

Tek-denekli Deneysel Araştırmalarda Etki Büyüklüğü Hesaplama: Örtüşmeyen Veriye Dayalı Yöntemlerin İncelenmesi

Salih Rakap¹

Ondokuz Mayıs Üniversitesi

Şerife Yücesoy-Özkan²

Anadolu Üniversitesi

Sinan Kalkan³

Çanakkale Onsekiz Mart Üniversitesi

Özet

Psikoloji, sosyal hizmetler ve özel eğitim gibi uygulamalı alanlarda yapılan araştırmalar için oldukça kullanışlı olan tek-denekli deneysel araştırma yöntemleri (single-case experimental research design), bağımlı ve bağımsız değişkenler arasındaki nedensel ya da işlevsel ilişkinin titiz bir şekilde incelenmesine olanak tanıyan deneysel yöntemlerdir. Tek-denekli deneysel araştırmalardan elde edilen bulguların, kanıta-dayalı uygulamaların belirlenebilmesine yönelik meta-analizler kapsamında kullanılabilmesi için geliştirilen birçok etki büyüklüğü hesaplama yöntemi bulunmaktadır. Bu bağlamda, uluslararası alanyazında, tek-denekli deneysel araştırmaları rapor eden makalelerin birçoğunda, etki büyüklüğü istatistiği, örtüşmeyen veriye dayalı yöntemlerden biri kullanılarak rapor edilmektedir. Türkiye’de son dönemlerde yürütülen tek-denekli deneysel araştırmalar kapsamında daha gelişmiş yöntemlerin kullanıldığı görülmekle birlikte, araştırmacıların büyük bir çoğunluğunun, ilk geliştirilen ve çeşitli sınırlılıkları olan yöntemleri tercih ettiği görülmektedir. Bu bağlamda, Türkiye’de ve uluslararası düzeyde tek-denekli deneysel araştırmalar yürüten araştırmacıların, örtüşmeyen veriye dayalı yöntemleri ve bu yöntemlerin aralarındaki farkları anlamaları ve hangi yöntemi ne zaman kullanmaları gerektiğini değerlendirebilecekleri rehber yayınlara duyulan ihtiyaç artmaktadır. Bu makalenin temel amacı, örtüşmeyen veriye dayalı 13 yöntemi tarihsel gelişim sırasına ve birbirleriyle olan ilişki durumuna göre tanıtmak ve yöntemlerin nasıl hesaplandığını uygulamalı bir örnek ile göstermektir. Bu temel amaca ek olarak, makale, 13 yöntemin güçlü ve zayıf yönlerini değerlendirmeyi, bu yöntemler kullanılarak elde edilen etki büyüklüğü değerlerinin yorumlanmasında kullanılan ölçütleri belirtmeyi ve etki büyüklüğü yöntemlerinin seçilmesinde dikkat edilmesi gereken hususlara dikkat çekmeyi amaçlamaktadır.

Anahtar kelimeler: Kanıta-dayalı uygulamalar, tek-denekli deneysel araştırmalar, meta-analiz, etki büyüklüğü, örtüşmeyen veriye dayalı yöntemler

Abstract

Single-case experimental research designs are rigorous and valid methods to investigate causal or functional relationships between dependent and independent variables. These designs are commonly used in applied fields of psychology, social work, and special education. A number of different methods have been developed to calculate effect size estimates to integrate results of single-case experimental research into meta-analysis of evidence-based practices. Many researchers in the international literature use nonoverlap methods to report effect sizes for single-case experimental research studies. In Turkey, although some researchers have recently used more sophisticated nonoverlap methods to calculate effect sizes for single-case experimental research studies, many still use early nonoverlap methods that have several limitations noted in the literature. In this respect, it is important to know the characteristics of nonoverlap methods, differences among these methods, and when to use which methods for researchers, conducting single-case experimental research studies. Therefore, the main purpose of this paper is to describe 13 nonoverlap methods to calculate effect size estimates in single-case experimental research in chronological order and explain their calculations using hypothetical data sets. Moreover, this paper aims to discuss advantages and disadvantages of these methods, offer guideline for interpreting effect size estimates, and provide considerations for selecting appropriate effect size methods based data obtained from single-case experimental research studies.

Keywords: Evidence-based interventions, single-case experimental design, meta-analysis, effect size, nonoverlap-based methods

Yazışma Adresi: ¹Doç. Dr. Salih Rakap, Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, 55139 Atakum / Samsun, salih.rakap@omu.edu.tr, ORC-ID: 0000-0001-7853-3825

²Doç. Dr., Anadolu Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, syucesoy@anadolu.edu.tr, ORC-ID: 0000-0002-0529-0639

³Öğr. Gör., Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, sinan.kalkan@omu.edu.tr, ORC-ID: 0000-0001-6890-6421

Gönderim Tarihi: 24.10.2017

Kabul Tarihi: 23.10.2018

Nedensellik çıkarımlarının yapılabileceği araştırma desenlerini kullanan çok sayıda çalışma tarafından desteklenen ve hedef kitle çıktıları üzerinde olumlu sonuçlarının olduğu gösterilen uygulamalar olarak tanımlanan kanıta-dayalı uygulamalar (Cook ve Cook, 2013), ilk olarak 1970'li yıllarda, İngiltere'de tıp alanında kullanılmaya başlanmış ve ilerleyen yıllarda psikoloji, sağlık bilimleri, hemşirelik ve eğitim bilimleri gibi alanlarda da önem kazanmaya başlamıştır (Odom ve ark., 2005; Rakap, 2016; Reichow, 2016). Bu bağlamda, özellikle 2000'li yılların ikinci yarısından itibaren özel gereksinimli bireylerle çalışan araştırmacılar da kanıta-dayalı uygulamaları belirlemek amacıyla meta-analiz çalışmalarına yönelmişlerdir (Odom ve ark., 2005; Olive ve Franco, 2007).

Kanıta-dayalı uygulamalar, belirli ölçütleri (örn., araştırma deseni, yöntemsel kalite, çalışma sayısı vb.) kullanarak bir uygulamanın etkililiğini inceleyen deneysel araştırmaların gözden geçirilmesi ve meta-analizin yapılması yoluyla belirlenmektedir (Cook ve Cook, 2013; Kratochwill, ve ark., 2010; Odom ve ark., 2005). Tek-denekli deneysel araştırma yöntemleri, bağımlı ve bağımsız değişkenler arasındaki nedensel ya da işlevsel ilişkinin (causal or functional relationship) bilimsel olarak titiz bir şekilde incelenmesine olanak tanınmasına karşın, sıklıkla kanıta-dayalı uygulamaların belirlenmesi için yürütülen meta-analiz çalışmalarının dışında tutulmaktadır (Maggin, Briesch ve Chafouleas, 2013; Rakap, Snyder ve Pasia, 2014; Wolery, Busick, Reichow ve Barton, 2010). Bunun temel nedenlerinden biri, tek-denekli deneysel araştırmalarda, uygulanan müdahalenin etkisini belirlemede birincil olarak kullanılan ve kabul gören yaklaşımın, bireyin müdahale öncesindeki, sırasındaki ve sonrasındaki performansına ait verilerin grafikleştirilerek incelenmesine ve yorumlanmasına dayanan görsel analiz olmasıdır (Rakap, 2015; Wolery ve ark., 2010). İkinci neden ise, 1980'li yıllardan itibaren tek-denekli deneysel araştırmalarda meta-analiz yapılmasına olanak tanıyacak etki büyüklüğü hesaplama yöntemlerinin geliştirilmesine yönelik çalışmalar yürütülmesine karşın henüz genel kabul gören bir yöntemin ya da yöntem grubunun belirlenememiş olmasıdır (Campbell, 2004; Parker ve Hagan-Burke, 2007; Rakap ve ark., 2014; Wolery ve ark., 2010).

Tek-denekli deneysel araştırma yöntemlerinde, uygulanan müdahalenin etkisini belirlemek üzere görsel analizin yanı sıra, etki büyüklüğü istatistiklerinin de kullanılıyor olmasının bazı üstünlükleri bulunmaktadır. Birincisi, etki büyüklüğü, müdahalenin etkisinin nesnel olarak değerlendirilmesine olanak tanır. İkincisi, müdahalenin etkisinin görsel analiz ile belirlenemeyecek kadar küçük olduğu durumlarda hassas bir ölçüm sağlar. Üçüncüsü, etki büyüklüğü, farklı çalışmaları karşı-

laştırarak, araştırmacıların, aynı ya da benzer bağımlı değişkenler üzerinde uygulanan farklı müdahale yöntemlerinden göreceli olarak daha etkili olan yöntemleri belirlemelerine yardımcı olur (Kazdin, 1982; Parker ve Hagan-Burke, 2007). Etki büyüklüğünün görsel analizle birlikte kullanılmasının üstünlüklerini gören araştırmacılar, tek-denekli deneysel araştırmalarda kullanılmak üzere, çoğu grup desenli deneysel araştırmalarda kullanılan etki büyüklüğü hesaplama yöntemlerinden uyarlanan, çeşitli hesaplama yöntemleri geliştirmişlerdir (Allison ve Gorman, 1993; Kratochwill ve Lewin, 1992; Parker, Vannest ve Davis, 2011; Scruggs, Mastropieri ve Casto, 1987). Bu doğrultuda, tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak için geliştirilmiş yöntemler temel olarak üç grupta toplanabilir. Bunlar; parametrik yöntemler (Center, Skiba ve Casey, 1985; Maggin ve ark., 2013; Van den Noortgate ve Onghena, 2003), standartlaştırılmış ortalama farkına dayalı yöntemler (Hedges, Pustejovsky ve Shadish, 2012) ve örtüşmeyen veriye dayalı yöntemlerdir (Parker, Vannest ve Davis, 2011; Scruggs ve ark., 1987).

Grup desenli deneysel araştırmalarda etki büyüklüğü hesaplamak için yaygın olarak kullanılan varyans analizi, regresyon ve çok düzeyli doğrusal model gibi parametrik yöntemlerin tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak için kullanılmasını sağlamak amacıyla yürütülen çalışmalar son yıllarda hız kazanmıştır (Yucesoy-Ozkan, Rakap ve Gulboy, 2020). Bu yöntemler, eğilim ve düzey değişimi gibi tek-denekli deneysel araştırmalarda karşılaşılan veri örüntülerinin modellenmesine olanak tanınmasına karşın, tek-denekli deneysel araştırmalarda elde edilen verilerin çoğunluğunun parametrik analizlerin yapılabilmesi için gerekli varsayımları (örn., normallik) karşılamaması ve parametrik analizlerin tek-denekli çalışmalarda sıklıkla karşılaşılan dışadüşen (outlier) veriden etkilenmesi nedeniyle sınırlı bir kullanım alanına sahiptirler (Parker ve Vannest, 2009; Wolery ve ark., 2010). Parametrik yöntemlerin kullanılması için gerekli olan istatistik bilgisi de araştırmacıların, bu yöntemleri kullanmalarını biraz daha güçleştirmektedir (Kratochwill ve ark., 2010; Rakap, 2015).

Cohen (1965) tarafından grup desenli deneysel araştırmalarda etki büyüklüğü hesaplamak için geliştirilen standartlaştırılmış ortalama farkı yöntemi ise, Busk ve Serlin (1992) tarafından tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak için uyarlanmıştır. Bu yöntemi kullanarak etki büyüklüğü hesaplanırken, başlama düzeyi evresinin veri ortalaması, uygulama evresi veri ortalamasından çıkarılır; ortaya çıkan sonuç, başlama düzeyi evresinin verilerine ait standart sapma değerine ya da bileşik (pooled) standart sapma değerine bölünür. Tek-denekli deneysel araştırmalarda, standart-

laştırılmış ortalama farkı kullanılarak etki büyüklüğü hesaplanırken, başlama düzeyi evresindeki ve uygulama evresindeki tüm veriler kullanılabilir. Hesaplanması ve sonuçlarının yorumlanması kolay olmasına rağmen, tek-denekli deneysel araştırmalarda sıklıkla gözlemlenen eğilime ve veriler arasındaki özilişkinin (autocorrelation) uygun bir şekilde ele alınmasına duyarısız olması nedeniyle, ileri versiyonları geliştirilene kadar bu yöntemin tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak için kullanılması sorunlu görünmektedir (Maggin ve ark., 2013). Tek-denekli deneysel araştırmalarda örtüşmeyen veriye dayalı etki büyüklüğü hesaplama yöntemlerinin (nonoverlap methods) geliştirilmesine yönelik çalışmalar 1980'li yılların sonlarına doğru başlamış ancak bu yaklaşımların kullanımı, kanıta-dayalı uygulamaların belirlenmesine yönelik çalışmalara paralel olarak son yıllarda artmıştır. Örtüşmeyen veriye dayalı etki büyüklüğü hesaplama yöntemleri, tek-denekli deneysel araştırmalardaki başlama düzeyi evresi verileri ile uygulama evresi verilerinin belirli ölçütler çerçevesinde karşılaştırılmasına dayanan yöntemlerdir. Bu yöntemlerin, parametrik ya da standartlaştırılmış ortalama farkına dayalı yöntemlere göre dört temel üstünlüğü bulunmaktadır. Birincisi, örtüşmeyen veriye dayalı yöntemlerin çoğunun, kalem ve cetvel kullanılarak grafik üzerinden kolayca hesaplanabilmesi (Ma, 2006; Parker, Vannest ve Davis, 2011); ikincisi, parametrik varsayımların karşılanmasının gerekmemesi (Parker, Vannest ve Davis, 2011; Yucesoy-Ozkan ve ark., 2020; Wolery ve ark., 2010); üçüncüsü, örtüşmeyen veriye dayalı yöntemlerin görsel analizle uyumlu olması (Yucesoy-Ozkan ve ark., 2020); dördüncüsü ise, özellikle son dönemlerde geliştirilmiş yöntemlerin başlama düzeyi evresindeki eğilimi tespit edecek duyarlılıkta olmasıdır (Parker, Vannest, Davis ve Sauber, 2011). Uluslararası alanyazın incelendiğinde, birbirine benzeyen; ancak ince ayrımlarla birbirinden farklılaşan ve son dönemlerde daha kapsamlı hale gelmiş olan örtüşmeyen veriye dayalı birçok yöntem olduğu görülmektedir. Tablo 1'de alanyazında sıklıkla kullanılan örtüşmeyen veriye dayalı etki büyüklüğü hesaplama yöntemlerinin yazarları, geliştirilme tarihleri, Türkçe ve İngilizce adları ile kısaltmaları sunulmuştur.

Günümüzde, Amerikan Psikoloji Birliği (American Psychological Association [APA]), grup desenli deneysel araştırmaların rapor edildiği yayınlarda etki büyüklüğünün mutlaka kullanılması gerektiğini belirtirken, tek-denekli deneysel araştırmaların yayımlandığı birçok dergide de etki büyüklüğünün rapor edilmesi gereklilik haline gelmiştir (Campbell, 2003; Kratochwill ve ark., 2013). Bu bağlamda, uluslararası alanyazında, tek-denekli deneysel araştırmaları rapor eden makalelerin birçoğunda, etki büyüklüğünün, örtüşmeyen veriye dayalı

yöntemlerden biri kullanılarak rapor edilmesi giderek yaygınlaşmaktadır. Türkiye'deki araştırmacıların son dönemlerde yürüttükleri tek-denekli deneysel araştırmalarda (Örn., Altun ve Yucesoy-Özkan, 2018; Rakap, 2017a; Rakap ve Balıkcı, 2017; Yucesoy-Ozkan, Gulboy ve Kaya, 2018) Tau-*U* gibi daha gelişmiş yöntemleri kullanarak hesaplanan etki büyüklüğü değerlerinin rapor edildiği görülmekle birlikte, araştırmacıların çoğunluğunun meta-analizlerde ilk geliştirilen Örtüşmeyen Veri Yüzdesi ya da Sıfır Veri Yüzdesi'ni tercih ettiği görülmektedir (Örn., Bozkus-Genc ve Yucesoy-Ozkan, 2016; Topper-Korkmaz ve Diken, 2010; Sönmez ve Diken, 2010; Sazak-Pınar ve Merdan, 2016; Tavil ve Karasu, 2013; Yucesoy-Ozkan ve Sonmez, 2011). Amerikan Eğitim Bilimleri Enstitüsü'ne bağlı What Works Clearinghouse (WWC), tek-denekli deneysel araştırma yöntemlerinin kullanıldığı araştırmalarda, tüm veri örüntülerine hassas olan bir yöntem geliştirilene kadar birden çok örtüşmeyen veriye dayalı yöntemin birlikte kullanılmasını ve rapor edilmesini önermektedir (Kratochwill ve ark., 2010; Maggin ve ark., 2013). Bu nedenle, Türkiye'de tek-denekli deneysel araştırmalar yapan araştırmacıların, örtüşmeyen veriye dayalı yöntemleri ve aralarındaki farkları anlamaları ve hangi yöntemi ne zaman kullanmaları gerektiğini bilmeleri oldukça önemlidir.

