

Bender - Geşalt Görsel Motor Koordinasyon Testi-II'nin (BGT-II) Global, Koppitz ve Geri Çağırma Puanlamaları İçin Yargıcılar Arası Güvenirliğin Değerlendirilmesi*

Inter-rater Reliability of the Bender Visual Motor Gestalt Coordination Test (Second Edition) for Global, Koppitz and Recall Scoring Systems

Sait Uluç¹, İ.Volkan Gülüm², Manolya Çalışır³

¹Yrd.Doç.Dr., ²Psk., ³Uz.Psk., Hacettepe Üniversitesi Psikoloji Bölümü, Ankara

ÖZET

Amaç: Günümüze kadar, beyindeki yapı ve işlev bozukluklarının belirlenmesine yönelik çok sayıda nöropsikolojik test geliştirilmiştir. Geşalt Görsel Motor Koordinasyon Testi (BGT), görsel algısal becerileri ölçmek için en sık kullanılan figür çizme görevlerinden biridir. Bu çalışma kapsamında BGT'nin yenilenmiş biçimi olan Bender Geşalt II'nin yargıcılar arası güvenirliliğinin saptanması ve yeni puanlama sisteminin geleneksel Koppitz puanlama sistemiyle yargıcılar arası güvenilirlik açısından karşılaştırılması hedeflenmiştir. **Yöntem:** Örneklem, 22 erkek ve 23 kız çocuk ve ergenden oluşturulmuştur. Çalışmada, orijinal Bender Geşalt Görsel Motor Koordinasyon Testi'nin yenilenmiş biçimi olan Geşalt Görsel Motor Koordinasyon Testi-II (BGT-II) kullanılmıştır. Protokoller, geleneksel Koppitz puanlama sistemi üzerinden eğitim almış ve alan deneyimine sahip 3 bağımsız yargıcı tarafından değerlendirilmiştir. **Bulgular:** BGT-II'nin Global Puanlama ve Koppitz Puanlamaları için 3 yargıcının puan ortalamaları üzerinden iç tutarlılık katsayıları hesaplanmış ve Global Puanlama setinin iç tutarlılık katsayısının Chronbach alfa = .93; Koppitz Puanlama setinin iç tutarlılık kat sayısının Chronbach alfa = .74 olduğunu görülmüştür. Ek olarak, tüm puanlama için yargıcılar arası uzlaşımın yüksek olduğu belirlenmiştir (Koppitz için ICC = .96, Global için ICC = .85 ve Geri çağırma için ICC = .84). **Sonuç:** Çalışmada, üç puanlama biçiminin de klinik uygulama için gerekli olan iç tutarlılık ve yargıcılar arası güvenilirlik değerlerine sahip oldukları görülmektedir.

Anahtar Sözcükler: Nöropsikolojik test, Bender-Geşalt Görsel Motor Koordinasyon Testi-II, yargıcılar arası güvenilirlik.

(*Klinik Psikiyatri* 2012; 15: 71-79)

SUMMARY

Objectives: In various studies which are made up to the present day, many neuropsychological tests are developed in order to determine structural and functional disorders in the brain. Bender Visual Motor Gestalt Coordination Test (BGT) is one of the most commonly used figure draw task that aimed to evaluate visual and perceptual skills. In that study it is aimed to establish inter-rater reliability of the Bender-Gestalt II, a new form of BGT, and to compare the new scoring system with the traditional Koppitz scoring system in terms of inter-rater reliability. **Method:** Sample is consisted of 22 male and 23 female child and adolescents. A newer form of the original Bender Visual Motor Gestalt Coordination Test (BGT), Gestalt Motor Coordination Test-II (BGT-II) is used in the study. Protocols are evaluated by three independent raters who are trained over Koppitz traditional scoring system and who have field experience. **Results:** For Global Scoring and Koppitz Scoring of BGT-II, internal consistency reliabilities of three raters over mean scores are calculated. It is found that internal consistency reliability for Global Scoring is, Cronbach alpha = .93, for Koppitz Scoring is, Cronbach alpha = .74. Results showed that two scoring systems have sufficiency (Cronbach alpha > .70) for clinic and research purposes. In addition, It is found that inter-rater agreement for all scoring system is very high (for Koppitz ICC = .96; for Global ICC = .85 and for Recall ICC = .84). **Conclusion:** It is seen that three scoring system has sufficient internal consistency and inter-rater reliability values for clinical practice.

Key Words: Neuropsychological test, Bender Visual Motor Gestalt Coordination Test-II, inter-rater reliability.

* Bu çalışma 110K533 nolu proje kapsamında TÜBİTAK tarafından desteklenmiştir.

GİRİŞ

1800'lerde nörologlar, herhangi bir görme bozukluğu olmadığı halde önceden karşılaşmış oldukları nesne ve kişileri anımsayamayan hastalar üzerinde sistematik çalışmalar yapmaya başlamıştır. Bu çalışmaların bir sonucu olarak, görsel bellek ve görsel algıdaki bozulmaların beyindeki ciddi yapı ya da işlev bozukluklarının göstergesi olabileceği kabul edilmiştir. Bu varsayımdan hareketle, beyindeki yapı ve işlev bozukluklarının belirlenmesine yönelik çok sayıda nöropsikolojik test geliştirilmiştir.

