

# Tek Denekli Araştırmalarda Etki Büyüklüğü Hesaplamada Kullanılan Örtüşmeyen Veriye Dayalı Yöntemlerin İncelenmesi

Şerife BALIKCI<sup>\*1</sup>, Sinan KALKAN<sup>2</sup>, Salih RAKAP<sup>3</sup>, Yusuf AKEMOĞLU<sup>4</sup>

**Öz:** Bilimsel dayanaklı uygulamalar (BDU), okullarda tüm çocukların öğrenme çıktılarına iyileştirmeyi amaçlayan öğretim yöntem ve stratejilerini içermektedir. Son yıllarda, eğitim alanında BDU'ların belirlenmesine yönelik hareket, meta-analiz uygulamalarını yaygınlaştırmış ve tek denekli deneysel araştırma (TDA) gibi alanlarda etki büyüklüğü hesaplama yöntemlerin geliştirilmesine neden olmuştur. Bu çalışmanın amacı, TDA bağlamında etki büyüklüğü hesaplamaları için geliştirilen beş yöntemi (TauÖrtüşmeme, Tau-U, Phi, robust Phi ve İLOF) karşılaştırmalı olarak incelemektir. Çalışma kapsamında TDA'lardan elde edilen 222 grafiğin her biri için beş yöntem kullanılarak etki büyüklüğü hesaplamaları ile görsel analizler gerçekleştirilmiştir. Bu analizler sonucu elde edilen veriler karşılaştırılarak incelenmiştir. Araştırmanın bulguları, incelenen beş yöntemden dördünün (İLOF, Tau-U, TauÖrtüşmeme ve Phi) birbirleriyle çok yüksek düzeyde korelasyona sahip olduğunu ve bu yöntemler kullanılarak hesaplanan etki büyüklüğü değerleri ile görsel analiz sonuçları arasında yüksek düzeyde bir uyum olduğunu ortaya koymuştur. Ayrıca, görsel analiz ile uyumları dikkate alındığından Tau-U ve İLOF yöntemlerinin diğer yöntemlere kıyasla daha iyi performans gösterdiği belirlenmiştir.

**Anahtar Sözcükler:** Etki Büyüklüğü, Örtüşmeyen Veriye Dayalı Yöntem, Tek Denekli Deneysel Araştırma, Tau-U, ÖRVEY, Phi, İLOF

## Investigation of Five Nonoverlap Methods to Calculate Effect Sizes in Single-Case Research

**Abstract:** Evidence-based practices (EBPs) include teaching methods and strategies aimed at improving learning outcomes of all children in schools. In recent years, movement to determine EBPs in education has expanded use of meta-analysis and led to development of effect size calculation methods in single-case experimental research (SCER). The purpose of this study was to examine five nonoverlap methods developed for effect size calculations in SCER. In this study, effect size calculations and visual analyzes were performed using five methods for each of 222 graphs obtained from SCER studies. Findings of the study revealed that four of the five methods (IRD, Tau-U, TauNovlap, and Phi) were highly correlated with each other and there were high levels of agreement between effect size estimates calculated using these methods and visual analysis. Considering their compatibility with visual analysis, it was determined that Tau-U and IRD performed slightly better than other methods.

**Keywords:** Effect Size, Nonoverlap Methods, Single-Case Experimental Research, Tau-U, PND, Phi, IRD

Geliş Tarihi: 29.11.2021

Kabul Tarihi: 11.03.2022

Makale Türü: Araştırma Makalesi

<sup>1</sup> Milli Eğitim Bakanlığı, Samsun, Türkiye; University of North Carolina Greensboro, School of Education, Department of Specialized Education Services, Greensboro, North Carolina, USA, e-posta: [s\\_rakap2@uncg.edu](mailto:s_rakap2@uncg.edu), ORCID: <https://orcid.org/0000-0003-3740-8037>

<sup>2</sup> Çanakkale Onsekiz Mart Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, Çanakkale, Türkiye, e-posta: [snnklkn35@gmail.com](mailto:snnklkn35@gmail.com), ORCID: <https://orcid.org/0000-0001-6890-6421>

<sup>3</sup> Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, Samsun, Türkiye, e-posta: [salih.rakap@omu.edu.tr](mailto:salih.rakap@omu.edu.tr); University of North Carolina Greensboro, School of Education, Department of Specialized Education Services, Greensboro, North Carolina, USA, e-posta: [s\\_rakap@uncg.edu](mailto:s_rakap@uncg.edu), ORCID: <https://orcid.org/0000-0001-7853-3825>

<sup>4</sup> Düzce Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, Düzce, Türkiye, e-posta: [yusufakemoglu@duzce.edu.tr](mailto:yusufakemoglu@duzce.edu.tr), ORCID: <https://orcid.org/0000-0001-6957-0085>

### Atf için/ To cite:

Balıkçı, Ş., Kalkan, S., Rakap, S. ve Akemoğlu, Y. (2022). Tek denekli araştırmalarda etki büyüklüğü hesaplamada kullanılan örtüşmeyen veriye dayalı yöntemlerin incelenmesi. *Yaşadıkça Eğitim*, 36(2), 291-306.

Son yıllarda öğretmenlerin çocuklarla çalışırken ve öğretim sırasında bilimsel dayanaklı uygulamaları kullanmaları tavsiye edilmekle birlikte, bazı ülkelerde bu uygulamaların kullanımı yasal olarak desteklenmektedir. Örneğin, Amerika Birleşik Devletleri'nde (ABD) Hiçbir Çocuk Geride Kalmasın Yasası (No Child Left Behind Act of 2001), Özel Gereksinimli Bireylerin Eğitim Yasası (Individuals with Disabilities Education Act of 2004) ve Her Öğrenci Başarılı Olsun Yasası (Every Student Succeeds Act of 2015) gibi federal eğitim yasaları, bilimsel araştırmalara dayanan etkili yöntemlerin ve öğretim stratejilerinin kullanımını vurgulamakta ve öğretmenlerin bu uygulamaları/stratejileri kullanımını zorunlu kılmaktadır (Yücesoy-Özkan ve diğerleri, 2020). Bilimsel dayanaklı uygulamalar (BDU'lar) olarak da bilinen bu yöntem ve stratejiler, okullarda tüm çocukların öğrenme çıktılarını iyileştirmeyi amaçlamaktadır (Cook ve diğerleri, 2009; Odom ve diğerleri, 2005; Parker ve diğerleri, 2007; Rakap, 2016). Öğretim sırasında BDU'ların kullanılmasına yönelik yasal zorunluluk, farklı alanlarda bu tür uygulamaların belirlenmesine yönelik çalışmaların başlamasına neden olmuştur. Bu bağlamda eğitim ve sosyal bilimler alanlarında BDU'ları belirlemek için ABD'de What Works Clearinghouse, Birleşik Krallık'ta Cochrane Collaboration ve Norveç'te Campbell Collaboration gibi merkezler kurulmuştur (Rakap ve diğerleri, 2014).

Yukarıda bahsi geçen ve benzeri diğer kuruluşlar, belirli uygulamalara odaklanan bir dizi bireysel çalışmadan elde edilen bulguları özetlemek ve belirlenen uygulamalar için kanıtın gücünü değerlendirmek amacıyla genellikle meta-analiz kullanmaktadır. Meta-analiz, karşılaştırılabilir çalışmaların sonuçlarını birleştirmek için tasarlanmış istatistiksel bir teknik olarak tanımlanmaktadır (Büyüköztürk ve diğerleri, 2020). Meta-analizlerin çeşitli çalışmalardan elde edilen sonuçları entegre etmesi ve müdahale etkisinin büyüklüğünü daha güvenilir bir şekilde belirlemesi gibi bazı avantajları bulunmaktadır (Rakap ve diğerleri, 2020). Meta-analizler bunun yanında bağımsız değişkenlerin göreceli etkisinin ve değişkenler arasındaki ilişkilerin gücünün objektif olarak değerlendirilmesine de izin verir. Ayrıca meta-analizlerde, çalışmalar arası varyasyon kontrol edilebilir ve varyasyonu açıklamak için aracı değişkenler dahil edilebilir (Büyüköztürk ve diğerleri, 2020; Van Den Noortgate ve Onghena, 2003).

BDU'ları belirlemek için yürütülen meta analiz çalışmalarının önemli özelliklerinden biri bu çalışmalara grup deneysel araştırma modellerini kullanan çalışmaların dahil edilmesi (örneğin, randomize kontrollü veya yarı deneysel modeller) ve tek denekli deneysel araştırma (TDA) desenlerini kullanan çalışmaların hariç tutulmasıdır. TDA desenlerini kullanan çalışmaların hariç tutulmasının nedeninin temel olarak metodolojik kaygılardan kaynaklandığı düşünülmektedir (Maggin ve diğerleri, 2013). Grup deneysel araştırmalarda meta-analizlerin nasıl gerçekleştirileceğini açıklayan çok sayıda araştırma ile farklı grup deneysel araştırma modellerini kullanan çalışmalar için etki büyüklüklerini hesaplama formülleri bulunmaktadır (Ugille ve diğerleri, 2012). Öte yandan, meta-analiz tekniklerinin TDA desenlerini kullanan çalışmalar için geçerli olup olmadığı konusunda araştırmacılar arasında bir fikir birliğinin de olmadığı görülmektedir (Campbell, 2013; Salzberg ve diğerleri, 1987; Wolery ve diğerleri, 2010). Ayrıca, uzmanlar arasında meta-analitik tekniklerin TDA çalışmaları için uygulanabilir olup olmadığı, uygulanabilir ise etki büyüklüklerinin nasıl hesaplanacağı ve TDA çalışmaları için müdahale etkisinin nasıl özetleneceği konusunda da fikir birliği yoktur (Campbell, 2004; Ugille ve diğerleri, 2012). Alanyazında, TDA çalışmalarından elde edilen verilere yönelik müdahale etkisini hesaplamak için bir dizi parametrik ve parametrik olmayan yöntemler önerilmektedir (Maggin ve diğerleri, 2011; Parker ve diğerleri, 2011).