Bu gereksinimden yola çıkarak çalışmanın amacı, (a) alanyazında sıklıkla kullanılan ve Tablo 1'de belirtilen, örtüşmeyen veriye dayalı 13 yöntemi tarihsel gelişim sırasına ve birbirleriyle olan ilişki durumuna göre tanıtmak, (b) yöntemlerin nasıl hesaplandığını uygulamalı bir örnekle göstermek, (c) yöntemlerin güçlü ve zayıf yönlerini değerlendirmek, (d) bu yöntemler kullanılarak elde edilen etki büyüklüğü değerlerinin yorumlanmasında kullanılan ölçütleri belirtmek ve (e) etki büyüklüğü yöntemlerinin seçilmesinde dikkat edilmesi gereken hususlara dikkat çekmektir. Bu bağlamda, 12 yöntem (Sıfır Veri Yüzdesi hariç) kullanılarak etki büyüklüğü hesaplamalarının nasıl gerçekleştirildiğini açıklamak amacıyla başlama düzeyi ve uygulama evrelerinde eğilim ve değişkenlik olan ve iki evre arasında örtüşen veriye sahip olan hipotetik bir veri seti (grafik) kullanılmıştır. Sıfır Veri Yüzdesi'nin yalnızca hedef davranışın azaltılmasını amaçlayan çalışmalarda kullanılması nedeniyle, bu yöntem kullanılarak etki büyüklüğü hesaplamalarının nasıl gerçekleştirileceğinin gösterilmesi için farklı bir grafik kullanılmıştır.

Etki Büyüklüğü Yöntemleri ve Hesaplamaları

Örtüşmeyen Veri Yüzdesi

Tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak amacıyla geliştirilen ilk yöntem, Örtüşmeyen Veri Yüzdesi'dir (Scruggs ve ark., 1987). Ulusal

Tablo 1. Örtüşmeyen Veriye Dayalı Yöntemlerinin Yazarları, Geliştirilme Tarihleri, Türkçe ve İngilizce Adları ve Kısaltmaları

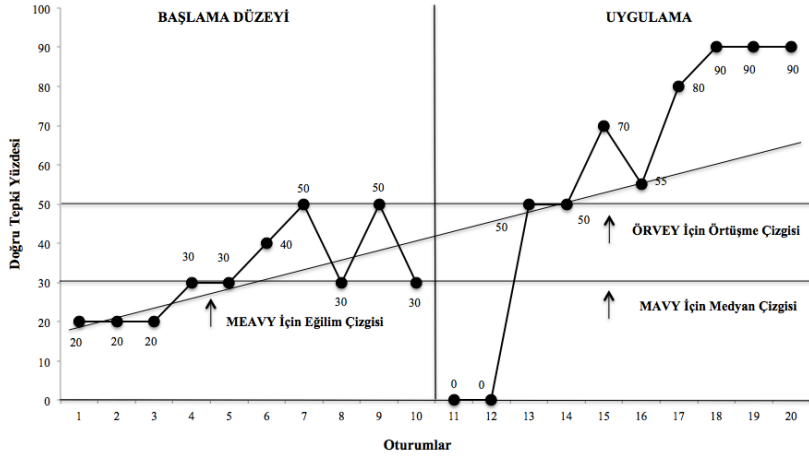
Yazar & Tarih	Türkçe		İngilizce	
	Ad	Kısaltma	Ad	Kısaltma
Scruggs ve ark., 1987	Örtüşmeyen Veri Yüzdesi	ÖRVEY	Percentage of Nonoverlapping Data	PND
Scotti ve ark., 1991	Sıfır Veri Yüzdesi	SVY	Percentage of Zero Data	PZD
Ma, 2006	Medyanı Aşan Veri Yüzdesi	MAVY	Percentage of Data Exceeding the Median	PEM
Parker ve Hagan-Burke, 2007	Tüm Örtüşmeyen Veri Yüzdesi	TÖVEY	Percentage of All Nonoverlapping Data	PAND
Manalov ve Solonas, 2009	Örtüşmeyen Düzeltilmiş Veri Yüzdesi	ÖDVEY	Percentage of Non-overlapping Corrected Data	PNCD
Parker ve Vannest, 2009	Tüm Örtüşmeyen Çiftler	TÖÇ	Nonoverlap of All Pairs	NAP
Parker, Vannest ve Brown, 2009	İlerleme Oranı Farkı	İLOF	Improvement Rate Difference	IRD
White ve Haring, 1980; Wolery, Busick, Reichow ve Barton, 2010	Medyan Eğilimini Aşan Veri Yüzdesi	MEAVY	Percentage of Data Exceeding a Median Trend	PEM-T
Parker ve Vannest, 2007	Örtüşen Veri Çiftleri	ÖVEÇ	Pairwise Data Overlap	PDO
Wolery ve ark., 2010	Örtüşen Veri Çiftleri-Karesi Alınmış	ÖVEÇ-KA	Pairwise Data Overlap Squared	PDO ²
Parker, Vannest ve Davis, 2011	Tau ^{Örtüşmeme} (Gruplar Arası Örtüşmeme İçin)	Tau ^{Örtüşmeme}	Kendall's Tau for nonoverlap between groups	Tau _{Novlap}
Parker ve ark., 2011	Tau-U (Başlama Düzeyi Eğilimi Kontrollü Tau)	Tau-U	Tau for nonoverlap with baseline trend control	Tau-U
Parker ve ark., 2011	Pearson'un Phi Değeri	Phi	Pearson's Phi	Phi

alanyazında ÖVY olarak da kısaltılan, Örtüşmeyen Veri Yüzdesi (ÖRVEY), uygulama evresinde yer alan ve sayısal değerleri başlama düzeyi evresindeki tüm veri noktalarının sayısal değerlerinden yüksek olan veri noktası sayısının, uygulama evresindeki tüm veri noktalarının sayısına oranı olarak kavramsallaştırılmaktadır (Parker, Vannest ve Davis, 2011; Scruggs ve ark., 1987; Scruggs ve Mastropieri, 2013). Davranışın azaltılmasını amaçlayan çalışmalarda, başlama düzeyi evresindeki en düşük değere sahip veri noktası temel alınarak hesaplanır. Şekil 1'de görüldüğü gibi, ÖRVEY hesaplanırken; (a) başlama düzeyi evresindeki en yüksek değere sahip veri noktası ya da noktaları belirlenir, (b) bu nokta(lar)dan, uygulama evresindeki son veri noktasına kadar, X eksenine paralel bir çizgi çizilir, (c) uygulama evresinde çizginin üzerinde kalan veri noktaları sayılır, (d) elde edilen sayı, uygulama evresindeki toplam veri noktası sayısına bölünür ve (e) sonuç 100 ile çarpılır (Scruggs ve Mastropieri, 1998;

Scruggs ve ark., 1987). ÖRVEY hesaplanırken başlama düzeyi evresine geri dönülmesi söz konusu olduğunda, birinci başlama düzeyi evresi veri setinin, birden fazla uygulama yapıldığıdaysa sonuncu uygulama evresi veri setinin kullanımı önerilmektedir (Olive ve Smith, 2005).

Şekil 1'deki örnekte, başlama düzeyi evresindeki 10 veri noktasından, sayısal değeri en yüksek veri noktasının (yedinci ve dokuzuncu) değeri 50'dir. Başlama düzeyi evresindeki bu değerden, X eksenine paralel, uygulama evresi boyunca uzanan örtüşme çizgisi çizildiğinde, uygulama evresinde çizginin üzerinde kalan veri noktası sayısının altı (veri değerleri = 70, 55, 80, 90, 90, 90) olduğu görülmektedir. Bu sayı, uygulama evresindeki toplam veri noktası sayısı olan 10'a bölünüp 100 ile çarpıldığında ÖRVEY değeri elde edilir. $ÖRVEY = (6 / 10) \times 100 = \%60$ 'dir.

ÖRVEY değerleri %0 ile %100 arasında değişmektedir. %0 etki büyüklüğü değeri, uygulama evresindeki



$$\text{ÖRVEY} = \left(\frac{\text{Uygulama Evresinde ÖRTÜŞME Çizgisinin Üzerinde Kalan Veri Noktası Sayısı}}{\text{Uygulama Evresindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$\text{ÖRVEY} = \left(\frac{6}{10} \right) \times 100 = 60$$

$$\text{MAVY} = \left(\frac{\text{Uygulama Evresinde MEDYAN Çizgisinin Üzerinde Kalan Veri Noktası Sayısı}}{\text{Uygulama Evresindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$\text{MAVY} = \left(\frac{8}{10} \right) \times 100 = 80$$

$$\text{MEAVY} = \left(\frac{\text{Uygulama Evresinde EĞİLİM Çizgisinin Üzerinde Kalan Veri Noktası Sayısı}}{\text{Uygulama Evresindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$\text{MEAVY} = \left(\frac{5}{10} \right) \times 100 = 50$$

Şekil 1. Örtüşmeyen Veri Yüzdesi (ÖRVEY), Medyanı/Ortancıyı Aşan Veri Yüzdesi (MAVY) ve Medyan Eğilimini Aşan Veri Yüzdesi (MEAVY) Yöntemlerinin Aynı Veri Seti Kullanılarak Hesaplanması.

tüm veri noktalarının sayısal değerlerinin başlama düzeyi evresindeki en yüksek veri noktasının sayısal değerine eşit ya da sayısal değerinden küçük olduğunda; %100 değeriye, uygulama evresindeki tüm veri noktalarının sayısal değerlerinin başlama düzeyi evresindeki en yüksek veri noktasının sayısal değerinden yüksek olduğunda elde edilir. ÖRVEY değerlerinin yorumlanmasında Scruggs ve arkadaşları (1987) tarafından önerilen ölçütler kullanılmaktadır. Buna göre, %50'den düşük ÖRVEY değeri etkisiz müdahaleye, %50 ile %69 arası değerler sorgulanabilir düzeyde etkili müdahaleye, %70 ile %89 arası değerler etkili müdahaleye, %90 ve üzeri değerler çok etkili müdahaleye işaret etmektedir (Scruggs ve Mastropieri, 1998; Scruggs ve Mastropieri, 2001).

ÖRVEY'in bilinen dört önemli üstünlüğü bulunmaktadır. Bunlardan birincisi, grafik üzerinden, cetvel ve kalem yardımıyla kolaylıkla hesaplanabilmesi ve karma-

şık formüllerin kullanımını gerektirmemesidir (Scruggs ve Mastropieri, 2001; Scruggs ve ark., 1987). İkincisi, hem artırılmak hem de azaltılmak istenen davranışlar için kullanılmasıdır (Olive ve Franco, 2007; Scruggs ve Mastropieri, 1998). Üçüncüsü, elde edilen değer yüzde olarak ifade edildiğinden, okuyucuların uygulamanın etkililiğine ilişkin fikir edinmelerini kolaylaştırmasıdır (Scruggs ve Mastropieri, 2001). Dördüncüsü ise, parametrik istatistik için gereken varsayımların karşılanmasını gerektirmemesidir (Ma, 2006).

ÖRVEY'in bazı sınırlılıkları da bulunmaktadır. Birincisi, hesaplanmanın, başlama düzeyindeki en yüksek değere sahip olan tek veri noktası temel alınarak yapılması ve başlama düzeyinde dışadüşen (outlier) bir veri noktası olması durumunda, uygulamanın gerçek etkisini tespit etme konusunda yeterince hassas olmamasıdır (Faith, Allison ve Gormann, 1996). Hesaplama, dışa-

düşen veri temel alınarak yapıldığında uygulama etkili olsa da etkisiz gibi görünebilir. İkincisi, başlama düzeyi evresinde bulunması muhtemel olan terapötik yöndeki eğilimle (uygulama evresinde müdahale sunulduktan sonra ortaya çıkması beklenen değişim) başa çıkmak için önlem alınmasına olanak tanımamasıdır. Üçüncüsü, birçok araştırmacı tarafından etki büyüklüğü hesaplama yöntemi olarak kabul edilmemesidir. Dördüncüsü, etki büyüklüğü değerinin, grafikteki toplam veri noktası sayısından etkilenmesi (Allison ve Gormann, 1993) ve beşincisi ise ÖRVEY için güven aralığı ve p değerlerinin hesaplanamıyor olmasıdır.

Medyanı Aşan Veri Yüzdesi

Ma (2006) tarafından geliştirilen Medyanı Aşan Veri Yüzdesi (MAVY), başlama düzeyi verileri temel alınarak belirlenen medyan çizgisini aşan uygulama evresi veri noktalarının sayısının, uygulama evresindeki tüm veri noktalarının sayısına oranı olarak kavramsallaştırılmaktadır. MAVY, ÖRVEY'in sahip olduğu temel sınırlılık olan, dışadüşen veriden etkilenme sorunuyla başa çıkmak için geliştirilmiştir ve medyanın başlama düzeyi verilerini iyi şekilde özetlediği varsayımına dayanmaktadır (Parker, Vannest ve Davis, 2011). Ma (2006; 2009), bir müdahale etkili olduğunda uygulama evresindeki verilerin büyük çoğunluğunun bağımlı değişkenin artırılmasının hedeflendiği çalışmalarda medyan çizgisinin üstünde, azaltılmasının amaçlandığı çalışmalarda medyan çizgisinin altında olacağını ve müdahale etkisiz olduğundaysa uygulama evresi verilerinin medyan çizgisinin çevresinde, altında ve üstünde dalgalanacağını belirtmektedir.

Şekil 1'de görüldüğü gibi, MAVY kullanılarak etki büyüklüğü hesaplanırken ÖRVEY hesaplamasında kullanılan basamaklar takip edilir. Ancak, ilk basamakta başlama düzeyi evresindeki en yüksek değere sahip olan veri noktası ya da noktalarının belirlenmesi yerine başlama düzeyi verilerinin medyanı (ortanca noktası) belirlenir (Parker, Vannest ve Davis, 2011). Şekil 1'de, başlama düzeyindeki 10 veri noktasının medyanı 30'dur. Bu düzeyden, X eksenine paralel, uygulama evresi boyunca uzanan medyan çizgisi çizildiğinde, uygulama evresindeki çizginin üzerinde kalan veri noktası sayısının sekiz (veri değerleri = 50, 50, 70, 55, 80, 90, 90, 90) olduğu görülmektedir. Bu sayı, uygulama evresindeki toplam veri noktası sayısı olan 10'a bölünüp 100 ile çarpıldığında MAVY değeri elde edilir. $MAVY = (8 / 10) \times 100 = \%80$ 'dir.

MAVY değerleri de %0 (uygulama evresindeki veri noktalarının tümü medyan çizgisinin altında olduğunda) ile %100 (uygulama evresindeki veri noktalarının tümü medyan çizgisinin üzerinde olduğunda) arasında değişmektedir. Ma (2006), MAVY değerlerinin yorumlan-

ması için herhangi bir ölçüt rapor etmemiş; ancak hem Ma (2006) hem de daha sonra diğer araştırmacılar, Scruggs, Mastropieri, Cook ve Escobar (1986) tarafından ÖRVEY için önerilen ölçütleri kullanmışlardır. ÖRVEY için geliştirilen ölçütleri MAVY değerlerini yorumlamak amacıyla kullanırken dikkatli olmak gerekir çünkü aynı veriler kullanılarak yürütülen karşılaştırmalı çalışmalarda, MAVY değerlerinin düzenli olarak ÖRVEY değerlerinden büyük çıktığı ve ÖRVEY ile MAVY değerleri arasındaki korelasyonun düşük olduğu görülmektedir (Vannest ve Ninci, 2015). Örneğin, Bozkus-Genc ve Yucesoy-Ozkan (2016), ÖRVEY ve MAVY arasındaki korelasyonun $r = .598$ olduğunu rapor ederken, Ma (2006) bu iki yöntem kullanılarak elde edilen etki büyüklüğü değerleri arasındaki korelasyonun $r = .690$ 'dan küçük olduğunu belirtmiştir.

