

Tek Denekli Deneysel Çalışmalarda Etki Büyüklüğü Hesaplaması: Regresyona Dayalı Olmayan Yöntemlerin İncelenmesi*

Nihal ŞEN**

Sedat ŞEN***

Öz

Tek denekli deneysel çalışmalarda meta-analiz yöntemlerine grup desenli çalışmalarda olduğu kadar yer verilmediği gözlenmektedir. Son yıllarda alanyazındaki bu eksikliği gidermek amacıyla tek denekli deneysel çalışmalarda etki büyüklüğü olarak kullanılabilir regresyona dayalı olan ve regresyona dayalı olmayan indeksler geliştirilmiştir. Regresyona dayalı olan indekslerin çoğu tek denekli deneysel araştırma verilerindeki zamansal bağımlılıktan etkilendiği için bu indekslerden daha az etkilenen ve alanyazında daha sıklıkla tercih edilen regresyona dayalı olmayan indeksler bu çalışmanın ana konusu olmuştur. Tek denekli deneysel çalışmalarda regresyona dayalı olmayan birçok indeks üretilmiş olmasına rağmen araştırmacıların çoğunun örtüşmeyen veri yüzdesi (ÖVY) ve sıfır veri yüzdesi (SVY) indekslerini tercih ettiği gözlenmiştir. Alanyazında, özellikle ÖVY indeksinin kullanılmasına ilişkin çok sayıda tartışma bulunmaktadır. Türkiye’deki alanyazında ÖVY ve SVY’ye alternatif olarak üretilen indeksleri açıklayan mevcut bir çalışma olmaması regresyona dayalı olmayan indekslerin incelendiği bu çalışmayı önemli kılmaktadır. Bu çalışmanın amacı tek denekli deneysel çalışmalarda etki büyüklüğü hesaplamak için kullanılan regresyona dayalı olmayan yöntemlerin incelenmesi ve bu yöntemlerin tek denekli deneysel araştırmalarda nasıl uygulanacağına örnek veriler üzerinden gösterilmesidir. Bu amacı gerçekleştirirken farklı yöntemlerin kullanımında verinin nasıl hazırlanacağı, nasıl analiz edileceği, birden fazla çalışma için nasıl sentezleneceği ve elde edilen sonuçların nasıl yorumlanacağından bahsedilmiştir. Bu çalışmada incelenen 10 farklı indeks temelinde araştırmacılar için önerilerde bulunulmuştur.

Anahtar Kelimeler: Tek denekli deneysel araştırmalar, meta-analiz, etki büyüklüğü, regresyona dayalı olmayan indeksler.

GİRİŞ

Son yıllarda eğitim alanında yapılan çalışmalarda yeni eğilimler bir alana ait çalışmalardan elde edilen verilerin sentezlenmesine daha çok önem verildiğini göstermektedir. Hattie (2009), eğitimde araştırma ve uygulama arasındaki farkın kapatılması için uygulayıcıların ve politikacıların, meta-analiz yoluyla elde edilen çeşitli kanıt türlerini özetlemeleri ve karşılaştırmaları gerektiğini belirtmiştir (Kavale, 2001). Meta-analiz (Glass, 1976), bir alanda yapılan çalışmaların istatistiksel olarak özetlenmesini sağlayan bir yöntemdir. Meta-analiz temel olarak spesifik bir konuda yapılan nicel çalışmalardan elde edilen etki büyüklüğü değerlerini kullanarak o konuya ait genel ortalama değerini hesaplamaya yardımcı olmaktadır.

Araştırmacılar için çok sayıda fayda sağlayan meta-analiz yöntemi, geleneksel meta-analiz çalışmalarında standartlaştırılmış ortalama farkı, korelasyon ve risk oranı gibi etki büyüklüğü değerleri ile uygulanabilmektedir (Lipsey ve Wilson, 2001). Deneysel çalışmalarda uygulanan müdahalenin etkililiğini belirlemek, araştırmacılar açısından büyük önem taşımaktadır. İstatistiksel anlamlılık yanında pratikteki anlamlılığa işaret eden etki büyüklüğü değerini hesaplamak araştırmacıya uygulama düzeyinin etkililiği hakkında yorumlanabilen (küçük etki, büyük etki gibi) bilgiler sunmaktadır.

* Bu çalışma 27. Uluslararası Eğitim Bilimleri (ICES/UEBK-2018) Kongresi’nde (18-22 Nisan 2018, Antalya) sözlü bildiri olarak sunulmuştur.

** Doktora öğrencisi, Bolu Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü, Özel Eğitim, Bolu, Türkiye, e-posta: nihallsenenn@gmail.com, ORCID ID: orcid.org/0000-0002-9511-8401

*** Dr. Öğr. Üyesi, Harran Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri, Şanlıurfa, Türkiye, e-posta: sedatsen@harran.edu.tr, ORCID ID: orcid.org/0000-0001-6962-4960

Bu makaleye atıfta bulunmak için:

Şen, N., & Şen, S. (2019). Calculation of effect size in single-subject experimental studies: Examination of non-regression-based methods. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 10(1), 30-48. DOI: 10.21031/epod.419625

Received: 30.04.2018

Accepted: 16.12.2018

Nitekim Amerikan Psikoloji Derneği, yayınlanan her çalışmanın mutlaka bir etki büyüklüğü değeri rapor etmesini tavsiye etmektedir (American Psychological Association, 2010).

Tek denekli deneysel araştırmalar (Kırcaali-İftar ve Tekin, 1997) uzun yıllar süren bir geçmişe sahip olmasına rağmen bu alandaki çalışmaları sentezlemeye yönelik meta-analiz uygulamalarının yapılması 1980'lerin sonuna doğru başlamıştır. Bu alanda yapılan çalışmaların çoğu tek denekli deneysel araştırmalardaki veri yapısından kaynaklanan problemlerden dolayı kabul edilebilir bir etki büyüklüğü geliştirmeyi başaramamıştır. Bu problemler arasında tek denekli deneysel çalışma verilerinde aynı birey üzerinde tekrarlı ölçümler yapılması, dolayısıyla artık değerlerin bağımlı olması (Huitema, 1985), ölçüm sayısının az olması ve bu ölçümlerin normal dağılım göstermemesi yer almaktadır. Bu problemler tek denekli deneysel çalışmalarda parametrik istatistiklerin bazı varsayımlarının sağlanmasını dolayısıyla bu istatistiklerin uygulanabilirliğini engellemiştir. Karşılaşılan bu sorunları çözebilmek adına tek denekli deneysel çalışmalarda uygulanabilecek meta-analizler için etki büyüklüğü değeri olarak kullanılacak birçok indeks önerilmiştir (Beretvas ve Chung, 2008). Bu indekslerin başında deney ve kontrol gruplarının (yani tek denekli deneysel araştırmalarda başlama ve uygulama düzeylerinin) standartlaştırılmış ortalama farkından hesaplanan tek göstergeye dayanan değerler (Busk ve Serlin, 1992) ve örtüşmeyen veri yüzdesi (Scruggs, Mastropieri ve Casto, 1987) gelmektedir. Bunlara ek olarak alanyazında özellikle yüzdeler hesabına dayanan çok sayıda indeks üretilmiştir (Ma, 2006; Parker ve Vannest, 2009; Parker, Vannest ve Brown, 2009; Parker, Vannest ve Davis, 2011). Bu parametrik olmayan yöntemlere ek olarak regresyon analizlerine dayanan çok sayıda istatistiksel yöntem de geliştirilmiştir (Allison ve Gorman, 1993; Huitema ve McKean, 2000; Swaminathan, Rogers, Horner, Sugai ve Smolkowski, 2014; van den Noortgate ve Onghena, 2003).

Alanyazında önerilen yöntemler tek denekli deneysel araştırmalardan elde edilecek farklı veri örüntüleri için uygundur. Özellikle regresyona dayalı olmayan yöntemler tek denekli deneysel alanyazında çok fazla kabul görmüştür ve hala kullanılmaktadır (Aslan, Yalçın ve Özdemir, 2016; Aydın, 2017; Karasu, 2009a, 2009b, 2011; Korkmaz ve Diken, 2010; Sönmez ve Diken, 2010; Tavail ve Karasu, 2013). Ayrıca ÖVY ve SVY indekslerinin Türkiye'deki araştırmacılar tarafından yaygın olarak kullanıldığı da gözlenmektedir. Özellikle ÖVY'nin eleştirilen sınırlılıklarına çözüm bulmak adına alanyazında regresyona dayalı olmayan pek çok alternatif etki büyüklüğü indeksi önerilmiştir. Bununla birlikte bu alternatif regresyona dayalı olmayan indekslerin Türkiye'deki araştırmacılar tarafından az da olsa kullanılmaya başlandığı gözlenmektedir (Bozkuş-Genç, 2017; Kaya, 2015; Uysal, 2017). Bu alternatif indekslerin uygulamalarda çok fazla yer almamasının nedenlerinden birisi Türkiye'de bu indeksleri açıklayan yöntemsel çalışmaların eksikliği olabilir. Yurt dışında birçok araştırmacı tek denekli deneysel çalışmalar için kullanılacak etki büyüklüğü değerlerini açıklayan ve karşılaştıran çalışmalar yürütmüştür (Alresheed, Holt ve Bano, 2013; Campbell, 2004; Heyvaert, Saenen, Campbell, Maes ve Onghena, 2014; Maggin, O'Keeffe ve Johnson, 2011; Maggin vd., 2011; Manolov ve Solanas, 2008; Manolov, Solanas ve Leiva, 2010; Olive ve Franco, 2008; Parker, Vannest ve Davis, 2011; Wolery, Busick, Reichow ve Barton, 2010). Alanyazındaki çalışmaların çoğu regresyona dayalı olmayan indekslerin karşılaştırılmasına dayalıdır. Parker, Vannest ve Davis'in (2011) dokuz indeksi karşılaştırdığı çalışması dışındakiler genel olarak birkaç indeks üzerinden karşılaştırma yapmıştır. Bu çalışmada tek denekli deneysel araştırmalarda kullanılan indekslerden 10 tanesinin açıklanması alanyazındaki çalışmalar arasında tek bir çalışmada ele alınan indeks sayısının fazlalığı açısından bu çalışmayı diğer çalışmalardan farklı kılmaktadır. Yapılmış olan geniş çaplı taramaya göre, Türkiye'de bu indeksleri karşılaştıran ya da açıklayan çalışma sayısı yok denecek kadar azdır. Bu anlamda verilebilecek tek örnek Karasu (2009a)'nın otizm spektrum bozukluğu olan çocukların iletişim ve sosyal becerilerini geliştirmek için doğal yaklaşımları temel alan sağaltım çalışmalarından seçilen bir grup çalışmanın sağladığı verileri etki büyüklüğü hesaplamalarında kullandığı çalışmadır. Bu çalışmada tek denekli deneysel araştırmalarda etki büyüklüğünü hesaplamak için geliştirilmiş yöntemlerden en kullanışlı olanını belirlemek amacıyla dört farklı yöntem; örtüşmeyen veri yüzdesi, sıfır veri yüzdesi, Swanson modeli ve ITSACORR (interrupted time-series analysis procedure) karşılaştırılmıştır. Dört farklı yöntemden elde edilen etki büyüklüğü değerleri ortalamalarının bir karşılaştırması ve bu yöntemlerin sonuçları arasındaki ilişkileri belirlemek adına korelasyon analizleri yapılmıştır.

