve bu varyans bileşenin ana etkiler içinde en düşük değere sahip varyans bileşeni olduğu görülmektedir. Elde edilen bulgular, kullanılan ölçme aracının, öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar verdiğini göstermektedir.

*Anahtar Kelimeler:* Vee diyagramı, Güvenirlik, Klasik Test Kuramı, Genellenebilirlik Kuramı

### Abstract

The aim of this study was to investigate reliability of score obtained from vee diagrams. Therefore in this study, the Classical test theory and generalizability theory were used. The study was carried out with descriptive research model. The participants of this research were composed of 32 students at 7<sup>th</sup> and 8<sup>th</sup> grades in Niğde. The responses given to the vee diagrams were scored through the holistic rubric. In CTT, Cronbach’s alpha reliability coefficient and Kendall’s concordance coefficient for interrater reliability; in G theory,  $E_p^2$  (G) and  $\Phi$  (Phi) coefficients with univariate model were computed. The Cronbach alpha reliability coefficients estimated were 0,77, 0,76, and 0,76 for each rater, respectively. For determination of interrater consistency, Kendall’s concordance coefficient was calculated as  $W=0,959$  for the first vee diagram,  $W=0,924$  for the second one,  $W=0,963$  for the third one,  $W=0,954$  for the fourth one. For the purpose of estimating variances and percent of variances in G study, fully crossed design was applied. Generalizability coefficient for vee diagram scores was 0,918 and phi coefficient was 0,899. As a result of the G study, it was determined that the variance component estimated for the student (s) main effect accounted for 42,4% of the total variance, and it was the variance component that had the highest value. The variance component estimated for raters (p) main effect, which is the lowest variance component in all main effect, accounted for 0,1% of the total variance.

\*Bu çalışma IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde (Hacettepe Üniversitesi, 9-13 Haziran 2014) sözlü bildiri olarak sunulmuştur.

\*\* Arş. Gör. Dr., Ömer Halisdemir Üniversitesi, Eğitim Fakültesi, Niğde-Türkiye, betul.polat6006@gmail.com

According to all results, it was seen that vee diagram was reliable for determination of students' mathematics achievement.

*Keywords:* Vee diagram, Reliability, Classical Test Theory, Generalizability Theory

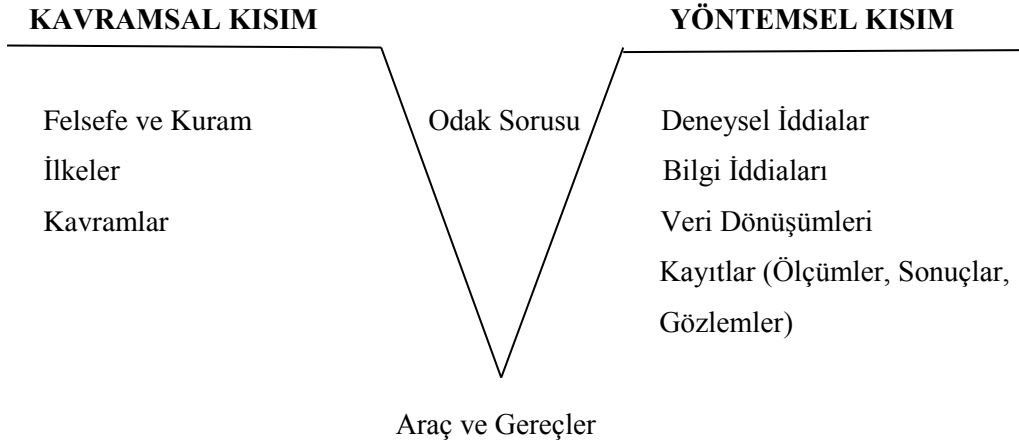
## GİRİŞ

21. yüzyıl teknoloji çağında bilginin öneminin hızla artması ile gelişmiş toplumlarda bireylerin sahip olması gereken özellikler değişmiştir. Günümüzde bilgileri olduğu gibi depolayan bireylere değil; yeni bilgilere doğru ve hızlı bir biçimde ulaşarak, öğrendikleri bilgi ve becerileri yaşamlarının her alanında etkin bir biçimde kullanabilen bireylere ihtiyaç duyulmaya başlanmıştır (Kutlu, Doğan ve Karakaya, 2010). Gereksinim duyulan bireylerin özelliklerinin değişmesi ile eğitim sistemleri yapılandırmacı yaklaşıma göre yeniden düzenlenmiştir. Bu kapsamda, 2005–2006 öğretim yılında ilköğretim 1–5 sınıflarının ve aşamalı olarak 2006–2007 öğretim yılından itibaren ise diğer sınıflar için hazırlanan yeni öğretim programlarında değişiklikler meydana gelmiştir (Gelbal ve Kelecioğlu, 2007).

Yapılandırmacı yaklaşım temele alınarak hazırlanan matematik programında, kavramsal yaklaşım benimsenerek kavramlara ve kavramlar arası ilişkilere ağırlık verilmiştir (Baykul, 2007). Bu bağlamda, öğrencilerin matematiği öğrenebilmeleri için sadece işlemsel bilginin değil kavramsal bilginin de önemli olduğuna vurgu yapılmıştır. Matematiksel anlama, öğrencinin sadece formülleri bilmesine, hesaplamaları doğru yapmasına bağlı değildir; aynı zamanda kavramları ve kavramlar arasındaki ilişkileri, işlemleri anlamasına ve matematiksel düşünmesinin gelişmesine dayanmaktadır. Bu doğrultuda, okul matematiğinde işlemsel çözüm yollarının öğretilmesinin yanında kavram ve bu kavramlar arasındaki ilişkilere önem verilmesinin gerekliliği ortaya konulmuştur (Baki ve Kartal, 2004).

Matematik problemlerinin çözümünde işlemsel bilginin yanında kavramsal bilginin de önemine vurgu yapılmasına rağmen TIMSS 2011 matematik sonuçları, Türkiye'deki 8. sınıf öğrencilerinin bilme düzeyi ortalamasının, uygulama ve muhakeme düzeylerinin ortalamasından daha düşük olduğunu göstermektedir. Bu durum, ülkemizdeki öğrencilerin bilme düzeyinde yeterliliğe sahip olmadan bile uygulama ve muhakeme yapabildiklerinin bir göstergesi olarak ele alınmaktadır. Bilme düzeyinin diğer düzeyler arasında en düşük ortalamaya sahip olması ise ülkemizde matematik öğretiminde temel kavramların öğretimine ve ölçülmesine yeterince ağırlık verilmemesi ile açıklanabilir (Yücel, Karadağ ve Turan, 2013). Bu nedenle ülkemizde matematik öğretiminde işlemsel bilginin yanında kavramsal bilgiyi sınavan ölçme ve değerlendirme araçlarına gereksinim duyulmaktadır.