MAVY'nin birkaç önemli üstünlüğü vardır. Birincisi, başlama düzeyi evresinde bulunması olası dışadüşen veriden etkilenmemesi; ikincisi, hesaplamının grafik üzerinde cetvel ve kalemle kolaylıkla yapılabilmesi; üçüncüsüyse, elde edilen değer yüzde olarak ifade edilmesi nedeniyle anlaşılmasının kolay olmasıdır. Üstünlüklerinin yanı sıra üç önemli sınırlılığı da söz konusudur. Birincisi, ÖRVEY gibi başlama düzeyi evresindeki muhtemel terapötik eğilimi dikkate almaması; ikincisi, aşırı düzeyde dalgalanma ve değişkenlik gösteren veri setlerinde medyanın veri setini temsil etme gücü azalacağından, MAVY değerinin güvenilir olmayabileceği; üçüncüsü, araştırmaların MAVY ile görsel analiz arasındaki uyuşmanın çok düşük düzeyde olduğunu göstermesidir (Parker ve Hagan-Burke, 2007; Qi, Barton, Collier, Lin ve Montoya, 2015; Wolery ve ark., 2010).

Medyan Eğilimini Aşan Veri Yüzdesi

Başlama düzeyi evresinde var olması muhtemel terapötik yöndeki eğilimi belirlemek üzere geliştirilen ilk etki büyüklüğü hesaplama yöntemi Medyan Eğilimini Aşan Veri Yüzdesi'dir (MEAVY). İlk olarak White ve Haring (1980) tarafından uzatılmış eğilim çizgisi (extended celeration line) yöntemi olarak geliştirilmiş ve son yıllarda bazı araştırmacılar tarafından gözden geçirilip düzenlenerek MEAVY olarak isimlendirilmiştir (Ma, 2006; Wolery ve ark., 2010). MEAVY, başlama düzeyi evresindeki veriler temel alınarak belirlenen medyan eğilim çizgisini aşan uygulama evresindeki veri noktalarının sayısının, uygulama evresindeki tüm veri noktaları sayısına oranı olarak kavramsallaştırılmaktadır (Wolery ve ark., 2010). Şekil 1'de görüldüğü üzere, MEAVY kullanılarak etki büyüklüğü hesaplanırken; (a) başlama düzeyi evresindeki veriler kullanılarak medyan eğilim çizgisi çizilir (başlama düzeyi evresindeki ilk iki ve son iki veri noktalarının orta noktalarını birleştiren çizgi) ve uygulama evresindeki son veri noktasına kadar bu çizgi

X eksenine üzerine uzatılır, (b) uygulama evresinde çizginin üzerinde kalan veri noktaları sayılır, (c) elde edilen sayı, uygulama evresindeki toplam veri noktası sayısına bölünür ve (d) sonuç 100 ile çarpılır (Rakap ve ark., 2014; Wolery ve ark., 2010). Başlama düzeyi evresinde terapötik yönde bir eğilim olmadığında, MAVY ile aynı sonucu verir.

Şekil 1'de, başlama düzeyi evresindeki 10 veri noktası için medyan eğilim çizgisi ilk iki veri noktasının (20 ile 20) orta noktası ile son iki veri noktasının (50 ile 30) orta noktasını birleştirecek şekilde çizilip uygulama evresi boyunca uzatıldığında, uygulama evresinde çizginin üzerinde kalan veri noktası sayısının beş (veri değerleri = 70, 80, 90, 90, 90) olduğu görülmektedir. Bu sayının, uygulama evresindeki toplam veri noktası sayısı olan 10'a bölünüp 100 ile çarpıldığında MEAVY değeri elde edilir. $MEAVY = (5 / 10) \times 100 = \%50$ 'dir.

MAVY'de olduğu gibi, MEAVY değerleri de %0 (uygulama evresindeki tüm veri noktaları medyan eğilimi çizgisinin altında olduğunda) ile %100 (uygulama evresindeki tüm veri noktaları medyan eğilimi çizgisinin üzerinde olduğunda) arasında yer almaktadır. Alanyazında, MEAVY değerlerinin yorumlanması için bir ölçüt bulunmamaktadır. Wolery ve arkadaşları (2010) MEAVY değerlerini yorumlamak üzere Scruggs ve Mastropieri (1998) tarafından önerilen ölçütleri önermişlerdir.

Diğer etki büyüklüğü hesaplama yöntemleriyle karşılaştırıldığında MEAVY'nin çeşitli üstünlükleri bulunmaktadır. Birincisi, başlama düzeyi evresinde yer alan ve terapötik yöndeki doğrusal eğilimi kontrol etme özelliğinin olmasıdır. İkincisi, başlama düzeyi evresindeki eğilimi kontrol eden Tau-U ile kıyaslandığında, hesaplamasının ve anlaşılmasının daha kolay olması; cetvel ve kalem kullanılarak grafik üzerinde kolayca hesaplanabilmesidir. Üçüncüsü, hesaplama yaparken başlama düzeyi evresindeki eğilim dikkate alındığından, daha kararlı sonuçlar verme olasılığının fazla olmasıdır. Dördüncüsü ise hem artırılmak istenen hem de azaltılmak istenen davranışlar için kullanılabilmesidir (Wolery ve ark., 2010).

MEAVY'nin en temel sınırlılığıysa, başlama düzeyi evresindeki yüksek değere sahip olan dışadüşen veri noktalarından olumsuz yönde etkilenmesi ve dolayısıyla, hesaplama sonuçları ile görsel analiz sonuçları arasındaki uyumun göreceli olarak düşük olmasıdır (Wolery ve ark., 2010). Yüksek değere sahip olan dışadüşen veri noktası ya da noktaları, eğilim çizgisinin terapötik yönde dikleşmesine neden olmaktadır. Bu durumda, uygulama evresindeki birçok veri noktası, eğilim çizgisinin altında kalmakta ve görsel analizde etkili bulunan müdahale, etki büyüklüğü sonuçlarına göre etkisiz görülebilmektedir (Rakap ve ark., 2014).

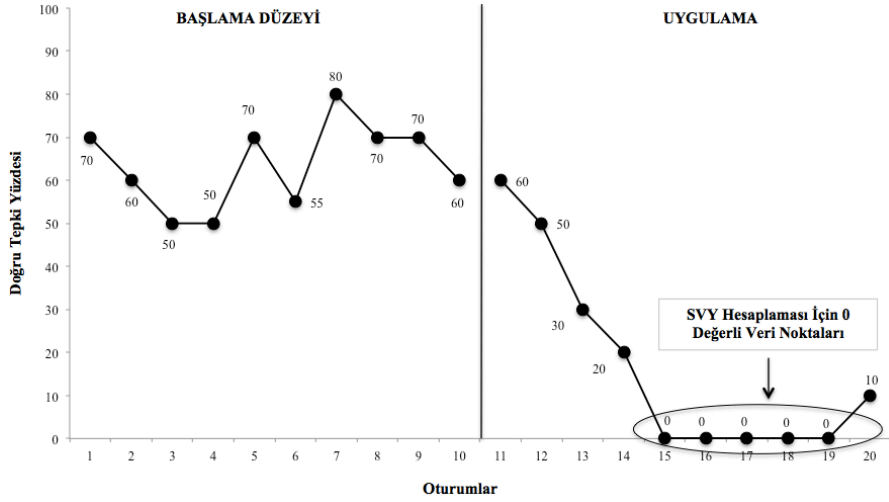
Sıfır Veri Yüzdesi

ÖRVEY, MAVY ve MEAVY yöntemleri, temel olarak, uygulama evresinde bağımlı değişken performansının artırılmasının hedeflendiği çalışmalarda etki büyüklüğü hesaplamak için geliştirilmiştir. Tek-denekli deneysel araştırma desenlerini kullanan ve uygulama sırasında hedef davranışın azaltılmasını amaçlayan çalışmalar için özel olarak geliştirilen örtüşmeyen veriye dayalı ilk etki büyüklüğü hesaplama yöntemi, Sıfır Veri Yüzdesi'dir (SVY; Scotti ve ark., 1991). Bu yöntem, uygulama evresinde 0 değerine sahip veri noktası sayısının, uygulama evresindeki ilk 0 değerine sahip veri noktası da dahil sonraki tüm veri noktaları sayısına oranı olarak kavramsallaştırılmaktadır (Campbell, 2003). Şekil 2'de görüldüğü gibi, SVY hesaplanırken; (a) uygulama evresindeki 0 değerine sahip ilk veri noktası belirlenir, (b) ilk 0 değerine sahip veri noktası da dahil olmak üzere takip eden tüm 0 değerine sahip veri noktalarının sayısı belirlenir, (c) 0 değerine sahip veri noktası sayısı, ilk 0 değerine sahip veri noktası da dahil olmak üzere bu noktadan sonra uygulama evresinde yer alan tüm veri noktalarının sayısına bölünür ve (d) sonuç 100 ile çarpılır (Scotti ve ark., 1991).

Şekil 2'de, uygulama evresinde 0 değerine sahip ilk veri 15. oturumda gözlemlenmiş ve bundan sonraki dört oturumda (16., 17., 18., ve 19.) veri değerleri 0 olarak kalmıştır. Son oturumdaysa (20. oturum) veri değeri 10'a yükselmiştir. Bu doğrultuda, uygulama evresinde 0 değerli beş veri yer alırken, ilk 0 değerli veri noktasından sonra, bu nokta da dahil, toplam altı veri noktası bulunmaktadır. Uygulama evresinde 0 sayısal değerine sahip veri noktası sayısının ilk 0 değerli veri de dahil olmak üzere bu noktadan sonraki toplam veri noktası sayısına bölünmesiyle SVY değeri elde edilir. $SVY = (5 / 6) \times 100 = \%83$ 'dür.

SVY değerlerinin yorumlanması için Scotti ve arkadaşları (1991) tarafından geliştirilen ölçütler kullanılmaktadır. Bu ölçütlere göre, %12'den küçük değer etkisiz, %12-42,9 arası değer sorgulanabilir, %43-%69,9 arası değer etkili, %70 ve üzeri değer çok etkili müdahaleye işaret etmektedir.

SVY'nin bazı üstünlükleri ve sınırlılıkları bulunmaktadır. Bu yöntemin en önemli üstünlüğü, azaltılmak istenen davranışlarda kolaylıkla kullanılması ve hesaplamasının kolay olmasıdır. SVY'nin iki temel sınırlılığından birincisi, hesaplama yapmak için uygulama evresinde 0 değerine sahip veri noktası olması gerekliliği ve 0 değerine sahip veri noktası olmadığında hesaplama yapılamaması (Scotti ve ark., 1991); ikincisi ise, azaltılmak istenen davranış çok düşük düzeylere kadar azalsa da 0 değerli veri noktası olmadığı için uygulamanın etkisiz gibi görünmesi ve dolayısıyla hatalı sonuçlar vermesidir. Bu konuya ilişkin Campbell (2004), SVY değerinin,



$$SVY = \left(\frac{\text{Uygulama Evresinde SIFIR Değerli Veri Noktası Sayısı}}{\text{Uygulama Evresindeki SIFIR Değerli İlk Veri Noktası da Dahil Sonraki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$SVY = \left(\frac{5}{6} \right) \times 100 = 83$$

Şekil 2. Sıfır Veri Yüzdesi (SVY) Yönteminin Hesaplanması.

azaltılmak istenen davranışın azaltıldığını değil, bastırıldığını gösteren bir değer olduğunu ifade etmektedir.

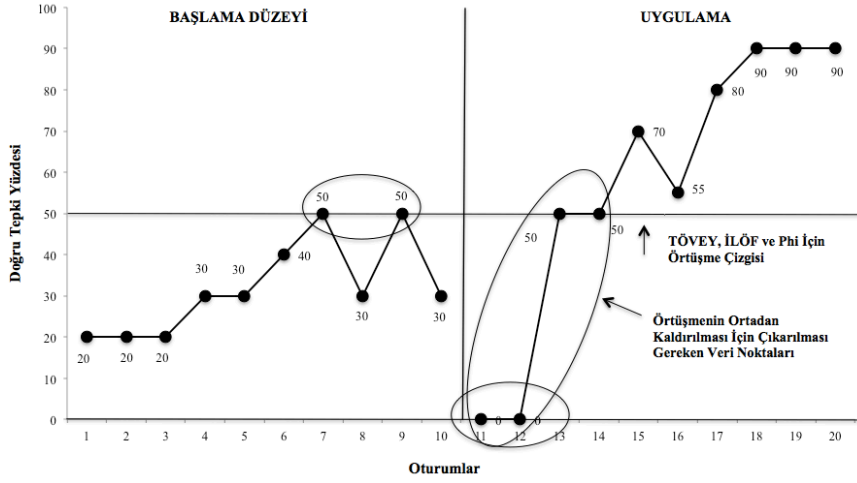
Tüm Örtüşmeyen Veri Yüzdesi

Parker ve Hagan-Burke (2007) tarafından geliştirilen Tüm Örtüşmeyen Veri Yüzdesi (TÖVEY), başlama düzeyi evresi ile uygulama evresi arasındaki örtüşmeyi ortadan kaldırmak amacıyla mümkün olan en az sayıda veri noktası çıkarıldıktan sonra başlama düzeyi ve uygulama evrelerinde kalan veri noktası sayısının, iki evredeki toplam veri noktası sayısına oranı olarak kavramsallaştırılmaktadır. ÖRVEY'in sahip olduğu en temel sınırlılık olan başlama düzeyi evresindeki tek bir veri noktasını temel alarak hesaplama özelliğine bir tepki olarak geliştirilen TÖVEY hem başlama düzeyi evresindeki hem de uygulama evresindeki tüm veri noktalarını kullanarak hesaplama yapmayı içermektedir. Şekil 3'te görüldüğü gibi, TÖVEY hesaplanırken; (a) başlama düzeyi ve uygulama evresindeki toplam veri noktası sayısı belirlenir, (b) başlama düzeyi ve uygulama evreleri arasındaki örtüşmeyi ortadan kaldırmak için gerekli olan en az veri noktası sayısı belirlenir, (c) örtüşmeyi ortadan kaldırmak için belirlenen veri noktası sayısı, başlama düzeyi ve uygulama evrelerindeki toplam veri noktası sayısından çıkarılır, (d) sonuç, başlama düzeyi ve uygulama evrelerindeki toplam veri noktası sayısına bölünür

ve (e) bölüm 100 ile çarpılır (Parker ve Hagan-Burke, 2007; Parker, Vannest ve Davis, 2011).

Şekil 3'te, başlama düzeyi evresi ile uygulama evresindeki, değerleri 0, 0, 50 ve 50 olan ilk dört veri noktasının ya da başlama düzeyi evresindeki, değeri 50 olan iki veri noktası ile (yedinci ve dokuzuncu oturumlar) uygulama evresindeki değeri 0 olan ilk iki veri noktasının çıkarılması gerekmektedir. Her iki durumda da toplam dört veri noktası çıkarılacağından etki büyüklüğü sonucunda bir değişiklik olmayacaktır. Bu dört veri noktası çıkarıldıktan sonra başlama düzeyi ve uygulama evresinde kalan veri noktası sayısının, bu iki evredeki toplam veri noktası sayısına bölünmesi ve çıkan sonucun 100 ile çarpılmasıyla TÖVEY elde edilir. $TÖVEY = (16 / 20) \times 100 = \%80$ 'dir.

Yukarıda açıklanan hesaplama ilave olarak, TÖVEY, Microsoft Office Excel programında sıralama işlemi kullanılarak da hesaplanabilmektedir (Parker, Vannest ve Davis, 2011; Schneider, Goldstein ve Parker, 2008). Ancak, bu yolla hesaplama, grafikteki verileri dijitalleşirmeyi gerektirdiğinden, elle hesaplanmanın tercih edilmesi önerilmektedir (Parker, Vannest ve Davis, 2011). TÖVEY değerleri %0 ile %100 arasında değişmektedir. Parker ve Hagan-Burke (2007) ile TÖVEY'i kullanan diğer araştırmacılar (örn., Parker, Vannest ve Davis,



$$TÖVEY = \left(\frac{\text{Başlama Düzeyi ve Uygulama Evrelerinde ÖRTÜŞMEYEN Veri Noktası Sayısı}}{\text{Başlama Düzeyi ve Uygulama Evrelerindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$TÖVEY = \left(\frac{16}{20} \right) \times 100 = 80$$

$$İLOF_1 = \left(\frac{\text{Uygulama Evresinde Kalan Veri Noktası Sayısı}}{\text{Uygulama Evresindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$İLOF_1 = \left(\frac{6}{10} \right) \times 100 = 60$$

$$İLOF_2 = (\text{Uygulama Evresi İlerleme Oranı} - \text{Başlama Düzeyi Evresi İlerleme Oranı}) \times 100$$

$$İLOF_2 = \left(\frac{8}{10} - \frac{2}{10} \right) \times 100 \Rightarrow \frac{6}{10} \times 100 = 60$$

Şekil 3. Tüm Örtüşmeyen Veri Yüzdesi (TÖVEY) ve İlerleme Oranı Farkı (İLOF) Yöntemlerinin Aynı Veri Seti Kullanılarak Hesaplanması.