### **Tek Denekli Araştırma Desenlerinde Görsel Analiz**

TDA desenlerini kullanan araştırmacılar, bir müdahalenin olumlu bir etkiye yol açıp açmadığını değerlendirmek için genel olarak grafikte gösterilen verilerin görsel analizine güvenmekte ve bir müdahalenin çıktılar üzerindeki etkilerini belirlemek için görsel analizin birincil yöntem olması gerektiğini savunmaktadırlar (Baer, 1977; Campbell, 2013; Wolery ve diğerleri, 2010). Görsel analiz kullanımının temelini, (a) araştırmacıların görsel analiz kullanarak önemli kabul edilebilecek kadar büyük herhangi bir müdahalenin etkisini tespit edebilmesi, (b) görsel analizin düşük hata oranı vermesi, (c) müdahalenin etkisini değerlendirebilmenin tutarlı bir yolu olması ve (d) araştırmacıların veri tabanlı kararlar almak için ham verileri sürekli incelemelerine olanak tanıyan dinamik bir süreç olması olmak üzere dört varsayım oluşturmaktadır (Baer, 1977; Parsonson ve Baer, 1978).

Araştırma bulguları, görsel analizin bir müdahale çok etkili olduğunda veya hiç etkili olmadığında güvenilir sonuçlar verdiğini, ancak zayıf ya da orta düzeyde etki gösteren araştırma sonuçları için yeterli güvenilirliğe sahip olmadığını ortaya koymaktadır (DeProspero ve Cohen, 1979; Parker ve Brossart, 2006). Alanyazında, araştırmacıların deneysel etki hakkındaki görsel analize dayalı yargılarının güvenilirliği konusunda şüphe uyandıran önemli kanıtlar bulunmaktadır. Verilerin görsel analizine bağlı yargıların güvenilirliğini araştıran çalışmalar, sıklıkla gözlemciler arası güvenilirlik katsayısını düşük ile orta düzeyde olduğunu rapor etmektedir (DeProspero ve Cohen, 1979; Park ve diğerleri, 1990). Bu nedenle bazı araştırmacılar, kurum ve kuruluşlar (örneğin, American Psychological Association [APA], Institute of Education Sciences [IES]) görsel analiz bulgularını desteklemek amacıyla istatistiksel yöntemlerin (örneğin, etki büyüklüğü hesaplama yöntemleri) kullanılması çağrısında bulunmuştur (Campbell, 2004; Campbell, 2013; Kratochwill ve diğerleri, 2010; Parker ve Brossart, 2006; Scruggs ve Mastropieri, 2013; Ugille ve diğerleri, 2012). Örneğin, ABD’de IES’in araştırma sentezi girişimi olan What Works Clearinghouse, TDA modellerini kullanan çalışmalarının bilimsel dayanaklı uygulamaların belirlenmesine yönelik meta analizlere dahil edilmesini desteklemek amacıyla TDA çalışmalarından elde edilen bulguların değerlendirilmesi ve özetlenmesi için standartlar geliştirmek üzere ulusal bir panel düzenlemiştir (Maggin ve diğerleri, 2013). Panel sonucunda, TDA çalışmalarının sonuçlarını raporlarken ve yorumlarken görsel analizi desteklemek için parametrik olmayan etki büyüklüğü yöntemlerinin yanı sıra regresyon temelli modellerin kullanılması tavsiye edilmiştir (Kratochwill ve diğerleri, 2010). Buna ek olarak, IES, TDA çalışmalarının meta analizleri ile ilişkili metodolojik sorunları araştıran projelere finansal destek sunmuştur (örneğin, Kratochwill ve Levin, 2006; Shadish ve Rindskopf, 2010; Van Den Noortgate ve diğerleri, 2011).

### **Tek Denekli Deneysel Araştırma Desenleri İçin İstatistiksel Yaklaşımlar**

Müdahale etkisini belirlemeye yönelik istatistiksel yaklaşımların görsel analizle birlikte kullanılmasının dört temel avantajı bulunmaktadır. Buna göre istatistiksel yaklaşımlar ilk olarak, güçlü iç geçerliliği olan araştırma desenleriyle eşleştirildiğinde müdahale etkisini nesnel olarak ölçer (Parker ve Brossart, 2003). İkincisi, başlama düzeyi verilerinin kararlılık göstermediği ya da başlama düzeyi ile uygulama düzeyi arasında nispeten küçük bir değişiklik meydana geldiği durumlarda ölçümlemeye hassasiyet sağlarlar (Kazdin, 1982). Üçüncüsü, TDA desenlerini kullanan farklı araştırmaların sonuçlarının birleştirilmesine olanak sağlarlar. Son olarak, benzer çalışmalarda müdahale etkisinin göreceli olarak karşılaştırılmasına olanak tanırırlar (Kazdin, 1982; Parker ve Hagan-Burke, 2007).

Bazı araştırmacılar (örneğin, Allison ve Gorman, 1993; Kratochwill ve Lewin, 1992; Scruggs ve diğerleri, 1987), uzun yıllar önce, TDA desenlerini kullanan çalışmaların meta analizlerini gerçekleştirmek amacıyla bu tür çalışmalardan elde edilen veriler için müdahale etkisini hesaplamaya yönelik yöntemleri tartışsa da TDA çalışmaları için müdahale etkisini hesaplama yöntemlerini araştıran çalışmaların yaygınlığı son 20 yılda önemli bir artış göstermiştir. Alanyazında, 1980’li yıllardan bu yana, TDA çalışmalarından elde edilen veriler temel alınarak müdahale etkisini hesaplamak amacıyla kullanılmak üzere 40’tan fazla yöntemin geliştirildiği, daha güncel araştırmaların, müdahale etkisini hesaplamak için daha gelişmiş yöntemler önerdiği görülmektedir (örneğin, Beretvas ve Chung, 2008; Gage ve Lewis, 2012; Hedges ve diğerleri, 2012; Maggin ve diğerleri, 2011; Parker ve diğerleri, 2007; Parker ve Vannest, 2009; Parker ve diğerleri, 2011; Rakap ve diğerleri, 2020; Ugille ve diğerleri, 2012; Yücesoy-Özkan ve diğerleri, 2020; Van Den Noortgate ve Onghena, 2003). Bu etki büyüklüğü yöntemleri, (a) standartlaştırılmış ortalama farkı (SOF; Hedges ve diğerleri, 2012), (b) parametrik yaklaşımlar (Center ve diğerleri, 1985-86; Maggin ve diğerleri, 2011; van den Noortgate ve Onghena, 2003) ve (c) parametrik olmayan yaklaşımlar (Parker ve diğerleri, 2011; Scruggs ve diğerleri, 1987) olmak üzere üç kategoride sınıflandırılabilir.

Bu yaklaşımların her birinin güçlü ve zayıf yönleri bulunmakla birlikte, doğası gereği TDA çalışmalarından elde edilen veriler de müdahale etkisini belirlemede bazı güçlüklerle neden olmaktadır (Maggin ve diğerleri, 2011). Örneğin, SOF’un kullanımı hesaplanan etki büyüklüğü değerinin grup deneysel araştırmalarda hesaplanan etki büyüklüğü değerlerine benzer olması nedeniyle oldukça kolaydır. Ancak, bu yöntem, TDA çalışmalarından sıklıkla görülen eğilime ve otokorelasyona karşı duyarlı değildir (Maggin ve diğerleri, 2011). Parametrik yaklaşımlar (örneğin, regresyon, hiyerarşik doğrusal model), eğilim ve seviye

dahil olmak üzere çeşitli veri örüntülerinin modellenmesine olanak tanır (Huitema ve McKean, 2000; Manolov ve Solanas, 2013). Ancak, TDA çalışmalarından elde edilen verilerin genellikle parametrik analizlerin yapılması için gerekli olan varsayımları karşılayamaması nedeniyle, varsayım ihlalleri ele alınmadan parametrik yaklaşımların kullanımını uygun değildir (Parker ve Brossart, 2003; Parker ve Vannest, 2009; Wolery ve diğerleri, 2010). Öte yandan, parametrik olmayan yaklaşımlar (örneğin, örtüşmeyen veriye dayalı yöntemler), parametrik istatistik varsayımlarının kısıtlamalarından muaftır ve yorumlanmaları oldukça kolaydır. Bu yöntemler yoluyla etki büyüklüğü değerleri, kalem ve cetvel kullanılarak verilerin sunulduğu grafikler üzerinden kolaylıkla hesaplanabilir (Ma, 2006; Parker ve diğerleri, 2011; Rakap ve diğerleri, 2014). Örtüşmeyen veriye dayalı yöntemler, TDA çalışmalarında başlama düzeyi ve uygulama aşamaları arasındaki seviye değişimini tespit etmek amacıyla tasarlanmıştır. Bu yöntemlerin birçoğu, TDA çalışmalarından elde edilen verilerde sıklıkla görülen eğilimi hesaba katmamaktadır (Parker ve diğerleri, 2011; Rakap ve diğerleri, 2014; Wolery ve diğerleri, 2010).