Kavramsal bilgiyi ölçebilen ölçme ve değerlendirme yaklaşımları, yapılandırmacı yaklaşımın temele alındığı programlarda yerini almıştır. Öğretim ve değerlendirme süreçlerini birbirine yakınlaştıran ve kavramsal bilgiyi de sınavan bu yeni yaklaşımlardan biri de vee diyagramlarıdır. Vee diyagramı, öğrencilerin bilgi yapılarını ve bilgi oluşturma süreçlerini anlamak amacıyla Gowin tarafından geliştirilmiştir (Alvarez ve Risko, 2007). Vee diyagramı "V" şeklinde bir diyagramdır ve iki temel kısımdan oluşur. Vee diyagramının sol tarafı kavramsal kısımdır. Bu kısım düşünme boyutunu içerir ve bu kısımda hipotezler geliştirmek için kullanılan kavramsal ve yapısal bilgi sunulur. Bu kısma kuram ve ilkeler ile kavramlar yazılır. Vee diyagramının sağ tarafı yöntemsel kısımdır. Vee diyagramının sağ kısmı ise uygulama boyutunu içerir ve burada yöntemsel ve işlemsel süreçler yer alır (Novak ve Gowin, 1984). "V" şeklinin ortasında odak sorusu ve şeklin sivri ucunda ise bilgi oluşturma sürecinde ihtiyaç duyulan araç ve gereçler yer alır. (Thiessen, 1993). Vee diyagramının genel gösterimi Şekil 1'de sunulmuştur (Novak ve Gowin, 1984).



Şekil 1. Vee Diyagramı

Vee diyagramı, problemin nasıl çözüldüğünü ya da bilginin nasıl ispat edildiğini, bu süreçte mevcut bilgilerle ihtiyaç duyulan yeni bilgilerin nasıl ilişkilendirildiğini ve bilgilerin nasıl yeniden yapılandırıldığını ortaya koymaktadır (Afamasaga-Fuata’i, 2004a). Bu bağlamda, vee diyagramı öğrencilerin problemleri çözerken var olan kavramsal bilgilerini sistemleştirmelerine, örgütlemelerine ve kullanmalarına yardımcı olmaktadır (Thiessen, 1993). Aynı zamanda vee diyagramı öğrencilerin kavram yanlışlarını belirlemeye ve gidermeye; kuram ve uygulama arasındaki boşluğu doldurmaya ve var olan kavramları organize etmeye de katkı sağlamaktadır. Vee diyagramı öğrenme sürecine sağladığı bu katkılardan dolayı hem öğretim hem de değerlendirme aracı olarak kullanılabilir (Afamasaga-Fuata’i, 2004b).

Matematik probleminin çözümü için gerekli olan işlemsel bilginin yanında kavramsal bilgiyi de sınavan vee diyagramlarından elde edilen sonuçlara dayalı olarak öğrenciler hakkında doğru kararlar verebilmek için bu puanların güvenilir olması gerekmektedir. Güvenirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesidir (Baykul, 2000). Güvenirlik, ölçme aracının ölçtüğü özelliği ya da özellikleri, ne derecede bir kararlılıkta ölçmekte olduğunun bir göstergesidir (Tekin, 2000). Bir ölçme aracıyla yapılan ölçmelerden elde edilen puanların güvenirliliği, ölçmenin üç temel kuramı olan klasik test kuramı (KTK), madde tepki kuramı (MTK) ve genellenebilirlik (G) kuramına dayalı yöntemlerle belirlenmeye çalışılmaktadır. Bu çalışmada KTK ve G kuramına dayalı yöntemler kullanılmıştır.

Güvenirlik, tesadüfi ölçme hatalarından etkilenen psikometrik bir özelliktir. Güvenirliliği etkileyen hataların kaynakları dört başlık altında toplanabilir; a) ölçmenin yapıldığı kişiler, b) ölçmeyi yapan kişi ya da kişiler, c) ölçme ortamı ve d) ölçme aracıdır (Baykul, 2000; Cohen ve Swerdlik, 2009; Crocker ve Algina, 1986; Erkuş, 2003; Turgut ve Baykul, 1992). Eğitim ve psikoloji alanlarındaki gerçek dünyada var olan bazı yapıları ölçme yoluyla kuramsal yapılara bağlanmasına ve bunların açıklanmasına yardımcı olan KTK’da güvenirlilik belirlenirken ölçme hatalarına karışan hatalar, sadece tek bir kaynaktan gelen hatalar olarak ele alınmaktadır. Ele alınan bu varyans (değişkenlik) kaynağına göre de güvenirlilik farklı adlar almakta ve farklı anlamlar taşımaktadır. Bu doğrultuda hata kaynaklarına göre güvenirlilik belirleme yöntemleri de farklılıklar göstermektedir (Baykul, 2000).

Vee diyagramlarıyla yapılan ölçmelerde, öğrencilerin farklı puanlayıcılar tarafından değerlendirmesi sonucunda puanlayıcılar ve vee diyagramları ölçme sonuçlarına karışan farklı hata kaynakları olarak ele alınmıştır. Bu nedenle çalışmada, vee diyagramlarından elde edilen puanların güvenirliliğini belirlemek için KTK’ya dayalı yöntemlerden Cronbach alfa iç tutarlık katsayısı ile puanlayıcılar arası güvenirlilik için Kendall’in uyuşum katsayısı hesaplanmıştır.

Cronbach alfa katsayısı, ölçme aracında bulunan maddelerin ya da görevlerin birbirleri ve testin tamamı ile ilişkisinin bir ölçüsüdür ve “iç tutarlılık” anlamında güvenirliliği vermektedir. Bu

anlamdaki güvenilirlikte hata kaynağı maddeler ya da görevlerdir ve zaman bağlı olarak ortaya çıkabilecek hata kaynağı göz ardı edilmektedir (Brennan, 2001; Crocker ve Algina, 1986; Güler, 2008). İki ya da daha fazla puanlayıcı ya da yargıcının ölçülen aynı özelliğe ilgili farklı bireylere ya da farklı maddelere ilişkin yaptıkları puanlamalar arasındaki tutarlılığın derecesi ise “puanlayıcı güvenilirliği” olarak adlandırılmaktadır ve bu durumda hata kaynağı puanlayıcıdır (Cohen ve Swerdlik, 2009; Güler, 2008). Kendall’ın uyum katsayısı, ikiden fazla puanlayıcının olduğu durumda güvenilirliğin hesaplanmasında kullanılan bir korelasyon katsayısıdır ve parametrik olmayan istatistiksel bir tekniktir (Martin ve Bateson, 1992; akt: Tavşancıl, 2002). Kendall’ın uyum katsayısında, puanlayıcılar arasındaki tutarlılık derecesinin en az 0,80 düzeyinde olması istenmektedir (Szymanski ve Linkowski, 1993).