Türkiye’deki alanyazında ÖVY ve SVY sonrası geliştirilen indeksleri açıklayan kapsamlı bir çalışma olmaması regresyona dayalı olmayan indekslerin incelendiği bu çalışmayı önemli kılmaktadır. Bu çalışmanın amacı tek denekli deneysel çalışmaların meta-analizinde uygulanan regresyona dayalı olmayan yöntemlerin incelenmesi ve bu yöntemlerin tek denekli deneysel araştırmalarda nasıl uygulanacağına örnek veriler üzerinde gösterilmesidir. Bu amacı gerçekleştirirken farklı yöntemlerin kullanımında verinin nasıl hazırlanacağı, nasıl analiz edileceği, birden fazla çalışma için nasıl sentezleneceği ve elde edilen sonuçların nasıl yorumlanacağından bahsedilecektir. Tek denekli deneysel yapıya uygun örnek veriler üzerinden elde edilen sonuçların bu indeksler arasında nasıl bir farklılık gösterdiği sorusu da bu çalışma ışığında cevaplanmaya çalışılmıştır. Bu çalışmada kullanılan tüm veriler yapay olarak üretilmiştir. SVY indeksi azalan davranışlardaki etkililiği belirlemede kullanıldığı için SVY hariç tüm yöntemler için aynı veri üzerinden örnek analiz yapılmıştır. Bu sebeple bu çalışmada aynı zamanda karşılaştırma da yapılmıştır. Bu çalışmada incelenen yöntemler alanyazında önerilen tüm indeksleri kapsamamaktadır. Tek denekli deneysel araştırma alanyazınındaki karşılaştırma çalışmaları taranarak en çok incelenen indeksler belirlenmiş ve bu çalışmaya dahil edilmiştir (Alresheed vd., 2013; Campbell, 2004; Heyvaert vd., 2014; Maggin vd., 2011; Maggin vd., 2011; Manolov ve Solanas, 2008; Manolov vd., 2010; Olive ve Franco, 2008; Parker, Vannest ve Davis, 2011; Wolery vd., 2010).

Örtüşmeyen Veri Yüzdesi (ÖVY)

Scruggs vd. (1987) tarafından önerilen örtüşmeyen veri yüzdesi (percentage of nonoverlapping data [PND]) başlama ve uygulama düzeylerinin karşılaştırılmasına dayanan ve parametrik olmayan bir indekstir. Başlama ve uygulama düzeylerindeki verilerden örtüşmeyenlerin hesaba katılmasına dayalı olduğundan “örtüşmeyen veri yüzdesi” olarak adlandırılmaktadır.

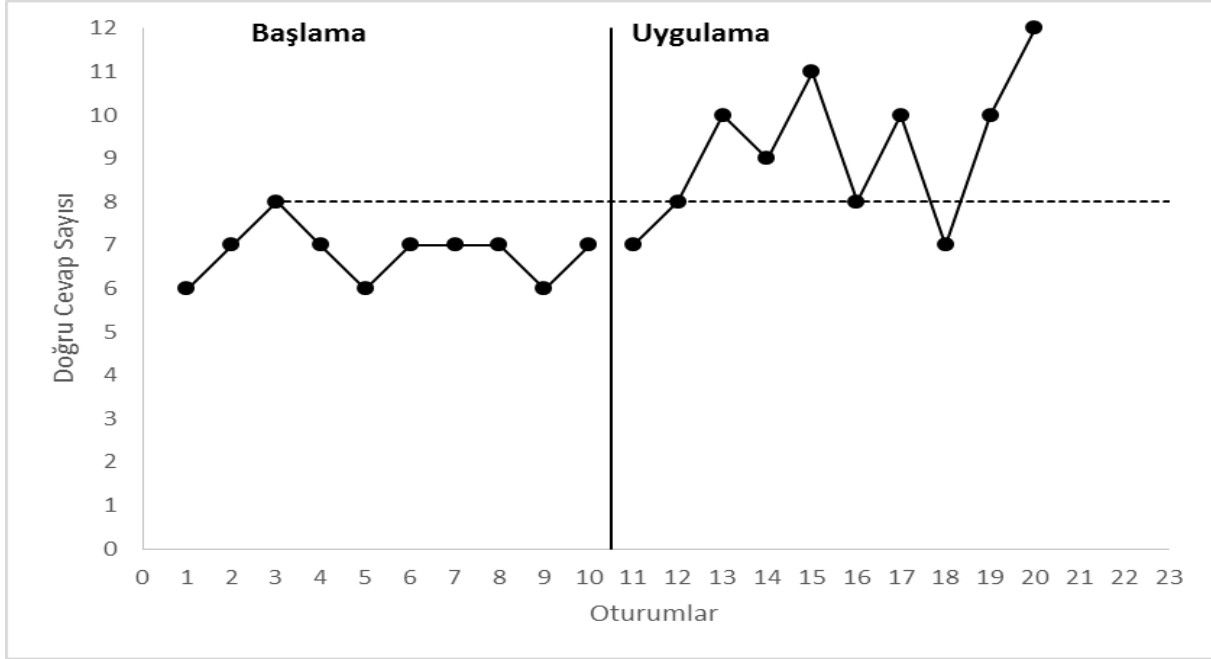
Bu indeksin hesaplanması, uygulama düzeyinde elde edilen değerlerden başlama düzeyinin maksimum değerini geçenlerin oranına karar verilmesi ile yapılmaktadır. Bir AB deseni için ÖVY değeri hesaplanırken şu adımlar takip edilir:

1. Grafikte başlama düzeyindeki en yüksek değer belirlenir,
2. Belirlenen bu maksimum değerden sağ tarafa (uygulama düzeyine) doğru yatay bir çizgi çekilir,
3. Çizilen bu yatay çizgi üzerinde olan uygulama düzeyi verileri belirlenir,
4. Üçüncü adımda elde edilen veri sayısı uygulama düzeyindeki toplam veri sayısına bölünür,
5. Dördüncü adımda elde edilen değer 100 ile çarpılarak ÖVY değeri hesaplanır.

Hedef davranışın artmasının beklendiği uygulama düzeyi durumlarında, ÖVY değeri hesaplanırken başlama düzeyinin en yüksek değerini geçen uygulama düzeyi değerleri hesaba katılır. Hedef davranışın azalması beklenen uygulama düzeyi durumlarında ise başlama düzeyinin en düşük değerinin altındaki uygulama düzeyi değerleri tespit edilir. Bu iki durumda da elde edilen değer sayısı uygulama düzeyindeki toplam veri sayısına bölünür. Eğer bir çalışma birden çok müdahale içerirse farklı müdahalelerden elde edilen ÖVY indeksleri birleştirilerek aritmetik ortalama değeri belirlenir ve bu değer tüm müdahalelerin genel etkililiğini değerlendirmek için kullanılır. Tek denekli deneysel araştırmalarda sıklıkla uygulanan ABAB deseni için de ÖVY değeri hesaplanabilmektedir. ABAB deseninde ilk olarak her iki AB deseni için ÖVY indeksleri hesaplanır. Daha sonra ABAB deseninin genel ÖVY değerini bulabilmek için bu iki ÖVY yüzdesinin aritmetik ortalaması alınır.

Şekil 1’de bir AB desenine ait başlama düzeyi (A) ve uygulama düzeyi (B) durumlarında bir bireyin bir bağımlı değişkene ait aldığı (minimum puan= 0, maksimum puan = 20) değerler verilmektedir. Her iki düzeyde de 10 ölçme yapılmıştır. Bu yapay veri hedeflenen davranışta artış beklenen bir uygulamayı temsil etmektedir. Şekil 1’deki başlama düzeyi ölçümlerine bakıldığında elde edilen en yüksek başlama düzeyi değeri 8 olarak görülmektedir. Bu değer üzerinden sağ tarafa doğru bir çizgi çekildiğinde uygulama düzeyinde 8’in üzerinde olan değerler belirlenir. Şekil 1’deki verilere göre uygulama düzeyinde altı ölçümün (10, 9, 11, 10, 10 ve 12) bu çizginin üzerinde olduğu görülmektedir. Maksimum değeri aşan veri sayısı uygulama düzeyindeki toplam veri sayısına (10’a) bölünürse 0.6

değeri elde edilir. Bu değeri 100 ile çarparsak Şekil 1'deki veride ÖVY değeri $0.6 \times 100 = 60$ olarak hesaplanır.



Şekil 1. Bir AB Deseninde Örtüşmeyen Veri Yüzdesi Değerinin Hesaplanması.

Scruggs ve Mastropieri'ye (1998) göre uygulama düzeyinin “çok etkili” olduğunu söyleyebilmek için bu yüzdenin 90'ın üzerinde olması gerekmektedir. Yüzde 71 ile 90 arası değerler “etkili” uygulama düzeyine işaret ederken 50 ile 70 arası değerler “tartışmalı” ya da “orta düzey” etkiye işaret etmektedir. Yüzde değerinin 50'nin altında olduğu durumlar uygulama düzeyindeki müdahalenin “etkisiz” olduğuna işaret etmektedir (Scruggs, Mastropieri, Cook ve Escobar, 1986; Strain, Kohler ve Gresham, 1998). Bu ölçütlere göre Şekil 1'deki müdahalenin “orta düzeyde” etkili olduğu söylenebilir.