1987 yılında Scruggs ve arkadaşları tarafından Örtüşmeyen Veri Yüzdesi (ÖRVEY) adıyla ilk yöntemin geliştirilmesinden bu yana, TDA çalışmalarında müdahale etkisini belirlemek amacıyla bir dizi örtüşmeyen veriye dayalı yöntem geliştirilmiştir (Rakap ve diğerleri, 2020). Bu çalışmada, son yıllarda bu amaçla geliştirilen ve alanyazında sıklıkla kullanılan örtüşmeyen veriye dayalı beş yöntem (TauÖrtüşmeme, Tau-U, Phi, robust Phi ve İLOF) karşılaştırmalı olarak incelenmiştir. Bu yöntemlerin karşılaştırılmasının birkaç nedeni bulunmaktadır. Birincisi, bu yöntemler son yıllarda geliştirilmiş ve TDA çalışmalarının meta-analizlerinde kullanılmış olmasına rağmen, TDA çalışmalarından elde edilen verilere uygunluklarını belirlemek için kapsamlı bir araştırmaya tabi tutulmamışlardır. İkincisi, bu yöntemlerle ilgili alanyazında bulunan araştırmalar, simülasyon verileri veya rastgele seçilen TDA çalışmalarından elde edilen verileri kullanarak yöntemleri geliştiren araştırmacılar tarafından yapılmıştır. Üçüncüsü, bu yöntemler daha önce alanyazında önerilen örtüşmeyen veriye dayalı yöntemleri geliştirmek amacıyla tasarlanmış, ancak performansları bu yöntemlerle kapsamlı bir şekilde karşılaştırılmamıştır. Dördüncüsü, bu yöntemlerin bir kısmının hesaplama biçimleri çok benzerdir ve belirli koşullar altında müdahale etkisine dair değerler benzer sonuçlar vermektedir. Bu bağlamda, TDA'larda müdahale etkisinin belirlenmesinde örtüşmeyen veriye dayalı yöntemlerin giderek yaygınlaştığı göz önüne alındığında, diğerlerinden daha üstün olan yöntemlerinin belirlenmesi, uygulayıcıların ve araştırmacıların kullandıkları yöntemler hakkında daha bilinçli kararlar vermelerine yardımcı olacaktır (Parker ve Hagan-Burke, 2007; Parker ve diğerleri, 2009; Parker ve diğerleri, 2011).

Alanyazında bu yöntemlerin diğer parametrik ve parametrik olmayan yöntemlerle karşılaştırıldığı birkaç çalışma yer almaktadır (örneğin, Parker ve Hagan-Burke, 2007; Parker ve diğerleri, 2009; Parker ve diğerleri, 2011; Rakap ve diğerleri, 2014; Rakap ve diğerleri, 2020; Yücesoy-Özkan ve diğerleri, 2020). Bu çalışmalardan elde edilen bulgular özetle şu şekildedir: (a) İLOF'un ÖRVEY, Medyanı Aşan Veri Yüzdesi (MAVY), Pearson r ve Kruskal-Wallis W ile karşılaştırıldığında görsel analiz sonuçları ile uyumu daha yüksektir; (b) İLOF kullanılarak hesaplanan etki büyüklüğü değerleri, ÖRVEY, MAVY, Pearson r ve Kruskal-Wallis W kullanılarak hesaplanan etki büyüklüğü değerleri ile orta - yüksek düzeyde korelasyona sahiptir; (c) ayırt edilebilirlik açısından İLOF, ÖRVEY'den, Pearson r ve Kruskal-Wallis W ise İLOF'den daha üstündür, (d) Tau-U, diğer yöntemler ile kıyaslandığında ayırt edicilik özelliği en iyi olan örtüşmeyen veriye dayalı yöntemdir. Bu çalışmaların bulguları umut verici olmakla birlikte araştırmacılar bu yöntemlerin performanslarının farklı veri setleri kullanılarak karşılaştırmalı olarak değerlendirilmesine olan gereksinimi vurgulamaktadır (Parker ve diğerleri, 2011; Yücesoy-Özkan ve diğerleri, 2020).

Bu çalışmanın amacı, TauÖrtüşmeme, Tau-U, Phi, robust Phi ve İLOF'un performanslarına ilişkin alanyazında yer alan kanıtları (a) ürettikleri müdahale etkisinin büyüklüğü, (b) müdahale etkisinin türüne ilişkin uyuma, (c) birbirleriyle ilişkileri ve (d) görsel analiz ile uyum olmak üzere dört başlık altında karşılaştırmalı olarak inceleyerek genişletmektir. Bu çalışmada dört araştırma sorusu ele alınmıştır:

1. Beş farklı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri nedir?
2. Beş farklı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri müdahale etkisinin türü bağlamında ne ölçüde uyumaktadır?

3. Beş farklı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri arasındaki ilişki nedir?
4. Beş farklı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri görsel analiz sonuçları ile ne ölçüde uyumaktadır?

## Yöntem

### Veri Seti

Bu çalışmada kullanılan A-B grafikleri ( $n = 222$ ), daha önce alanyazında doğal öğretim yaklaşımlarının incelendiği sistematik tarama ve meta-analiz çalışmalarına dahil edilen (Gülboy ve diğerleri, 2021; Rakap ve Parlak-Rakap, 2011; Snyder ve diğerleri, 2015) 38 çalışmadan alınmıştır. Grafiklerin elde edildiği bu çalışmalar 1981 ve 2019 yılları arasında 19 farklı dergide yayınlanmıştır. İncelenen çalışmalarda denekler arası çoklu başlama deseni ( $n = 23$ ), davranışlar arası çoklu başlama deseni ( $n = 9$ ), ortamlar arası çoklu başlama deseni ( $n = 1$ ), dönüşümlü uygulamalar deseni ( $n = 3$ ) ve AB deseni ( $n = 2$ ) olmak üzere farklı TDA desenleri kullanılmıştır. Bir grafiğin bu çalışmaya dahil edilebilmesi için, her birinde en az üç veri noktası olan başlama düzeyi (A) ve uygulama (B) aşamalarına sahip olması gerekmektedir. Bu çalışmaya, dönüşümlü uygulamalar modelini kullanan ya da genelleme veya kalıcılık aşamalarını içeren grafiklerde dahil edilmiştir, ancak bu grafiklerden yalnızca ilk başlama düzeyi ve ilk uygulama aşamalarından elde edilen veriler kullanılmıştır. Bir grafiğin, çıktılar bağlamında birden fazla değişken içermesi durumunda, her değişken için ayrı bir grafik oluşturulmuş ve çalışmaya dahil edilmiştir.

Bu çalışmada kullanılan grafikler, orijinal makalelerden taranmış ya da dijital olarak kopyalanmış ve Ungraph (Biosoft, 2004) programı kullanılarak dijitalleştirilmiştir. Elde edilen veriler, Microsoft™ Excel™ kullanılarak grafiğe dönüştürülmüş ve yeni grafikler, güvenilirliği kontrol etmek amacıyla orijinal grafiklerle karşılaştırılmıştır. Verileri dijitalleştirmek ve grafikleri yeniden oluşturmak için kullanılan prosedürler Rakap ve diğerleri, (2014) tarafından ayrıntılı olarak açıklanmıştır. Bu çalışmada yer alan AB grafikleri ortalama 24 (ranj = 5 - 78) veri noktası içermektedir. Bu bağlamda, başlama düzeyinde ortalama 11 (ranj = 3 - 66), müdahale aşamasında ise 13 (ranj = 3 - 78) veri noktası yer almaktadır.

### Görsel Analiz

Bu çalışmaya dahil edilen 222 A-B grafiğinin görsel analizleri, TDA'larda grafiksel gösterim ve görsel analiz konusunda deneyime sahip iki araştırmacı tarafından gerçekleştirilmiştir. Görsel analizde her bir araştırmacı, "Başlama düzeyi aşaması ile uygulama aşaması karşılaştırıldığında, verilerde uygulanan müdahale lehine bir değişiklik görülüyor mu?" sorusunu "evet" veya "hayır" seçeneklerinden birini kullanarak yanıtlamıştır. İki araştırmacının görsel analiz sonuçlarına dayanarak, her bir grafik üç kategoriden biri altında sınıflandırılmıştır. Bunlar: (a) değişimin olduğu grafikler (diğer bir ifadeyle iki araştırmacının müdahale lehine bir değişimin olduğu konusunda hem fikir olduğu grafikler;  $n = 186$ ), (b) değişimin olmadığı grafikler (diğer bir ifadeyle iki araştırmacının müdahale lehine bir değişimin olmadığı konusunda hem fikir olduğu grafikler;  $n = 26$ ) ve (c) fikir birliği sağlanmayan grafikler (diğer bir ifadeyle iki araştırmacının müdahale lehine bir değişimin olup olmadığı konusunda fikir birliğine varamadıkları grafikler;  $n = 15$ ). Görsel analiz bağlamında değerlendiriciler arası güvenilirlik hesaplaması yapmak amacıyla iki araştırmacı 222 grafiğin tamamını ilk görsel analizden yaklaşık altı ay sonra ayrı ayrı tekrar görsel analiz yoluyla incelemişlerdir. İlk ve ikinci görsel analiz sonuçları arasındaki uyum, her iki araştırmacı için %100 olarak belirlenmiştir.