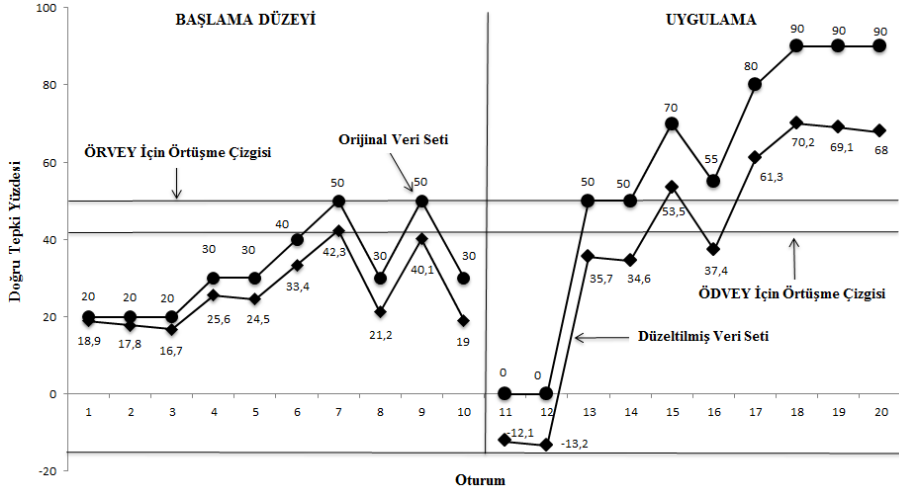
2011), değerlerin yorumlanması için herhangi bir ölçüt önermezlerken, TÖVEY’i kullanan araştırmacıların da ÖRVEY için geliştirilen ölçütleri kullandığı görülmektedir (örn. Lenz, 2013).

TÖVEY’in bazı üstünlükleri vardır ve bunlardan birincisi, grafik üzerinden bir cetvel ve kalem yardımıyla kolaylıkla hesaplanabilmesidir (Parker, Hagan-Burke ve Vannest, 2007). İkincisi, hem başlama düzeyi evresindeki hem de uygulama evresindeki tüm veri noktalarının kullanılmasıdır (Wendt, 2009). Üçüncüsü ise, etki büyüklüğü değerlerinin, grup desenli deneysel çalışmalarda kullanılan etki büyüklüğü değerleri ile karşılaştırılabilecek şekilde bir indekse dönüştürülebilmesidir (Parker ve ark., 2007). TÖVEY’in birtakım sınırlılıkları da vardır ve bunlardan en önemlisi, hesaplamada üst noktalara duyarsızlık olması ve iki evredeki veri noktaları arasında bulunan uzaklığı göz önüne almaksızın hesaplama ya-

pılmasıdır. Diğer bir ifadeyle, müdahale çok etkili olmasına karşın iki evre arasında hiç örtüşme olmadığında, etki büyüklüğünün %100 bulunması ve müdahale etkili olmamasına karşın etkili gibi görünmesidir. İkincisi, bu yöntem ile etki büyüklüğü hesaplanırken başlama düzeyi evresinde terapötik yöndeki eğilimlerin kontrol edilmemesidir. Üçüncüsü ise, hesaplama yapabilmek için çok sayıda veri noktasına (en az 20 veri noktası) gereksinim duyulmasıdır (Wendt, 2009).

Örtüşmeyen Düzeltilmiş Veri Yüzdesi

Örtüşmeyen Düzeltilmiş Veri Yüzdesi (ÖDVEY) Manolov ve Solonas (2009) tarafından önerilmiş, ÖRVEY’in sınırlılıklarının üstesinden gelebilmek üzere ÖRVEY hesaplamasından önce bir veri düzeltme süreci içeren etki büyüklüğü hesaplama yöntemidir. Düzeltmenin amacı, uygulama ile ilişkili olmayan ve uygulama



$$\text{ÖDVEY} = \left(\frac{\text{Düzeltilmiş Veri Setinin Uygulama Evresinde ÖRTÜŞME Çizgisinin Üzerinde Kalan Veri Noktası Sayısı}}{\text{Uygulama Evresindeki Toplam Veri Noktası Sayısı}} \right) \times 100$$

$$\text{ÖRVEY} = \left(\frac{5}{10} \right) \times 100 = 50$$

Şekil 4. Örtüşmeyen Düzeltmiş Veri Yüzdesi (ÖDVEY) Yönteminin Hesaplanması.

öncesindeki olası eğilimi verilerden ayırmaktır. Şekil 4'te görüldüğü gibi, ÖDVEY ile etki büyüklüğü hesaplanırken; (a) başlama düzeyindeki veri noktalarına ilişkin bir fark dizisi oluşturulur, (b) elde edilen bu fark dizisindeki yeni değerlerin aritmetik ortalaması hesaplanır, (c) hem başlama düzeyi evresindeki hem de uygulama evresindeki her bir veri noktası için fark dizisinin aritmetik ortalaması ile verinin sıra sayısı çarpılarak eğilim düzeltme faktörü hesaplanır, (d) düzeltme faktör puanları, orijinal veri noktalarından çıkarılarak veri düzeltmesi işlemi yapılır, (e) düzeltme işlemi sonrasında başlama düzeyi evresindeki en yüksek değere sahip veri noktası belirlenir, (f) bu noktadan, uygulama evresindeki son veri noktasına kadar, X eksenine paralel bir çizgi çizilir, (g) uygulama evresinde çizginin üzerinde kalan veri noktaları sayılır, (h) elde edilen sayı, uygulama evresindeki toplam veri noktası sayısına bölünür ve (i) sonuç 100 ile çarpılır (Manolov ve Solonas, 2009). Fark dizisini oluşturmak için grafikteki başlama düzeyi evresinde yer alan her bir veri noktasından bir önceki veri noktası çıkarılarak, başka bir deyişle, ikinci veri noktasından birinci, üçüncü veri noktasından ikinci veri noktası çıkarılarak yeni bir değer elde edilir. Bu işlem, başlama düzeyi evresindeki tüm veri noktaları tamamlanmaya kadar devam ettirilir. Böylece $[n_A - 1]$ sayıda yeni veri noktası elde edilmiş olur.

Şekil 4'te orijinal veri seti ile aynı veri setine ait düzeltilmiş veri seti yer almaktadır. Düzeltilmiş veri setini elde etmek için öncelikle başlama düzeyi evresindeki verilere ilişkin fark dizisi oluşturulmuştur [(20 - 20 = 0), (20 - 20 = 0), (30 - 20 = 10), (30 - 30 = 0), (40 - 30 = 10), [(50 - 40 = 10), (30 - 50 = -20), (50 - 30 = 20), (30 - 50 = -20)]. Daha sonra elde edilen bu fark dizisinin aritmetik ortalaması hesaplanmıştır: [(0 + 0) + (10 + 0) + (10) + (10) + (-20) + (20) + (-20)] / 9 = 1,1]. Sonra, hem başlama düzeyi hem de uygulama evresi için eğilim düzeltme faktörleri hesaplanmıştır. Başlama düzeyi evresi için eğilim düzeltme faktörü (1,1 x 1 = 1,1), (1,1 x 2 = 2,2), (1,1 x 3 = 3,3), (1,1 x 4 = 4,4), (1,1 x 5 = 5,5), (1,1 x 6 = 6,6), (1,1 x 7 = 7,7), (1,1 x 8 = 8,8), (1,1 x 9 = 9,9), (1,1 x 10 = 11) şeklinde; uygulama evresi için eğilim düzeltme faktörleri ise (1,1 x 11 = 12,1), (1,1 x 12 = 13,2), (1,1 x 13 = 14,3), (1,1 x 14 = 15,4), (1,1 x 15 = 16,5), (1,1 x 16 = 17,6), (1,1 x 17 = 18,7), (1,1 x 18 = 19,8), (1,1 x 19 = 20,9), (1,1 x 20 = 22) şeklindedir. Daha sonra başlama düzeyi evresinde yer alan yeni veri seti (20 - 1,1 = 18,9), (20 - 2,2 = 17,8), (20 - 3,3 = 16,7), (30 - 4,4 = 25,6), (30 - 5,5 = 24,5), (40 - 6,6 = 33,4), (50 - 7,7 = 42,3), (30 - 8,8 = 21,2), (50 - 9,9 = 40,1), (30 - 11 = 19) şeklinde; uygulama evresinde yer alan veri seti (0 - 12,1 = -12,1), (0 - 13,2 = -13,2), (50 - 14,3 = 35,7), (50 - 15,4 = 34,6), (70 - 16,5 = 53,5), (55 - 17,6 = 37,4), (80 - 18,7

= 61,3), (90 - 19,8 = 70,2), (90 - 20,9 = 69,1), (90 - 22 = 68) şeklinde düzeltilmiştir. Düzeltme sonrasında ÖRVEY hesaplamasında kullanılan adımlar takip edilmiştir. Şekil 4'teki başlama düzeyi evresinde en yüksek değeri olan veri noktası 42,3'tür. Bu noktadan uygulama evresindeki son veri noktasına kadar, X eksenine paralel bir çizgi çizildiğinde uygulama evresinde çizginin üzerinde kalan veri noktası sayısı beştir (veri değerleri = 53,5 - 61,3 - 70,2 - 69,1 - 68). Bu sayı, uygulama evresindeki toplam veri noktası sayısı olan 10'a bölünüp 100 ile çarpıldığında ÖDVEY değeri elde edilir. $\text{ÖDVEY} = (5 / 10) \times 100 = \%50$ 'dir.

ÖDVEY'in önemli üstünlüklerinden birincisi, veriler sıralı olarak birbirine bağımlı olduğunda, diğer bir deyişle, ardışık gözlemler arasında özilişki bulunduğunda, rahatlıkla kullanılmalıdır. İkincisi ise hem başlama düzeyi hem de uygulama evresindeki tüm veri noktalarının kullanılmasıdır. ÖDVEY'in çeşitli sınırlılıkları da vardır. Bunlardan en önemlisi, veriler sıralı olarak birbirine bağımlı olmadığına ya da herhangi bir eğilim söz konusu olmadığında ÖDVEY'in kullanılmaması tavsiye edilir çünkü bu gibi durumlarda hatalı sonuç verebilir. İkincisi, hesaplama yaparken veri noktalarının değerlerini belirlemek üzere grafikleri dijitalleştirmek gerekmektedir. Üçüncüsü, yöntemin etkililiğini belirlemek üzere ÖDVEY'i yorumlayacak bir ölçüt bulunmamaktadır ve ÖRVEY'de kullanılan yorumlama ölçütlerinin kullanımı için çok uygun değildir (Manolov ve Solonas, 2009; Bozkus-Genc ve Yucesoy-Ozkan, 2016).

İlerleme Oranı Farkı

İlk olarak tıp alanında, risk-azaltma ya da risk farkı yöntemi adıyla geliştirilmiş bir etki büyüklüğü hesaplama yöntemi olan İlerleme Oranı Farkı (İLOF), Parker ve arkadaşları (2009) tarafından tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamak için uyarlanmıştır. İLOF, başlama düzeyi evresi ile uygulama evresindeki ilerleme oranlarının farkı olarak kavramsallaştırılmaktadır. Bir evre için ilerleme oranıysa, o evrede ilerleme gösteren veri noktaları sayısının, tüm veri noktalarının sayısına bölünmesi olarak tanımlanır. Başlama düzeyi evresinde ilerleme gösteren veri noktası, uygulama evresindeki en az bir veri ile eşit ya da daha yüksek sayısal değere sahip veri noktası anlamına gelirken; uygulama evresinde ilerleme gösteren veri noktası, başlama düzeyi evresindeki tüm veri noktalarından sayısal değer olarak daha yüksek olan veri noktası olarak tanımlanır.

İLOF kullanılarak etki büyüklüğü hesaplanırken (Şekil 3'e bakınız); (a) başlama düzeyi ve uygulama evresindeki örtüşen veriler belirlenir ve (b) başlama düzeyi evresi ile uygulama evresindeki örtüşmeyi ortadan kaldırmak için çıkarılması gereken en az sayıda veri noktasına karar verilir. Bu aşamadan sonra, verilerin hangi

evreden çıkarıldığına bağlı olarak, hesaplama üç farklı şekilde yapılabilir. Buna göre örtüşmeyi ortadan kaldırmak için çıkarılan veri noktaları yalnızca uygulama evresinden çıkarıldığında İLOF; uygulama evresinde kalan veri noktası sayısının, uygulama evresindeki toplam veri noktası sayısına bölünmesi ve sonucun 100 ile çarpılmasıyla hesaplanır. Örtüşmeyi ortadan kaldırmak için çıkarılan veri noktaları yalnızca başlama düzeyi evresinden çıkarıldığında İLOF; başlama düzeyi evresinden çıkarılan veri noktası sayısının, başlama düzeyi evresindeki toplam veri noktası sayısına bölünmesi, elde edilen bölümün 1'den çıkarılması ve sonucun 100 ile çarpılmasıyla elde edilir. Örtüşmeyi ortadan kaldırmak için çıkarılan veri noktaları hem başlama düzeyi hem de uygulama evrelerinden çıkarıldığında İLOF; uygulama evresinde kalan veri noktası sayısının bu evredeki toplam veri noktası sayısına, başlama düzeyi evresinden çıkarılan veri noktası sayısının da bu evredeki toplam veri noktası sayısına bölünmesi (ilerleme oranlarının belirlenmesi), başlama düzeyi evresine ait ilerleme oranının, uygulama evresine ait ilerleme oranından çıkarılması ve sonucun 100 ile çarpılması ile hesaplanır (Parker ve ark., 2009; Vannest ve Ninci, 2015). Veri örüntülerinin özelliklerinden dolayı hesaplamalar sonucunda kimi zaman aynı sonuçlar elde edilecek olsa da her zaman aynı sonuçlara ulaşılmadığını akılda tutmak gerekir.

Şekil 3'te, başlama düzeyi ile uygulama evresi arasındaki örtüşmeyi ortadan kaldırmak için uygulama evresinden dört veri noktasının çıkarıldığı durumu ele aldığımızda, İLOF, uygulama evresinde kalan veri noktası sayısının, bu evredeki toplam veri noktası sayısına bölünmesi ve sonucun 100 ile çarpılması yoluyla hesaplanır. $\text{İLOF}_1 = (6 / 10) \times 100 = \%60$ 'dir. Hem başlama düzeyi hem de uygulama evresinden ikişer veri noktasının çıkarıldığı durumu ele aldığımızda ise, $\text{İLOF}_2 = [(8 / 10) - (2 / 10)] \times 100 = \%60$ 'tır.

İLOF kullanılarak hesaplanan etki büyüklüğü %0 ile %100 arasında değerler almaktadır. %50'nin altındaki İLOF değerleri, çok küçük ya da sorgulanabilir etkiye sahip müdahaleleri, %50-69 arası değerler orta düzey etkiye sahip müdahaleleri, %70-74 arası değerler etkili müdahaleleri ve %75 üzeri değerler çok etkili müdahaleleri işaret etmektedir (Parker ve ark., 2009; Rakap, 2015; Vannest ve Ninci, 2015).

İLOF'un bazı üstünlükleri olduğunu söylemek mümkündür. Birincisi, elle basit biçimde hesaplanması ve başlama düzeyi evresi ile uygulama evresindeki ilerleme oranının farkı olarak kavramsallaştırıldığı için kolayca yorumlanabilmesidir. İkincisi, büyük veri setleri ile hesaplama yapmaya olanak sağlayan internet tabanlı (<http://www.singlecaseresearch.org/calculators/ird>) bir hesap makinası bulunmasıdır. Üçüncüsü, karmaşık ya da çok fazla evresi olan tek-denekli deneysel desenlerde de

kolaylıkla kullanılmalıdır. İLOF'un en temel sınırlılığıysa, başlama düzeyi evresinde yer alan terapötik yöndeki eğilimi dikkate almamasıdır (Vannest ve Ninci, 2015).