Parametrik istatistiklerdeki varsayımları karşılamak zorunda olmamasının yanında hesaplanmasının ve yorumlanmasının kolay olması ÖVY'nin araştırmacılar tarafından sıklıkla tercih edilmesini sağlamıştır (Ma, 2006). Bu avantajlarının yanında birçok sınırlılığa sahip olan ÖVY indeksi bazı araştırmacılar tarafından eleştirilmektedir (Allison ve Gorman, 1994; Strain vd., 1998). Bu eleştirilerin başında ÖVY'nin teknik olarak bir etki büyüklüğü değeri sayılamaması gelmektedir. Bu konu ile ilgili Scruggs ve Mastropieri (2013) tarafından yapılan ÖVY ve etki büyüklüğü arasında ilişki olduğunu gösteren açıklamalar da mevcuttur. Bu değer ile ilgili yapılan eleştirilerden bir diğeri de örtüşmeyen veri yüzdesi hesabının başlama durumundaki tek bir değer dışında diğer tüm değerleri göz ardı etmesidir. ÖVY'ye yöneltilen başka bir eleştiri de doğrusal artış ya da azalıştan oluşan başlama düzeyindeki eğilimi hesaba katmaması ve eğitim değişimlerini tespit edememesidir. Bu eleştiriler ışığında ÖVY indeksinden tek başına güvenilir bir sonuç elde edilemeyeceği de unutulmamalıdır. (Scruggs ve Mastropieri, 1998). Bunun sebebi ÖVY indeksinin bilinen bir örnekleme dağılımına sahip olmaması ve bu indeks için bir p değerinin hesaplanamamasıdır (Parker, Hagan-Burke ve Vannest, 2007). Dolayısıyla bir p değerinin hesaplanamaması bu indeks üzerinde çıkarım yapma imkânını ortadan kaldırmaktadır. Başlama düzeyindeki değerlerin arasında 0 ya da 100 olması durumunda ÖVY indeks değerinin uygulama düzeyini doğru yansıtmaması da eleştirilen başka bir durumdur (Ma, 2006). Scruggs ve Mastropieri (1998) başlama düzeyindeki veri noktalarından birinin tabanda (0) veya tavanda (100) olması durumlarında ÖVY hesaplamayı önermemektedir. Son olarak, diğer örtüşme indeklerinde de olduğu gibi, ÖVY değerinin uygulama düzeyi veri sayısı arttıkça artması da bir sınırlılık olarak görülmektedir (Allison ve Gorman, 1993).

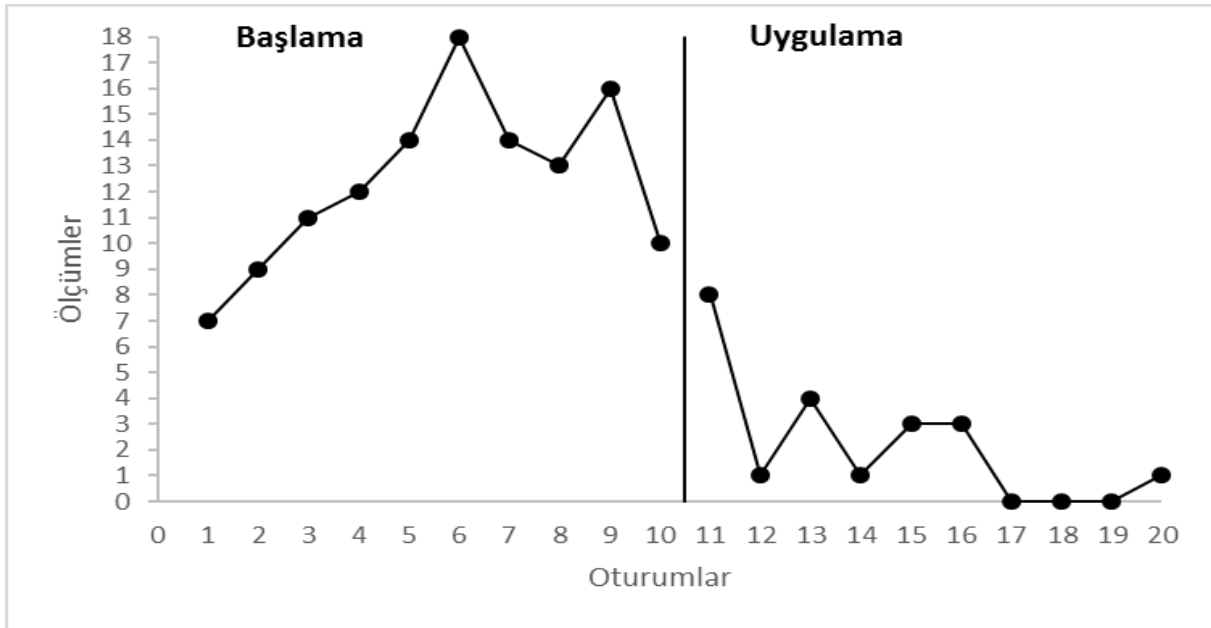
Sıfır Veri Yüzdesi (SVY)

Sıfır veri yüzdesi (Percentage of Zero Data [PZD]) Scotti, Evans, Meyer ve Walker (1991) tarafından tek denekli deneysel çalışmalarda uygulama düzeyinin etkililiğini göstermek için geliştirilen bir indekstir. SVY davranış azalma derecesine karşı davranış baskılanma derecesini temsil etmektedir ve Campbell (2004, s. 235) tarafından etki büyüklüğünün tutucu (stringent) bir ölçüsü olarak görülmektedir. Yani hedeflenen davranışın sıfıra ulaşmasını ve sıfır düzeyinde kalmasını gerektiren bir ölçüdür.

SVY değerini bir AB deseninde elde edebilmek için şu adımlar takip edilir:

1. Öncelikle uygulama düzeyinde sıfır olan ilk veri noktası tespit edilir,
2. Birinci adımda tespit edilen sıfır noktası da dâhil sıfır noktasında kalan uygulama düzeyi veri sayısı belirlenir,
3. İkinci adımda belirlenen sıfır değerleri sayısı uygulama düzeyinde ilk sıfırdan sonraki veri sayısına bölünür,
4. Üçüncü adımda elde edilen değer 100 ile çarpılarak SVY değeri elde edilir.

Şekil 2’de bir AB desenine ait başlama düzeyi (A) ve uygulama düzeyi (B) durumlarında bir bireyin bağımlı değişkene ait aldığı (maksimum puan = 20) değerler verilmektedir. Bu örnek veride uygulama düzeyindeki müdahalenin katılımcıdaki davranışı azaltması beklenmektedir. Her iki durumda da 10 ölçme yapılmıştır. Uygulama düzeyindeki ölçümlere bakıldığında katılımcının sıfır puan aldığı nokta yedinci ölçme noktasıdır. Bu nokta sonrasında katılımcı üzerinde üç ölçme daha yapılmış ve bu katılımcı iki kere daha sıfır puan elde etmiştir. Bu veriler ışığında ilk sıfır da dâhil bu nokta sonrası yapılan ölçme sayısı dördtür ve katılımcı bu dört ölçmenin üçünde sıfır puan elde etmiştir. SVY değerini hesaplayabilmek için elde edilen sıfır değerlerini bu ölçme sayısına bölmemiz yeterlidir. Böylelikle SVY değeri $3/4 \times 100 = 75$ olarak elde edilir. Birden fazla uygulama düzeyi olduğu durumlarda (ör. ABCD) bazı araştırmacılar (Reichle, 2007) sadece en sonuncu uygulama düzeyi (ör. D) üzerinden SVY değerini hesaplamışlardır. Denekler arası çoklu başlama düzeyi desenlerinde her bir denek için ayrı ayrı SVY değeri hesaplanır (Schlosser ve Koul, 2015). ABAB desenlerini içeren çalışmalarında Wehmeyer vd. (2006) her AB düzey çifti için bir SVY değeri hesapladığını ve hesaplanan her SVY’nin ayrı değerler olarak ele alındığını belirtmişlerdir.



Şekil 2. Bir AB Deseninde Sıfır Veri Yüzdesi Hesaplanması.

SVY değeri 0 ile 100 arası değerler alır ve yüksek değerler etkili uygulamaya işaret eder. SVY'nin yorumlanmasında Scotti vd. (1991) tarafından belirlenen ölçütler kullanılmaktadır. Scotti ve diğerlerine (1991) göre %55–80 arası değerler “orta düzey” etkiye işaret ederken %80 üzerindeki değerler “yüksek” etkideki müdahaleyi göstermektedir. Yüzde 18'in altındaki SVY değerleri “etkisiz” müdahale anlamına gelirken %18-54 arası değerler “şüpheli etki” düzeyine işaret etmektedir. Yukarıda verilen ölçütlere göre Şekil 2'deki müdahalenin “orta düzeyde” etkiye sahip olduğu söylenebilir.

SVY indeksinde ilk sıfır noktasından önceki tüm veriler göz ardı edildiği için ÖVY ile karşılaştırıldığında bazı durumlarda veri kaybı olmaktadır. Aynı zamanda uygulama düzeyinde sıfırı göstermeyen değerler de SVY değerinin hesaplanmasına dâhil edilmemektedir. Ayrıca ÖVY'de olduğu gibi aykırı değerlerden ve zamandan kaynaklanan eğilimden kolay etkilenebilmektedir (Allison ve Gorman, 1993). SVY indeksinin diğer bir dezavantajı da sadece uygulama düzeyindeki verileri hesaba katmasıdır. Sadece hedef davranışın ortadan kalkmasına odaklandığı için her uygulama durumuna uygun olmamaktadır.