### Etki Büyüklüğü Değerlerinin Hesaplanması

Beş örtüşmeyen veriye dayalı yöntem kullanılarak 222 grafiğin her biri için ayrı ayrı etki büyüklüğü değerleri hesaplanmıştır. Phi ve robust Phi veri grafikleri kullanılarak elle hesaplanırken, TauÖrtüşmeme, Tau-U ve İLOF etki büyüklüğü hesaplamaları için tasarlanmış bir çevrimiçi hesap makinesi kullanılarak hesaplanmıştır (www.singlecaseresearch.org; Vannest ve diğerleri, 2011). Bu yöntemleri kullanarak etki büyüklüğü hesaplama formülleri Parker ve diğerleri (2011) ile Rakap ve diğerleri (2020) tarafından ayrıntılı olarak açıklanmıştır. Etki büyüklüğü hesaplamaları bağlamında değerlendiriciler arası güvenilirlik değerlendirmesi, TauÖrtüşmeme, Tau-U ve İLOF için 222 grafiğin tamamında, Phi ve robust Phi için ise

grafiklerin %50'sinde ( $n = 111$ ) iki araştırmacının etki büyüklüğü hesaplamalarını gerçekleştirmesi ve elde edilen değerlerin karşılaştırılması yoluyla yapılmıştır. Birinci ve ikinci araştırmacı tarafından hesaplanan etki büyüklüğü değerleri arasındaki korelasyon değeri  $r = .98$  olarak bulunmuştur. Değerlendiriciler arası güvenilirlik katsayısı Tau-U, TauÖrtüşmeme ve İLOF için %100, Phi ve robust Phi için %95 olarak hesaplanmıştır.

### Verilerin Analizi

Bu çalışmada veriler, SPSS 28 istatistik programları kullanılarak analiz edilmiştir. Beş yöntem kullanılarak 222 grafik için etki büyüklüğü değerleri hesaplandıktan sonra, veriler görsel analiz sonuçları ile birlikte bu istatistik programına girilmiştir. Her bir örtüşmeyen veriye dayalı yöntem için ortalama etki büyüklükleri ve standart sapmalar ile beş yöntem arasındaki korelasyonlar hesaplanmıştır. Ayrıca, bu analizler, değişimin olduğu grafikler ( $n = 181$ ), değişimin olmadığı grafikler ( $n = 26$ ) ve fikir birliği sağlanmayan grafikler ( $n = 15$ ) için de ayrı ayrı yapılmıştır.

### Bulgular

#### Etki Büyüklüğü Değer Aralıkları

Bu çalışmaya dahil edilen 222 grafiğin her biri için beş örtüşmeyen veriye dayalı yöntem kullanılarak etki büyüklüğü değerleri hesaplanmıştır. Bu bağlamda, 222 grafik temel alınarak hesaplanan ortalama etki büyüklüğü değerleri %73.7 (robust Phi) ile %79.2 (İLOF) arasında değişmektedir. Görsel analiz sonuçlarında göre değişim olduğu tespit edilen 181 grafik için ortalama etki büyüklüğü değerleri %81 (robust Phi) ile %86.9 (TauÖrtüşmeme) arasında değişirken, değişim olmadığına karar verilen 26 grafik için ortalama etki büyüklüğü değerleri %36.8 (Tau-U) ile %43.1 (Phi) arasında değişmektedir. Değerlendiricilerin fikir birliğine ulaşamadıkları 15 grafik için hesaplanan ortalama etki büyüklüğü değerler ise %49.4 (robust Phi) ile %60.6 (İLOF) arasında değişmektedir. Beş örtüşmeyen veriye dayalı yöntem kullanılarak hesaplanan etki büyüklüğü değerlerine ilişkin betimleyici istatistikler, grafik türlerine göre Tablo 1'de gösterilmektedir.

**Tablo 1.** Grafik Türüne Göre Etki Büyüklüğü Değerlerine İlişkin Tanımlayıcı İstatistikler

Yöntem	Grafikler			
	Değişimin Olduğu ( $n = 181$ )	Değişimin Olmadığı ( $n = 26$ )	Fikir Birliği Sağlanmayan ( $n = 15$ )	Genel ( $n = 222$ )
TauÖrtüşmeme	86.9 (17.6)	38.3 (26.5)	53.9 (28.8)	79.0 (25.9)
Tau-U	86.3 (18.3)	36.8 (24.3)	53.6 (28.3)	78.3 (26.2)
Phi	85.6 (17.7)	43.1 (24.3)	59.5 (20.8)	78.8 (23.7)
Robust Phi	81.0 (26.7)	37.5 (28.9)	49.4 (29.1)	73.7 (31.0)
İLOF	86.7 (16.8)	37.8 (25.0)	60.6 (19.4)	79.2 (24.4)

#### Etki Büyüklüğü Değerlerinin Sınıflandırılması

Beş yöntem kullanılarak hesaplanan etki büyüklüğü değerleri Parker ve diğerleri (2009) tarafından belirlenen kriterlere göre “etkili” veya “etkisiz” olarak sınıflandırılmıştır. Bu kriterlere göre, %50’den büyük etki büyüklüğü değerleri etkili müdahaleye işaret ederken, %50 veya daha düşük etki büyüklüğü değerleri etkisiz bir müdahaleye işaret etmektedir. Bulgular, 222 grafiğin 188’inde (%85), beş örtüşmeyen veriye dayalı yöntem kullanılarak hesaplanan etki büyüklüğü değerlerinin müdahale etkisinin türü bağlamında uyumlu olduğunu göstermiştir. Bu bağlamda, en az dört yöntem 205 grafikte (%92) uyumluluk göstermiştir. Görsel analiz sonucunda bir değişimin var olduğu tespit edilen 181 grafik bağlamında, 170 grafikte (%94) en az üç yöntem, 167 grafikte (%92) en az dört yöntem ve 155 grafikte (%86) beş yöntemin tümü tarafından hesaplanan etki büyüklüğü değerlerinin etkili müdahale aralığında (%50’den yüksek) olduğu belirlenmiştir. Ek olarak, bu 181 grafikten 6’sı için beş yöntem tarafından hesaplanan etki büyüklüğü değerlerinin etkisiz müdahale aralığında (%50 ve daha düşük) olduğu belirlenmiştir.

Görsel analiz sonucunda bir değişimin var olmadığı tespit edilen 26 grafik bağlamında, 19 grafikte (%73) en az üç yöntem, 14 grafikte (%54) en az dört yöntem ve 13 grafikte (%50) beş yöntemin tümü tarafından hesaplanan etki büyüklüğü değerlerinin etkisiz müdahale aralığında (%50 ve daha düşük) olduğu tespit

edilmiştir. Bunun yanında, 26 grafiğin 5'i için beş yöntem kullanılarak hesaplanan etki büyüklüğü değerlerinin etkili müdahale aralığında (%50'den yüksek) olduğu görülmüştür. Görsel analiz sonucunda değerlendiriciler arası fikir birliğinin sağlanamadığı 15 grafik bağlamında, beş yöntemin tamamı tarafından hesaplanan etki büyüklüğü değerleri 5 grafikte %50'den yüksek, 5 grafikte ise %50 ve daha düşük olarak belirlenmiştir. Kalan 5 grafik için ise beş yöntem tarafından hesaplanan etki büyüklüğü değerleri bir uyum göstermemiştir.

### Korelasyon

Tablo 2, 222 grafik temel alınarak beş örtüşmeyen veriye dayalı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri arasındaki korelasyona ilişkin katsayıları göstermektedir. Genel olarak, Pearson momentler çarpımı korelasyon katsayıları .84 (Tau-U ve robust Phi) ile .98 (TauÖrtüşmeme ve Tau-U) arasında değişmektedir. Görsel analiz sonucunda bir değişimin var olduğu tespit edilen 181 grafik (ranj = .82 - .98), bir değişimin var olmadığı tespit edilen 26 grafik (ranj = .66 - .98) ve fikir birliğinin sağlanamadığı 15 grafik (ranj = .79 - .99) temel alınarak beş örtüşmeyen veriye dayalı yöntem kullanılarak hesaplanan etki büyüklüğü değerleri arasında da orta ile yüksek düzeyde korelasyon olduğu belirlenmiştir. Korelasyon matrislerinin taranması, beş örtüşmeyen veriye dayalı yöntemden dördünün (TauÖrtüşmeme, Tau-U, Phi ve İLOF) tüm grafik türlerinde yüksek korelasyonlara sahip olduğunu ve en az ilişkili yöntemin ise robust Phi olduğunu göstermektedir.