G kuramı davranış ölçümlerinde; genellenebilirlik (G) ve karar (K) çalışmalarıyla güvenilir gözlemlerin tasarımını, araştırılmasını sağlayan ve temelinde varyans analizine (ANOVA) dayanan istatistiksel bir kuramdır. G kuramı, birçok varyans kaynağını tek bir analizde ele alır ve dolayısıyla her bir varyans kaynağının büyüklüğünün hesaplanmasını sağlar. Genellenebilirlik (G) kuramında KTK’nın aksine aynı anda bütün hata kaynakları dikkate alınarak tek bir hesaplama ile kapsamlı bir güvenilirlik elde edilmektedir (Brennan, 2001; Shavelson ve Webb, 1991).

G kuramında puanlayıcı, zaman, test formu, madde, görev gibi bir ölçme içerisinde yer alabilen bütün hata kaynakları göz önüne tutularak güvenilirlik için genellenebilirlik (G veya  $E\rho^2$ ) ve güvenilirlik (Phi veya  $\Phi$ ) katsayıları elde edilmektedir (Atılğan, 2004). G katsayısı göreceli modellerde ve bağlı değerlendirmeler için kullanılır. G katsayısı, bir bireyin gözlenen puanından, o bireyin evren puanının ne derece doğru genellendiğinin bir göstergesidir. G katsayısı gerçek varyansın, gerçek varyans ile göreceli varyansın toplamına bölünmesiyle hesaplanır. KTK’ndaki güvenilirlik katsayısı gibi G katsayısı da bireylerin puanlarındaki çeşitliliğin oranını yansıtır. Phi katsayısı ise mutlak modellerde ve mutlak değerlendirmeler için kullanılır. Phi katsayısı gerçek varyansın gerçek varyans ile mutlak varyansın toplamına bölünmesiyle hesaplanır (Atılğan, 2005; Güler, 2008). Dolayısıyla, bu iki katsayı hatanın nasıl ele alındığına bağlı olarak değişiklik gösterir (Güler, 2008). Brennan (2001), G ve  $\Phi$  katsayılarının kabul edilebilir yeterlik sınırlarının isteğe bağlı olarak değişebildiğini ancak bazı araştırmacıların G ve  $\Phi$  katsayılarının 0,80’den büyük olduğu durumda güvenilirliğin “yüksek” olarak değerlendirilebileceğini belirtmektedir. Shavelson ve Webb (1991) ise G ve  $\Phi$  katsayılarının en az 0,80 olması gerektiğini vurgulamaktadırlar.

Alanyazın incelendiğinde, vee diyagramlarının bir ölçme aracı olarak güvenilirliğinin ortaya konduğu sadece bir çalışmanın olduğu göze çarpmaktadır. Çalışkan (2014) çalışmasında puanlayıcılar arası güvenilirliği kestirmek amacıyla iki puanlayıcıdan elde edilen puanlar arasındaki korelasyon katsayısını hesaplamıştır. Çalışmada vee diyagramlarının iki puanlayıcı tarafından puanlanması sonucunda 0,87 güvenilirlik katsayısı elde edilmiştir. Bu nedenle işlemsel bilginin yanında kavramsal bilginin öğretiminde ve ölçülmesinde önemli bir yere sahip olan vee diyagramından elde edilen puanların güvenilirliğinin belirlenmesi gerekli görülmüştür.

### ***Araştırmanın Amacı***

Araştırmanın amacı, vee diyagramlarından elde edilen puanların güvenilirliğinin KTK ve G kuramı ile incelemektir. Bu amaç doğrultusunda KTK’ya dayalı yöntemlerden Croanbach alfa iç tutarlık katsayısı ile puanlayıcılar arası güvenilirlik için Kendall’ın uyum katsayısı hesaplanmıştır. G kuramında ise tek değişkenli modelle  $E\rho^2$  (G) ve  $\Phi$  (Phi) katsayıları kestirilmiştir. Bu doğrultuda araştırmada aşağıda verilen araştırma sorularına yanıt aranmıştır.

1. Vee diyagramlarından elde edilen puanlara ait iç tutarlığın derecesi nedir?
2. Vee diyagramlarına ait yapılan puanlamada puanlayıcılar arası tutarlılığın derecesi nedir?
3. Öğrenci, puanlayıcı, vee diyagramı ve bunların etkileşimlerine ait kestirilen varyans bileşenleri nasıldır?
4. Üç puanlayıcının 32 öğrencinin her bir vee diyagramlarına verdikleri yanıtları puanlanmasıyla elde edilen güvenilirlik katsayıları nasıldır?

5. Puanlayıcı sayısının bir artırılıp bir azaltıldığı ve vee diyagramı sayısının 2, 4, 6 ve 8 olduğu durumlarda elde edilen güvenilirlik katsayıları nasıldır?

## YÖNTEM

Bu bölümde araştırmanın modeli, çalışma grubu, veri toplama aracı ve verilerin analizi ile ilgili bilgiler sunulmuştur.

### *Araştırmanın Modeli*

Vee diyagramlarından elde edilen puanların güvenirliğinin klasik test kuramına ve genellenebilirlik kuramına dayalı yöntemlere göre belirlemeyi amaçlayan bu araştırma betimsel araştırma niteliğindedir. Betimsel araştırmalarda amaç, bir olgunun ya da durumun kesin tanımlamalarını ortaya koymaktır (Johnson ve Christensen, 2004).

### *Çalışma Grubu*

Araştırmanın çalışma grubunu Niğde iline bağlı bir devlet ortaokulunun 7. ve 8. sınıfına giden toplam 32 öğrenci oluşturmuştur. Puanlayıcı sayısının üç olmasından dolayı aracın puanlama açısından kullanılabilirliği göz önüne alınarak 32 öğrenci araştırma kapsamına alınmıştır. Araştırmanın amacına daha iyi hizmet edebilecek bir çalışma grubu oluşturmak için matematik başarısı açısından farklı seviyelerde olan farklı sınıf düzeylerinden öğrenciler araştırmaya dâhil edilmiştir. Öğrencilerin tümü vee diyagramları hakkında bilgilendirilmiştir. Öğrencilerin sınıf ve cinsiyetlerine ilişkin dağılımları Tablo 1’de sunulmuştur.