Başlama Düzeyinin Ortancasını Aşan Veri Yüzdesi (BOAVY)

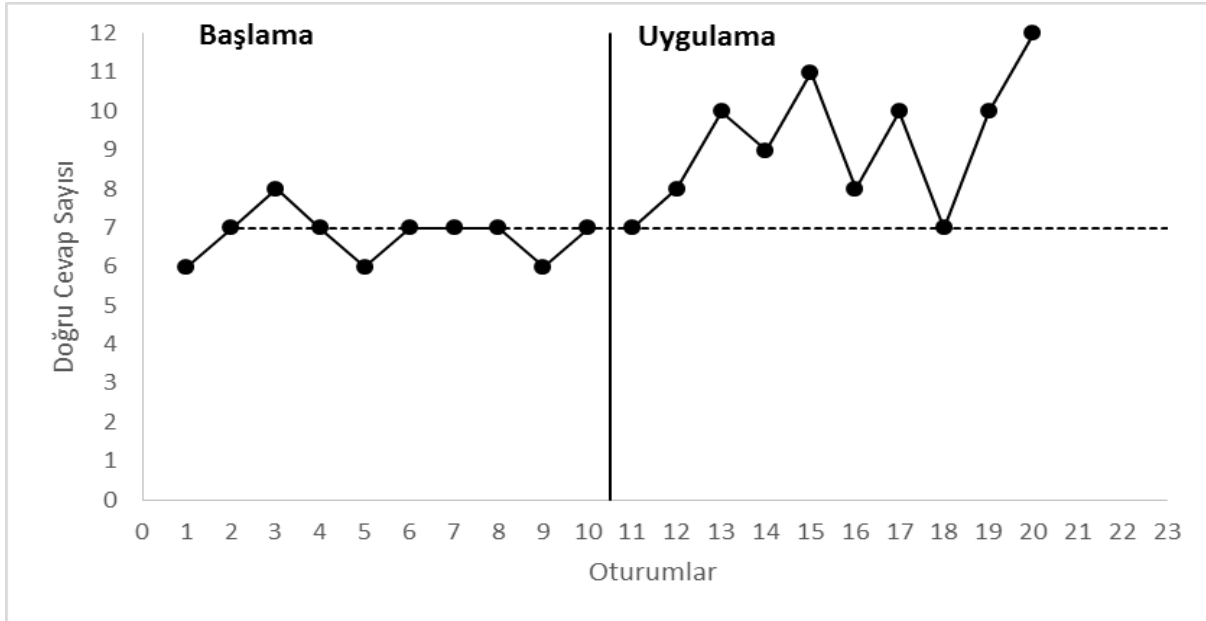
Ma (2006, s.600)'nın belirttiği üzere ÖVY'de başlama düzeyi verilerindeki 0 ya da 100 değerinin grafik üzerinde oluşturulan yatay çizgi ile (tavan ve taban [ceiling and floor] etkisi) ÖVY değerinin hesaplanmasını anlamsız kılması ya da 0 olarak hesaplanması bu yöntemin zayıflıklarından biri olarak görülmüştür. Düzeyler arasında ani olmayan azalış ve artıştan kaynaklanan bir eğilim etkisi bulunabilmektedir. Bu eğilim etkisi ÖVY'de göz ardı edilmektedir. Bu sebeplerden dolayı Ma (2006) ÖVY yönteminde II. tür hata yapma riskinin fazla olduğunu ileri sürmüş ve bu yöneme alternatif olarak başlama düzeyinin ortanca değerini aşan veri yüzdesi (percentage of data points exceeding median [PEM]) yöntemini önermiştir.

Bir AB deseninde BOAVY değerinin hesaplanmasında şu adımlar takip edilir:

1. Grafik üzerinde başlama düzeyinin ortanca değeri belirlenir,
2. Ortanca değer üzerinden sağ tarafa doğru yatay bir çizgi çekilir,
3. Bu yatay çizgi üzerinde kalan uygulama düzeyi verileri işaretlenir,
4. Üçüncü adımda belirlenen verilerin sayısı uygulama düzeyindeki toplam veri sayısına bölünür,
5. Dördüncü adımda elde edilen değer 100 ile çarpılarak BOAVY değeri hesaplanır.

BOAVY değeri hesaplanırken uygulanan müdahalenin hedef davranışı arttırması bekleniyorsa başlama düzeyinin ortanca değeri üzerindeki uygulama düzeyi değerleri hesaba katılırken, uygulamadaki müdahale etkisinin bireyden elde edilen değerlerde azalmaya yol açacağı beklentisi var ise uygulama düzeyindeki verilerden ortanca değer altındaki kalanlar hesaplamaya katılır (Ma, 2006).

Şekil 3'teki başlama düzeyi ölçümlerine bakıldığında bu düzeyin ortanca değerinin 7 olduğu görülür. Bu değer üzerinden sağ tarafa doğru bir çizgi çekildiğinde uygulama düzeyinde 7'nin üzerinde olan değerler belirlenir. Şekil 3'teki verilere göre uygulama düzeyinde sekiz ölçümün (8, 10, 9, 11, 8, 10, 10 ve 12) bu çizginin üzerinde olduğu görülmektedir. Uygulama düzeyinde ortanca değeri aşan veri sayısının uygulama düzeyindeki toplam veri sayısına (10'a) bölünmesiyle 0.8 değeri elde edilir. Bu değeri 100 ile çarparsak Şekil 3'teki veride BOAVY değeri $0.8 \times 100 = 80$ olarak hesaplanır.



Şekil 3. Bir AB Deseninde Başlama Düzeyinin Ortancasını Aşan Veri Yüzdesi Değerinin Hesaplanması.

Hesaplanan BOAVY yüzdeler için önerilen ölçütlere bakılacak olursa %91 ile %100 arası değerler “çok etkili” bir uygulamaya işaret ederken %70 ile %90 arası değerler “orta düzey” uygulama etkisine işaret etmektedir. BOAVY değeri %70’in altında olan çalışmalarda uygulama etkisinin “tartışmalı” ya da “etkisiz” olduğu söylenebilir (Ma, 2006). Heyvaert ve diğerlerine (2014) göre %50–70 arası “şüpheli” durumlara, %50’nin altında ise “etkisiz” uygulama düzeyine işaret etmektedir. Etkisiz deney durumlarında veri noktaları ortanca değerinin etrafında aşağı yukarı sürekli dalgalanmalar göstermektedir. Şekil 3’teki veri kullanarak elde edilen BOAVY değeri yukarıda verilen ölçütlere göre “orta düzey” etkiyi işaret eder.

BOAVY yönteminin verideki otokorelasyondan çok az etkilendiği ve etkisiz müdahale ile etkili müdahaleyi diğer indekslere göre daha iyi ayırt ettiği bulunmuştur (Manolov vd., 2010). BOAVY indeksi başlama düzeyi ile uygulama düzeyi arasındaki eğilim değişimini değil de hedef davranışın bu düzeyler arası değişimini hesaplamada kullanılmaktadır. Ayrıca bu indeks uygulama düzeyindeki değişimi ve eğilimi hesaba katmamaktadır. İlk uygulama düzeyinden sonraki başlama ve uygulama düzeyi çiftinde dik (orthogonal) eğim olduğunda BOAVY değeri ÖVY’deki eğim değeri hesaplanamaması durumuna kısmi çözüm sunmaktadır. Ma (2006)’ya göre BOAVY indeksinin sınırlılıklarından birisi de bu indeksin hesaplanmasında ortanca üzerindeki veri değerlerinin büyüklüklerinin göz ardı edilmesi yani bu indeksin ortanca üzerindeki veri noktalarına karşı hassas olmamasıdır. Ayrıca BOAVY değerinin AB desenleri dışındaki veri durumlarında nasıl hesaplanacağı da orijinal çalışmada belirtilmeyen durumlar arasındadır.

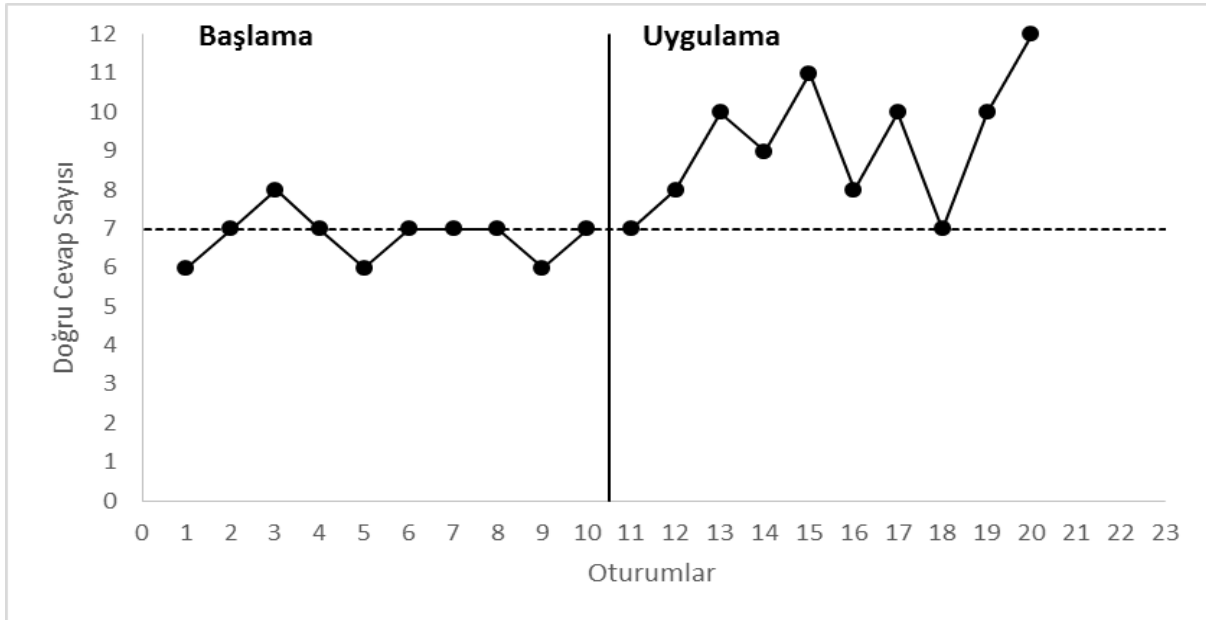
Başlama Düzeyinin Ortanca Eğilimini Aşan Veri Yüzdesi (BOEAVY)

Başlama düzeyinin ortanca eğilimini aşan veri yüzdesi (Percentage of data exceeding a median trend [PEM-T]) tek denekli deneysel çalışmalarda kullanılan ortadan bölme (split-middle) tekniğini (White ve Haring, 1980) kullanarak eğilimi göz önünde bulunduran bir etki büyüklüğü indeksidir (Wolery vd., 2010). BOAVY değerinin bir versiyonu olan BOEAVY indeksi de aynı amaçla kullanılan nonparametrik bir istatistiktir.