**Tablo 2.** Beş Örtüşmeyen Veriye Dayalı Yöntem için Korelasyon Matrisi (N = 222)

Yöntem	TauÖrtüşmeme	Tau-U	Phi	Robust Phi	İLOF
TauÖrtüşmeme	1				
Tau-U	.978	1			
Phi	.923	.901	1		
Robust Phi	.861	.840	.964	1	
İLOF	.951	.932	.958	.876	1

Parker ve diğerleri (2009), yeni etki büyüklüğü yöntemlerinin incelendiği çalışmalarda, bu yöntemlerin alanda yaygın bir şekilde kullanılan etki büyüklüğü yöntemleriyle karşılaştırılmasını önermektedir. Bu bağlamda, bu çalışmada incelenen beş örtüşmeyen veriye dayalı yöntem, TDA desenlerini kullanan araştırmalarda müdahale etkisini belirlemek için en yaygın olarak kullanılan ve ilk etki büyüklüğü hesaplama yöntemi olarak kabul edilen Örtüşmeyen Veri Yüzdesi (ÖRVEY) ile karşılaştırılmıştır. İncelenen her bir örtüşmeyen veriye dayalı yöntem ile ÖRVEY arasındaki korelasyonlar grafik türüne göre Tablo 3'te gösterilmektedir. Grafiklerin tümü göz önüne alındığında, ÖRVEY ile en yüksek korelasyona İLOF (.87), en düşük korelasyona Robust Phi (.75) yöntemlerinin sahip olduğu görülmektedir. Benzer şekilde, görsel analiz sonucunda bir değişimin var olduğu ya da olmadığı tespit edilen grafiklerde ÖRVEY ile İLOF (.83 ve .66) en yüksek korelasyona, Robust Phi (.71 ve .42) ise en düşük korelasyona sahiptir. Görsel analiz sonucunda bir değişimin var olup olmadığına ilişkin fikir birliği sağlanamayan grafiklerde ise ÖRVEY ile Phi (.66) en yüksek, TauÖrtüşmeme (.39) ise en düşük korelasyona sahiptir.

**Tablo 3.** Grafik Türüne Göre ÖRVEY'in TauÖrtüşmeme, Tau-U, Phi, Robust Phi ve İLOF ile Korelasyonu

Grafik Türü	N	TauÖrtüşmeme	Tau-U	Phi	Robust Phi	İLOF
Değişimin Olduğu	181	.74	.79	.78	.71	.83
Değişimin Olmadığı	26	.59	.64	.58	.42	.66
Fikir Birliği Sağlanmayan	15	.39	.40	.66	.57	.61
Genel	222	.82	.84	.84	.75	.87

### Görsel Analiz ile Uyuşma

Görsel analiz sonucunda bir değişimin var olduğu tespit edilen 181 grafik için İLOF kullanılarak hesaplanan etki büyüklüğü değerlerinin %95'i, Tau-U ve Phi kullanılarak hesaplanan etki büyüklüğü değerlerinin %94.5'i, TauÖrtüşmeme kullanılarak hesaplanan etki büyüklüğü değerlerinin %93.4'ü ve Robust Phi kullanılarak hesaplanan etki büyüklüğü değerlerinin ise %87.3'ü, Parker ve diğerleri (2009) tarafından belirtilen kriterler göre etkili uygulama kategorisinde yer almıştır. Görsel analiz sonucunda bir değişimin var olmadığı tespit edilen 26 grafik için TauÖrtüşmeme ve Tau-U kullanılarak hesaplanan etki büyüklüğü değerlerinin %65.4'ü, İLOF kullanılarak hesaplanan etki büyüklüğü değerlerinin %57.7'si, Robust Phi kullanılarak

hesaplanan etki büyüklüğü değerlerinin %53.9'u ve Phi kullanılarak hesaplanan etki büyüklüğü değerlerinin ise %50'si Parker ve diğerleri (2009) tarafından belirtilen kriterler göre etkisiz uygulama kategorisinde yer almıştır.

Beş yöntemden dördü (Tau-U, İLOF, TauÖrtüşmeme ve Phi) kullanılarak hesaplanan etki büyüklüğü değerleri bir değişimin var olduğu tespit edilen grafiklerde görsel analiz sonuçları ile çok yüksek düzeyde uyuşurken, bir değişimin var olmadığı tespit edilen grafiklerde görsel analiz sonuçları ile etki büyüklüğü değerleri arasındaki uyuşma oldukça düşüktür (ranj =%50 - %65.4). Bu bağlamda, her yöntem için toplam uyumsuzluk yüzdeleri hesaplanmış ve Tablo 4'te gösterilmiştir. Görsel analiz sonuçları ile en az uyumsuzluk gösteren yöntem Tau-U (%9.2), en yüksek uyumsuzluk gösteren yöntem ise Robust Phi (%16.9) olarak bulunmuştur. Görsel analiz sonuçları ile uyumsuzluk düzeyleri İLOF için %9.7, TauÖrtüşmeme için %10.1 ve Phi için ise %11.1 olarak hesaplanmıştır. Görsel analiz sonucunda değerlendiriciler arası fikir birliğinin sağlanamadığı 15 grafik bağlamında, İLOF değerlerinin %73'ü, Phi değerlerinin %60'ı, Tau-U ve TauÖrtüşmeme değerlerinin ise %53'ü etkili olarak sınıflandırılmıştır. Bunun aksine, Robust Phi değerlerinin %53'ü etkisiz olarak sınıflandırılmıştır.

**Tablo 4.** Görsel Analiz Sonuçları ile Örtüşmeyen Veriye Dayalı Yöntemler Kullanılarak Hesaplanan Etki Büyüklüğü Değerleri Arasındaki Uyuşma

Yöntem	Grafikler Analiz (N = 207)		Toplam Uyuşma		Toplam Uyuşmazlık	
	Etkili (N = 181)	Etkisiz (N = 26)	N (N = 207)	%	N (N = 207)	%
TauÖrtüşmeme	169	17	186	89.9	21	10.1
Tau-U	171	17	188	90.8	19	9.2
Phi	171	13	184	88.9	23	11.1
robust Phi	158	14	172	83.1	35	16.9
İLOF	172	15	187	90.3	20	9.7

## Tartışma

Bu çalışmada, TDA çalışmaları bağlamında etki büyüklüğü hesaplamaları için kullanılmak üzere son dönemlerde geliştirilmiş beş farklı örtüşmeyen veriye dayalı yöntem karşılaştırmalı olarak incelenmiş ve bu yöntemlerin nasıl ve ne zaman kullanılabileceğine yönelik önerilerde bulunulmuştur. İlerleyen bölümde, çalışmanın ana bulguları araştırma soruları temelinde tartışılmış; bu çalışmanın ve ilişkili diğer çalışmaların bulguları ışığında örtüşmeyen veriye dayalı yöntemlerin kullanımına yönelik önerilerde bulunulmuş ve çalışmanın sınırlılıkları ile ileriki araştırmalar için önerilere yer verilmiştir.

İlk araştırma sorusu, her bir yöntem kullanılarak elde edilen etki büyüklüğü değerlerinin büyüklüklerinin karşılaştırılmasına odaklanmıştır. Bu çalışmada yer alan 222 grafik bağlamında, İLOF, TauÖrtüşmeme, Tau-U ve Phi kullanılarak hesaplanan ortalama etki büyüklüğü değerleri (ranj = 78.3 – 79.2) birbirine benzer nitelikteyken, robust Phi kullanılarak hesaplanan etki büyüklüğü değerleri ortalaması ise diğer yöntemlerden yaklaşık %4 daha düşük çıkmıştır. Benzer şekilde, değişimin olduğu tespit edilen grafikler (n = 181) kullanılarak hesaplanan etki büyüklüğü değerleri bağlamında, İLOF, TauÖrtüşmeme, Tau-U ve Phi ortalamaları benzerlik gösterirken, robust Phi için hesaplanan ortalama etki büyüklüğü değeri, diğer yöntemlerden yaklaşık %4 daha düşüktür. Değişimin olmadığı grafikler (n = 26) bağlamında ise Phi yöntemi, diğer yöntemler tarafından hesaplanan etki büyüklüğü değerleri ortalamalarından daha yüksek bir ortalama değere sahiptir. Değerlendiricilerin müdahale etkisinin varlığı konusunda fikir birliği sağlayamadıkları grafikler (n = 15) bağlamında, ortalama İLOF ve Phi etki büyüklüğü değerleri diğer yöntemlere nispeten daha yüksekken, ortalama robust Phi etki büyüklüğü değeri diğer yöntemlere göre daha düşük bulunmuştur.

Etkili müdahale rapor eden grafikler kullanılarak hesaplanan etki büyüklüğü değerlerinin %50'den büyük olması beklenirken, etkisiz müdahale rapor eden grafikler kullanılarak hesaplanan etki büyüklüğü değerlerinin %50 veya daha düşük olması beklenmektedir. Beş yöntemin tümü kullanılarak hesaplanan ortalama etki büyüklüğü değerlerinin, veri örüntülerinde değişimin olduğu ve değişimin olmadığı grafikler



bağlamında bu aralıklar içinde yer aldığı görülmektedir. Değerlendiricilerin müdahalenin etkililiği hakkında fikir birliği sağlayamadıkları grafikler için etki büyüklüğü değerlerinin %40 ile %60 arasında olması beklenmektedir (Yücesoy-Özkan ve diğerleri, 2020). Bu bağlamda, İLOF ve Phi için ortalama etki büyüklüğü değerleri %60 veya biraz üzerindeyken, diğer yöntemler kullanılarak hesaplanan ortalama etki büyüklüğü değerlerinin %49 ile %54 arasında olduğu görülmüştür.