Tablo 1. Öğrencilerin Sınıf ve Cinsiyete Göre Dağılımları

Sınıf	Cinsiyet	N	%	Toplam
7. Sınıf	Kız	9	56,3	16
	Erkek	7	43,7	
8. Sınıf	Kız	10	62,5	16
	Erkek	6	37,5	

### *Veri Toplama Aracı*

Vee diyagramları “Geometri ve Ölçme” alt öğrenme alanlarında yer alan “Doğruda Açılar” ve “Çokgenlerde Açılar” konularında yer alan kazanımlara dayalı olarak hazırlanmıştır. Bu doğrultuda vee diyagramlarının yanıtlama süresi de göz önüne alınarak dört farklı vee diyagramı hazırlanmıştır. Hazırlanan vee diyagramları kazanıma ve öğrenci seviyelerine uygunluğu, ifadelerin anlaşılabilirliği açısından matematik öğretmenin görüşüne sunulmuş ve dönütler doğrultusunda gerekli düzeltmeler yapılmıştır. Öğrencilerin vee diyagramlarına verdikleri yanıtlar vee diyagramı puanlama anahtarıyla puanlanmıştır. Vee diyagramı puanlama anahtarı hazırlanırken Novak ve Gowin (1984) tarafından önerilen puanlama anahtarı dikkate alınmıştır. Vee diyagramı puanlama anahtarı, vee diyagramlarının kavramsal kısmı ile yöntemsel kısmı için iki farklı bütünsel dereceli puanlama anahtarının hazırlanmasıyla oluşturulmuştur. Vee diyagramının her bir kısmı için hazırlanan bütünsel puanlama anahtarları 0 ile 4 arasında derecelendirilmiştir. Öğrencilerin vee diyagramlarından aldıkları puanlar, vee diyagramının her iki kısımdaki öğrenci yanıtlarının ayrı ayrı puanlanması ve puanların toplanmasıyla elde edilmiştir.

### *Verilerin Analizi*

Verilerin analizi iki aşamada gerçekleştirilmiştir. İlk aşamada, üç puanlayıcıdan elde edilen puanlar için açımlayıcı faktör analizi yapılmıştır. Açımlayıcı faktör analizi (AFA) sonucunda, her bir

puanlayıcıdan elde edilen puanlarda öz değeri birden büyük tek bir faktörün olduğu gözlenmiştir. Diğer bir deyişle, her bir puanlayıcıdan elde edilen puanlarda vee diyagramlarının tek bir faktör altında toplandığı ve bu puanlarda benzer sonuçların elde edildiği gözlenmiştir. Bu bulgular vee diyagramlarının ölçmek istenilen tek bir ortak yapıyı ölçtüğünü ortaya koymaktadır. Tablo 2’de AFA sonuçları sunulmuştur.

Tablo 2. Açımlayıcı Faktör Analizi Sonuçları

Vee diyagramı	Faktör yükleri			Ortalama puan
	Puanlayıcı 1	Puanlayıcı 2	Puanlayıcı 3	
VD1	,927	,907	,931	,932
VD2	,830	,842	,844	,852
VD3	,820	,803	,821	,827
VD4	,457	,458	,376	,443
Öz değer	2,429	2,386	2,395	2,474
Açıklanan varyans	60,722	59,652	59,872	61,838

Vee diyagramlarından elde edilen puanların güvenilirliğinin kestirildiği ikinci aşamada, KTK’dan ve G kuramından faydalanılmıştır. KTK kapsamında, Cronbach alfa katsayısı ve Kendall’in uyum katsayısı (W) kestirilmiş ve verilerin analizinde SPSS 13 programından yararlanılmıştır. G kuramına göre, ana ve ortak etkiler için varyans bileşenlerinin kestirilmesinde, puanların güvenilirliği için G ve Phi katsayılarının hesaplanmasında ve K çalışmasında EduG programı kullanılmıştır. Çalışmada farklı ilköğretim okullarından üç matematik öğretmeni puanlayıcı olarak görev almıştır. Vee diyagramı ve puanlanması hakkında öğretmenlere bilgi verilmiştir. Her bir öğrenci tüm vee diyagramlarını yanıtlamış ve tüm öğrencilerin yanıtları üç puanlayıcı tarafından aynı koşullar altında bağımsız olarak puanlanmış olduğu için değişkenlik kaynaklarına ilişkin tümüyle çaprazlanmış (öğrenci x vee diyagramı x puanlayıcı) desen kullanılmıştır.

## BULGULAR VE YORUMLAR

Bu bölümde öncelikle vee diyagramlarından alınan puanlara ait betimsel istatistikler hesaplanmıştır. Puanlara ait betimsel istatistikler Tablo 3’te sunulmuştur.

Tablo 3. Vee Diyagramından Alınan Puanlara İlişkin Betimsel İstatistikler

Vee Diyagramı	Puanlayıcı	$\bar{X}$	Sx	En küçük değer	En büyük değer	Çarpıklık	Basıklık
VD1	P1	2,91	2,44	1,00	8,00	1,024	-,612
	P2	2,78	2,35	1,00	8,00	1,137	-,278
	P3	2,88	2,42	1,00	8,00	1,132	-,351
VD2	P1	1,72	2,23	0,00	7,00	1,194	,303
	P2	2,07	2,15	0,00	7,00	1,153	,192
	P3	2,16	2,34	0,00	7,00	,939	-,557
VD3	P1	2,72	1,58	1,00	7,00	1,299	1,802
	P2	2,91	1,72	1,00	7,00	1,062	,637
	P3	2,88	1,76	1,00	7,00	1,081	,416
VD 4	P1	1,66	1,86	0,00	6,00	,796	-,724
	P2	1,50	1,75	0,00	5,00	,684	-1,163
	P3	1,56	1,78	0,00	6,00	,939	-,225
Toplam	P1	9,00	6,29	2,00	25,00	,976	-,052
	P2	9,25	6,15	2,00	24,00	1,083	,005
	P3	9,47	6,39	2,00	25,00	,937	-,208

Her bir vee diyagramından alınabilecek en düşük puan “0”, en yüksek puan ise “8” dir. Öğrencilerin hem her bir vee diyagramında hem de tamamından aldıkları puanların düşük olduğu ve en düşük ortalamanın dördüncü vee diyagramına ait olduğu görülmektedir. Puanlayıcıların verdikleri puanlara

ilişkin ortalamalar incelendiğinde ise puanlayıcılar arasında Vee Diyagramı 2’de daha çok farklılıkların olduğu gözlenmektedir. Puanlara ilişkin çarpıklık değerleri incelendiğinde ise, puanların sağa çarpık bir dağılım gösterdiği görülmektedir.

Araştırmanın amacı doğrultusunda öncelikle KTK’ya ilişkin güvenilirlik analizi bulguları, daha sonra G Kuramına ilişkin analiz sonuçları aşağıda sunulmuştur.

### **Klasik Test Kuramı**

Vee diyagramından elde edilen puanların güvenilirliğini kestirmek için KTK’ya dayalı yöntemlerden Croanbach alfa iç tutarlık katsayısı ile Kendall’in uyuşum katsayısı hesaplanmıştır. Cronbach alfa katsayısı her bir puanlayıcının öğrencilere verdikleri puanlar için ayrı ayrı hesaplanmıştır. Ayrıca üç puanlayıcının öğrencilere verdikleri puanların ortalaması alınarak Cronbach alfa değeri hesaplanmıştır. Birinci, ikinci ve üçüncü puanlayıcıların vee diyagramlarını değerlendirmeleri ile elde edilen Cronbach alfa katsayıları Tablo 4’te sunulmuştur.