Bir AB deseninde BOEAVY değerinin hesaplanmasında şu adımlar takip edilir:

1. Grafik üzerinde White and Haring'de (1980) tarif edildiği gibi ortadan bölme tekniği kullanılarak başlama düzeyi verilerinden uygulama düzeyi verilerine doğru bir ortanca eğilim çizgisi çizilir,
2. Ortanca eğilim çizgisi üzerinde kalan uygulama düzeyi verileri işaretlenir,
3. İkinci adımda belirlenen verilerin sayısı uygulama düzeyindeki toplam veri sayısına bölünür,
4. Üçüncü adımda elde edilen değer 100 ile çarpılarak BOEAVY değeri hesaplanır.

Şekil 4'teki başlama ve uygulama düzeyi verilerine göre BOEAVY değeri hesaplamak istersek, başlama düzeyindeki verileri kullanarak bir ortanca eğilim çizgisi oluşturmamız gerekir. Manolov, Sierra, Solanas ve Botella (2014) tarafından hazırlanan R sözdizimi kullanılarak Şekil 4'teki eğim çizgisi elde edilmiştir. Şekil 4'te uygulama düzeyi değerlerinin sekiz tanesinin oluşturulan bu eğilim çizgisinin üzerinde kaldığı görülür. Buradan BOEAVY değerini $8/10 \times 100 = 80$ olarak hesaplayabiliriz. Birden çok AB deseni bulunan bir deneysel araştırmada BOEAVY değeri her bir AB deseni için hesaplanır ve elde edilen değerlerin ortalaması alınır. BOEAVY'yi geliştiren araştırmacılar bu indeksin daha karmaşık desenler için nasıl hesaplanacağı konusunda bir öneride bulunmamışlardır.



Şekil 4. Bir AB Deseninde Başlama Düzeyinin Ortanca Eğilimini Aşan Veri Yüzdesi Değerinin Hesaplanması.

Alanyazında BOEAVY değerlerinin yorumlanması konusunda herhangi bir net ölçüt bulunmamaktadır. Bu indeks için Ma (2006)'da sunulan ölçütler kullanılabilir. Bu ölçütlere göre %90 ve üzeri değerler "çok etkili" bir uygulamaya işaret ederken %70 ile %89 arasındaki değerler "orta düzey" uygulama etkisine işaret etmektedir. %50 ile %69 arasındaki değerler "tartışmalı" ya da "etkisiz" uygulamaya işaret eder. BOEAVY değeri %50'nin altında ise uygulama "etkisiz" olarak değerlendirilir. Bu ölçütler spesifik olarak BOEAVY için geliştirilmediğinden ölçütlerin kullanılmasında ihtiyatlı davranılması gerekmektedir. Başlama düzeyi evresinde eğilim olan grafiklerde etki büyüklüğü değeri daha küçük çıkabileceğinden bu ölçütlerin daha ihtiyatlı bir şekilde kullanılması gerekmektedir. BOEAVY indeksi de bazı avantajlara ve sınırlılıklara sahiptir. Yorumlanmasının kolay olması ve başlama düzeyindeki doğrusal eğilimi hesaba katabilmesi bu indeksin başlıca avantajlarından. Başlama düzeyinde ortanca eğilim çizgisinin aykırı değerlerden etkilenmesi de bu indeksin hesaplanmasını etkileyen durumlar arasında görülmektedir.

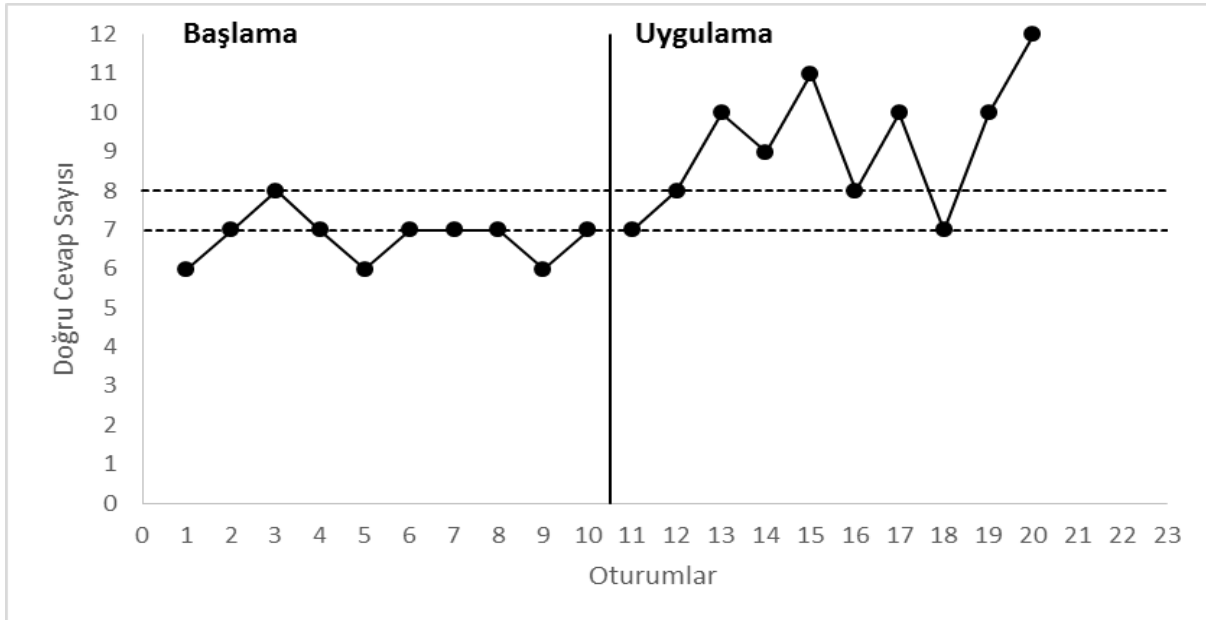
Örtüşmeyen Tüm Verilerin Yüzdesi (ÖTVY) ve Phi Katsayısı

Parker vd. (2007) tarafından önerilen örtüşmeyen tüm verilerin yüzdesi (Percentage of All Nonoverlapping Data [PAND]) indeksi diğer örtüşmeyen veri yüzdesi indekslerinde olduğu gibi örtüşmeme yüzdesini bulmaya yardımcı olmaktadır. ÖTVY değeri iki düzey arasındaki örtüşmeyi ortadan kaldıracak minimum sayıdaki veri noktasının toplam veriden çıkarılmasından sonra geriye kalan veri yüzdesi olarak tanımlanmaktadır (Parker, Vannest ve Davis, 2011).

Bir AB deseninde ÖTVY değerini hesaplayabilmek için şu adımlar takip edilir:

1. Öncelikle iki düzey arasında kesişime (örtüşmeye) neden olan veri noktaları belirlenir,
2. Bu iki düzey arasındaki kesişimi ortadan kaldıracak minimum sayıdaki veri sayısı tespit edilir,
3. İkinci adım takip edilerek kesişen/örtüşen veri noktaları sayılır,
4. Üçüncü adımda belirlenen örtüşen veri sayısı toplam veri sayısına bölünerek örtüşen veri yüzdesi hesaplanır,
5. Dördüncü adımda elde edilen örtüşme yüzdesi 100'den çıkarılarak tüm örtüşmeyen veri yüzdesi (ÖTVY) elde edilir.

Şekil 5'teki başlama ve uygulama düzeyi verilerine göre ÖTVY değerini hesaplamak için başlama ve uygulama düzeyleri arasında örtüşen bölge ve bu örtüşen bölgeye neden olan veri noktaları belirlenir. Şekil 5'te gösterildiği gibi başlama düzeyinin en üst noktasından bir çizgi ve uygulama düzeyinin en alt noktasından bir çizgi çekilerek iki düzey arasında örtüşen bölge belirlenir. Daha sonra bu örtüşmeyi ortadan kaldıracak minimum sayıdaki veri noktasına karar verilir. Şekil 5'teki grafikte başlama düzeyindeki bir veri noktasını (8) ve uygulama düzeyindeki iki veri noktasını (7 ve 7) çıkarırsak bu iki düzey arasındaki örtüşmeyi ortadan kaldırarak örtüşmeyen tüm veri noktalarının sayısını elde ederiz. Daha sonra tüm veri sayısından (20) örtüşen veri sayısını (2+1=3) çıkarırsak tüm örtüşmeyen veri sayısını (17) elde ederiz. Böylelikle ÖTVY değerini $(20-3)/20=0.85$ buluruz. Yüzdeler olarak ifade edebilmek için bu değeri 100 ile çarparsak tüm örtüşmeyen veri yüzdesini elde ederiz.



Şekil 5. Bir AB Deseninde Örtüşmeyen Tüm Verilerin Yüzdesi Değerinin Hesaplanması.

ÖTVY'yi geliştiren Parker vd. (2007) sadece grafiğe bakılarak hesaplanabilen ÖTVY'nin kolay olduğunu fakat çok fazla ölçümün yapıldığı (karmaşık tek denekli deneysel desenler) durumlarda

araştırmacılar için grafiğe bakmanın doğru sonuçlar bulmayı zorlaştıracağını belirtmektedir. ÖVY gibi örtüşmeyen veri değerlerine dayanan ÖTVY indeksi ÖVY'nin aksine başlama düzeyindeki tüm verileri kullanarak hesaplanır. ÖTVY indeksinin diğer bir avantajı da bu indeksin Pearson Phi korelasyonuna dönüştürülebilir olmasıdır. Parker vd. (2007)'nin belirttiği üzere Phi değeri bilindik bir örnekleme dağılımına sahiptir ve Phi için p değeri elde edilebilmektedir. Ayrıca güven aralığı ve güç hesaplamalarına imkân vermektedir. ÖTVY indeksi Phi değeri vasıtasıyla Cohen d değerine dönüştürülerek yorumlanabilmektedir (Parker vd., 2007). Bu özellikler sayesinde ÖTVY değeri ÖVY indeksi ile elde edilemeyen birçok bilgiyi elde edebilmek imkânı sunmaktadır. Sadece ortalama düzey değişimlerini ölçebilen ÖTVY, pozitif başlama düzeyi eğilimini kontrol edememektedir (Parker vd., 2007, s.196).