Beyin işlevlerindeki bozulmalara yüksek duyarlılıkları nedeniyle şekil çizme görevleri, nöropsikolojik testlerde merkezi bir öneme sahiptir (Lezak 1995). Sözel uyarıların tersine görsel uyarıların kullanıldığı bu tür görevler, hem araştırma hem de tanılama amacıyla tüm dünyada yaygın olarak kullanılmaktadır. Bu tür araçların beyin hasarına duyarlılığı kadar kullanım kolaylıklarına sahip olmaları da yaygın olarak tercih edilmelerine neden olmaktadır (Sivan 1992). Bu araçların en önemli avantajlarından biri, bu tür görevleri uygulamak için gerekli olan sürenin oldukça kısa olmasıdır (Örneğin, Bender Geşalt Testi için ortalama 15 dakika ve Benton Görsel Bellek Testi'nin benzer formları için ortalama 10 dakika gibi). Bir diğer avantaj, bu tür görevlerin özel araç ve donanımlara gerek duymadan uygulanabilmesidir. Uygulama için, genellikle, standart uyarı kartları, kalem, kağıt ve silgi yeterli olmaktadır. Bu tür görevlerin sözel bir bilgi birikimi gerektirmemesi de önemli bir avantajdır ve düşük eğitimli ya da okuma yazma bilmeyen katılımcılara rahatlıkla uygulanabilmektedir. Uygulama kolaylığına rağmen, tüm klinik testlerde olduğu gibi, sonuçların puanlanması ve yorumlanması klinik uzmanlık ve birikim gerektirmektedir. Şekil çizme görevlerinde puanlama, genel olarak şeklin yeniden üretimi (kopyalanması ya da hatırlanması) sırasında katılımcı tarafından yapılan hata sayısı, hata türü ya da tepkinin gecikme süresi üzerinden yapılmaktadır. Elde edilen puanlar norm örnekleminden ilgili yaş dilimi ya da gelişim düzeyi için elde edilmiş olan puanlarla karşılaştırılarak anlamlandırılmaktadır. Bu tür ölçüm araçlarında, uygulamacılar arası tutarlılığı arttırabilmek için iyi yapılandırılmış puanlama rehberleri kullanılmaktadır. Ancak, hata sayısı ve türlerinin belirlenmesinde

uygulamacıların eğitim düzeyleri ve klinik deneyim miktarına bağlı olarak tutarsızlıklar ortaya çıkabilmektedir. Bu nedenle, bu tür araçlar için en iyi güvenilirlik ölçütlerinden birinin 'Yargıcılar-arası Güvenirlik Katsayıları'nın hesaplanması olduğu kabul edilmektedir.

Geşalt Görsel Motor Koordinasyon Testi, diğer adıyla Bender-Geşalt Testi (BGT; Bender 1938) görsel algısal (visuo-perceptual) becerileri ölçmek için en sık kullanılan figür çizme görevlerinden biridir (Piotrowski 1995, Costenbader ve ark. 2000, Cashel 2002). BGT, özgül öğrenme güçlüğüne yanı sıra, bilişsel bozuklukların lateralizasyonunun belirlenmesinde oldukça yararlıdır (Lezak 1995). Ek olarak, kısa süreli bellek ve çalışma belleği kapasitelerinin, görsel algısal becerilerin ve görsel yapılandırma becerilerinin ölçülmesinde bu araca başvurulabilmektedir. Bu çalışma kapsamında, BGT'nin yenilenmiş biçimi olan Bender Geşalt-II'nin yargıcılar arası güvenirliliğinin saptanması ve yeni puanlama sisteminin, geleneksel Koppitz puanlama sistemiyle yargıcılar arası güvenirlilik açısından karşılaştırılması hedeflenmiştir. Aşağıda BGT-II'nin yenilenmiş biçimleriyle ilgili tanımlayıcı bilgiler sunulmuştur.

Bender Geşalt II Testi

Bender Geşalt Testi ilk olarak 1938'de Lauretta Bender tarafından bir araya getirilmiştir. İlgili testte, Wertheimer (1923) tarafından geşalt algı kurallarını örneklemek için oluşturulan 30 figürlük koleksiyondan seçilen 9 figür kullanılmıştır. Orijinal Bender Geşalt Testi 8.5-11 inçlik kartlar üzerine basılı olan ve zorluk derecesi giderek artan bu 9 figürden oluşmaktadır. Aracın standart uygulaması, kartların belirli bir sırayla katılımcıya gösterilerek, kart üzerindeki figürün, A4 boyutundaki kağıda katılımcı tarafından kopyalanması biçimindedir. Bender'in yayınından yıllar sonra çok farklı uygulama ve değerlendirme yöntemleri ortaya çıkmıştır. En yaygın kullanılan yaklaşımlardan biri, katılımcının çizdiği kopyaların görece doğrulukları ve oluşturdukları bütünlük açısından genel olarak puanlanmasıdır. Yaygın olarak kullanılan bir başka puanlama yöntemi de Koppitz'in (1963, 1975) çocuklar için Gelişimsel Puanlama Sistemidir. Koppitz'in sisteminin test-tekrar test güvenirliliği (4 aylık zaman aralığında) düşük