Tablo 4. Cronbach Alfa Güvenirlik Katsayıları

Puanlayıcı	Cronbach alfa
P1	0,77
P2	0,76
P3	0,76
Ortalama puan	0,78

Her bir puanlayıcının vee diyagramlarına verdiği puanlar için hesaplanan Croanbach alfa güvenilirlik katsayıları birinci puanlayıcı için 0,77, ikinci puanlayıcı için 0,76 ve üçüncü puanlayıcı için 0,76 bulunmuştur. Puanlayıcıların öğrencilerin yanıtlarına verdikleri puanların ortalaması alınarak hesaplanan Cronbach alfa güvenilirlik katsayısı ise 0,78 bulunmuştur. Cronbach alfa katsayıları incelendiğinde bu değerlerin birbirine yakın olduğu gözlenmektedir. Nunnally (1978)’ a göre 0,70 ve üzeri güvenilirlik katsayısı kabul edilebilir olarak görülmektedir. Bu doğrultuda elde edilen bu katsayılar göre, vee diyagramlarının öğrenci matematik başarısını birbirleriyle tutarlı bir şekilde ölçtüğü sonucuna ulaşılabilir.

Üç puanlayıcının aynı koşullar altında her bir öğrenciyi bağımsız olarak puanlamasına ilişkin elde edilen puanlar arasındaki tutarlılık derecesi parametrik olmayan bir teknik olan Kendall’in uyuşum katsayısı ile hesaplanmıştır. Bulgular Tablo 5’te sunulmuştur.

Tablo 5. Kendall’in Uyuşum Katsayıları

Vee diyagramı	W	$\chi^2$	p
VD1	0,959	89,222	,000*
VD2	0,924	88,754	,000*
VD3	0,963	89,511	,000*
VD4	0,954	88,754	,000*
Toplam	0.987	91.795	.000*

\*p<0,05

Tablo 5 incelendiğinde dört vee diyagramından elde edilen puanlar için uyuşum katsayısı W=0,987 ( $\chi^2=91,795$ ; p<0,05) bulunmuştur. Birinci vee diyagramı için W=0,959 ( $\chi^2=89,222$ ; p<0,05), ikinci vee diyagramı için W=0,924 ( $\chi^2=85,972$ ; p<0,05), üçüncü vee diyagramı için W=0,963 ( $\chi^2=89,511$ ; p<0,05) ve dördüncü vee diyagramı için W=0,954 ( $\chi^2=88,754$ ; p<0,05) olduğu ve bu katsayıların oldukça yüksek olduğu görülmektedir. Bu bulgulara dayalı olarak her bir puanlayıcının aynı koşullar altında vee diyagramlarını değerlendirmeleri arasında uyumun yüksek olduğu, dolayısıyla

puanlayıcıların öğrencilere verdikleri puanlarda katılık/cömertlik düzeylerinin birbirine benzer olduğu sonucuna ulaşılabilir.

### Genellenebilirlik Kuramı

Üç puanlayıcıdan her birinin öğrencilerin her bir vee diyagramına vermiş oldukları yanıtları değerlendirmeleri doğrultusunda çaprazlanmış (öxvxp) desen ile yapılan G çalışması sonucunda kestirilen varyans bileşenleri ve açıklama yüzdeleri Tablo 6'da verilmiştir. Çaprazlanmış (öxvxp) desen Tablo 6'da da görüldüğü üzere sekiz varyans kaynağına ayrılmaktadır. Her bir varyans kaynağına ait kestirilen varyans bileşenleri aşağıda açıklanmıştır.

Tablo 6. Vee Diyagramlarına İlişkin Puanlamada öxvxp Modeli İçin Kestirilen Varyans Bileşenleri

Varyans kaynağı	Kareler toplamı	sd	Kareler ortalaması	Varyans	%
Öğrenci(ö)	902,372396	31	29,108787	2,567484	42,4
Vee diyagramı (v)	6,320312	3	2,106771	0,000262	0,3
Puanlayıcı (p)	2,255208	2	1,127604	0,000152	0,1
öxv	93,429688	93	1,004620	2,173787	38,3
öxp	17,244792	62	0,278142	0,000178	0,2
vxp	113,515625	6	18,919271	0,000437	0,5
öxvxp,e	554,984375	186	2,983787	1,165230	18,2
Toplam	1690,122396	383			100

G çalışması sonucunda kestirilen varyans bileşenleriyle bu bileşenlerin toplam varyansı açıklama yüzdeleri incelendiğinde, öğrenci (ö) ana etkisi için kestirilen varyans bileşenin toplam varyansın %42,4'ünü açıkladığı görülmektedir. Öğrenciler için kestirilen varyans bileşeni, toplam varyans içinde en yüksek değere sahip varyans bileşenidir. Genellenebilirlik çalışmalarında, öğrenci (yanıtlayıcı) ana etkisi evren puanı varyansı olarak değerlendirilmekte ve ölçülen özellik bakımından öğrenciler arasındaki farklılığı ifade etmektedir (Brennan, 2001). Öğrenciler için kestirilen varyansın toplam varyans içindeki oranının en büyük olması istenilen bir durum olduğundan, araştırma sonucunda elde edilen bu bulgu, vee diyagramlarının ölçülen özellik bakımından öğrenciler arasındaki farklılıkları ortaya çıkarılabildiğini ortaya koymaktadır.

G çalışması sonucunda kestirilen vee diyagramı (v) ana etkisi varyans bileşeni toplam varyansın %0,3'ünü açıklamaktadır. Vee diyagramı ana etkisi varyans bileşeninin büyüklüğünün, ana etkiler içinde ikinci sırada yer aldığı görülmektedir. Bu bulgu, vee diyagramlarının güçlük düzeylerinin birbirinden çok fazla farklılaşmadığını ortaya koymaktadır. Puanlayıcı (p) ana etkisi için kestirilen varyans bileşeninin ise toplam varyansın %0,1'ini açıkladığı ve bu varyans bileşeninin ana etkiler içinde en düşük değere sahip varyans bileşeni olduğu görülmektedir. G çalışması sonucunda kestirilen puanlayıcı ana etkisinin toplam varyans içindeki oranının çok düşük çıkması, puanlayıcıların tüm öğrenciler için yaptıkları puanlamalar arasında büyük farklılıkların bulunmadığını, yapılan puanlamalar arasında bir uyumun olduğunu ortaya koymaktadır.