Tüm Çiftlerin Örtüşmemesi (TÇÖ)

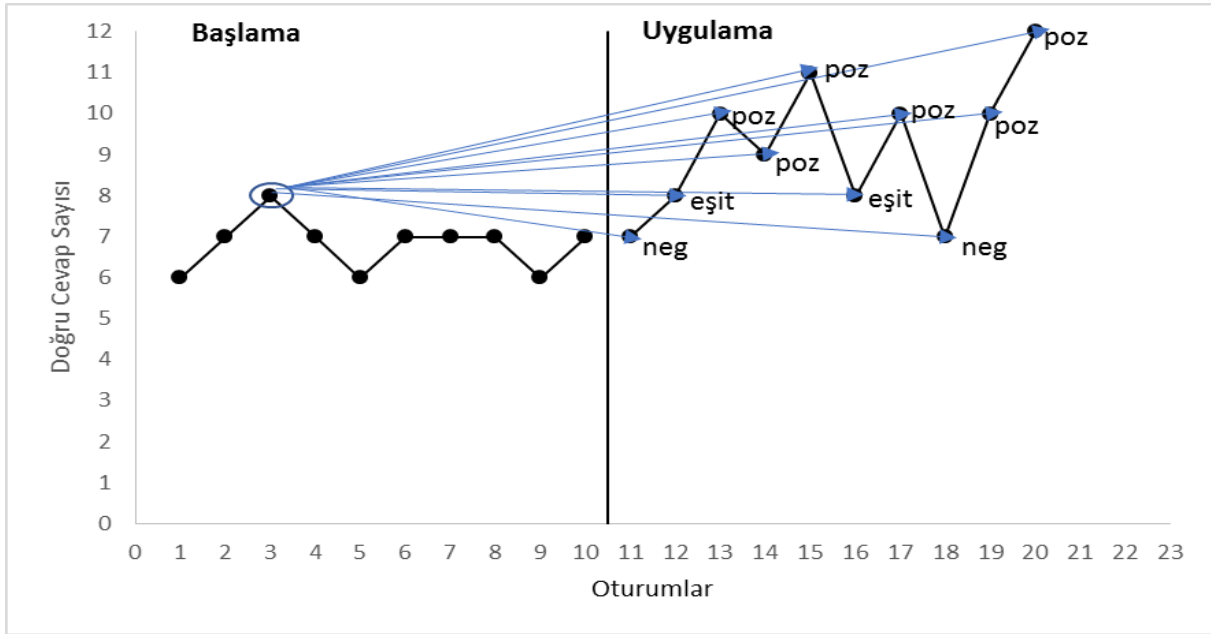
Parker ve Vannest (2009) tarafından geliştirilen tüm çiftlerin örtüşmemesi (nonoverlap of all pairs [NAP]) indeksi tamamıyla ikili karşılaştırmalara dayanarak hesaplanan bir değerdir.

Bir AB deseninde TÇÖ değerini hesaplayabilmek için şu adımlar takip edilir:

1. Öncelikle başlama düzeyindeki her bir değer uygulama düzeyindeki her bir değer ile karşılaştırılır. Bir AB deseninde başlama düzeyi veri noktalarından bir tanesini uygulama düzeyi veri noktalarının hepsi ile karşılaştırdığımızda her bir veri çifti karşılaştırması için karşımıza üç durum çıkar. Bu durumlardan birincisi, başlama düzeyinde ele alınan veri noktasının uygulama düzeyinde karşılaştırmaya tabi tutulan noktaya doğru gelişim göstermesidir (pozitif değer). Başka bir ifadeyle, uygulama düzeyinde karşılaştırma yaptığımız veri noktasının bu veri noktasından büyük olmasıdır. İkincisi, ele alınan başlama düzeyi verisinin bu düzeyden uygulama düzeyine hiçbir değişime uğramama (eşit değer) durumudur. Üçüncüsü, başlama düzeyinde karşılaştırma yaptığımız veri noktasının uygulama düzeyindeki veri noktalarından büyük olduğu yani azalma gösterdiği (negatif) durumları içerir.
2. Karşılaştırma yapılan her bir başlama düzeyi veri noktası için bir toplam puan hesaplanır. Bu puan hesaplanırken karşılaştırmadaki durumlara göre pozitif çıkan ($A_n < B_n$) durumlar için +1, negatif çıkan ($A_n > B_n$) durumlar için 0 ve eşit çıkan ($A_n = B_n$) durumlar için 0.5 değeri verilir.
3. İkinci adımda elde edilen bu değerler toplanarak karşılaştırmaya tabi tutulan başlama düzeyi veri noktası için toplam gelişim puanı elde edilir.
4. Aynı işlemi başlama düzeyindeki diğer veri noktaları için de yaparız ve toplam gelişim puanlarını hesaplarız.
5. Önceki adımlarda elde edilen bu toplam gelişim puanını olabilecek tüm veri çiftlerinin sayısına bölerek TÇÖ değerini elde ederiz.

Bir AB deseninde tüm veri çiftlerinin sayısı, başlama ve uygulama düzeylerindeki veri sayılarının çarpımına ($A \times B$) eşittir. Şekil 6'daki veri için tüm veri çiftlerinin sayısı (10×10) 100'dür. Şekil 6'daki 10 başlama düzeyi verisinin her birini uygulama düzeyi veri noktaları ile karşılaştırma yaparak gelişim gösteren (+1), eşit değere sahip olan (0.5) ve düşüş gösteren (0) veri çifti karşılaştırmalarını not ederiz. Şekil 6'da sunulan başlama düzeyindeki ilk veri noktasının (6) her bir uygulama düzeyi veri noktası ile karşılaştırılması (6-7, 6-8, 6-10, 6-9, 6-11, 6-8, 6-10, 6-7, 6-10, 6-12) ile elde edilen durumlara göre başlama düzeyinin ilk veri noktası için gelişim puanı 10 (10 pozitif durum) olarak elde edilir. Benzer şekilde başlama düzeyindeki ikinci veri noktası olan 7 değeri için de gelişim puanı 9 (8 pozitif ve 2 eşit durum) olarak elde edilir. Aynı şekilde başlama düzeyinin üçüncü veri noktasının (8) uygulama düzeyinin tüm veri noktaları ile karşılaştırılması yapılmış ve örnek teşkil etmesi için Şekil 6'da gösterilmiştir. Şekil 6'daki karşılaştırma durumlarına göre başlama düzeyinin üçüncü veri noktası için gelişim puanı 7 (6 pozitif, 2 eşit durum, 3 negatif durum) olarak elde edilir. Her bir veri noktası için puanlar sırasıyla (ölçme noktası 1'den 10'a doğru) 10, 9, 7, 9, 10, 9, 9, 9, 10 ve 9 olarak hesaplanır.

Bulunan bu değerler toplandığında toplam gelişim veri çifti sayısı 91 olarak bulunur ve bu sayıyı toplam veri çifti sayısına (100'e) bölersek TÇÖ değerini 91 (%91) olarak elde ederiz.



Şekil 6. TÇÖ Hesaplanmasında Kullanılan Veri Çifti Karşılaştırmasının 3. Veri Noktası İçin Gösterilmesi (poz=pozitif; neg=negatif).

TÇÖ indeksi her bir başlama düzeyi değeri ile her bir uygulama düzeyi değeri arasındaki örtüşmemeyi özetlemeye çalışmaktadır. Kısaca A ile B düzeyleri arasında gelişim gösteren verilerin yüzdesi olarak tanımlanabilir (Parker vd., 2011). TÇÖ tüm veri çiftlerini hesaba kattığı için bütüncül bir örtüşmeyen veri istatistiğidir ve Cohen *d* etki büyüklüğüne dönüştürülebilmektedir. Diğer örtüşmeyen veri indeksleri gibi TÇÖ değeri de grafik kullanılarak hesaplanabilir fakat büyük veri setlerinde tüm veri çiftlerinin karşılaştırmasını yapmak kolay olmamaktadır. Grafikten elle hesaplamaların yanında TÇÖ değeri SPSS gibi bilindik istatistik programları vasıtasıyla alıcı işletim karakteristikleri (Receiver Operator Characteristics [ROC]) veya Mann-Whitney U testi analizlerini kullanarak elde edilebilir. ROC analizinde eğri altındaki alanı (Area Under the Curve; AUC) bularak TÇÖ hesaplanabilir. Çok fazla veri noktasının bulunması ya da birden fazla başlama ve uygulama düzeyleri olması durumunda, diğer örtüşmeyen veri yüzdelerini grafikten hesaplamak zor olabilmektedir. TÇÖ değeri bu durumlarda bile ROC analizinde eğri altındaki alanı bularak kolaylıkla elde edilebilmektedir. Ayrıca TÇÖ değeri <http://www.singlecaseresearch.org/calculators/nap> adresi üzerinden veri girişi yapılarak hesaplanabilir. Parker ve Vannest'te (2009, s.359) bahsedildiği üzere eğri altında kalan değer (AUC değeri) 0.5 ile 1 arasında değişen bir olasılık değeridir. Parker ve Vannest'in (2009) TÇÖ değerlerinin yorumlanabilmesi için vermiş olduğu ölçütlere göre 0-0.65 arası "zayıf etki", 0.66-0.92 arası "orta düzey etki" ve 0.92 üzeri "yüksek etki" olarak yorumlanmaktadır.

Parker ve Vannest'e (2009) göre TÇÖ'nün, diğer örtüşmeyen veri indeksleri ile karşılaştırıldığında beş avantajı öne çıkmaktadır. Bunların başında çok sayıda yayınlanmış çalışmanın karşılaştırılmasında sonuçları daha iyi ayırt edebilmesi gelmektedir. İkinci avantajı TÇÖ hesaplamasının SPSS gibi yaygın olarak kullanılan bilgisayar programları yardımıyla yapılabilmesinden dolayı elle hesaplanan diğer indekslere göre daha az hata yapıyor olmasıdır. Üçüncü ve dördüncü avantajları ise verinin tamamının kullanılıyor olmasından dolayı R^2 ile yüksek ilişkisinin olması ve görsel analizler tarafından geçerliğinin sağlanmasıdır. Ayrıca TÇÖ indeksi R^2 değerine ve dolayısıyla Cohen *d* değerine dönüştürülebilmektedir (Parker ve Vannest, 2009). Parker ve Vannest (2009) tarafından belirtilen beşinci avantaj ise düşük güven aralıklarına sahip olan TÇÖ değerinin daha kesin sonuçlar

sunuyor olmasıdır. Ayrıca boylamsal verideki otokorelasyon durumunda da TÇÖ iyi derecede performans göstermektedir (Manolov, Solanas, Sierra ve Evans, 2011).