İLOF, Tau-U ve TauÖrtüşmeme, son zamanlarda birçok bireysel TDA çalışmasında ya da meta-analiz çalışmasında etki büyüklüğü rapor etmek için kullanılmıştır (örneğin, Bouck ve diğerleri, 2021; Klingbeil ve diğerleri, 2019; Pyle ve Fabiano, 2017). Alanyazında yer alan bu çalışmalar tarafından rapor edilen etki büyüklüğü değerleri ile bu çalışma kapsamında elde edilen ortalama etki büyüklüğü değerlerinin karşılaştırılması, bazı farklılıklar olmasına rağmen, ortalama etki büyüklüğü değerlerinin kabaca karşılaştırılabilir nitelikte olduğunu göstermektedir. Dahası, bu çalışma kapsamında, İLOF, Tau-U ve TauÖrtüşmeme yöntemleri kullanılarak elde edilen ortalama etki büyüklüğü yüzdeleri Yücesoy-Özkan ve diğerleri (2020) tarafından bu yöntemler için rapor edilen etki büyüklüğü değerleri (%80 - %83) ile benzerlik göstermektedir.

İkinci araştırma sorusu, beş farklı yöntem kullanılarak hesaplanan etki büyüklüğü değerlerinin müdahale etkisinin türü bağlamında uyuma düzeylerine odaklanmıştır. Genel olarak, grafiklerin %85'inde beş örtüşmeyen veriye dayalı yöntemin tamamı değişimin olduğu ya da olmadığı konusunda aynı sonucu vermiştir. Bunun yanında, grafiklerin %92'sinde en az dört etki büyüklüğü hesaplama yöntemi, %100'ünde ise en az üç yöntem değişimin olduğu ya da olmadığı konusunda uyum göstermiştir. Dört etki büyüklüğü hesaplama yönteminin uyduğu durumların 13'ünde hatanın robust Phi kullanılarak hesaplanan etki büyüklüğü değerinden kaynaklandığı 1'inde ise Tau-U kullanılarak hesaplanan etki büyüklüğü değerinden kaynaklandığı belirlenmiştir. Üç etki büyüklüğü hesaplama yönteminin uyduğu durumlarda ise hataların TauÖrtüşmeme - Tau-U (N = 7), Phi - robust Phi (N = 7) ve Phi - İLOF (N = 3) çiftlerinden kaynaklı olduğu tespit edilmiştir. Bu bulgular, İLOF kullanılarak hesaplanan etki büyüklüklerinin hata oranının diğer yöntemlere kıyasla az da olsa daha düşük olduğunu göstermektedir. İLOF ile ilgili bu bulgu, alanyazında yer alan ve İLOF'ün diğer etki büyüklüğü hesaplama yöntemleri ile karşılaştırıldığı çalışmaların bulguları ile benzerlik göstermektedir. Örneğin, Yücesoy-Özkan ve diğerleri (2020), İLOF'ün hata oranının bu çalışma kapsamında incelenen dört yönteme kıyasla daha düşük olduğunu rapor etmektedir.

Üçüncü araştırma sorusu, beş yöntem kullanılarak elde edilen etki büyüklüğü değerlerinin birbiriyle olan ilişkilerine odaklanmıştır. Bu çalışmada, 222 grafik temel alınarak, dört örtüşmeyen veriye dayalı yöntem (TauÖrtüşmeme, Tau-U, Phi ve İLOF) kullanılarak hesaplanan etki büyüklüğü değerleri arasında çok yüksek düzeyde korelasyonlar (her bir çift arasındaki korelasyon >.90) olduğu görülmüştür. Ayrıca, bu dört yöntemin, alanyazında önerilen ilk örtüşmeyen veriye dayalı yöntem olan ÖRVEY ile yüksek düzeyde korelasyona (.82 - .87) sahip olduğu belirlenmiştir. Bu çalışmanın bulguları, Yücesoy-Özkan ve diğerleri (2020) tarafından yürütülen ve bu dört yöntem tarafından hesaplanan etki büyüklüğü değerleri arasında yüksek korelasyon (.71 - .90) rapor eden çalışmanın bulgularıyla örtüşmektedir. Yücesoy-Özkan ve diğerleri (2020) tarafından rapor edilen ÖRVEY ile TauÖrtüşmeme, Tau-U, Phi ve İLOF yöntemleri arasındaki korelasyon değerleri (.46 - .83) bu çalışmada rapor edilen bulgulara kıyasla daha düşüktür.

Bu çalışmada, Phi hariç diğer yöntemler ile robust Phi arasındaki korelasyonun .88'dan düşük olduğu tespit edilirken Phi ile robust Phi arasındaki korelasyon .96 olarak belirlenmiştir. En yüksek düzeyde korelasyon ise TauÖrtüşmeme - Tau-U (.98) ile İLOF - Phi (.96) yöntemleri arasında gözlenmiştir. Değişimin olduğu ve olmadığı grafikler ile değerlendiricilerin fikir birliğine ulaşamadıkları grafikler ayrı ayrı incelendiğinde de benzer korelasyon örüntüleri gözlemlenmiştir. Etki büyüklüğü değerlerinin hesaplanması bağlamında TauÖrtüşmeme - Tau-U ve İLOF - Phi çiftleri arasındaki benzerlikler dikkate alındığında, bu çiftler arasında yüksek korelasyonun olması beklenen bir sonuçtur. Esasen, TauÖrtüşmeme ve Tau-U, başlama düzeyinde terapötik yönde bir eğilim olmadığında aynı etki büyüklüğü değeri ile sonuçlanmaktadır. Benzer şekilde, bir grafik için 2 x 2 tablosu dengeli marjinal değerlere sahip olduğunda, İLOF kullanılarak hesaplanan etki büyüklüğü değeri Phi kullanılarak hesaplanan etki büyüklüğü değerine eşittir (Parker ve diğerleri, 2011). Phi yöntemi ile kıyaslandığında, robust Phi ile diğer üç yöntem arasındaki korelasyonun

göreceli olarak daha düşük olmasının temel nedeninin robust Phi yönteminin etki büyüklüğü hesaplamaları için daha konservatif bir yaklaşım kullanması olduğu düşünülmektedir. Bu bağlamda, robust Phi hesaplanırken örtüşmeyi ortadan kaldırmak için gerekli olan en az veri noktası sayısının yarısı kullanılarak başlama düzeyi ve uygulama evreleri için ayrı ayrı oranlar oluşturulurken, Phi hesaplamalarında oranlar örtüşmeyi ortadan kaldırmak için gerekli olan en az veri noktası sayısının, çıkarıldıkları evdeki toplam veri noktası sayısından çıkarılması yoluyla elde edilmektedir (Rakap ve diğerleri, 2020).

Son araştırma sorusu, her bir yöntem kullanılarak hesaplanan etki büyüklüğü değerleri ve görsel analiz sonuçları arasındaki uyuma odaklanmıştır. Veri örüntülerinde değişimin olduğu 181 grafik bağlamında, görsel analistlerin değerlendirmeleri ile etki büyüklüğü değerleri arasındaki uyum, İLOF (%95), TauÖrtüşmeme (%94), Phi (%94) ve Tau-U (%93) yöntemleri için benzer düzeydedir. Robust Phi, bu grafikler bağlamında görsel analistlerin değerlendirmeleri ile nispeten düşük düzeyde (%87) uyum göstermiştir. Görsel analistler tarafından veri örüntülerinde değişimin olmadığı değerlendirilen 26 grafik bağlamında, TauÖrtüşmeme ve Tau-U görsel analiz sonuçları ile en yüksek (%65) uyumu elde ederken, onu İLOF (%58), robust Phi (%54) ve Phi (50 %) bu yöntemleri izlemiştir. Tau-U ve İLOF kullanılarak hesaplanan etki büyüklüğü değerleri görsel analistlerin grafiklerde değişim olduğu ya da olmadığı yönündeki değerlendirmeleri ile en yüksek uyuma sahipken (%90.8 ve %90.3), bu iki yöntemi TauÖrtüşmeme (%89.9) ve Phi (% 88.9) takip etmiştir. İncelenen beş yöntem arasında robust Phi (%83.1) görsel analiz sonuçları ile en düşük uyuma sahip olan yöntem olarak bulunmuştur. Yücesoy-Özkan ve diğerleri (2020) tarafından yürütülen çalışmada TauÖrtüşmeme, Tau-U, Phi ve İLOF yöntemlerinin görsel analiz sonuçları ile uyum düzeylerinin %88.8 ile % 93.3 arasında olduğu belirlenmiştir. Bu yönüyle iki çalışma bulgularının benzerlik gösterdiği söylenebilir.