Öğrenci-vee diyagramı ortak etkisi (öxv) için kestirilen varyans bileşeni, toplam varyans içinde en büyük ikinci değere sahiptir. Bu varyans bileşeni toplam varyansın %38,3'ünü açıklamaktadır. Bu bulgu, öğrencilerin bir vee diyagramından diğerine bağlı konularının değiştiğini ortaya koymaktadır. Öğrenci-puanlayıcı ortak etkisi (öxp) için kestirilen varyans bileşeni ise toplam varyansın %0,2'sini açıklamaktadır. Bu oranın sifra çok yakın olması bir puanlayıcının belli bir öğrenciyi diğer puanlayıcılara göre daha katı/cömert puanlamadıklarını göstermektedir. Vee diyagramı-puanlayıcı ortak etkisi (vxp) varyans bileşeni ise, toplam varyansın %0,5'ini açıklamaktadır. Bu doğrultuda, bu oranın da sifra çok yakın olması puanlayıcıların bir vee diyagramını diğerleriyle aynı kararlılıkta puanladıklarını göstermektedir. Kalan etki varyans bileşeninin (öxvxp) toplam varyansı açıklama oranı %18,2'dir ve toplam varyans içinde en büyük



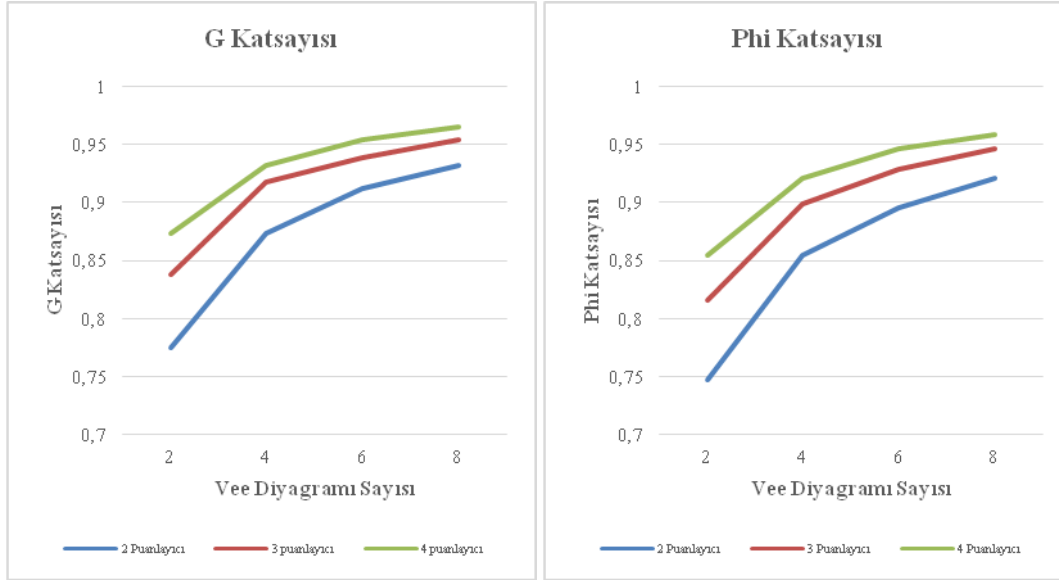
üçüncü değere sahiptir. Bu bulguya göre öğrenci-vee diyagramı-puanlayıcı ortak etkisi ve/veya tesadüfi hataların yüksek olduğu sonucuna ulaşılabilir.

Vee diyagramlarından elde edilen verilere göre orijinal puanlayıcı ve vee diyagramı sayılarına göre G ve Phi( $\Phi$ ) katsayıları kestirilmiştir. Ayrıca puanlayıcı sayısının bir artırılıp bir azaltılması ve vee diyagramı sayılarının iki arttırılıp iki azaltılması senaryolarına göre yapılan karar çalışmaları sonucunda G ve Phi ( $\Phi$ ) katsayıları kestirilmiştir. Bulgular Tablo 7’de sunulmuştur.

Tablo 7. Vee Diyagramlarından Elde Edilen Verilere Göre Kestirilen G ve Phi Katsayıları

Puanlayıcı sayısı	Vee diyagramı sayıları							
	2		4		6		8	
	$E\rho^2$	$\Phi$	$E\rho^2$	$\Phi$	$E\rho^2$	$\Phi$	$E\rho^2$	$\Phi$
2	0,775	0,747	0,873	0,855	0,912	0,895	0,932	0,921
3	0,838	0,816	<b>0,918</b>	<b>0,899</b>	0,939	0,929	0,954	0,947
4	0,873	0,855	0,932	0,921	0,954	0,947	0,965	0,959

Dört vee diyagramının üç puanlayıcı tarafından değerlendirilmesi sonucunda elde edilen puanlardan kestirilen G ( $E\rho^2$ ) katsayısı 0,918; Phi ( $\Phi$ ) katsayısı 0,899 bulunmuştur. Bu değerler incelendiğinde değerlerin birbirine çok yakın ve oldukça yüksek olduğu görülmektedir. Vee diyagramı sayısının sabit tutularak (4) puanlayıcı sayısının bir azaltılması (2) durumunda kestirilen G katsayısı 0,873 ve  $\Phi$  katsayısı 0,855, puanlayıcı sayısının bir arttırılması (4) durumunda ise G katsayısı 0,932 ve  $\Phi$  katsayısı 0,921 olarak kestirilmiştir. Puanlayıcı sayısının sabit tutularak (3) vee diyagramı sayısının iki azaltılması (2) durumunda kestirilen G katsayısı 0,838 ve  $\Phi$  katsayısı 0,816, iki arttırılması (6) durumunda ise G katsayısı 0,939 ve  $\Phi$  katsayısı 0,929 olarak hesaplanmıştır. Tablo 6’dan görüleceği üzere, puanlayıcı sayısının sabit tutularak vee diyagramı sayısının iki artması durumunda, G katsayısındaki artış 0,021 iken, vee diyagramı sayısının sabit tutularak puanlayıcı sayısının bir arttırılması durumunda G katsayısındaki artış 0,014 olmaktadır. Benzer durum,  $\Phi$  katsayısı için de incelendiğinde, puanlayıcı sayısının sabit tutularak vee diyagramı sayısı iki arttırıldığında meydana gelen artış 0,030 iken, vee diyagramı sayısının sabit tutularak puanlayıcı sayısı bir arttırıldığında gözlenen artış 0,021 olduğu gözlenmektedir. Tablo 7 ve Şekil 2 incelendiğinde, vee diyagramı sayısı sabit tutulduğunda puanlayıcı sayısındaki artışın ya da azalmanın G ve  $\Phi$  katsayılarını fazla etkilemediği görülmektedir. Puanlayıcı sayısı sabit tutulduğunda vee diyagramı sayısının arttırılması durumunda da G ve  $\Phi$  katsayılarında daha fazla bir artışın olduğu fakat bu artışın giderek azaldığı gözlenmektedir. Diğer bir deyişle, hem G katsayısı, hem de  $\Phi$  katsayısı için vee diyagramı sayılarını iki arttırmanın, puanlayıcı sayısını bir arttırmaya göre daha büyük artışa sebep olduğu söylenebilir. Dolayısıyla bir ölçme aracının kullanışlılığı dikkate alındığında, puanlayıcı sayısını arttırmanın emek ve zaman açısından ekonomik olmadığı söylenebilir. Vee diyagramları işlemsel bilginin yanında kavramsal bilgiyi de değerlendirildiği için yanıtlama süresinin göz önüne alınarak vee diyagramı sayısının arttırılması ölçme aracının güvenilirliği, kapsam geçerliği ve kullanışlılığı açısından daha uygun olacağı söylenebilir.