olmasına karşın (.58-.66) değerlendirmeciler arası güvenirliliği mükemmeldir (.88-.96). Araştırmalar zihinsel gerilik ve Bender'de yapılan hata sayısı arasında anlamlı bir ilişki olduğunu göstermiştir (Andert ve ark. 1978). Benzer bir şekilde, birinci sınıf öğrencilerinin okuma ve aritmetik becerileriyle, Bender puanları arasında anlamlı ilişkinin olduğu belirlenmiştir (Ackerman ve ark. 1971). Koppitz (1975) beyin hasarlı çocukların puanlarıyla normal çocukların puanları arasında anlamlı bir fark olduğunu göstermiştir. Koppitz'in sistemi 10 yaşa kadar olan çocuklar için hazırlanmıştır. 10 yaştan sonra Bender puanları ve zeka testleri arasındaki ilişki ortadan kalkmaktadır. Daha önemlisi, 10 yaşın üzerindeki bireylerin çoğu testi hatasız olarak tamamlayabilmektedir. Bender Gestalt Testi'nin ilk sürümü Yalın (1980) ve Somer (1988) tarafından Türkiye'de çalışılmıştır. Ancak, bu çalışmaların sınırlılıkları nedeniyle Amerika normlarının kullanımı yaygın olarak devam etmiştir (Özer 2011). Ülkemizde, yakın zamanda Özer (2007) tarafından kapsamlı bir norm güncelleme çalışması yapılmıştır. 5-12 yaş aralığında 253 katılımcının yer aldığı bu çalışmada klasik dokuz kartlık dizin ve Koppitz gelişimsel puanlama sistemi kullanılmıştır. Araştırmanın sonuçları 5 ve 6 yaş gruplarının Amerika örnekleminde daha yüksek ortalama puanlara sahip olduğuna, diğer yaş gruplarının ise benzer ortalamalar verdiğine işaret etmiştir. Yine Özer (2011) tarafından yapılan ve sosyo-kültürel etkilerin BGT performansı üzerindeki etkilerini değerlendirmeye yönelik bir diğer çalışmada özel okul öğrencilerinin devlet okuluna devam eden yaşlılarına oranla görece yüksek başarımlar gösterdikleri belirlenmiştir. Bu sonuç eğitim kalitesinin BGT performansı üzerinde etkili bir değişken olduğu biçiminde yorumlanmıştır. Ek olarak, BGT performansını diğer şekil çizme görevleriyle karşılaştıran çalışmalar da bulunmaktadır (Özer 2009). Ancak, aracın yargıcılar arası güvenirliliğine ilişkin Türkiye'de yapılmış bir çalışma bulunmamaktadır. Yurt dışında yapılan çalışmalarda, Bender Geşalt Testi'nin ilk sürümüne ilişkin klinik deneyimler ve yıllar içinde oluşan bilgi birikimi temel alınarak yenileme çalışmasına gidilmiştir. Son sürüm olan Bender Geşalt II, gerek uyaran kartları gerekse uygulama ve puanlama işlemleri açısından büyük ölçüde yenilenmiştir. BGT'nin yenileme çalışması disiplinler arası bir bakış açısıyla gerçekleştirilmiştir.

Söz konusu testte, Psikoloji, Nöroloji, Eğitim Bilimleri gibi farklı alanlara sahip uzmanların danışmanlığından yararlanılmış ve yenileme için dört ana hat belirlenmiştir (Brannigan ve Decker 2006). Bu ana hatlar şöyle özetlenebilir:

1. Asıl 9 figürü koruyarak desen sayısının artırılması ve böylece testin yaş aralığının genişletilmesi: Aracın yaş aralığının genişletilebilmesi için güçlük derecelerine göre asıl 9 şeklin önüne ve ardına gelecek şekilde yeni şekiller eklenmiştir. İlk aşamada geniş bir figür havuzu oluşturulmuş ve uzman yargıcılardan bu havuzdaki figürleri hem güçlük derecesi hem de orijinal figürlerle benzerlikleri açısından puanlamaları istenmiştir. Bu değerlendirme sonunda teste eklenebilecek 16 figür belirlenmiştir. İzleyen ön çalışma sonucunda madde analizleri temel alınarak 4 kolay figür (orijinalin önüne gelecek) ve 3 zor figür (orijinalin sonuna gelecek) seçilerek yeni Bender Geşalt II Test dizini oluşturulmuştur. Bender Geşalt II, 2 ayrı alt test içinde toplam 16 figür içermektedir. İlk alt test 8 yaş ve altına uygulanmakta, 9'u orijinal ve 4'ü güçlük seviyesi düşük (yeni) olmak üzere 13 figürü kapsamaktadır. İkinci alt test ise 8 yaş ve üzerine uygulanmakta, 9'u orijinal testten gelen ve 3'ü güçlük seviyesi yüksek (yeni) olmak üzere toplam 12 figür içermektedir. Böylece aracın uygulama yaş aralığı 3 yaş ile 80 yaşı kapsayacak biçimde genişletilmiştir.
2. Önceki formda yer almayan hatırlama görevinin eklenerek bellek süreçlerinin de değerlendirilmesine olanak sağlanması: Pek çok klinik uygulamada bellekle ilişkili bilgilerin büyük önem taşıdığı göz önünde bulundurularak Bender Gestalt II'ye geri çağırma (recall) görevi eklenmiştir. Orijinal Bender ile yapılan çalışmalarda uzmanların standart uygulama işleminin dışına çıkarak, geri çağırma görevleri kullandıkları bilimsel paylaşımlar aracılığıyla bilinen bir durumdur (Brannigan ve Decker 2006). Ancak, standart uygulama yönergesi ve norm çalışmasında bu tür görevlere yer verilmemiş olması, elde edilen sonuçların kişisel değerlendirmelerle sınırlı kalmasına yol açmıştır. Uygulama yönergesi ve norm çalışmasına geri çağırma görevinin eklenmesiyle Bender Geşalt II'de bu eksiklik giderilmiştir.
3. Puanlamaya hata sayısı kadar çizim kalitesinin de