Phi (Pearson'ın Phi Değeri)

Phi, Parker, Vannest ve Davis (2011) tarafından TÖVEY'in bir uzantısı olarak geliştirilmiştir. Phi hesaplanırken; (a) başlama düzeyi ve uygulama evreleri arasındaki örtüşmeyi ortadan kaldırmak için gerekli olan en az veri noktası sayısı belirlenir, (b) örtüşmeyi ortadan kaldırmak için gerekli olan en az veri noktası sayısının yarısı kullanılarak başlama düzeyi ve uygulama evreleri için ayrı ayrı oranlar oluşturulur, (c) bu oranlar 2×2 tabloya girilerek çapraz tablo (crosstab) analizi yapılır ve (d) sonuç 100 ile çarpılır (Parker, Vannest ve Davis, 2011). Şekil 3'te, uygulama evresinden dört veri noktasının çıkarıldığı durumda, başlama düzeyi ve uygulama evresi için oran $2/8$ olacaktır. Bu oranlar ile çapraz tablo analizi yaparak Phi değeri, Şekil 5'teki gibi hesaplanır.

	B	U	T	
Alt	a	b	e	$\text{Phi} = \frac{ad - bc}{\sqrt{efgh}}$
Üst	c	d	f	
T	g	h	n	

	B	U	T	
Alt	8	2	10	$\text{Phi} = \left[\frac{(8 \times 8) - (2 \times 2)}{\sqrt{10 \times 10 \times 10 \times 10}} \right] \times 100$
Üst	2	8	10	
T	10	10	20	$\text{Phi} = \left[\frac{64 - 4}{100} \right] \times 100 = \%60$

Not. B = Başlama düzeyi evresi; U = Uygulama evresi; T = Toplam.

Şekil 5. Phi Etki Büyüklüğü Değerinin Hesaplanması.

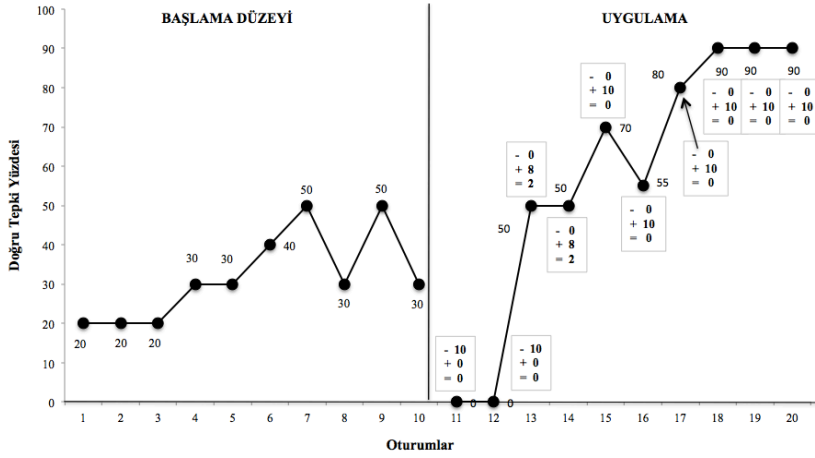
Phi değerleri de 0 ile 100 arasında değişmektedir. Alanyazında Phi değerlerinin yorumlanması için geliştirilmiş bir ölçüt bulunmadığından, Phi'yi kullanan araştırmacıların Scruggs ve arkadaşları (1987) tarafından ÖRVEY için geliştirilen ölçütleri kullandıkları görülmektedir. Phi'nin en temel üstünlüğü, TÖVEY'de olduğu gibi, elde edilen etki büyüklüğü değerinin grup desenli deneysel çalışmalarda kullanılan etki büyüklüğü değerleri ile karşılaştırılabilecek bir indekse dönüştürülebilmesidir. Diğer üstünlüğüyse, hesaplama yapılırken hem başlama düzeyi hem de uygulama evresindeki tüm veri noktalarının kullanılmasıdır (Parker, Vannest ve Davis, 2011). Phi'nin temel sınırlılığıysa, hesaplamasının diğer yöntemlere göre biraz daha karmaşık olmasıdır.

Tüm Örtüşmeyen Çiftler

Parker ve Vannest (2009) tarafından geliştirilen Tüm Örtüşmeyen Çiftler (TÖÇ), başlama düzeyi evresindeki her bir veri noktasının, uygulama evresindeki her bir veri noktası ile karşılaştırılması sonucu elde edilen örtüşmeyen veri çifti sayısının, karşılaştırılan toplam çift sayısına oranı olarak kavramsallaştırılmaktadır. TÖVEY, İLOF ve Phi gibi yöntemlerde hem başlama düzeyi hem de uygulama evresindeki tüm veriler hesaplama sırasında kullanılmasına rağmen bu iki evredeki veriler bireysel olarak birbiriyle karşılaştırılmamaktadır. Bu bağlamda, TÖÇ daha fazla bilgi kullanılarak etki büyüklüğü hesaplamasına olanak tanımaktadır. Şekil 6'da görüldüğü üzere TÖÇ hesaplanırken; (a) başlama düzeyi ve uygulama evresindeki veri noktası sayıları çarpılarak karşılaştırmada kullanılacak olan toplam çift sayısı belirlenir, (b) başlama düzeyi evresindeki her bir veri noktası, uygulama evresindeki her bir veri noktasıyla karşılaştırılarak uygulama evresi veri değerlerinin büyük olduğu çift sayısı belirlenir, (c) başlama düzeyi evresindeki her bir veri noktası, uygulama evresindeki her bir veri noktasıyla karşılaştırılarak eşit sayısal değere sahip olan çift sayısı belirlenir, (d) uygulama evresindeki veri değerlerinin büyük olduğu çift sayısının tamamı ile eşit sayısal değere sahip çift sayısının yarısı toplanır, (e) elde edilen değer, karşılaştırılan toplam çift sayısına bölünür ve (f) sonuç 100 ile çarpılır.

Şekil 6'da, karşılaştırılacak çift sayısı $10 \times 10 = 100$ 'dür. Uygulama evresindeki 0 değerine sahip ilk iki veri, sayısal değer olarak başlama düzeyi evresindeki tüm veri noktalarından küçüktür (Büyük = 0; Eşit = 0; Küçük = 10). Uygulama evresindeki, 50 değerine sahip üçüncü ve dördüncü veri noktaları ise başlama düzeyi evresindeki iki veri noktası ile eşit sayısal değere sahipken, başlama düzeyi evresinde geri kalan sekiz veri noktasından daha büyüktür (Büyük = $8 \times 2 = 16$; Eşit = $2 \times 2 = 4$; Küçük = 0). Uygulama evresindeki son altı veri noktasının sayısal değerleri ise başlama düzeyi evresindeki tüm veri noktalarının sayısal değerlerinden büyüktür (Büyük = $10 \times 6 = 60$; Eşit = 0; Küçük = 0). Bu bağlamda, uygulama evresi veri noktalarının, başlama düzeyi evresi veri noktalarından büyük olduğu toplam çift sayısı 76 iken, eşit değerli çift sayısı 4'tür. Buna göre, TÖÇ = $[(76 + 4 / 2) / 100] \times 100 = \%78$ 'tir.

TÖÇ değerleri %0 ile %100 arasında yer almaktadır. TÖÇ değerlerinin yorumlanmasında, %50 ve altındaki değerler etkisiz müdahaleye, %86 ve üzerindeki değerler çok etkili müdahaleye işaret etmektedir (Petersen-Brown, Karich ve Symos, 2012; Vannest ve Ninci, 2015). Bu araştırmacılar, %5 ile %86 arasındaki değerlere yönelik bir etkililik düzeyi önermemişlerdir. Ancak, bu aralıkta, etki büyüklüğü değeri arttıkça, müdahalenin etkililiğinin de arttığı kabul edilebilir.



$$TÖÇ = \left[\frac{(\text{Toplam "+" Sayısı}) + \left(\frac{\text{Toplam "=" Sayısı}}{2} \right)}{\text{Karşılaştırılan Toplam Çift Sayısı}} \right] \times 100$$

$$TÖÇ = \left[\frac{76 + \left(\frac{4}{2} \right)}{100} \right] \times 100 = 78$$

$$ÖVEÇ = \left[\frac{(\text{Toplam "+" Sayısı})}{\text{Karşılaştırılan Toplam Çift Sayısı}} \right] \times 100$$

$$ÖVEÇ = \frac{76}{100} \times 100 = 76$$

$$ÖVEÇ - KA = \frac{[ÖVEÇ]^2}{100} \quad \Rightarrow \quad ÖVEÇ - KA = \frac{[76]^2}{100} = 58$$

$$Tau_{Örtüşmeme} = \left(\frac{\text{Toplam "+" Sayısı} - \text{Toplam "-" Sayısı}}{\text{Karşılaştırılan Toplam Veri Çifti Sayısı}} \right) \times 100$$

$$Tau_{Örtüşmeme} = \left(\frac{76 - 20}{100} \right) \times 100 = 56$$

Şekil 6. Tüm Örtüşmeyen Çiftler (TÖÇ), Örtüşen Veri Çiftleri (ÖVEÇ), Örtüşen Veri Çiftleri-Karesi Alınmış (ÖVEÇ-KA) ve TauÖrtüşmeme Yöntemlerinin Aynı Veri Seti Kullanılarak Hesaplanması.

TÖÇ'ün iki temel üstünlüğü bulunmaktadır. Birincisi, hesaplama yapılırken hem başlama düzeyi hem de uygulama evresindeki tüm veri noktalarının kullanılması; ikincisi ise, diğer yöntemlere göre, başlama düzeyi ve uygulama evresindeki tüm veri noktalarını karşılaştırması nedeniyle daha güvenilir sonuçlar elde edilmesini sağlamasıdır (Wendt, 2009). TÖÇ'ün en temel sınırlılığı ise,

hesaplamanın uzun zaman alması ve özellikle karmaşık ve kalabalık verileri olan grafiklerde hesaplamanın zor olmasıdır (Wendt, 2009). Bu sorunu çözmek için internet tabanlı bir hesap makinası geliştirilmiştir (<http://www.singlecaseresearch.org/calculators/nap>). Ayrıca, TÖÇ de diğer birçok hesaplama yöntemi gibi, başlama düzeyi evresinde yer alan terapötik yondaki eğilimi hesaba katmamaktadır.

Örtüşen Veri Çiftleri ve Örtüşen Veri Çiftleri-Karesi Alınmış

Parker ve Vannest (2007) tarafından önerilen Örtüşen Veri Çiftleri (ÖVEÇ) yöntemi, TÖÇ (Parker ve Vannest, 2009) yönteminde olduğu gibi, başlama düzeyi evresindeki her bir veri noktasının, uygulama evresindeki her bir veri noktası ile karşılaştırılması sonucu elde edilen örtüşmeyen veri çifti sayısının, karşılaştırılan toplam çift sayısına oranı olarak kavramsallaştırılmaktadır. Bu yöntemin TÖÇ'ten temel farkı, karşılaştırma sonucunda sayısal değer olarak eşit olan çiftleri hesaplama dışı bırakması, başka bir deyişle, hesaplamada yalnızca uygulama evresindeki veri değerlerinin büyük olduğu çift sayısını kullanmasıdır. Dolayısıyla bu yöntemin TÖÇ'e göre daha katı olduğu söylenebilir. Wolery ve arkadaşları (2010), ÖVEÇ yöntemini daha sağlam hale getirmek için karesinin alınmasını önermişlerdir (ÖVEÇ-KA). Bu doğrultuda ÖVEÇ hesaplanırken; (a) başlama düzeyi ve uygulama evresindeki veri noktası sayıları çarpılarak karşılaştırmada kullanılacak toplam çift sayısı belirlenir, (b) başlama düzeyi evresindeki her bir veri noktası, uygulama evresindeki her bir veri noktasıyla karşılaştırılarak uygulama evresinde veri değerlerinin büyük olduğu çift sayısı belirlenir, (c) uygulama evresindeki veri değerlerinin büyük olduğu çift sayısı, karşılaştırılan toplam çift sayısına bölünür ve (d) sonuç 100 ile çarpılır. Çıkan ÖVEÇ değerinin karesi alınarak ÖVEÇ-KA elde edilir. Şekil 6'da hem başlama düzeyi hem de uygulama evresinde 10'ar veri noktası olduğu için karşılaştırılacak toplam çift sayısı 100'dür (10 x 10). Uygulama evresindeki veri noktalarının sayısal değerlerinin, başlama düzeyi evresindeki veri noktalarından büyük olduğu çift sayısı 76'dır. Bu durumda, $\text{ÖVEÇ} = (76 / 100) \times 100 = \%76$ ve $\text{ÖVEÇ-KA} = (76 \times 76) / 100 = \%58$ 'dir.

ÖVEÇ değerleri de %0 ile %100 arasında değişmektedir. Alanyazında, ÖVEÇ ve ÖVEÇ-KA ile elde edilen etki büyüklüğü değerlerini yorumlamak için herhangi bir ölçütün bulunmadığı görülmektedir. Hesaplama benzerlikten dolayı TÖÇ (Petersen-Brown ve ark., 2012; Vannest ve Ninci, 2015) için geliştirilen ve daha önce belirtilen ölçütler ÖVEÇ ve ÖVEÇ-KA değerlerini yorumlarken kullanılabilir. Bu noktada göz önünde bulundurulması gereken önemli bir nokta, bir grafikte başlama düzeyi ve uygulama evresinde eşit değere sahip veri olmadığında, TÖÇ ve ÖVEÇ değerlerinin aynı olacağı, eşit değerler bulunması durumundaysa, TÖÇ değerinin, ÖVEÇ değerinden daha büyük olacaktır. Genel olarak, ÖVEÇ-KA değeri, TÖÇ ve ÖVEÇ değerlerinden küçüktür. Bu durumun iki istisnası, TÖÇ ve ÖVEÇ değerlerinin 0 ya da 1 olmasıdır ki, bu durumda, ÖVEÇ-KA değeri de 0 ya da 1 olacaktır.

TÖÇ'ün sahip olduğu tüm üstünlüklere ve sınırlılıklara sahip olan ÖVEÇ ve ÖVEÇ-KA, TÖÇ'e göre

daha katıdır. Her iki yöntem de başlama düzeyi evresindeki dışadüşen veriden ciddi düzeyde etkilenmezken, başlama düzeyi evresindeki eğilimi dikkate almamaktadır. ÖVEÇ-KA değerleri ise, ÖVEÇ değerlerinden daha katıdır.

Tau_{Örtüşmeme} (Gruplar Arası Örtüşmeme İçin Kendall'in Tau Değeri)

Parker ve arkadaşları (2011) tarafından geliştirilen Tau_{Örtüşmeme}, başlama düzeyi evresindeki her bir veri noktasının, uygulama evresindeki her bir veri noktası ile karşılaştırılması sonucu elde edilen örtüşen veri çifti sayısının, örtüşmeyen veri çifti sayısından çıkarılması ve sonucun, karşılaştırılan toplam çift sayısına oranı olarak kavramsallaştırılmaktadır. Bu ifadeden anlaşılacağı gibi Tau_{Örtüşmeme}, TÖÇ ve ÖVEÇ ile benzerlik göstermektedir. TÖÇ ve ÖVEÇ, örtüşmeyen veri çifti yüzdesi olarak ifade edilirken, Tau_{Örtüşmeme} (örtüşmeyen veri çifti yüzdesi – örtüşen veri çifti yüzdesi) olarak ifade edilmektedir (Parker Vannest, Davis ve ark., 2011). Şekil 6'da olduğu gibi, Tau_{Örtüşmeme} hesaplanırken; (a) başlama düzeyi ve uygulama evresindeki veri noktası sayıları çarpılarak karşılaştırmada kullanılacak toplam çift sayısı belirlenir, (b) başlama düzeyi evresindeki her bir veri noktası, uygulama evresindeki her bir veri noktasıyla karşılaştırılarak uygulama evresinde veri değerlerinin büyük olduğu çift sayısı belirlenir, (c) başlama düzeyi evresindeki her bir veri noktası, uygulama evresindeki her bir veri noktasıyla karşılaştırılarak başlama düzeyi evresinde veri değerlerinin büyük olduğu çift sayısı belirlenir, (d) başlama düzeyi evresinde veri değerlerinin büyük olduğu çift sayısı, uygulama evresinde veri değerlerinin büyük olduğu çift sayısından çıkarılır, (e) elde edilen değer karşılaştırılan toplam çift sayısına bölünür ve (f) sonuç 100 ile çarpılır. Şekil 6'da, karşılaştırılacak toplam çift sayısı 100, uygulama evresindeki veri noktalarının sayısal değer olarak büyük olduğu çift sayısı 76, başlama düzeyi evresindeki veri noktalarının sayısal değer olarak büyük olduğu ya da eşit olan çift sayısı 20'dir. Bu durumda, $\text{Tau}_{\text{Örtüşmeme}} = \left(\frac{76 - 20}{100} \right) \times 100 = \%56$ olarak hesaplanır.