Şekil 2. K Çalışması ile Vee Diyagramı ve Puanlayıcı Sayıları Senaryolarına Göre G ve  $\Phi$  Katsayıları

## SONUÇ, TARTIŞMA VE ÖNERİLER

Öğrencilerin matematik başarısının ölçülmesinde kullanılan dört vee diyagramının KTK'ya göre iç tutarlılık düzeyinin belirlenmesi amacıyla Cronbach alfa güvenilirlik katsayısı hesaplanmıştır. Üç puanlayıcı için ayrı ayrı ve tüm puanlayıcıların vee diyagramlarına verdikleri puanların ortalamalarından elde edilen güvenilirlik katsayıları 0,76 ile 0,78 arasına bulunmuştur. Bu katsayılar göre, matematik başarısının ölçülmesinde kullanılan vee diyagramlarının matematik başarısını birbirleriyle tutarlı bir şekilde ölçtüğünü ortaya koymaktadır. Üç puanlayıcı arasında tutarlılığın olup olmadığını belirlemek için KTK'ya göre hesaplanan Kendall'in uyum katsayısı sonuçları puanlayıcıların vee diyagramlarına verilen yanıtları puanlamadaki katılık/cömertlik düzeylerinin birbirine benzer olduğunu göstermektedir.

Araştırmada yer alan dört vee diyagramına tümüyle çaprazlanmış övxp modeli uygulanarak yapılan G çalışması sonucunda öğrenci (ö) ana etkisi için kestirilen varyans bileşeninin toplam varyansın büyük bir kısmını açıkladığı gözlenmektedir. Bu bulgu, vee diyagramlarının öğrenciler arasındaki farklılıkları ortaya çıkarabildiğini ortaya koymaktadır. Puanlayıcı (p) ana etkisi için kestirilen varyans değişkeninin açıklama oranının ise en düşük değere sahip varyans bileşeni olduğu ve sıfıra oldukça yakın olduğu gözlenmektedir. Bu bulgu ise puanlayıcıların yaptıkları puanlamalar arasında bir tutarlılığın olduğunu ortaya koymaktadır. G çalışması sonucunda dört vee diyagramı sayısına ve üç puanlayıcıya göre elde edilen G katsayısı ile  $\Phi$  katsayısı birbirlerine çok yakın ve çok yüksek bulunmuştur. Yapılan K çalışması sonucunda puanlayıcı sayısındaki artış ya da azalma vee diyagramlarındaki artış ya da azalmaya göre G ve  $\Phi$  katsayılarını daha az etkilemektedir. Dolayısıyla zaman ve emek açısından ölçme aracının kullanışlılığı ile aracın kapsam geçerliği dikkate alınarak vee diyagramlarının sayısı artırılarak daha güvenilir ölçmeler yapılabilir.

Yapılan güvenilirlik çalışmaları sonucunda elde edilen tüm bu bulgular, vee diyagramının öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar ortaya koyduğunu göstermektedir. Bu açıdan vee diyagramlarından elde edilen puanlarla öğrenciler hakkında alınan kararların güvenilir olduğu sonucuna varılabilir. Bu bağlamda, öğretmenlere öğrencilerinin hem kavramsal hem de işlemsel bilgilerinin ölçmelerinde vee diyagramlarını kullanmaları önerilmektedir. Araştırmacılar ise matematik dersinden farklı bir disiplinde vee diyagramları hazırlayarak benzer çalışmalar yürüterek sonuçları karşılaştırabilirler.

## KAYNAKÇA

- Afamasaga-Fuata'i, K. (2004a, September). *An undergraduate's understanding of differential equations through concept maps and vee diagrams*. In A. J. Canäs, J. D. Novak & Gonázales (eds). *Concept Maps: Theory, Methodology, Technology*. Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain.
- Afamasaga-Fuata'i, K. (2004b, September). *Concept maps and vee diagrams as tools for learning new mathematics topics*, In A. J. Canäs, J. D. Novak and Gonázales (eds). *Concept Maps: Theory, Methodology, Technology*. Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain.
- Alvarez, M.C. & Risko, V.J. (2007). The use of vee diagrams with third graders as a metacognitive tool for learning science concepts. Retrieved from: <http://e-research.tnstate.edu/pres/5>
- Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- Atılğan, H. (2005). Genellenebilirlik kuramı ve puanlayıcılar arası güvenirlilik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama*, 4(7), 95-108.
- Baki, A., ve Kartal, T. (2004). Kavramsal ve İşlemsel bilgi bağlamında lise öğrencilerinin cebir bilgilerinin değerlendirilmesi. *Türk Eğitim Bilimleri Dergisi*, 2(1), 13-26.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme Klasik Test Teorisi ve uygulaması*. Ankara: ÖSYM Yayınları
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Cohen, J. R., & Swerdlik, E. M. (2009). *Psychological testing and assessment: An Introduction to Tests and Measurement*. (7th Ed.). New York: McGraw-Hill Book Co.
- Crocker, L. M., & Algina, L. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Çalışkan, İ. (2014). The perceptions of pre-service science teachers' about using vee diagrams and electronic portfolios in physics laboratory course. *Educational Research and Reviews*, 9(6), 173-182.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği
- Gelbal S., ve Kelecioğlu H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. NY: Pearson/Allyn & Bacon.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A primer*. Newbury Park, CA: Sage Publications.
- Szymanski, E. M., & Linkowski, D. C. (1993). Human resource development: An examination of perceived training needs of certified rehabilitation counselors. *Rehabilitation Counseling Bulletin*, 37 (2), 163-176.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*, Ankara: Nobel Yayıncılık.
- Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınları
- Turgut, M. F., ve Baykul, Y. (1992). *Ölçme teknikleri*. Ankara: ÖSYM Yayınları.
- Thiessen, R. (1993). The Vee Diagram: A Guide for Problem Solving, *Aims Newsletter*, May/June.
- Yücel, C., Karadağ, E., ve Turan, S. (2013, Şubat). *TIMSS 2011 ulusal ön değerlendirme raporu*. Eskişehir Osmangazi Üniversitesi Eğitim Fakültesi Eğitimde Politika Analizi Raporlar Serisi I, Eskişehir.