dâhil edilebilmesi: "Global Puanlama Sistemi'nde" her bir şeklin kopyası, çizimin kalitesi temel alınarak, 0 ile 4 arasında değişen 5 puanlık Likert tipi bir ölçek üzerinden değerlendirilmektedir. Puanlama sisteminde her bir katılımcı için üç tür puan elde edilmektedir. İlk aşamada her bir figür için ayrı ayrı puanlama yapılmaktadır. İkinci aşamada tüm ölçek performansı için kopyalama ve geri çağırma puanları hesaplanmaktadır. Farklı puanlama sistemleri, geçerlik, güvenilirlik ve kullanım kolaylığı açısından standardizasyon grubu temelinde birbirleri ile karşılaştırılmışlardır. Bu karşılaştırmalar sonucunda Global Puanlama Sistemi'nin diğer bir çok sisteme göre daha avantajlı olduğu görülmüştür (Brannigan ve Decker 2006).

4. Norm gruplarının zenginleştirilerek güncellenmesi: Bir ölçüm aracının kalitesinin en önemli göstergelerinden biri, yeterince geniş, temsili ve güncel bir norm grubuna sahip olmasıdır. Bu nedenle, yaş ve diğer gelişimsel özelliklere duyarlı bir norm grubunun oluşturulması yenileme çalışmasının en önemli parçası olmuştur. Bender Geşalt II testinde göze çarpan diğer yeniliklerse, pakete yanıtların çoktan seçmeli formunun eklenmiş olması ve ek bir gözlem formunun konmuş olmasıdır. Uygulamacı, bu form aracılığıyla testi alan kişinin duyuşsal ve motor işlevleri, test alma tutumları gibi gözlem bilgilerini kontrol listesi üzerinden rapor edebilmektedir. Ek olarak, uygulamacının figür kopyalama ile ilgili klinik değeri olduğunu düşündüğü bilgileri kaydedebilmesi için de alan ayrılmıştır.

Bu çalışma kapsamında, Bender Geşalt-II'nin yeni kart dizini kullanılarak 9 kartlık geleneksel Koppitz puanlama sistemi, 16 kartlık Global puanlama ve geri çağırma puanlama sistemlerinin yargıcılar arası güvenilirliklerinin saptanması ve karşılaştırılması hedeflenmiştir. Koppitz puanlaması açısından, bu çalışma aracılığıyla kliniklerde yaygın olarak kullanılan bir ölçüm aracının güvenilirliğine ilişkin önemli bir eksikliğin giderileceği düşünülmüştür. Ek olarak, hata sayısı yerine çizim kalitesini temel alan Global puanlama açısından yargıcıların ne ölçüde uyumlu olduğunun belirlenmesinin, bu puan türüne yönelik ileride yapılacak psikometrik çalışmalar ve norm çalışmaları için bir öngörü oluşturabileceği varsayılmıştır.

GEREÇ VE YÖNTEM

Örneklem

Katılımcı Örneklemi: Çalışmanın örnekleme 22'si erkek ve 23'ü kız olmak üzere toplam 45 çocuk ve ergenden oluşturulmuştur. Katılımcıların yaş aralığı 6 yaş 0 ay ile 16 yaş 11 ay arasında değişmektedir. 6 yaş 0 aydan itibaren 12 aylık her bir yaş dilimi 2'si kız 2'si erkek olmak üzere dörder katılımcıyla temsil edilmiştir. Katılımcıların tamamında sağ el baskındır. Katılımcıların daha önceden tanılanmış psikiyatrik bir sorunu ya da fiziksel bir engeli bulunmamaktadır.

Bu çalışmanın katılımcı evrenini 110K383 no'lu ve 'WÇZÖ-IV Türkiye norm Çalışması' başlıklı TÜBİTAK projesinin örneklem havuzu oluşturmaktadır. Öncelikli olarak 110K383 nolu projenin 800 kişilik Ankara örnekleme 11 yaş dilimi ve 2 cinsiyete göre 22 katmana ayrılmıştır. Ardından her bir katmandan 2 denek seçkisiz atama yoluyla belirlenerek bu çalışmanın örnekleme oluşturulmuştur (Not: Bu çalışmanın 110K383 projesi ile katılımcı havuzunun kullanılması dışında herhangi bir ilişkisi bulunmamaktadır. BGT-II uygulamaları 110K383 projesinin veri toplama işlemleri tamamlandıktan yaklaşık 3 ay sonra bağımsız olarak yapılmıştır.).

Yargıcı Örneklemi: Protokoller 3 bağımsız yargıcı tarafından değerlendirilmiştir. Yargıcılar geleneksel Koppitz puanlama sistemi üzerinden eğitim almış ve alan deneyimine sahip, yeterlik aşamasını geçmiş klinik psikoloji doktora öğrencileridir. Yargıcılara Hacettepe Üniversitesi Klinik Psikoloji Doktora Programı eğitiminde kullanılan Koppitz Puanlama Rehberi ve Global Puanlama Rehberi üzerinden kısa bir eğitim verilmiştir. Yargıcılar birbirinin puanlamalarına ve protokollerin sahiplerine ilişkin demografik bilgilere kördür.