Tau

Parker, Vannest, Davis ve Sauber (2011) tarafından geliştirilen bir diğer etki büyüklüğü indeksi Tau (Kendall's Tau non-overlap) değeridir. Tau değerinin hesaplanmasında da TÇÖ değerinin hesaplanmasında olduğu gibi örtüşmeyen veri çiftleri kullanılmaktadır. TÇÖ'de örtüşmeyen veri çiftleri yüzdesi bulunurken Tau indeksinde örtüşmeme yüzdesinden örtüşme yüzdesi çıkarılmaktadır (Parker, Vannest, Davis ve Sauber, 2011). PVÇS başlama düzeyinden uygulama düzeyine artış gösteren (pozitif) veri çifti sayısını, NVÇS ise başlama düzeyinden uygulama düzeyine değer kaybeden (negatif) veri çifti sayısını ve TVÇS de tüm veri çifti sayısını temsil ettiği düşünülürse, başlama ve uygulama düzeyleri arasındaki veri çiftleri kullanılarak Tau değeri Eşitlik 1 ile hesaplanabilir.

$$Tau = \frac{PVÇS - NVÇS}{TVÇS} \quad (1)$$

TÇÖ'de olduğu gibi bir AB deseninde tüm veri çiftlerinin sayısı başlama ve uygulama düzeylerindeki veri sayısının çarpımına (A×B) eşittir. Burada pozitif veri çifti sayısından negatif veri çiftlerinin sayısını çıkararak elde ettiğimiz değeri (örtüşen veri çiftleri sayısını) tüm veri çifti sayısına böldüğümüzde Tau değerini elde etmiş oluruz.

Şekil 6'da sunulan veride daha önce yapılan hesaplamalara benzer bir şekilde veri çiftlerinin (100 veri çiftinin) her birinin karşılaştırılmasıyla pozitif çift değerlerinin sayısı 84, negatif veri çiftlerinin sayısı 2 ve eşit olan veri çiftlerinin sayısı da 14 olarak bulunmuştur. Bu değerler ışığında Tau değeri $(84 - 2)/100 = 0.82$ olarak hesaplanır. Bu ondalık değeri yüzdelik ölçeğe çevirebilmek için 100 ile çarpmak gerekir. TÇÖ'de olduğu gibi Tau değeri de %50 ile %100 arasında değişen değerler alabilmektedir. Parker ve diğerlerinin (2011) belirttiği gibi Tau değeri bilindik istatistik programları kullanılarak Kendall sıra korelasyonu ya da Mann-Whitney U testi analizleriyle elde edilebilir (Parker, Vannest, Davis ve Sauber, 2011). Kendall sıra korelasyonunda elde edilen S değeri toplam veri çifti sayısına (A×B) bölünerek Tau değeri hesaplanabilir.

Bir AB deseninde Tau değerini Kendall sıra korelasyonundan elde etmek için şu adımlar takip edilir:

1. Düzey değişkeni oluşturulur: Başlama düzeyindeki değerlerin sayısı kadar 0 değeri, uygulama düzeyindeki değerlerin sayısı kadar da 1 değeri girilerek bir Düzey değişkeni oluşturulur (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1),
2. Birinci adımda oluşturulan Düzey değişkeni ile başlama ve uygulama düzeylerinin ham verisi (6, 7, 8, 7, 6, 7, 7, 7, 6, 7, 7, 8, 10, 9, 11, 8, 10, 7, 10, 12) arasında Kendall sıra korelasyonu hesaplanır,
3. Kendall sıra korelasyonundan S değeri elde edilir (S=82),
4. Üçüncü adımda bulunan S değeri toplam veri çifti sayısına (100) bölünür ve Tau değeri .82 olarak elde edilir. Ayrıca bu hesaplama <http://www.singlecaseresearch.org/calculators/tau-u> adresi üzerinden ham veri girilerek de yapılabilir.

Tau-U

Parker, Vannest, Davis ve Sauber (2011) tarafından istenmeyen başlama düzeyini kontrol edebilen bir Tau değeri (Tau-U) daha önerilmiştir. Tau-U indeksi Tau'da olduğu gibi örtüşmeyen veriyi hesaba katmanın yanında başlama düzeyinin eğilimini de kontrol edebilmektedir. Bu sayede Tau-U indeksi müdahale öncesinde (yani başlama düzeyi sırasında) başlayan potansiyel bir artışın ötesinde sadece müdahale aşamasında ortaya çıkan artış miktarının nicelleştirilmesini sağlar.

Bir AB deseninde Tau-U değerini Kendall sıra korelasyonundan elde etmek için şu adımlar takip edilir:

1. Düzey değişkeni oluşturulur: Başlama düzeyindeki verilere en büyük değerdeki sayı 1 olacak şekilde sıra numaraları verilir ((6, 7, 8, 7, 6, 7, 7, 7, 6, 7) için (3, 2, 1, 2, 3, 2, 2, 2, 3, 2)) ve oluşturulan veri Düzey A verisi olarak girilir. Düzey A'daki en son sayıdan devam edilerek sıradaki sayı Düzey B'deki her bir eleman için atanır ((7, 8, 10, 9, 11, 8, 10, 7, 10, 12) için (4, 4, 4, 4, 4, 4, 4, 4, 4, 4)). Bu iki düzeyin verisi birleştirilerek ortak bir Düzey değişkeni oluşturulur (3, 2, 1, 2, 3, 2, 2, 2, 3, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4),
2. Birinci adımda oluşturulan Düzey değişkeni ile başlama ve uygulama düzeylerinin ham verisi (6, 7, 8, 7, 6, 7, 7, 6, 7, 7, 8, 10, 9, 11, 8, 10, 7, 10, 12) arasında Kendall sıra korelasyonu hesaplanır,
3. Kendall sıra korelasyonundan S değeri elde edilir (S=83),
4. Üçüncü adımda bulunan S değeri toplam veri çifti sayısına (100) bölünür ve Tau-U değeri .83 olarak elde edilir. Bu değeri 100'lük ölçekte ifade edebilmek için 100 ile çarpılır. Ayrıca bu hesaplama <http://www.singlecaseresearch.org/calculators/tau-u> adresi üzerinden veri girilerek de yapılabilir.

Parker vd. (2011) Tau-U değerinin AB ve ABA desenleri için nasıl hesaplanabileceğini göstermişlerdir. Bu çalışmada daha karmaşık desen durumlarında sadece ilk baştaki başlama (A) ve uygulama (B) düzeylerinin hesaplamaya katıldığından bahsedilmiştir. Alanyazında geliştirilen en güncel indekslerden biri olan Tau-U indeksinin diğer tek denekli deneysel desenlerde nasıl kullanılacağına dair Rakap (2015, s.26) çalışmasında bazı açıklamalara yer verilmiştir. Çoklu başlama düzeyine sahip desenlerde, Tau-U, her başlama-uygulama (AB) müdahale karşılaştırması için ayrıca hesaplanır ve elde edilen tüm Tau-U değerlerinin ortalaması alınarak genel Tau-U değeri bulunur. ABAB deseni kullanıldığında, her başlama ve uygulama çifti (yani A₁B₁ ve A₂B₂) için Tau-U değeri hesaplanır. Genel Tau-U değeri de bu çiftlerden elde edilen Tau-U değerlerinin ortalaması alınarak elde edilir. ABCD deseninde, Rakap'a (2015) göre her uygulama düzeyi için uygulama düzeyi ile başlama düzeyi karşılaştırılarak (örneğin A başlama düzeyi olmak üzere AB, AC, AD) Tau-U değerleri elde edilmelidir. Tam modelin genel etki büyüklüğü son müdahale aşaması (AD karşılaştırması) kullanılarak hesaplanmalıdır. Bu desen durumunda genel Tau-U değeri başlama düzeyi ile son uygulama düzeyinin karşılaştırılmasından elde edilen değerdir.

Parker ve Vannest'te (2009, s.364) TÇÖ için verilen ölçütler Tau-U için de kullanılabilir. Bu ölçütlere göre 0-65 arası yüzdeler "zayıf veya küçük" etkiye, 66-92 arası yüzdeler "orta düzey" etkiye, 93-100 arası yüzdeler de "yüksek" etkiye işaret eder. Bu ölçütler spesifik olarak Tau-U için geliştirilmediğinden ölçütlerin kullanılmasında ihtiyatlı davranılması gerekmektedir. Şekil 6'daki veriyi kullanarak elde edilen %82 değerindeki Tau-U değeri "orta düzeyde" bir müdahale etkisine işaret eder.

Tau-U ayrı ayrı eğilimi ve örtüşmemeyi ya da her ikisini de aynı anda analiz edebilme esnekliğine sahiptir. Bu esnekliğin yanında Tau-U indeksi diğer örtüşme indekslerine göre daha fazla istatistiksel güce sahiptir. Tau-U indeksinin ayırt ediciliği ve görsel analizlerle tutarlı oluşu da bu indeksin güçlü yanlarından. Tau-U indeksinde eğilimin kontrol edilmeye çalışılmasından dolayı elde edilen değerlerde bir düşüş oluşmaktadır. Bu düşüş Parker vd. (2011) tarafından bu indeksin bir sınırlılığı olarak görülmektedir. Tau-U indeksinin AB ve ABA desenlerinden daha karmaşık tek denekli deneysel desen durumlarında hesaplanamaması da bir sınırlılık olarak görülmektedir.

Buraya kadar açıklanan örtüşme yöntemlerinin grup desenli çalışmalarda parametrik olmayan başat istatistikler (nonparametric dominance statistics) olarak bilinen yöntemlerden adapte edildiği göz önünde bulundurulmalıdır. Başat istatistik, bir gruptan rasgele seçilen bir puanın ikinci bir gruptaki bir puanı geçme olasılığı olarak tanımlanabilir (Parker, Vannest ve Davis, 2011). Başat istatistiklerin eşdeğerleri olan örtüşme istatistiklerinin tek denekli deneysel araştırmalarda kullanılması bu alandaki yayınlara daha fazla güvenilirlik kazandırmıştır.

Düşüş Veri Yüzdesi (DVY)

Campbell (2003) tarafından başlama düzeyi ortalama düşüş yüzdesi (Mean Baseline Reduction [MBR]) olarak adlandırılan düşüş veri yüzdesi (Percentage Reduction Data) O'Brien ve Repp (1990) tarafından geliştirilmiştir. Hedeflenen davranıştaki uygulama düzeyinden kaynaklanan azalmayı hesaplamaya yaramaktadır. DVY indeksi başlama düzeyinden düşüş yüzdesi olarak da bilinir ve davranışın ne kadar azaldığını tespit etmek için uygulanır.