Tau_{Örtüşmeme} değerleri %0 ile %100 arasındadır. Alanyazında, Tau_{Örtüşmeme} değerlerini yorumlamak için belirlenmiş bir ölçüt bulunmadığından ve hesaplama benzer olduğundan, TÖÇ (Petersen-Brown ve ark., 2012; Vannest ve Ninci, 2015) için geliştirilen ölçütler Tau_{Örtüşmeme} değerini yorumlarken de kullanılabilir. Burada dikkat edilmesi gereken nokta, Tau_{Örtüşmeme} değerlerinin, TÖÇ değerlerinden küçük olacağı ve başlama düzeyi evresi ile uygulama evresi arasındaki örtüşme arttıkça, bu iki etki büyüklüğü sonucu arasındaki farkın da artacağıdır. Tau_{Örtüşmeme}, TÖÇ ve ÖVEÇ'in sahip olduğu üstünlüklere ve sınırlılıklara sahipken, bu yöntemlere göre

daha katı bir yöntemdir. $Tau_{Ortüşme}$ değerleri, ÖVEÇ-KA değerleri ile benzerlik göstermektedir. Ayrıca, $Tau_{Ortüşme}$ değerlerini hesaplamak üzere internet tabanlı (<http://www.singlecaseresearch.org/calculators/Tau-U>) bir hesap makinası bulunmaktadır.

Tau-U (Başlama Düzeyi Eğilimi Kontrollü)

Başlama düzeyinde yer alan terapötik yöndeki eğilimi kontrol eden ve başlama düzeyi ile uygulama evrelerindeki tüm veri noktalarının karşılaştırılmasına dayalı olarak hesaplanan etki büyüklüğü yöntemi olan $Tau-U$, örtüşmeyen veri temelli etki büyüklüğü hesaplama yöntemleri arasında en gelişmiş olanıdır (Parker ve ark., 2011). Bu yöntem kullanılarak etki büyüklüğü hesaplaması diğer yöntemlere göre daha karmaşıktır. $Tau-U$, Kendall Sıralı Korelasyon ve Mann-Whitney U testinin bir uzantısı olarak düşünülebilir (Corder ve Foreman, 2009). $Tau-U$ hesaplanırken; (a) başlama düzeyi ve uygulama evresindeki veri noktası sayıları çarpılarak karşılaştırmada kullanılacak toplam çift sayısı belirlenir, (b) Kendall Sıralı Korelasyon analizi yapılarak S değeri hesaplanır, (c) S değeri, karşılaştırılan toplam çift sayısı bölünür ve (d) sonuç 100 ile çarpılır.

Şekil 6'da, karşılaştırılacak toplam çift sayısı 100'dür. Bu grafikteki verileri kullanarak S değerini hesaplamak için, Tablo 2'deki gibi düzey ve veri başlıklı iki satırlı (ya da sütunlu) bir tablo oluşturulur. Düzey satırına, başlama düzeyi evresindeki veri noktalarının sıra sayıları ($n = 10$) tersten başlanarak yazılırken, uygulama evresindeki tüm verilerin karşısına, bir sonraki zaman kodu olan 11 yazılır. Veri satırına ise, başlama düzeyi evresindeki ilk veri noktasından başlayarak grafikte yer alan tüm veri noktalarının sayısal değerleri yazılır. Bu değerler kullanılarak Kendall Sıralı Korelasyon değeri olan S hesaplanır ($S = 31$). Bu durumda, $Tau - U = \left(\frac{31}{100}\right) \times 100 = \%31$ olarak hesaplanır.

$Tau-U$ değerleri %0 ile %100 arasında değişmektedir. Bu bağlamda, %20'den küçük değerler müdahalenin etkisinin çok küçük olduğunu, %20 ile %59 arasındaki değerler müdahalenin orta düzeyde etkiye sahip olduğunu, %60 ile %79 arasındaki değerler müdahalenin etkisinin büyük olduğunu, %80 ve üzeri değerler ise müdahalenin etkisinin çok büyük olduğunu göstermektedir (Vannest ve Ninci, 2015).

$Tau-U$ diğer yöntemler ile karşılaştırıldığında bazı üstünlükleri bulunmaktadır. Birincisi, başlama düzeyi ve uygulama evresindeki tüm veri noktalarını karşılaştırma-

sı ve bunu yaparken de başlama düzeyi evresindeki eğilimi dikkate alması nedeniyle diğer yöntemlere göre daha güvenilir ve hassas sonuçlar elde edilmesini sağlaması; ikincisi ise, başlama düzeyi ve uygulama evrelerinde çok az sayıda veri olduğu durumlarda dahi hesaplamaların etkili sonuçlar vermesidir (Brossart, Vannest, Davis ve Patience, 2014; Parker ve ark., 2011; Vannest ve Ninci, 2015). Bu yöntemin en temel sınırlılığı ise, hesaplama yapılırken bir istatistik paketinin kullanılmasına gerek duyulmasıdır. $Tau-U$ hesaplamalarının el ile yapılması mümkün olmakla birlikte özellikle veri noktası sayısı çok fazla olan grafiklerde kalem-kâğıt kullanılarak yapılan hesaplama hataya oldukça açıktır (Rakap, 2015). Bu nedenle, $Tau-U$ değerlerini hesaplamak üzere internet-tabanlı bir hesap makinası geliştirilmiştir (<http://www.singlecaseresearch.org/calculators/Tau-U>).

Etki Büyüklüğü Değerlerinin Hesaplanması ve Yorumlanması

Tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplamalarının yapılabilmesi için ham veriyeye gereksinim duyulmakta; ancak, bu yöntemleri kullanan araştırmalarda veriler grafik haline getirilerek sunulmaktadır. Bu durum ise, etki büyüklüğü hesaplamaları yapılmadan önce grafikteki verilerin dijitalleştirilmesini gerekli kullmaktadır. X ve Y eksenlerindeki değerlerin net olarak belli olduğu grafiklerde, veri noktalarının sayısal değerlerinin cetvel ve kalem kullanılarak belirlenmesi mümkün olmakla birlikte, eksenlerdeki değerlerin net olmadığı, veri noktası sayısının çok olduğu ve bir grafikte birden fazla bağımlı değişkene ilişkin veri sunulduğu durumlarda, verileri dijitalleştirmek için geliştirilmiş programlar kullanılmalıdır. Alanyazında, grafiklerdeki verilerin sayısal değerlerinin tespit edilmesi (dijitalleştirilmesi) için geliştirilmiş, geçerliği ve güvenilirliği incelenmiş bazı programlar yer almaktadır: DataThief (Moeyaert, Maggin ve Verkuilen, 2016), Digitizelt (Rakap, Rakap, Evran ve Cig, 2016), GraphClick (Boyle, Samaha, Rodewald ve Hoffmann, 2013; Rakap ve ark., 2016), UnGraph (Moeyaert ve ark., 2016; Rakap ve ark., 2016; Shadish ve ark., 2009), WebPlotDigitizer (Drevon, Fursa ve Malcolm, 2017; Moeyaert ve ark., 2016) ve XYit (Moeyaert ve ark., 2016) bunlardan bazılarıdır. Etki büyüklüğü hesaplamalarında verilerin sayısal değerlerinin belirlenmesinde bu programların kullanılması tercih edilmelidir.

Tablo 2. Şekil 6'daki Grafiğe Göre Kendall Sıralı Korelasyon Hesaplamasında Kullanılacak Değerler

Düzey	10	9	8	7	6	5	4	3	2	1	11	11	11	11	11	11	11	11	11	
Veri	20	20	20	30	30	40	50	30	50	30	0	0	50	50	70	55	80	90	90	90

Etki büyüklüğü hesaplamaları sonucu elde edilen değerler yorumlanırken, sadece etkinin büyüklüğünü sınıflamak için geliştirilen ölçütleri kullanmak, elde edilen sonucun uygulamadaki önemini göz ardı edilmesine neden olmaktadır ve grup desenli deneysel çalışmalarda da eleştirilen bir konudur (Yucesoy-Ozkan ve ark., 2020). Tek-denekli deneysel araştırmalarda etki büyüklüğü değerleri yorumlanırken, hesaplama sonucu elde edilen sayısal değer yanında değiştirilmesi hedeflenen davranışın ya da becerinin klinik önemi de göz önünde bulundurulmalıdır (Brossart ve ark., 2014; Lenz, 2013; Vannest ve Ninci, 2015; Wolery ve ark., 2010). Örneğin, otizmi olan bir çocuğun dil ve konuşma becerilerini geliştirmek üzere uygulanan bir müdahale sonucu elde edilen etki büyüklüğünün %55 olduğunu düşünelim. Bu değer, birçok etki büyüklüğü ölçütüne göre küçük bir etki olarak görülecek ve uygulanan müdahalenin işe yararlığının sorgulanmasına neden olacaktır. Ancak, bu %55 değeri, çocuğun, müdahale öncesinde sahip olmadığı bazı becerilere sahip olmaya başladığının bir göstergesi olabilir (örn., isteklerini tek sözcük kullanarak belirtmek) ya da daha önce uygulanan müdahale sonucunda elde edilen etki büyüklüğü değerine göre daha büyük olabilir ve çocuğa ileride uygulanması muhtemel müdahalelere yön verebilir. Bu nedenle, bir etki büyüklüğü değerini yorumlarken, etkililiği incelenen müdahalenin bağlamı da dikkate alınmalıdır (Vannest ve Ninci, 2015).

Etki büyüklüğü değerlerini yorumlamak için geliştirilen ölçütler, araştırmacılar tarafından özellikle meta-analizler kapsamında ya da farklı müdahale yöntemlerinden etkili olanın belirlenmesinde sıklıkla kullanılmaktadır. Yukarıda belirtildiği ve Tablo 3'te de görüldüğü gibi, örtüşmeyen veriye dayalı etki büyüklüğü hesaplama yöntemlerinin birçoğu için özel olarak geliştirilmiş ölçütler bulunmamaktadır. Bu noktada, birçok araştırmacının bir etki büyüklüğü değerini yorumlamak için geliştirilmiş ölçütleri, başka etki büyüklüğü

yöntemleri ile elde edilen değerleri yorumlamak için kullandığı görülmektedir. Etki büyüklüğü hesaplamalarında kullanılan süreçlerin birbirinden çok farklı olduğu durumlarda bu tür bir uygulamanın, etkililiği incelenen müdahale hakkında sağlıklı bilgi vermeyi güçleştireceği unutulmamalıdır.

Etki Büyüklüğü Sonuçları ile Görsel Analiz Arasındaki İlişki

Tek-denekli deneysel araştırmalar için etki büyüklüğü hesaplama yöntemlerini geliştirmek üzere çalışan araştırmacılar da dahil olmak üzere bu alanda uygulamalı araştırmalar yapan birçok araştırmacı, tek-denekli deneysel araştırmalarda bağımsız değişken ile bağımlı değişken arasındaki nedensel ya da işlevsel ilişkinin belirlenmesinde birincil olarak kullanılması gereken yöntemin görsel analiz olduğu konusunda hemfikirlerdir (Kratichwill ve ark., 2010; 2013; Rakap, 2017b; Wolery ve ark., 2010). Görsel analiz, bağımsız değişken ile bağımlı değişken arasındaki nedensel ya da işlevsel ilişkiyi ortaya koymak amacıyla, bir evredeki ve/veya ardışık evrelerdeki verilerin; düzey (level), eğilim (trend), kararlılık (stability), etkinin anılığı (immediacy of effect) ve örtüşme (overlap) gibi değişkenler açısından incelenmesidir (Alberto ve Troutman, 2009; Genc-Tosun, 2017). Düzey, bir evredeki verilerin medyanı/ortancası ya da ortalama değeridir. Bir evreden diğerine geçildiğinde bağımlı değişken düzeyinde ani ve büyük bir değişiklik oluyorsa, müdahalenin etkili olduğu kabul edilir (Kazdin, 1982; Rakap, 2017b). Eğilim, bir evredeki verilerin yönünün, artması ya da azalmasıdır. Bir evreden diğerine geçildiğinde bağımlı değişkene ilişkin dik ve aniden yükselen/azalan bir eğilim ortaya çıkıyorsa, müdahalenin etkili olduğu kabul edilir (Kazdin, 1982; Rakap, 2017b). Kararlılık, bir evredeki verilerin değişkenliği, dalgalanmalar şeklindeki iniş-çıkışlarıdır (Cooper, Heron ve Heward, 2014; Rakap, 2017b). Uygulama evresindeki verilerde herhangi bir dalgalanma görülüyor ve veriler karar-

Tablo 3. Etki Büyüklüğü Değerlerini Yorumlama Ölçütleri

Yöntem	Etkisiz	Sorgulanabilir	Etkili	Çok Etkili
Örtüşmeyen Veri Yüzdesi (ÖRVEY)	$EB \leq 50$	$50 < EB < 70$	$70 \leq EB < 90$	$EB \geq 90$
Sıfır Veri Yüzdesi (SVY)	$EB \leq 12$	$12 < EB < 43$	$43 \leq EB < 70$	$EB \geq 70$
İlerleme Oranı Farkı (İLÖF)	$EB \leq 50$		$70 \leq EB < 75$	$EB \geq 75$
Tüm Örtüşmeyen Çiftler (TÖÇ)	$EB \leq 50$	$50 < EB < 86$	$EB \geq 86$	
Tau-U	$EB \leq 20$	$20 < EB < 60$	$60 \leq EB < 80$	$EB \geq 80$

Not. Tabloda belirtilen değerler yüzdelerdir. Medyanı/Ortancayı Aşan Veri Yüzdesi (MAVY), Medyan Eğilimini Aşan Veri Yüzdesi (MEAVY), Tüm Örtüşmeyen Veri Yüzdesi (TÖVEY), Örtüşmeyen Düzeltilmiş Veri Yüzdesi (ÖDVEY), Phi, Örtüşen Veri Çiftleri (ÖVEÇ), Örtüşen Veri Çiftleri-Karesi Alınmış (ÖVEÇ-KA) ve Tau_{Örtüşmeyen} yöntemleri kullanılarak elde edilen etki büyüklüğü değerlerinin yorumlanması için literatürde belirtilmiş ölçütler bulunmadığı için bu yöntemler tabloya eklenmemiştir.

lı seyrediyorsa, müdahalenin etkili olduğu kabul edilir. Acil etki, bir evreden diğer evreye geçildiğinde bağımlı değişkende gözlenen değişikliğin hızıdır. Değişikliğin hızı arttıkça, müdahalenin etkisinin de arttığı kabul edilir (Alberto ve Troutman, 2009; Rakap, 2017b). Örtüşme ise, ardışık evrelerde yer alan verilerin birbirlerine benzer olması, diğer bir deyişle, verilerin sayısal değerlerinin aynı aralık içinde yer almalarıdır (Kennedy, 2005; Rakap, 2017b). İki evre arasında örtüşme olmadığında ya da az olduğunda müdahalenin etkili olduğu varsayılır. Bu değişkenlerden hiçbiri müdahalenin etkililiği konusunda karar vermek için tek başına yeterli olmadığından, tüm değişkenlerin birlikte ele alınması gerekir.