Araçlar

Geşalt Görsel Motor Koordinasyon Testi-II (BGT-II): BGT-II, orijinal Bender Geşalt Görsel Motor Koordinasyon Testinin yenilenmiş biçimidir. BGT-II, 2 ayrı alt test içinde toplam 16 figür içermektedir. İlk alt test 8 yaş ve altına uygulanmakta, 9'u orijinal ve 4'ü güçlük seviyesi düşük (yeni) olmak üzere 13 figürü kapsamaktadır. İkinci alt test ise 8 yaş ve üzerine uygulanmakta, 9'u orijinal test-

ten gelen ve 3'ü güçlük seviyesi yüksek (yeni) olmak üzere toplam 12 figür içermektedir. Böylece aracın uygulama yaş aralığı 3 yaş ile 80 yaşı kapsayacak biçimde genişletilmiştir. Araçtan çizim kalitesini temel alan Global Puan ve Uygulama sonunda anımsanan figür sayısını temel alan Geri Çağırma Puanı olmak üzere iki temel puan elde edilmiştir. Global puanlama için yargıcılar arası güvenirliliğin .83 ile .94, geri çağırma puanları için ise .94 ile .97 arasında değiştiği belirlenmiştir (Brannigan ve Deker 2003). Global puanlama için yarı test güvenirliliği .91 ve test tekrar test güvenirliliği .85 olarak hesaplanmıştır (Brannigan ve Deker 2003). Bu çalışmada ek olarak 9 kartlık klasik dizin için Koppitz puanları da hesaplanmıştır. Aracın yenilenme süreci ve puanlamasıyla ilgili daha geniş bilgi giriş bölümünde aktarılmıştır.

İşlem

Katılımcı havuzu örneklem bölümünde aktarıldığı gibi katmanlara ayrıldıktan sonra seçkisiz olarak belirlenen katılımcılara telefonla ulaşılmış ve BGT-II çalışması hakkında bilgi verilerek çalışmaya davet edilmişlerdir. Katılmayı kabul eden gönüllülere evlerinde ulaşılmış, ardından ebeveyn ve çocuk bilgilendirilmiş, onam formları okunarak onayları alınmıştır. BGT-II uygulamaları tek bir oturumda yapılmıştır. Protokoller 3 bağımsız yargıcı tarafından değerlendirilmiştir. Yargıcıların hepsine orijinal protokoller teslim edilmiş ve ilk yargıcının puanlamasının ardından geri alınarak sonraki yargıcıya verilmiştir. Böylece, yanıtların fotokopi ile çoğaltılmasından kaynaklanabilecek çizim kalitesindeki bozulmalar kontrol edilmiştir. Yargıcılar birbirinin puanlamalarına ve protokollerin sahiplerine ilişkin demografik bilgilere kördür. Yargıcı değerlendirmeleri SPSS paket programına girilmiş ve toplu halde analiz edilmiştir.

BULGULAR

Öncelikli olarak BGT-II'nin Global Puanlama ve Koppitz Puanlamaları için 3 yargıcının puan ortalamaları üzerinden iç tutarlılık katsayıları hesaplanmıştır. Analiz sonuçları 16 karttan oluşan Global Puanlama setinin iç tutarlılık katsayısının Chronbach alfa= .93 ve 9 karttan oluşan Koppitz

Tablo 1. BGT-II'nin farklı puanlama yaklaşımları arasındaki ilişkiler

Puan Türü	Geri çağırma (recall)	Global
Global	.38*	
Koppitz	-.33*	-.50**

*p<.05, **p<.01

Puanlama setinin iç tutarlılık katsayısının Chronbach alfa=.74 olduğunu göstermiştir. Analiz sonuçları iki puanlama türünün de klinik ve araştırma amaçlı kullanım için yeterli (Chronbach alfa > .70) olduğuna işaret etmiştir. Global, Koppitz ve geri çağırma puanlamalarının arasındaki ilişkileri belirleyebilmek için Pearson korelasyon katsayıları hesaplanmış ve ulaşılan sonuçlar Tablo 1'de sunulmuştur.

Bu ön hesaplamaların ardından çalışmanın asıl konusu olan yargıcılar arası güvenirliliğe ilişkin hesaplamalar gerçekleştirilmiştir. Sırasıyla (1) toplam puanlar açısından yargıcılar arası uzlaşım katsayıları, (2) Global puanlamanın her bir kartı için yargıcılar arası uzlaşım katsayıları ve (3) Koppitz sistemi için 9 kartın her birindeki toplam hata sayısı açısından yargıcılar arası uzlaşım katsayıları hesaplanmıştır ve ilgili tablolarda sunulmuştur. Tüm yargıcılar arası güvenirlilik katsayıları Sınıf-içi Korelasyon Katsayısı-Mutlak Uzlaşım Yöntemi (ICC) kullanılarak hesaplanmıştır.