Bu çalışmanın bulguları, genel olarak, incelenen beş yöntemden dördünün (İLOF, Tau-U, TauÖrtüşmeme ve Phi) birbirleriyle çok yüksek düzeyde korelasyona sahip olduğunu ve bu yöntemler kullanılarak hesaplanan etki büyüklüğü değerleri ile görsel analiz sonuçları arasında yüksek düzeyde bir uyum olduğunu ortaya koymuştur. Görsel analiz ile uyumları dikkate alındığından Tau-U ve İLOF yöntemlerinin diğer yöntemlere kıyasla az da olsa daha iyi performans gösterdiği söylenebilir. Bu çalışmada incelenen beş örtüşmeyen veriye dayalı yöntem, göreceli olarak yakın zamanda geliştirilmiştir ve göreceli performansları kapsamlı bir şekilde incelenmemiştir. Bu nedenle alanyazında bu çalışmanın bulgularını karşılaştırmak için sınırlı sayıda çalışma bulunmaktadır (örneğin, Yücesoy-Özkan ve diğerleri, 2020). Yücesoy-Özkan ve diğerleri (2020), bu çalışma kapsamında incelenen beş yöntemden dördünü de (İLOF, Tau-U, TauÖrtüşmeme ve Phi) içeren 12 örtüşmeyen veriye dayalı yöntemi 101 A-B grafiği temel alarak incelemiş ve ÖRVEY, Tau-U ve MAVY yöntemlerinden birinin TauÖrtüşmeme, İLOF, Tüm Örtüşmeyen Çiftler ve Tüm Örtüşmeyen Veri Yüzdesi yöntemlerinden biriyle birlikte kullanılmasını önermişlerdir. Bu bağlamda, Yücesoy-Özkan ve diğerlerinin (2020) bulguları bu çalışmanın bulgularıyla benzerlik göstermektedir.

### Sınırlılıklar

Bu çalışmanın iki sınırlılığı bulunmaktadır. Bu çalışmada, etki büyüklüğü hesaplama yöntemlerini karşılaştırmalı olarak incelemek amacıyla 222 grafikte yer alan ilk başlama düzeyi (A) ile ilk uygulama aşamalarından (B) elde edilen veriler kullanılmıştır. Her ne kadar benzer sonuçlar ortaya çıkaracağı düşünülse de bir grafikte yer alan tüm verilerin kullanılmamış olması bir sınırlılık olarak değerlendirilebilir. Bu tür daha karmaşık TDA desenlerinde etki büyüklüğü hesaplamalarının nasıl yapılacağına dair bilgiler alanyazında mevcuttur (örneğin, Parker ve diğerleri, 2009; Parker ve diğerleri, 2011). Bu çalışmanın ikinci sınırlılığını ise ver setinde yer alan değişimin olmadığı tespit edilen grafik sayısının 26 olmasıdır. Bu nedenle, örtüşmeyen veriye dayalı yöntemlerin veri örüntülerinde bir değişimin olmadığı belirlenen grafikler bağlamında karşılaştırılması, nispeten az sayıda grafiğe dayanmaktadır.

### İleri Araştırmalar için Öneriler

Diğer yayınlanmış makalelerden elde edilen gerçek veri setlerini kullanarak umut vadeden örtüşmeyen veriye dayalı yöntemlerin göreceli performanslarını inceleyen ve bu çalışmanın bulgularını doğrulayan araştırmalar yapılmalıdır. Ayrıca, umut vadeden örtüşmeyen veriye dayalı yöntemler, TDA çalışmalarından

elde edilen verilerle kullanımlarının uygunluęunu belirlemek iin alanyazında nerilen parametrik etki byklğ hesaplama yntemleriyle karřılařtırılabilir. Buna ek olarak, gelecekteki arařtırmalarda, rtşmeyen veriye dayalı yntemlerin performansını incelemek iin veri rntlerinde deęiřimin olmadıęı ok sayıda grafik kullanılabilir.

#### Yazar(lar)ın Beyanı

**Arařtırmacıların katkı oranı beyanı:** Arařtırmacılar alıřmaya eřit oranda katkı yapmıřlardır.

**Etik Kurul Kararı:** Bu arařtırmada, mevcut yayınlanmış arařtırmalar kapsamında toplanan veriler kullanıldıęı iin etik kurul iznine gerek duyulmamaktadır.

**atıřma beyanı:** Arařtırmada yazarlar arasında ya da dięer kiři/kurum/kuruluřlarla herhangi bir ıkar atıřması bulunmamaktadır.

**Destek ve teřekkr:** Bu arařtırma iin herhangi bir kurumdan finansal destek alınmamıřtır.

#### Kaynaklar

- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavior Research and Therapy*, 31(6), 621–631. [https://doi.org/10.1016/0005-7967\(93\)90115-B](https://doi.org/10.1016/0005-7967(93)90115-B)
- Baer, D. (1977). Perhaps it would be better not to know everything. *Journal of Applied Behavior Analysis*, 10(1), 167-172. <https://doi.org/10.1901/jaba.1977.10-167>
- Beretvas, S. N., & Chung, H. (2008). A review of single-subject experimental design meta-analyses: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, 2(3), 129-141. <https://doi.org/10.1080/17489530802446302>
- BioSoft. (2004). *UnGraph for Windows (Version 5.0)*. Author.
- Bouck, E. C., Long, H., & Park, J. (2021). Using a virtual number line and corrective feedback to teach addition of integers to middle school students with developmental disabilities. *Journal of Developmental & Physical Disabilities*, 33(1), 99-116. <https://doi.org/10.1007/s10882-020-09735-z>
- Bykztrk, ř., Kılı akmak, E., Akgn, . E., Karadeniz, ř., & Demirel, F. (2020). Eęitimde blmsel arařtırma yntemleri. Pegem Akademi. <https://doi.org/10.14527/9789944919289>
- Campbell, J. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28(2), 234-246. <https://doi.org/10.1177/0145445503259264>
- Campbell, J. (2013). Commentary on PND at 25. *Remedial and Special Education*, 34(1), 20-25. <https://doi.org/10.1177/0741932512454725>
- Center, B. J., Skiba, R. J., & Casey, A. (1985-86). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education*, 19(4), 387-400. <https://doi.org/10.1177/002246698501900404>
- Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Exceptional Children*, 75(3), 365-383. <https://doi.org/10.1177/001440290907500306>
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis*, 12(4), 573-579. <https://doi.org/10.1901/jaba.1979.12-573>
- Gage, N. A. & Lewis, T. J. (2012). HLM-meta-analysis of single-subject design research. *Journal of Special Education*. <https://doi.org/10.1177/0022466912443894>
- Gulboy, E., Yucesoy-Ozkan, S., & Rakap, S. (2021). *Embedded instruction in early childhood: A systematic review and meta-analysis of single-case experimental research studies*. Under Review.
- Hedges, L. G., Pustejovsky, J., & Shadish, W. R. (2012). A standardized mean difference effect size for single-case designs. *Research Synthesis Methods*, 3(3), 224-239. <https://doi.org/10.1002/jrsm.1052>

- Huitema, B. E., & McKean, J. W. (2000). A simple and powerful test for autocorrelated errors in OLS intervention models. *Psychological Reports*, 87(1), 3-20. <https://doi.org/10.2466/pr0.2000.87.1.3>
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.
- Klingbeil, D. A., Van Norman, E. R., McLendon, K. E., Ross, S. G., & Begeny, J. C. (2019). Evaluating Tau-U with oral reading fluency data and the impact of measurement error. *Behavior Modification*, 43(3), 413-438. <https://doi.org/10.1177/0145445518760174>
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_scd.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf).
- Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education*. Erlbaum.
- Kratochwill, T., & Levin, J. (2006). *Single-case research design and analysis: Applications in educational intervention research*. <http://ies.ed.gov/funding/grantsearch/details.asp?ID=554>
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification*, 30(5), 598-617. <https://doi.org/10.1177/0145445504272974>
- Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature-base. *Remedial and Special Education*, 34(1), 44-58. <https://doi.org/10.1177/0741932511435176>
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, 49(3), 301-321. <https://doi.org/10.1016/j.jsp.2011.03.004>
- Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, 51(2), 201-215. <https://doi.org/10.1016/j.jsp.2012.12.005>
- Odom, S., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practice. *Exceptional Children*, 71(2), 137-148. <https://doi.org/10.1177/001440290507100201>
- Park, H., Marascuilo, L., & Gaylord-Ross, R. (1990). Visual inspection and statistical analysis in single-case designs. *Journal of Experimental Education*, 58(4), 311-320. <https://doi.org/10.1080/00220973.1990.10806545>
- Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case data: A comparison of seven statistical methods. *Behavior Therapy*, 34(2), 189-211. [https://doi.org/10.1016/S0005-7894\(03\)80013-8](https://doi.org/10.1016/S0005-7894(03)80013-8)
- Parker, R. I., & Brossart, D. F. (2006). Phase contrasts for multi-phase single case intervention designs. *School Psychology Quarterly*, 21(1), 124-136. <https://doi.org/10.1521/scpq.2006.21.1.46>
- Parker, R. I., & Hagan-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification*, 31(6), 919-936. <https://doi.org/10.1177/0145445507303452>
- Parker, R. I., Hagan-Burke, S., & Vannest, S. (2007). Percentage of all nonoverlapping data (PAND): An alternative to PND. *Journal of Special Education*, 40(4), 194-204. <https://doi.org/10.1177/00224669070400040101>
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Nonoverlap of all pairs (NAP). *Behavior Therapy*, 40(4), 357-367. <https://doi.org/10.1016/j.beth.2008.10.006>
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, 75(2), 135-150. <https://doi.org/10.1177/001440290907500201>

- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303-322. <https://doi.org/10.1177/0145445511399147>
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy*, 42(2), 284-299. <https://doi.org/10.1016/j.beth.2010.08.006>
- Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. T. Kratchowill (Ed.), *Single subject research* içinde. Academic Press.
- Pyle, K., & Fabiano, G. A. (2017). Daily report card intervention and attention deficit hyperactivity disorder: A meta-analysis of single-case studies. *Exceptional Children*, 83(4), 378-395. <https://doi.org/10.1177/0014402917706370>
- Rakap, S. (2016).  zel e itimde bilimsel dayanaklı uygulamalar. V. Aksoy (Ed.), * zel e itim* (ss. 181-211) içinde. Pegem Akademi.
- Rakap, S., & Parlak-Rakap, A. (2011). Effectiveness of embedded instruction in early childhood special education: A literature review. *European Early Childhood Education Research Journal*, 19, 79-96. <https://doi.org/10.1080/1350293X.2011.548946>
- Rakap, S., Snyder, P., & Pasia, C. (2014). Comparison of nonoverlap methods for identifying treatment effect in single-subject experimental research. *Behavioral Disorders*, 39(3), 128-145. <https://doi.org/10.1177/019874291303900303>
- Rakap, S., Y cesoy- zkan,  ., & Kalkan, S. (2020). Tek-denekli Deneysel Arařtırmalarda Etki B y kl    Hesaplama:  rt şmeyen Veriye Dayalı Y ntemlerin İncelenmesi. *T rk Psikoloji Dergisi*, 35(85), 40-60. <https://doi.org/10.31828/tpd1300443320181023m000015>
- Salzberg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for single-subject research: When does it clarify, when does it obscure? *Remedial and Special Education*, 8(2), 43-48. <https://doi.org/10.1177/074193258700800209>
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, 34(1), 9-19. <https://doi.org/10.1177/0741932512440730>
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, 8(2), 24-33. <https://doi.org/10.1177/074193258700800206>
- Shadish, W., & Rindskopf, D. (2010). A d-estimator for single case designs. <http://ies.ed.gov/funding/grantsearch/details.asp?ID=953>
- Snyder, P., Rakap, S., Hemmeter, M. L., McLaughlin, T., Sandall, S., & McLean, M. (2015). Naturalistic instructional approaches in early learning: A systematic review of the empirical literature. *Journal of Early Intervention*, 37, 69-97. <https://doi.org/10.1177/1053815115595461>
- Ugille, M., Moeyaert, M., Beretvas, S., Ferron, J., & Van Den Noortgate, W. (2012). Multilevel meta-analysis of single-subject experimental designs: A simulation study. *Behavior Research Methods*, 44(4), 1244-1254. <https://doi.org/10.3758/s13428-012-0213-1>
- Van Den Noortgate, W., Beretvas, N., & Ferron, J. (2011). Multilevel synthesis of single-case experimental data: Further developments and empirical validation. <http://ies.ed.gov/funding/grantsearch/details.asp?ID=1121>
- Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single case research. *Behavior Research Methods, Instruments & Computers*, 35(1), 1-10. <https://doi.org/10.3758/BF03195492>
- Vannest, K. J., Parker, R. I., & Gonen, O. (2011). *Single case research: Web based calculators for SCR analysis (Version 1.0) [Web-based application]*. College Station: Texas A&M University. [www.singlecaseresearch.org](http://www.singlecaseresearch.org)

- Wolery, M., Busick, M., Reichow, B., & Barton, E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education*, 44(1), 18-28. <https://doi.org/10.1177/0022466908328009>
- Yucesoy-Ozkan, S., Rakap, S., & Gulboy, E. (2020). Evaluation of treatment effect estimates in single-case experimental research: comparison of twelve overlap methods and visual analysis. *British Journal of Special Education*, 47(1), 67-87. <https://doi.org/10.1111/1467-8578.12294>

## EXTENDED ABSTRACT

### Introduction

Federal education policies including the No Child Left Behind Act of 2001, the Individuals with Disabilities Education Act of 2004 and Every Student Succeeds Act of 2015 suggest the use of instructional practices that are developed based on scientifically based research to improve outcomes for all children in schools (Yucesoy-Ozkan et al., 2020). These legal documents also call for identifying effective, evidence-based practices (EBPs). Several different organizations across various countries are established to identify EBPs in education and other fields (Rakap et al., 2014). These organizations have used meta-analysis to determine evidence-based status of specific instructional approaches. Meta-analysis conducted by these organizations often exclude studies using single-case experimental research (SCER) designs mainly due to methodological concerns (Maggin, et al., 2013) as there has been lack of consensus among SCER researchers on how to calculate effect sizes for the data obtained from such studies (Campbell, 2004; Maggin et al., 2013; Ugille et al., 2012). In recent years, therefore, several parametric and non-parametric methods have been developed to calculate effect size estimates in SCER studies (Maggin et al., 2011; Parker et al., 2011). The purpose of the present study was to investigate relative performances of five non-parametric nonoverlap methods, TauNovlap, Tau-U, Phi, robust Phi, and IRD, in calculating effect size estimates for data obtained from SCER studies. Four research questions guided the present study: (a) What are the ranges of effect size estimates calculated using five methods? (b) To what extent do five methods agree on the types of treatment effect? (c) What are the correlations among effect size estimates calculated using the five methods? and (d) To what extent do these five methods agree with the results of visual analysis?

### Method

A-B graphs ( $n = 222$ ) used in this study were obtained from 38 studies previously identified in systematic reviews of literature focused naturalistic instruction (Gulboy et al., 2021; Rakap & Parlak-Rakap, 2011; Snyder et al., 2015). Visual analyses of 222 A-B graphs were conducted by two researchers with prior experience in visual analysis of SCER data. As part of the visual analysis, each researcher answered the question of “Did a change exist in the data from baseline phase to intervention phase in favor of the intervention?” using “yes” or “no” options. Based on these ratings, each graph was categorized under one of the three groups: (a) graphs with change (i.e., when two researchers agreed on the existence of a change;  $n = 186$ ), (b) graphs with no change (i.e., when two researchers agreed on the non-existence of a change;  $n = 26$ ), and (c) graphs without consensus (i.e., when two researchers disagreed on whether a change occurred;  $n = 15$ ). Effect size estimates using five nonoverlap methods were calculated for each of the 222 graphs. Phi and robust Phi was calculated by hand using the graphed data, while TauNovlap, Tau-U, and IRD were calculated using an online calculator designed to calculate effect size estimates for SCER data (Vannest et al., 2011). SPSS 28 statistical package was used to conduct data analyses for the present study. Mean effect size estimates and standard deviations for each nonoverlap method, and correlations among the five methods were calculated separately for all 222 graphs, graphs with change, graphs with no change, and graphs without consensus disagreement.

### Results

The mean effect size scores for 222 graphs across the five nonoverlap methods ranged between 73.7% (robust Phi) and 79.2% (IRD). For the 181 graphs for which two judges agreed on the change in the data patterns, the mean effect size scores ranged from 81.0% (robust Phi) to 86.9% (TauNovlap). The mean effect size scores for the 26 graphs for which judges agreed on no change in the data patterns ranged between 36.8% (Tau-U) and 43.1% (Phi). For the 15 graphs judges disagreed about the change in data patterns, the mean effect size scores ranged from 49.4% (robust Phi) to 60.6% (IRD). Effect size scores calculated using five methods were classified as effective or not effective based on the guidelines provided by Parker et al. (2009). According to these guidelines, an intervention with effect size score of 51% or higher is considered to be an effective treatment and an intervention with effect size score of 50% or less considered to be an ineffective treatment. Results showed agreement among all five methods on the type of treatment effect for 188 graphs (85%). In 205 graphs, there was agreement among at least 4 methods (92%). Overall, correlations ranged between .84 (Tau-

U and robust Phi) and .98 (TauNovlap and Tau-U). Moderate to high correlations among five nonoverlap methods were also observed for graphs two judges agreed on the change in the data patterns (range = .82 - .98), graphs for which judges agreed on no change in the data patterns (range = .66 - .98), and graphs judges disagreed about the occurrence of change in data patterns (range = .79 - .99). The most remotely connected method appears to be the robust Phi. For the 181 graphs judged by visual analysts to have a change in data patterns, 95.0% of the IRD estimates, 94.5% of the Tau-U and Phi estimates, 93.4% of the TauNovlap estimates, and 87.3% of the robust Phi estimates fell into the effective treatment category. For the 26 graphs visual analysts agreed on no change occurred in data patterns, 65.4% of the TauNovlap and Tau-U estimates, 57.7% of the IRD estimates, 53.9% of the robust Phi estimates, and 50.0% of the Phi estimates fell into ineffective treatment category.

### **Conclusion**

The present study compared five nonoverlap methods for determining effect size estimates in SCER designs. Overall, four of the five methods, IRD, Tau-U, TauNovlap, and Phi, had very high levels of correlation with each other, and were comparable with respect to mean effect size estimates they produced and their agreement with judgments of visual analysts. From the findings of this study, IRD and Tau-U appear to be superior to the other methods investigated.