Tek-denekli deneysel araştırma desenlerini kullanan çalışmalarda etki büyüklüğü hesaplamalarının yapılması ve rapor edilmesi bağlamında, önemli konulardan biri, görsel analiz sonuçları ile etki büyüklüğü hesaplamaları arasındaki ilişkinin incelenmesidir. ÖRVEY'in ortaya çıkışından itibaren etki büyüklüğü yöntemlerini birbiriyle ve görsel analiz sonuçları ile kıyaslayan birçok çalışma bulunmaktadır. Örneğin, Ma (2006), MAVY ile görsel analiz sonuçları arasındaki korelasyonun ÖRVEY ile görsel analiz sonuçları arasındaki korelasyondan daha yüksek olduğunu ortaya koymuştur. Parker ve Hagan-Burke (2007), MAVY, ÖRVEY ve İLOF yöntemlerini inceledikleri çalışmada, bu yöntemler arasında İLOF'un görsel analiz sonuçları ile en uyumlu yöntem olduğunu, görsel analiz sonuçları ile en az uyumlu olan yöntemin de MAVY olduğunu rapor etmiştir. Başka bir çalışmada Parker ve Vannest (2009), TÖÇ yöntemini ÖRVEY, MAVY ve TÖVEY ile karşılaştırmış ve TÖÇ ile TÖVEY'in diğer iki yönteme oranla görsel analiz sonuçları ile daha uyumlu olduğunu ortaya koymuşlardır. Wolery ve diğerleri (2010), ÖRVEY, ÖVEÇ-KA, MAVY ve MEAVY yöntemlerini karşılaştırdıkları çalışmada, MAVY ve MEAVY yöntemlerinin diğer iki yönteme göre görsel analiz bulguları ile daha uyumlu olduğunu bulmuşlardır.

Bu çalışmaların sonuçları, bireysel olarak etki büyüklüğü değerleri ile görsel analiz bulguları arasındaki ilişki hakkında bilgi verici nitelikte olmasına rağmen farklı veri setleri ve grafikler kullanılarak elde edildikleri için etki büyüklüğü yöntemlerinin kapsamlı ve karşılaştırmalı olarak değerlendirilmesine olanak tanımamaktadır. Bu bağlamda, Yucesoy-Ozkan ve arkadaşları (2020), aynı veri ve grafik setini kullanarak bu makalede incelenen, örtüşmeyen veriye dayalı 12 yöntemi birbiriyle ve görsel analiz bulguları ile karşılaştırmıştır. Bu çalışmanın bulguları, TÖÇ, ÖRVEY, MEAVY ve İLOF yöntemlerinin görsel analiz ile en uyumlu yöntemler olduğunu ortaya koymuştur. Yucesoy-Ozkan ve arkadaşları (2020) tarafından yürütülen çalışmanın başka veri ve grafik setleri kullanılarak tekrarlanması ve sonuçların

karşılaştırılması, örtüşmeyen veriye dayalı etki büyüklüğü yöntemlerinin işleyişi hakkında daha kapsamlı bilgi elde edilmesine olanak sağlayacaktır.

Etki Büyüklüğü Yöntemini Seçerken Dikkat Edilmesi Gereken Noktalar

Kullanılacak yöntemin seçilmesinde tek-denekli deneysel araştırmalardan elde edilen verilerin özellikleri (örn., veri noktası sayısı, başlama düzeyinde istenmeyen yönde eğilimin bulunması ve örtüşme oranı) dikkate alınmalıdır (Brossart ve ark., 2014). Örtüşmeyen veriye dayalı yöntemler, başlama düzeyi ve uygulama evrelerindeki verilerin oranları dikkate alınarak hesaplandığından, az sayıda veri noktası sayısı, etki büyüklüğü değerinin gereğinden küçük ya da büyük çıkmasına neden olabilir (Lenz, 2013). Başlama düzeyi evresinde en az üç ya da beş veri noktasının bulunması bir gereklilik olmasına rağmen tutarlı bir başlama düzeyi verisi elde edilmeden uygulama evresine geçilmemelidir. Bu bağlamda, başlama düzeyi evresinde istenmeyen yönde bir eğilimin varlığı gözlemlenmişse (ya da bekleniyorsa), yeterli düzeyde veri toplamak daha önemli hale gelmektedir (Brossart ve ark., 2014). Başlama düzeyi ve uygulama evrelerinde örtüşmenin çok olduğu veri setlerinde, hesaplama sırasında bu iki evredeki verilerin bireysel olarak karşılaştırılmasına olanak tanıyan yöntemler tercih edilmelidir. Bu bağlamda, görsel analizle uyum ile ilgili çalışma bulguları da göz önünde bulundurularak, Tau-*U* ya da MEAVY ile ÖRVEY, TÖÇ ya da İLOF yöntemlerinin bir ya da birkaçının birlikte kullanılması önerilebilir.

Sonuç

Kanıt-dayalı uygulamaların belirlenmesine yönelik çalışmalar doğrultusunda tek-denekli deneysel araştırma desenlerini kullanan çalışmalarda etki büyüklüğü kullanımına yönelik gereksinim giderek artarken, etki büyüklüğü hesaplamak için kullanılacak yöntemler üzerinde bir fikir birliği henüz oluşmamıştır. Bu bağlamda, parametrik yöntemleri, standartlaştırılmış ortalama farkına dayalı yöntemleri ve örtüşmeyen veriye dayalı yöntemleri karşılaştırmalı olarak inceleyen çalışmaların yürütülmesine gereksinim duyulmaktadır. Etkili ve tek-denekli deneysel araştırmalardan elde edilen verilerdeki tüm örüntüleri (düzey, eğilim ve dalgalanma) dikkate alan bir yöntem geliştirilinceye kadar tek-denekli deneysel araştırmalarda etki büyüklüğü hesaplanırken, birden fazla yöntemin bir arada kullanılması ve rapor edilmesi gerekmektedir (Brossart ve ark., 2014; Kratochwill ve ark., 2010; Rakap ve ark., 2014). Ancak, her şart ve durumda unutulmaması gereken en önemli nokta, tek-denekli deneysel araştırmalarda bağımlı ve bağımsız

değişkenler arasındaki nedensel ya da işlevsel ilişkinin belirlenmesinde kullanılacak birincil yöntemin görsel analiz olduğu ve etki büyüklüğü hesaplamalarının görsel analizi destekleyici ve tamamlayıcı olarak kullanılması gerektiğidir. Dahası, görsel analiz ve etki büyüklüğü hesaplamaları yapıldıktan sonra araştırma bulguları yorumlanırken araştırmanın yürütüldüğü bağlam ve değiştirilmesi hedeflenen davranışın klinik önemi de göz önünde bulundurulmalıdır.

Kaynaklar

- Alberto, P. A. ve Troutman, A. C. (2009). *Applied behavior analysis for teacher: (8. baskı)*. Upper Saddle River, NJ: Pearson Education.
- Allison, D. B. ve Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavioral Research*, 31(6), 621-631.
- Altun, D. E. ve Yucesoy-Ozkan, S. (2018). Serbest zaman becerilerinin öğretiminde hata düzeltmesi yapılarak ve yapılmadan sunulan video ipucunun karşılaştırılması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 37(2), 15-33.
- Bozkus-Genc, G. ve Yucesoy-Ozkan, S. (2016). Meta-analysis of pivotal response training for children with autism spectrum disorder. *Education and Training in Autism and Developmental Disabilities*, 51(1), 13-26.
- Boyle, M. A., Samaha, A. L., Rodewald, A. M. ve Hofmann, A. N. (2013). Evaluation of the reliability and validity of GraphClick as a data extraction program. *Computers in Human Behavior*, 29(3), 1023-1027.
- Brossart, D. F., Vannest, K. J., Davis, J. L. ve Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation*, 24(3-4), 464-491.
- Busk, P. L. ve Serlin, R. C. (1992). Meta-analysis for single-case research. R. Kratochwil ve J. R. Levin (Ed.), *Single-case research design and analysis: New directions for psychology and education* içinde (ss. 133-157). Hillsdale, NJ: Lawrence Erlbaum.
- Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 24(2), 120-138.
- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject design. *Behavior Modification*, 28(2), 234-246.
- Center, B. A., Skiba, R. J. ve Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education*, 19(4), 387-400.
- Cohen, J. (1965). Some statistical issues in psychological research. B. B. Wolman (Ed.), *Handbook of clinical psychology* içinde (ss. 95-121). New York, NY: McGraw-Hill.
- Cohen, R. (1988). *Statistical power analysis for the behavioral sciences* (2. baskı). New York, NY: Routledge.

- Cook, B. G. ve Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education, 47*(2), 71-82.
- Cooper, J. O., Heron, T. E. ve Heward, W. L. (2014). *Applied behavior analysis*. (2. baskı). Harlow, Essex: Pearson Education Limited.
- Corder, G. W. ve Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: A step-by-step approach*. Hoboken, NJ: Wiley.
- Drevon, D., Fursa, S. R. ve Malcolm, A. L. (2017). Inter-coder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification, 41*(2), 323-339.
- Faith, M. S., Allison, D. B. ve Gorman, B. S. (1996). Meta-analysis of single-case research. R. D. Franklin, D. B. Allison ve B. S. Gorman (Ed.), *Design and analysis of single-case research* içinde (ss. 245-277). Hillsdale, NJ: Lawrence Erlbaum.
- Hedges, L. V., Pustejovsky, J. E. ve Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3*(3), 224-239.
- Hinton P. R. (2004). *Statistics explained: A guide for social science students* (2 baskı). New York, NY: Routledge.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson Education.
- Kratochwill, T. R. ve Levin, J. R. (Ed.). (1992). *Single-case research design and analysis: New developments for psychology and education*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., ve Shadish, W. R. (2010). *Single-case designs technical documentation*. (19.10.2018 tarihinde http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf adresinden alınmıştır.)
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. ve Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38.
- Lenz, A. S. (2013). Calculating effect size in single-case research: A comparison of nonoverlap methods. *Measurement and Evaluation in Counseling and Development, 46*(1), 64-73.
- 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.
- Ma, H. H. (2009). The effectiveness of intervention on the behavior of individuals with autism: A meta-analysis using percentage of data points exceeding the median of baseline Phase (PEM). *Behavior Modification, 3*(3), 339-359.
- Maggin, D. M., Briesch, A. M. ve 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.
- Manalov, R. ve Solonas, A. (2009). Percentage of non-overlapping corrected data. *Behavior Research Methods, 41*(4), 1262-1271.
- Moeyaert, M., Maggin, D. ve Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification, 40*(6), 874-900.
- NCLB (2002). No Child Left Behind, U.S. Department of Education, Washington D. C.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B. ve Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137-149.
- Olive, M. L. ve Franco, J. H. (2007). (Effect) size matters: And so does the calculation. *The Behavior Analyst Today, 8*(2), 76-86.
- Olive, M. L. ve Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology, 25*(2-3), 313-324.
- Parker, R. I. ve Hagen-Burke, S. (2007). Median-based overlap analysis for single case data: A second study. *Behavior Modification, 31*(6), 919-936.
- Parker, R. I., Hagan-Burke, S. ve Vannest, K. (2007). Percentage of all non-overlapping data (PAND) an alternative to PND. *The Journal of Special Education, 40*(4), 194-204.
- Parker, R. I. ve Vannest, K. J. (2007). *Pairwise data overlap for single case research*. Yayınlanmamış makale.
- Parker, R. I. ve Vannest, K. J. (2009). An improved effect size for single case research: Nonoverlap of all pairs (NAP). *Behavior Therapy, 40*(4), 357-367.
- Parker, R. I., Vannest, K. J. ve Brown, L. (2009). The improvement rate difference for single case research. *Exceptional Children, 75*(2), 135-150.
- Parker, R. I., Vannest, K. J. ve Davis, J. L. (2011). Effect size in single-case research: A review of nine non-overlap techniques. *Behavioral Modification, 35*(4), 303-322.
- Parker, R. I., Vannest, K. J., Davis, J. L. ve Sauber, S. B. (2011). Combining nonoverlap and trend for sing-

- le-case research: Tau-U. *Behavior Therapy*, 42(2), 284-299.
- Petersen-Brown, S., Karich, A. C. ve Symons, F. J. (2012). Examining estimates of effect using non-overlap of all pairs in multiple baseline studies of academic intervention. *Journal of Behavioral Education*, 21(3), 203-216.
- Qi, C. H., Barton, E. E., Collier, M., Lin, Y. L. ve Montoya, C. (2015). A systematic review of effects of social stories interventions for individuals with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 33(1), 25-34.
- Rakap, S. (2015). Effect sizes as result interpretation aids in single-subject experimental research: description and application of four nonoverlap methods. *British Journal of Special Education*, 42(1), 11-33.
- Rakap, S. (2016). Özel eğitimde bilimsel dayanaklı uygulamalar. V. Aksoy (Ed.), *Özel eğitim içinde* (ss. 181- 211). Ankara: Pegem Akademi.
- Rakap, S. (2017a). Impact of coaching on preservice teachers' use of embedded instruction in inclusive preschool classrooms. *Journal of Teacher Education*, 68(2), 125-139.
- Rakap, S. (2017b). Tek denekli deneysel araştırma yöntemleri. D. Erbas ve S. Yucesoy-Ozkan (Ed.), *Uygulamalı davranış analizi içinde* (ss. 153-212). Ankara: Pegem Akademi.
- Rakap, S. ve Balıkcı, S. (2017). Using embedded instruction to teach functional skills to a preschool child with autism. *International Journal of Developmental Disabilities*, 63(1), 17-26.
- Rakap, S., Rakap, S., Evran, D. ve Cig, O. (2016). Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphC-lick, and DigitizeIt. *Computers in Human Behavior*, 55, 159-166.
- Rakap, S., Snyder, P. ve Pasia, C. (2014). Comparison of nonoverlap methods for identifying treatment effect in single-subject experimental research. *Behavioral Disorders*, 39(3), 128-145.
- Reichow, B. (2016). Evidence-based practice in the context of early childhood special education. B. Reichow, B. A. Boyd, E. E. Barton ve S. L. Odom (Ed.), *Handbook of early childhood special education içinde* (ss. 107-124). New York, NY: Springer.
- Sazak-Pınar, E. ve Merdan, F. (2016). Grafik düzenleyicilerin otizmliler öğrencilere fen bilgisi kavramlarının öğretimindeki etkililiği. *Ahi Evren Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 17(1), 111-131.
- Schneider, N., Goldstein, H. ve Parker, R. (2008). Social skills interventions for children with autism: A meta-analytic application of percentage of all non-overlapping data (PAND). *Evidence-Based Communication Assessment and Intervention*, 2(3), 152-162.
- Scotti, J. R., Evans, I. M., Meyer, L. H. ve Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation*, 96(3), 233-256.
- Scruggs T. E. ve Mastropieri M. A. (1994). The effectiveness of generalization training: A quantitative synthesis of single subject research. T. E. Scruggs ve M. A. Mastropieri (Ed.), *Advances in learning and behavioral disabilities içinde* (ss. 259-280). Bingley, UK: Emerald.
- Scruggs, T. E. ve Mastropieri, M. A. (1998). Synthesizing single subject research: Issues and applications. *Behavior Modification*, 22(3), 221-242.
- Scruggs, T. E. ve Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality*, 9(4), 227-244.
- Scruggs, T. E. ve 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.
- Scruggs, T. E., Mastropieri, M. A. ve Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education*, 8(2), 24-33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. ve Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders*, 11(4), 260-271.
- Scruggs, T. E., Mastropieri, M. A., Forness, S. R. ve Kavale, K. A. (1988). Early language intervention: A quantitative synthesis of single-subject research. *Journal of Special Education*, 22, 259-283.
- Shadish, W. R., Brasil, I. C., Illingworth, D. A., White, K. D., Galindo, R., Nagler, E. D. ve Rindskopf, D. M. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods*, 41(1), 177-183.
- Sönmez, M. ve Diken, İ. H. (2010). Problem davranışların azaltılmasında işlevsel iletişim öğretiminin etkililiği: Betimsel ve meta-analiz çalışması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 11(1), 1-16.
- Tavil, Y. Z. ve Karasu, N. (2013). Aile eğitim çalışmaları: Bir gözden geçirme ve meta-analiz örneği. *Eğitim ve Bilim*, 38(168), 85-95.
- Toper-Kormaz, Ö. ve Diken, İ. (2010). Stereotipik davranışların azaltılmasında kullanılan yöntemlerin etkililiği: Betimsel ve meta analiz. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 11(2), 1-12.

- Van den Noortgate, W. ve Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods*, 35(1), 1-10.
- Vannest, K. J. ve Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling and Development*, 93(4), 403-411.
- Wendt, O. (2009, May). Calculating effect sizes for single subject experimental designs: An overview and comparison. Dokuzuncu Campbell Collaboration Colloquium'da sunulan bildiri, Oslo, Norway.
- White, O. R. ve Haring, N. G. (1980). *Exceptional teaching: A multimedia training package*. Columbus, OH: Merrill.
- Wolery, M. Busick, M., Reichow, B. ve Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44(1), 18-28.
- Yucesoy-Ozkan, S., Gulboy, E. ve Kaya, F. (2018). Teaching children with intellectual disabilities through video prompting: Smartphone vs. tablet. *International Journal of Early Childhood Special Education*, 10(1), 33-49.
- 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.
- Yucesoy-Ozkan, S. ve Sonmez, M. (2011). Examination of single subject studies conducted on individuals with disabilities by using self management strategies: A meta analysis study. *Educational Sciences: Theory and Practice*, 11(2), 809-821.