Toplam Puanlar Açısından Yargıcılar Arası Uzlaşım

Üç yargıcının toplam puanlar üzerindeki uzlaşım düzeyleri her bir puanlama türü için Sınıf-İçi Korelasyon Katsayısı-Mutlak Uzlaşım Yöntemi kullanılarak hesaplanmıştır. Analiz sonuçları Tablo 2'de sunulmuştur. Buna göre, toplam puanlar için üç puanlama sisteminde de yargıcılar arasında istatistiksel olarak anlamlı düzeyde uzlaşım olduğu görülmektedir. En yüksek uzlaşımın Koppitz puanlama sisteminde (ICC=.96) elde edildiği, Global Puanlama sistemi (ICC=.85) ve Geri Çağırma puanlamalarının (ICC=.84) uzlaşım düzeylerinin oldukça yakın olduğu bulunmuştur.

Tablo 2. Toplam puanlar açısından yargıcılar arası uzlaşım

Puanlama	Cronbach Alpha	Sınıfıçı Korelasyon	95% Güven Aralığı			F Test		
			Alt sınır	Üst sınır	Değer	sd1	sd2	p
Global	.95	.85	.75	.91	20.13	44	88	.000
Koppitz	.99	.96	.94	.98	85.04	44	88	.000
Geri Çağırma	.94	.84	.75	.90	16.63	44	88	.000

Tablo 3. Global puanlamadaki her bir kart açısından yargıcılar arası uzlaşım

Kart	Cronbach Alpha	Sınıfıçı Korelasyon	95% Güven Aralığı			F Test		
			Alt sınır	Üst sınır	Değer	sd1	sd2	p
1	.99	.94	.99	.99	432.86	44	88	.000
2	.99	.99	.98	.99	250.50	44	88	.000
3	.99	.98	.96	.99	123.96	44	88	.000
4	.99	.98	.96	.98	118.42	44	88	.000
5	.94	.84	.74	.90	17.90	44	88	.000
6	.79	.52	.33	.69	4.761	44	88	.000
7	.93	.80	.70	.88	14.43	44	88	.000
8	.94	.84	.74	.90	17.58	44	88	.000
9	.89	.72	.59	.82	9.33	44	88	.000
10	.84	.63	.49	.76	6.33	44	88	.000
11	.89	.73	.60	.83	9.38	44	88	.000
12	.89	.70	.53	.81	9.24	44	88	.000
13	.96	.86	.75	.92	23.06	44	88	.000
14	.97	.92	.87	.95	34.24	44	88	.000
15	.96	.90	.85	.94	28.08	44	88	.000
16	.98	.92	.87	.96	42.94	44	88	.000

Global Puanlamada Yargıcılar Arası Uzlaşım

Global Puanlama sistemi için yargıcıların tüm kartlarda istatistiksel olarak anlamlı bir uzlaşım sergiledikleri görülmüştür. Kartlar arasındaki uzlaşım miktarının ICC=.52 ile ICC=.99 arasında değiştiği belirlenmiştir. Analiz sonuçları Tablo 3'te sunulmuştur.

Koppitz Puanlamadaki Yargıcılar Arası Uzlaşım

Koppitz Puanlama sistemi için de yargıcıların tüm

kartlarda istatistiksel olarak anlamlı bir uzlaşım sergiledikleri görülmüştür. Kartlar arasındaki uzlaşım miktarının ICC=.85 ile ICC=.98 arasında değiştiği belirlenmiştir. Analiz sonuçları Tablo 4'te sunulmuştur.

TARTIŞMA

Şekil çizme görevleri, diğer psikolojik ölçüm araçlarıyla karşılaştırıldığında gerek uygulamacı gerekse katılımcı açısından uygulaması oldukça pratik ölçüm araçlarıdır. Ek olarak, katılımcı

Tablo 4. Koppitz puanlamasındaki her bir kart açısından yargıcılar arası uzlaşım

Kart	Cronbach Alpha	Sınıfçı Korelasyon	95% Güven aralığı		Değer	sd1	F Test	
			Lower Bound	Upper Bound			sd2	p
A	.96	.88	.81	.93	23.08	44	88	.000
1	.94	.85	.76	.91	17.60	44	88	.000
2	.98	.93	.89	.96	40.28	44	88	.000
3	.99	.98	.96	.99	133.30	44	88	.000
4	.97	.92	.87	.95	34.55	44	88	.000
5	.97	.90	.85	.94	29.05	44	88	.000
6	.99	.96	.94	.98	75.10	44	88	.000
7	.96	.87	.79	.92	20.59	44	88	.000
8	.99	.98	.96	.98	126.46	44	88	.000