## EXTENDED ABSTRACT

### Introduction

With the increase in the importance of knowledge in the technology period of 21<sup>st</sup> century, the characteristics of individuals needed by societies have changed. Today, not the individuals storing information but the ones using the information and skills learnt at school in real life are needed. In this context, education systems have been reconstructed in accordance with the constructivist approach. Accordingly, a variety of changes were made in the curriculums depending on this approach. In the math curriculum, by adapting a conceptual approach, the concepts and the relationships between concepts were emphasized. Moreover, not only the procedural knowledge but also the conceptual knowledge was emphasized in the teaching of math.

Despite the emphasis on the conceptual knowledge along with the procedural knowledge, TIMSS 2011 mathematic results showed that eighth graders' mean of the level of knowing is lower than the mean of the level of applying and reasoning in Turkey. This situation can be explained as not enough importance was given to teaching and measuring of basic concepts in mathematics education in Turkey. Therefore, measurement and evaluation instruments testing the conceptual knowledge along with the procedural knowledge are required in our country.

Measurement and evaluation methods testing the conceptual knowledge took place in these curriculums. One of these methods is Vee diagrams. Vee diagrams reveal how information is proved or how a problem is solved, how the current information and the required new information is associated. Depending on the results obtained from vee diagrams, which test the conceptual knowledge along with the procedural knowledge needed for the solution of a mathematics problem, it can be stated that these scores should be reliable in order to make accurate decisions about students.

When the literature is reviewed, no study was observed about the reliability of scores obtained from vee diagrams. Therefore, it was considered that investigating the reliability of the scores obtained from vee diagram is necessary. In this direction, the reliability of the scores obtained from vee diagram is aimed to predict by computing Cronbach alpha internal consistency coefficients and Kendall's concordance coefficient for interrater reliability based on Classical test theory (CTT) and the  $E\rho^2$  (G) and  $\Phi$  (Phi) coefficients with univariate model based on generalizability (G) theory.

### **Method**

This research aiming at determining the reliability of the scores obtained from vee diagrams based on the techniques of CTT and G theory is a descriptive study. The participants of this research were composed of 32 students at seventh or eighth grades of a state school in Niğde. Four different vee diagrams based on the objectives in the topics of "Angles of a line" and "Angles of a polygon" in "Geometry and Measurement" learning domain were prepared. A mathematic teacher was asked to assess these vee diagrams in terms of students' levels and objectives and the comprehensibility of the statements. The responses given to the vee diagrams were scored through a holistic rubric,

The analysis of the data was conducted in two steps. In the first step, an exploratory factor analysis was conducted on the scores obtained from three raters. As a result of the exploratory factor analysis, it was observed that the scores given by each rater on vee diagrams gathered under one factor. In the second step, reliability of scores obtained from vee diagrams were predicted. SPSS 13 Software was used for estimating the Cronbach Alpha coefficient and Kendall's Concordance Coefficient (W). EduG Software was used for estimating the variance components, G and Phi coefficients, and Decision (D) study. The fully crossed design was used for source of variation since each student responded to all items and three raters rated these responses.

### **Results and Discussion**

The estimated Cronbach alpha coefficients were 0,77, 0,76, and 0,76 for each rater, respectively. Cronbach alpha reliability coefficient was calculated as 0,78 by using the mean scores of grades given by three raters. These values were determined to be close. According to coefficients obtained, vee diagrams were determined to measure the mathematics achievement consistently. The degree of consistency among the scores given by three raters under the same conditions was estimated using Kendall's Concordance coefficient. Kendall's concordance coefficient for four vee diagrams was estimated to be  $W=0,987$ . Kendall's concordance coefficient were estimated as  $W=0,959$  for the first vee diagram,  $W=0,924$  for the second one,  $W=0,963$  for the third one,  $W=0,954$  for the fourth one. These coefficients were observed to be high. Depending on these findings, it can be stated that the consistency among the assessments made by three raters under the same conditions was high.

As a result of the G study, it was determined that the variance component estimated for the student (s) main effect accounted for 42,4% of the total variance, and it was the variance component that had the highest value. This finding indicates that vee diagrams have a capacity to reveal the differences among the students in terms of the characteristics which are measured. The variance component estimated for the vee diagram (v) main effect accounted for 0,3% of the total variance. This is the second among the main effect. This finding shows that the difficulty levels of vee diagrams were similar. The variance component estimated for raters (p) main effect accounted for 0,1% of the total variance. This is the lowest value among the main effects. This finding shows that there was consistency between raters' scores.

The variance component estimated for the student-vee diagram common effect accounted for 38,3% of the total variance. This value is the second highest among the total variance. This finding shows that relative position of the students varied from one vee diagram to the other. The variance component estimated for the student-rater common effect accounted for 0,2% of the total variance. This finding shows that all raters are same level of lenient/severe across all students. The variance component estimated for the vee diagram-rater common effect accounted for 0,5% of the total variance. This rate is close to zero, which shows that the raters graded the vee diagrams with the same stability. The residual effect variance component accounted for 18,2% of the total variance. This is the third highest value in the total variance. This finding shows that student-vee diyagram-rater common effect and/or random error had a high value.

As a result of the scenarios where the number of raters were added and deducted one and the number of vee diagrams were added and deducted two, G and Phi ( $\Phi$ ) coefficients were estimated. Estimated from the assessments of four diagrams by three raters, G ( $E\rho^2$ ) coefficient was determined to be 0,918, and Phi ( $\Phi$ ) coefficient was determined to be 0,899. As a result of the D study, it was observed that an increase or decrease in the number of raters barely had an effect on G and  $\Phi$  coefficients when the number of vee diagrams was fixed. On the other hand, when the number of raters was fixed, an increase in the number of vee diagrams caused a higher increase in G and  $\Phi$  coefficients; however, this increase diminished gradually. In other words, it can be stated that increasing the number of vee diagrams by two causes a bigger increase in both G and  $\Phi$  coefficients compared to increasing the number of raters by one. Therefore, when the practicality of a measurement tool is considered, it can be expressed that increasing the number of raters is not economical in terms of time and effort. Increasing the number of vee diagrams by considering the response time since the vee diagrams assess the conceptual knowledge along with the procedural knowledge is better in terms of reliability, content validity, and practicality.

According to the findings, it can be stated that vee diagrams yield reliable results in assessment of mathematics achievement. In this direction, the teachers are recommended to use vee diagrams to measure the students' both conceptual and procedural knowledge. The researchers can conduct similar studies by preparing vee diagrams for courses other than mathematics and compare the results.