Summary

Effect Size Calculations in Single-Case Experimental Research Design: An Investigation of Nonoverlap Methods

Salih Rakap

Ondokuz Mayıs University

Şerife Yücesoy-Özkan

Anadolu University

Sinan Kalkan

Çanakkale Onsekiz Mart University

Initially used in the medical field in the 1970s, evidence-based practices have become important in many areas including psychology, health sciences, nursing, and educational sciences (Odom et al. 2005; Rakap, 2016; Reichow, 2016). No Child Left Behind (NCLB, 2002) law amended in 2002 brought radical changes in education in the United States. NCLB required all schools and teachers to use evidence-based practices while teaching. Following the passage of this law, many researchers and organizations in the United States and around the world began to work on determining evidence-based practices in general and special education through rigorous meta-analysis of previous research (Odom et al., 2005; Olive & Franco, 2007).

An important limitation of these meta-analyses has been the exclusion of studies employing single-case research (SCER) designs (Maggin, Briesch, & Chafouleas, 2013; Rakap, Snyder, & Pasia, 2014; Wolery, Busick, Reichow, & Barton, 2010). SCER designs are commonly used in psychology, social work, special education, and applied behavior analysis disciplines and allow rigorous investigation of the causal or functional relationship between dependent and independent variables. The main reason for the exclusion of SCER studies from meta-analyses of evidence-based practices is the lack of agreed-upon methods to calculate effect size estimates for data obtained from SCER studies (Campbell, 2004; Parker & Hagan-Burke, 2007; Rakap et al., 2014; Wolery et al., 2010). To overcome this problem, several research groups have been working on developing effect size calculation methods for SCER. These methods can be grouped under three main categories: parametric methods (Center, Skiba, & Casey, 1985; Maggin et al., 2013; Van den Noortgate & Onghena, 2003), standardized mean difference (Hedges, Pustejovsky, & Shadish, 2012), and nonoverlap methods (Parker et al., 2011; Scruggs et al., 1987).

International literature includes a number of differ-

ent nonoverlap methods that are similar but have small differences. These include Percentage of Nonoverlapping Data (PND; Scruggs et al., 1987), Percentage of Zero Data (PZD; Scotti, Evans, Meyer, & Walker, 1991), Percentage of Data Exceeding the Median (PEM; Ma, 2006), Percentage of All Nonoverlapping Data (PAND; Parker & Hagan-Burke, 2007), Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009), Improvement Rate Difference (IRD; Parker, Vannest, & Brown, 2009), Percentage of Data Exceeding a Median Trend (PEM-T; White & Haring, 1980; Wolery et al., 2010), Pairwise Data Overlap (PDO; Parker & Vannest, 2007), Pairwise Data Overlap Squared (PDO²; Wolery et al., 2010), Tau_{Novlap} (Parker, Vannest, Davis et al., 2011), Tau-U (Parker, Vannest, Davis et al., 2011) and Phi (Parker, Vannest, & Davis, 2011). Although methods developed more recently are more sophisticated, many national and international researchers still use early methods, such as PND or PZD when calculating effect size estimates in SCER (e.g., Bozkus-Genc & Yucesoy-Ozkan, 2016; Sonmez & Diken, 2010; Sazak-Pinar & Merdan, 2016; Tavit & Karasu, 2013; Toper-Korkmaz & Diken, 2010; Yucesoy-Ozkan & Sonmez, 2011).

Considering the emphasize of the American Psychology Association (APA) and What Works Clearinghouse (WWC) on the use of effect size estimates to supplement the findings of studies using group or SCER, it has become important for national researchers to know characteristics of nonoverlap methods, their advantages and limitations, and when to use which methods based on data patterns. The purpose of the present study is to describe 13 nonoverlap methods used to calculate effect size estimates in SCER, explain their calculations using hypothetical data sets, discuss their advantages and disadvantages, and provide a guideline for interpreting effect size estimates and considerations for selecting appropriate effect size methods based data obtained from SCER studies.

Address for Correspondence: Asst. Prof. Salih Rakap, Ondokuz Mayıs University, Faculty of Education, Special Education, Atakum / Samsun

E-mail: salih.rakap@omu.edu.tr

Effect Size Methods and Calculations

Percentage of Nonoverlapping Data

Percentage of nonoverlapping data (PND; Scruggs et al., 1987) is the first method developed to calculate effect size estimates in SCER. PND is conceptualized as the percentage of data points in the intervention phase that have higher (or lower for studies aiming to reduce behavior) numeric value than the highest datum point in the baseline phase (Parker, Vannest, & Davis, 2011; Scruggs et al., 1987; Scruggs & Mastropieri, 2013). PND effect size estimate is calculated using the following steps: (a) determining the datum point with the highest numeric value in the baseline phase, (b) drawing a line (nonoverlap line) that is parallel to x-axis from this datum point towards the intervention phase, (c) counting the number of data points in the intervention phase that remain above the nonoverlap line, (d) dividing the sum obtain in step c to total number of data points in the intervention phase, and (e) multiplying the result obtained in step d by 100 (Scruggs & Mastropieri, 1998; Scruggs et al., 1987).

Percentage of Data Exceeding the Median

Percentage of data exceeding the median (PEM; Ma, 2006) is conceptualized as the percentage of data points in the intervention phase that is above the median line drawn based on the baseline phase data. PEM effect size estimate is calculated by (a) determining the median of the baseline phase data, (b) drawing a line (median line) that is parallel to x-axis from this datum point towards the intervention phase, (c) counting the number of data points in the intervention phase that remain above the median line, (d) dividing the sum obtain in step c to total number of data points in the intervention phase (Ma, 2006), and (e) multiplying the result obtained in step d by 100 (Parker, Vannest, & Davis, 2011).

Percentage of Data Exceeding a Median Trend

Percentage of data exceeding a median trend (PEM-T; White & Haring, 1980; Wolery et al., 2010) is conceptualized as the percentage of data in the intervention phase that is above the trend line drawn based on data in the baseline phase. PEM-T effect size estimate is calculated by (a) drawing a line from the middle point of the first two data points to the last two data points in the baseline phase (trend line) and extending the line across the intervention phase, (b) counting the number of data points in the intervention phase that remain above the trend line, (c) dividing the sum obtain in step b to total number of data points in the intervention phase, and (d) multiplying the result obtained in step c by 100 (Rakap et al., 2014; Wolery et al., 2010).

Percentage of Zero Data

Percentage of zero data (PZD; (Scotti et al., 1991) is the only nonoverlap method specifically developed for calculating effect size estimates for SSER studies aiming to decrease behaviors. PZD is calculated by finding the first data point in the intervention phase that reaches zero and calculating the percentage of data points obtained in the intervention phase, including the first zero point, that remains at zero (Scotti et al., 1991).

Percentage of All Nonoverlapping Data

Percentage of all nonoverlapping data (PAND; Parker & Hagan-Burke, 2007) is conceptualized as the percentage of data remaining in the baseline and intervention phases after the minimum number of data points is removed to eliminate the overlap between the baseline and intervention phases. PAND is calculated by (a) determining the total number of data points across the baseline and intervention phases, (b) determining minimum number of data points that eliminates the overlap between the baseline and intervention phases, (c) subtracting the number obtained in step b from the sum obtained in step a, (d) dividing the result obtained in step c by the sum obtained in step a, and (e) multiplied the quotient obtained in step d by 100 (Parker & Hagan-Burke, 2007; Parker, Vannest, & Davis, 2011).

Percentage of Non-overlapping Corrected Data

Percentage of non-overlapping corrected data (PNCD; Manalov & Solonas, 2009) is developed to overcome the limitations of PND by offering a correction procedure implemented prior to the application of PND. PNCD aims to control for unwanted trend observed in baseline phase. PNCD is calculated by (a) subtracting each datum point in baseline phase from the consecutive data point to create a difference series with $n_A - 1$ data points, (b) calculating mean of the difference series, (c) computing the trend-correction factor for each datum point across phases by multiplying the mean calculated in step b with the sequence number of each datum point, (d) subtracting the trend-correction factor calculated for each datum point from the original value of the datum point, and (e) applying the PND procedure to calculate PNCD.

Improvement Rate Difference

Improvement rate difference (IRD; Parker, Vannest, & Brown, 2009) was initially developed in the medical field as risk-reduction or risk difference method. IRD is conceptualized as the difference between the improvement rates of the baseline and intervention phases. To calculate IRD, first, the total number of data points across baseline and intervention phases is determined,

and then the minimum number of data points that eliminate the overlap between the baseline and intervention phases is determined. When all data points are removed from the intervention phase to eliminate the overlap, the IRD is calculated by (a) dividing the number of remaining data points in the intervention phase after removal to the total number of data points in the intervention phase, and (b) multiplying the quotient by 100. When all data points are removed from the baseline phase, the IRD is calculated by (a) dividing the number of remaining data points in the baseline phase after removal to the total number of data points in the baseline phase, (b) subtracting the quotient from 1, and (c) multiplying the result by 100. When the data points are removed both from baseline and intervention phases to eliminate the overlap, IRD is calculated by (a) dividing the number of remaining data points in the intervention phase by the total number of data points in the intervention phase, (b) dividing the number of data points removed from the baseline phase by the total number of data points in the baseline phase, (c) subtracting the quotient from step a from step b, and (d) multiplying the result by 100 (Parker et al., 2009; Vannest & Ninci, 2015).

Pearson's Phi

Pearson's Phi (Parker, Vannest, & Davis, 2011) is developed as an extension to PAND. Phi is calculated by (a) determining the minimum number of data points removed from the baseline and intervention phases to eliminate the overlap between phases, (b) using half of the number of data points removed to create ratios for the baseline and intervention phases, (c) submitting these two ratios in 2×2 table to cross-tab analysis, and (d) multiplying the result by 100 (Parker, Vannest, & Davis, 2011).

Nonoverlap of All Pairs

Nonoverlap of all pairs (NAP; Parker & Vannest, 2009) is conceptualized as the percentage of all pairwise comparisons across the baseline and intervention phases that show improvement from the baseline phase to the intervention phase. NAP is calculated in four steps: (a) determining the total number of pairs by multiplying the number of data points in the baseline and intervention phases, (b) comparing each datum point in the baseline phase with each datum point in the intervention phase to determine the number of pairs in which the intervention phase datum point has a higher numeric value than the datum point in the baseline phase (improving data points) and the number of pairs in which numeric values of the data points across phases are equal (tie), (c) summing the total number of improved pairs and half of the pairs with tie, (d) dividing the sum obtained in step

c by the total number of pairs, and (e) multiplying the quotient by 100 (Parker & Vannest, 2009).

Pairwise Data Overlap and Pairwise Data Overlap Squared

Pairwise data overlap (PDO; Parker & Vannest, 2007) is also conceptualized as the percentage of all pairwise comparisons across the baseline and intervention phases that show improvement from the baseline phase to the intervention phase. Calculation of effect size estimates using PDO is very similar to the calculation of effect size estimates using NAP; the only difference between the two methods is that PDO does not take ties (pairs with equal values) into account. PDO^2 (Wolery et al., 2010) is calculated by squaring the effect size estimate obtained from PDO analysis.

τ_{Novlap}

τ_{Novlap} (Parker, Vannest, Davis et al., 2011) is also based on pairwise comparisons of data as in NAP and PDO. To calculate effect size estimates using τ_{Novlap} , each datum point in the baseline phase is compared to each datum point in the intervention phase. Next, the total number of decreasing pairs (i.e., when a baseline datum point > an intervention datum point) is subtracted from the total number of pairs with improvement (i.e., when an intervention datum point > a baseline datum point). The result is divided by the total number of pairs, and resulting quotient is multiplied by 100 (Parker, Vannest, Davis et al., 2011).

τ_{U}

τ_{U} (Parker, Vannest, Davis et al., 2011) is the second nonoverlap method that can control for a baseline trend in a therapeutic direction. τ_{U} is the most sophisticated nonoverlap method to calculate effect size estimates in SCER. Calculation of τ_{U} statistics is more complex than the calculations of effect size estimates using the other nonoverlap methods. τ_{U} can be considered as an extension to Kendall's rank order correlation and Mann-Whitney U test. To calculate effect size estimates using τ_{U} , (a) the total number of pairs is determined by multiplying the number of data points in baseline and intervention phases, (b) Kendall's rank order correlation analysis is conducted to obtain S value, (c) S value is divided by the total number of pairs calculated in step a, and (d) the quotient is multiplied by 100 (Parker, Vannest, Davis et al., 2011).

Interpreting Effect Size Estimates

Researchers have developed benchmarks for interpreting effect size estimates calculated using different

methods. Effect size calculation methods such as PND, PZD, IRD, NAP, and Tau-*U* have their own benchmarks specifically developed for these methods. However, methods such as PEM, PEM-T, PAND, PDO, PDO², and Tau_{Novlap} do not have their own criteria to evaluate obtained effect sizes, and many researchers use benchmarks developed for PND while interpreting effect size estimates calculated using these methods. This may be problematic especially when the characteristics of effect size calculation methods are considerably different from PND. Researchers must be cautious when using benchmarks developed for another method in their research. Moreover, effect size estimates should never be used alone to evaluate the effectiveness of an intervention in SSER studies; they should always be used in conjunction with visual analysis.

Effect Size Estimates and Visual Analysis

Numerous studies have been conducted to investigate the relationships between the magnitude of effect size estimates and results of visual analyses since the inception of the PND (e.g., Ma, 2006; Parker & Hagan-Burke, 2007; Parker & Vannest, 2009; Rakap et al., 2014; Rakap et al., 2017; Wolery et al., 2010). For example, Ma (2006) reported that the correlation between PEM and visual analysis results were higher than the correlation between PND and visual analysis. Parker and Hagan-Burke (2007) found that IRD had the highest correlation with visual analysis while PEM had the lowest correlation with the results of the visual analysis. In two different studies, Rakap and colleagues (2014; 2017) compared the effect size estimates calculated by 13 different nonoverlap methods and visual analysis results using the same data set and reported that PND was the most compatible effect size methods with visual analysis followed by PEM-T, NAP, Tau_{NOVLAP}, IRD, and PAND, while PDO² was the least compatible method. More research is needed to determine the most compatible effect size methods with visual analysis using different datasets.

Considerations for Selecting Effect Size Methods

When selecting the appropriate effect size method for a SCER study, several characteristics of data and data patterns (e.g., number of data points in each phase, a baseline trend in therapeutic direction, and overlap level) should be considered (Brossart et al., 2014). A number of data points in each phase significantly influence the effect size estimates calculated using nonoverlap methods. Although three or five data points in the baseline phase are acceptable in SCER, a stable data patterns

should be obtained before an intervention is implemented (Brossart et al., 2014; Lenz, 2013). Moreover, when a baseline trend in therapeutic direction exists, baseline phase should be extended until a stable pattern of data is observed and an effect size estimate (e.g., Tau-*U* or PEM-T) that controls for the baseline trend should be used to calculate effect sizes (Brossart et al., 2014). In data sets where there is a lot of overlapping data between the baseline and intervention phases, one of the effect size methods that allows comparison of individual data points across phases should be used. Overall, considering their consistency with the results of visual analysis and ability to control for some of the characteristics of data obtained from SCER studies, it can be recommended that PEM-T or Tau-*U*, along with PND, NAP or IRD should be used together to calculate and report effect sizes estimates in SCER.

Conclusions

The need for using effect size estimates to determine the experimental effect in SCER is increasing within the context of identifying evidence-based practices; however, there is no consensus about which effect size methods should be used to calculate effect sizes in SCER. In this respect, there is a need for more studies comparing and contrasting parametric methods, methods based standardized mean difference, and nonoverlap methods as well as their relationships with visual analysis. Until a method that controls for all characteristics and data patterns (e.g., level, trend and variability) of SCER is developed, more than one method to calculate and report effect size estimates should be used (Brossart et al., 2014; Kratochwill et al., 2010; Rakap et al., 2014). Nevertheless, an important point that should always be remembered under any circumstances is that the primary method for determining causal or functional relationships between dependent and independent variables in SCER is visual analysis and effect size estimates must be used to supplement and support findings of the visual analysis. Moreover, when interpreting the results of visual analysis and effect size calculations, researchers should consider the context in which the study is conducted and clinical importance of the target behavior to be changed.