hakkında güvenilir ve zengin bilgiler sağlayabilmektedir. Kullanım kolaylığının bu tür araçların yaygınlığını arttırmasına karşın temel psikometrik özelliklerinin belirlenmesine yönelik çalışmaların sıklıkla ihmal edilebildiği görülmektedir. Bu çalışma kapsamında, yeni geliştirilen BGT-II'nin en temel psikometrik ölçütlerinden biri olan yargıcılar arası güvenilirlik katsayısı, geleneksel Koppitz sistemiyle karşılaştırmaya olanak sağlayacak biçimde gözden geçirilmiştir. Yargıcılar arası uyum hesaplanırken en sık kullanılan yöntem ikili yargıcı setleri arasındaki ilişkinin uygun korelasyon katsayısı ile hesaplanarak bu ikili setlerin ortalamalarının rapor edilmesidir. Bu yöntemsel yaklaşımla ilişkili en önemli sıkıntı, yargıcı puanlamalarındaki eş zamanlı düşme ve yükselmeleri göz önünde bulundurmasına karşın yargıcıların puanlama eğilimleri açısından ortaya çıkan farkların göz ardı edilmesidir. Bir başka ifadeyle, çok yüksek puanlar verme eğilimindeki bir yargıcı ile çok düşük puanlar verme eğilimindeki bir yargıcının puanları eşgüdümlü olarak değişiyorsa puan farkları ne kadar yüksek olursa olsun yargıcıların uzlaşımı çok yüksek çıkmaktadır. Bu yargıcılar arası uzlaşım hakkında test kullanıcılarını yanıltacak kadar yüksek güvenilirlik katsayılarının rapor edilmesine yol açmaktadır. Bu sorunu gidermek için bu çalışmada daha çağdaş olan, ANOVA temelli ICC yöntemi kullanılmıştır. Böylece, puanlar arasındaki eş zamanlı değişim kadar yargıcıların puanlama eği-

limleri de analize katılabilmektedir. Ek olarak, ICC yönteminin aynı anda ikiden fazla yargıcıyı analize katmaya olanak tanınması analiz sonuçlarının güvenilirliğini arttırmaktadır. Hesaplanan ICC değerlerini ANOVA tablolarında yer alan Sınıf-İç Korelasyon sütunundan izlemek mümkündür. Basit korelasyon ortalamalarına dayanan uzlaşım değerlerini merak eden okuyucular aynı tablolarda yer alan Cronbach Alfa değerlerini kullanabilirler. Koppitz'inki de dahil olmak üzere eski çalışmalarda rapor edilen yargıcı uyumları sıklıkla Cronbach Alfa sütununda yer alan ikili korelasyon ortalamalarıdır. Bu çalışma, Cronbach Alfa değerlerinin ve Sınıf içi korelasyon değerlerinin göz önünde bulundurulmasını gerektirecek kadar önemli farklar gösterdiğine işaret etmiştir. Buna dayanarak, gerek Koppitz gerekse Global puanlamada farklı uzmanların puanlamaları arasındaki uyumun önceden belirtildiği gibi (Koppitz 1975) mükemmel ya da mutlak olmadığı görülmektedir. Bu çalışma sonucu elde edilen bulguların da bu görüşü desteklediği görülmektedir. Önceki çalışmalarda, Global puanlama için yargıcılar arası güvenirlığın .83 ile .94 arasında değiştiği belirlenmiştir (Brannigan ve Deker 2003). Bu çalışmada, basit korelasyon ortalamalarını yansıtan alfa değerinin .95, ICC'ye dayalı uzlaşımın ise .85 olduğu görülmüştür. Buna dayanarak, göz önünde bulundurulması gereken gerçek uzlaşım değerinin ICC=.85 olduğu kabul edilebilir. Önceki çalışmalarda geri çağırma puanları için

yargıcılar arası güvenilirliğin .94 ile .97 arasında değiştiği belirlenmiştir (Brannigan ve Dekker 2003). Bu çalışmada ise alfa değerinin .94, ICC'ye dayalı uzlaşımın ise .84 olduğu görülmüştür. Alfa değeri geçmiş çalışmalarla daha tutarlı görünse de daha güvenilir olan ICC=.84 değerinin temel alınmasının daha yararlı olacağı düşünülmüştür.

Bulgular bölümünde aktarılan hesaplamalar genel olarak değerlendirildiğinde, üç puanlama biçiminin de klinik uygulama için gerekli olan iç tutarlılık ve yargıcılar arası güvenilirlik değerlerine sahip oldukları görülmektedir. İç tutarlılık katsayıları bu çalışma bağlamında incelendiğinde Global puanlama yaklaşımının (.93) Koppitz yaklaşımına göre (.74) daha yüksek bir iç tutarlılığa sahip olduğu görülmektedir. Bu durum, kart sayısındaki artışın iç tutarlılık kat sayısını arttırdığı biçiminde açıklanabileceği gibi, Global puanlama yaklaşımının genel değerlendirme açısından daha yüksek bir bütünlüğe sahip olduğu ve genel çizim kalitesini daha iyi yansıttığı biçiminde de yorumlanabilir. Testin psikometrik açıdan güvenilirliğinin bir diğer göstergesi de yargıcılar arası uzlaşım katsayılarıdır. Bu ölçüt çerçevesinde Koppitz puanlamasının hem toplam puan düzeyinde hem de tek tek kartlar düzeyinde görece olarak Global puanlamadan daha iyi olduğu görülmektedir. Bu durumun, en önemli gerekçesi Global puanlamada genel çizim kalitesine göre puanlama yapılmasına karşın, Koppitz'te puan alacak hata tiplerinin detaylı olarak tanımlanmış olması olarak açıklanabilir. Koppitz'in hata temelli ve detaylandırılmış yönergesi yargıcıların insiyatif kullanma olasılığını azaltmakta ve benzer puanlamalar yapma olasılığını arttırmaktadır. Ek olarak, Global puanlamada hata tür ve sıklığı rapor edilmemesine karşın Koppitz puanlaması bu tür bilgilere ulaşılmasına olanak sağlamaktadır.

SONUÇ

Bu çalışmadan elde edilen sonuçlar çerçevesinde, geleneksel Koppitz tarzı ve yeni geliştirilen Global puanlamanın kendi avantaj ve dezavantajlarına sahip olduğu görülmektedir. Global puanlama çizim kalitesini puanlanabilir hale getirmekte ve toplam değerlendirmede daha yüksek bir bütünlük sergilemektedir. Bu, katılımcının genel değerlendirmesine yönelik (genel gelişim düzeyi ya da okula hazırlıklı olma düzeyi gibi) önemli katkılar sağlama potansiyeline işaret etmektedir. Öte yandan, Koppitz puanlaması hata temelli ve daha detaylı bir sonuç sunmaktadır. Dolayısıyla, katılımcının güçlüğünün daha özgün yönleri hakkında (hata türleri ve belirli işlev bozuklukları arasında ilişkiler kurma gibi) bilgi sağlama potansiyeline sahip görünmektedir. Bu nedenle, test geliştirme grubunun önerdiği gibi Global puanlama temel alınsa bile alan uygulamalarında Koppitz puanlamasının da rapor edilmeye devam edilmesi ve yeni kart dizinleri için hata temelli puanlama ölçütlerinin geliştirilmesi yararlı gözükmektedir.

Bu çalışmada, yeni geliştirilen BGT II'nin güvenilirliğine odaklanılmıştır. Ancak, testin geçerliği de en az güvenilirliği kadar önem taşımaktadır. Bu çalışmada yer verilmeyen geçerlilik olgusunun alternatif testlerle (Benton Görsel Bellek Testi gibi) ya da kriter grupları üzerinden (organik beyin hasarı ya da diğer tanı grupları gibi) yapılacak çalışmalarla gösterilmesi gerekmektedir. Bu güvenilirlik çalışması gibi temel çalışmaların, ileride yapılacak geçerlik çalışmalarına önemli katkılar sağlayabileceği düşünülmektedir.

Yazışma adresi: Dr. Sait Uluç, Hacettepe Üniversitesi Psikoloji Bölümü Psikoterapi Araştırmaları Laboratuvarı, Beytepe, Ankara, psysait@hacettepe.edu.tr

KAYNAKLAR

Ackerman PT, Peters JE, Dykman RA (1971) Children with specific learning disabilities: WISC profiles. *Learn Disabil*, 4:150.

Andert JN, Hustak TL, Dinning WD (1978) Bender-Gestalt reproduction times for retarded adults. *J Clin Psychol*, 34: 927-929.

Bender L (1938) A visual motor gestalt test and its clinical use. American Orthopsychiatric Association, Research Monographs (No. 3), New York, American Orthopsychiatric Association.

Benton AL (1946) A Visual Retention Test for Clinical Use. New York, The Psychological Corporation.

Benton AL (1955) The Visual Retention Test. New York, The Psychological Corporation.

Benton AL (1963) The Revised Visual Retention Test: Clinical and Experimental Applications. 3. Baskı, New York, The Psychological Corporation.

Benton AL (1974) The Revised Visual Retention Test: Clinical

- and experimental applications. 4. Baskı, San Antonio TX, The Psychological Corporation.
- Brannigan GG, Decker SL (2003) Bender-Gestalt II Examiner's Manual. Rolling Meadows, IL, Riverside Publishing.
- Brannigan GG, Decker SL (2006) The Bender-Gestalt II. Am J Orthopsychiatry, 76:10-12.
- Bronner AF (1917) Psychology of Special Abilities and Disabilities. Boston, MA: Little, Brown.
- Cashel ML (2002) Child and adolescent psychological assessment: current clinical practices and the impact of managed care. Prof Psychol Res Pr, 33:446-453.
- Costenbader V, Allison MR, DiFonzo N (2000) Kindergarten screening: a survey of current practice. Psychol Sch, 37:323-332.
- Goldstein DJ, Britt TW (1994) Visual-motor coordination and intelligence as predictors of reading, mathematics and written language ability. Percept Mot Skills, 78:819-823.
- Koppitz EM (1963) The Bender-Gestalt Test for young children. 2. Baskı, New York, Grune & Stratton.
- Koppitz EM (1975) The Bender-Gestalt Test for young children: II Research and Application, 1963-1973. New York, Grune & Stratton.
- Lezak MD (1995) Neuropsychological assessment. 3. Baskı, New York, Oxford University Press.
- Özer S (2007) Turkish children's Bender Gestalt Test performance : A pilot study and preliminary norms. Percep Mot Skills, 105:872-882.
- Özer S (2009) Relationship of bender gestalt developmental scores and human drawing developmental scores in a sample of Turkish preschool children. Sch Psychol Int, 30:137-147.
- Özer S (2011) Turkish children's Bender-Gestalt Test performance: Differences in public and private school children. Psychol Rep, 108:169-181.
- Piotrowski C, Keller JW (1989) Psychological testing in out-patient mental health facilities: a national study. Psychol Sch, 20:423-425.
- Sivan AB (1992) Benton Visual Retention Test: 5. Baskı, (Manual). London, The Psychological Corporation.
- Somer O (1988) Çocuklar için gelişimsel Bender-Gestalt motor testi üzerine bir çalışma Basılmamış Doktora Tezi, Ege Üniversitesi, İzmir.
- Yalın A (1980) Epileptik çocukların tanısında Bender-Geşalt testinin kullanımı. Basılmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.