DVY değerini hesaplayabilmek için uygulama ve başlama düzeylerindeki verilerin aritmetik ortalamalarını hesaplamak yeterlidir. Bir AB deseninde DVY hesaplanırken şu adımlar takip edilebilir:

1. Başlama düzeyinin aritmetik ortalaması hesaplanır,
2. Uygulama düzeyinin aritmetik ortalaması hesaplanır,
3. Başlama düzeyinin aritmetik ortalamasından uygulama düzeyinin aritmetik ortalaması çıkartılır,
4. Üçüncü adımda bulunan değer başlama düzeyi aritmetik ortalamasına bölünür (Campbell, 2003),
5. Dördüncü adımda elde edilen değer 100 ile çarpılması ile DVY yüzdesi elde edilir. Başka bir şekilde göstermek gerekirse; M_B başlama düzeyi ortalaması ve M_U uygulama düzeyi ortalaması olmak üzere

$$DVY = \frac{M_B - M_U}{M_B} \times 100 \quad (2)$$

eşitliği ile DVY değeri hesaplanır.

Eğer uygulama düzeyindeki işlemin amacı bir davranışı azaltmak değil de artırmak ise Eşitlik 2'de pay kısmındaki başlama düzeyi ortalaması (M_B) ile uygulama düzeyi ortalaması (M_U) yer değiştirilerek DVY değeri hesaplanır. Yani uygulama düzeyi ortalamasından başlama düzeyi ortalaması çıkarılıp başlama düzeyi ortalamasına bölünmesi gerekir. Şekil 6'da sunulan veri üzerinden DVY değerini hesaplayabilmek için öncelikle her iki düzey için de aritmetik ortalama değerini hesaplamamız gerekir. Şekil 6'daki veriye göre başlama düzeyinin aritmetik ortalaması 6.8 ve uygulama düzeyinin aritmetik ortalaması 9.2 olarak hesaplanır. Bu değerler ışığında DVY yüzdesi $((9.2 - 6.8)/6.8) \times 100 = 35.3$ olarak hesaplanır.

DVY değeri %100 olarak bulunduğunda problemlili davranışın tamamen uzaklaştırıldığına, %0 DVY değeri ise başlama düzeyine göre hiçbir değişiklik gözlenmediğine işaret eder. Negatif DVY değeri de problemlili davranışın uygulama düzeyinde arttığını göstermektedir. Bazı araştırmacılar, DVY değerini hesaplarken başlama düzeyi ve uygulama düzeyindeki verilerin son üçünün (Olive ve Franco, 2008) ya da son beşinin (Lydon, Healy, O'Reilly ve McCoy, 2013) kullanılmasını önermektedir. Eğer başlama düzeyindeki değerler sıfır olursa DVY değeri hesaplamak mümkün olmamaktadır. Bu durum DVY'nin bir sınırlılığıdır. DVY değeri için açıkça belirtilmiş bir yorumlama ölçütü bulunmamakla beraber Bell, Skinner ve Fisher (2009, s. 5) DVY değerleri için küçük (.20), orta (.50) ve yüksek (.80) olacak şekilde ölçütler kullanmıştır. Alanyazında DVY indeksinin farklı tek deneysel desenlerde nasıl hesaplanacağına dair çok fazla çalışma bulunmamaktadır. Carr, Severtson ve Lepper (2009) ABA tasarımları için DVY değerini hesaplarken sonucu başlama düzeyindeki ve uygulama düzeyindeki verileri kullanmıştır. Ayrıca Carr vd. (2009) çoklu başlama deseni için, her bir katılımcının başlama ve uygulama düzeyi verilerini, DVY değerini hesaplamada kullanmıştır.

Standartlaştırılmış Ortalama Farkı (SOF)

Grup desenli çalışmaların incelendiği meta-analizlerde genel olarak standartlaştırılmış ortalama farkı (Cohen's d , Hedges' g ve Glass' delta) indeksleri etki büyüklüğü hesaplamak için kullanılmaktadır. Tek denekli deneysel çalışmalarda da standartlaştırılmış ortalama farkına benzer bir değer kullanılabileceği araştırmacılar tarafından önerilmiştir (Busk ve Serlin, 1992; Shadish, Hedges ve

Pustejovsky, 2014). Tek denekli deneysel araştırmalardaki deneyin etkililiğini göstermek için geliştirilen (Busk ve Serlin, 1992) SOF değeri hesaplanırken şu adımlar takip edilebilir:

1. Başlama düzeyinin aritmetik ortalaması hesaplanır,
2. Uygulama düzeyinin aritmetik ortalaması hesaplanır,
3. Başlama düzeyinin standart sapması hesaplanır,
4. Uygulama düzeyinin aritmetik ortalamasından başlama düzeyinin aritmetik ortalaması çıkartılır,
5. Dördüncü adımda bulunan değer başlama düzeyinin standart sapmasına bölünür (Campbell, 2003).

Başka bir şekilde göstermek gerekirse; M_U uygulama düzeyi ortalaması, M_B başlama düzeyi ortalaması ve S_B başlama düzeyinin standart sapması olmak üzere

$$SOF = \frac{M_U - M_B}{S_B} \quad (3)$$

eşitliği ile elde edilebilir.

Şekil 6'da sunulan veri üzerinden SOF değerini hesaplayabilmek için her iki düzey için bir aritmetik ortalama değeri ve sadece başlama düzeyi için standart sapma değerini hesaplamamız gerekir. Şekil 6'daki veriye göre başlama düzeyinin aritmetik ortalaması 6.8 ve uygulama düzeyinin aritmetik ortalaması 9.2 olarak hesaplanır. Ayrıca SOF değerini elde edebilmemiz için gerekli olan başlama düzeyi standart sapması 0.632 olarak bulunur. Bu değerler ışığında SOF değeri $(9.2 - 6.8) / 0.632 = 3.79$ olarak hesaplanır. Şekil 6'daki müdahalenin etkisinin çok büyük olmadığı görsel analiz ve örtüşmeyen veri temelli indeksler ile de gösterilmektedir. Ancak SOF sonucu, uygulanan müdahalenin çok etkili bir müdahale olarak görünmesine neden olmaktadır. Bu örnek ışığında SOF'un müdahale etkililiği konusunda yanıltıcı sonuçlar sunması ihtimaldir.

Ortak ölçüğe dönüştürülebilmenin yanında önerilen diğer indekslerin çoğunun aksine standartlaştırılmış ortalama farkı (SOF) bilinen bir örnekleme dağılımına sahiptir ve regresyon, ANOVA gibi istatistiklerin uygulanabilmesine izin vermektedir. Binom (binomial) işaret testi yardımıyla bu değer için güven aralıkları da hesaplanabilmektedir (Busk ve Serlin, 1992). Olive ve Franco'ya (2008, s.8) göre SOF yönteminin birçok güçlü yanı bulunmaktadır. Bunlardan birincisi, hesaplamalarda aritmetik ortalamasının kullanılması ve bu ortalama değerinin hem artan hem de azalan uygulama düzeyi etki durumlarında hesaplanabilmesidir. Örtüşmeyen veri yüzdesine dayanan yaklaşımların aksine SOF yönteminde hiçbir veri değeri göz ardı edilmez. SOF yönteminin başka bir güçlü yanı ise, bilindik etki büyüklüğü değerleri (Cohen d) ile aynı şekilde yorumlanabilmesidir. Olive ve Franco'ya (2008, s.7) göre tek denekli deneysel çalışmalarda elde edilen SOF değeri Cohen (1988) tarafından önerilen ölçütlere göre yorumlanabilir. Bu ölçütlere göre $d = 0.2$ küçük, $d = 0.5$ orta ve $d = 0.8$ büyük etki büyüklüğü olarak yorumlanmaktadır. Araştırmacıların bu ölçütleri kullanırken ihtiyatlı davranmaları önerilmektedir. Olive ve Smith (2005) birçok farklı tek denekli deneysel desenin bulunduğunu ve bu desenlerin her biri için SOF değerinin hesaplanmasında farklı veriler kullanarak aynı yolun izlenmesi gerektiğini belirtmektedir. Örneğin, farklı deneklerin kullanıldığı çoklu başlama düzeyi desenlerinde her bir denek için ayrı ayrı SOF değeri hesaplanması gerektiği belirtilmektedir. Geri dönüşlü desenlerde de orijinal başlama düzeyi ve en son uygulanan uygulama düzeyi SOF hesaplamasında kullanılmalıdır. Bir ABA'B' deseni için SOF değeri hesaplaması da mümkündür. Bu durumlarda araştırmacının her iki AB deseni (AB ve A'B') için SOF değerlerini (SOF_1 ve SOF_2) hesaplayıp bu iki değer aritmetik ortalamasını alması gerekir. SOF değerinin eleştirilen yanları arasında zamandan kaynaklı olası bir eğilim etkisini ve düzeyler arasındaki eğim değişimini hesaba katamaması sayılabilir. Standartlaştırılmış ortalama farkı, otokorelasyondan etkilenmesi ve ortaya çıkan etki büyüklüğü değerlerinin genellikle çok büyük olması nedeniyle sorunlu bir etki büyüklüğü hesaplama yöntemi olarak görülmektedir. Bu bağlamda, son yıllarda birçok araştırmacı bu indeks geliştirilmesi için çaba göstermektedir. Birçok araştırmacı tarafından, ÖVY'de olduğu gibi, sıklıkla kullanılmasına karşın yukarıda bahsedilen sınırlılıklar giderilmeden meta-analizler kapsamında

kullanılmasının sorunlu olduğu düşünülmektedir. Ayrıca SOF istatistiği diğer regresyona dayalı olmayan yaklaşımların çoğunda olduğu gibi boylamsal veri yapısından kaynaklanan bağımlılığı hesaba katmamaktadır. Bu durum standart hata oranlarının yanlış tahmin edilmesine yol açmaktadır. Dolayısıyla I. tür veya II. tür hata oranlarında artışa neden olabilmektedir.

Etki Büyüklüğü İndekslerinin Raporlanması

Farklı çalışmalardan elde edilen etki büyüklüğü değerlerinin bir meta-analiz çalışmasında kullanılabilmesi için bir araya getirilmesi (özetlenmesi) ve yorumlanması gerekmektedir. Maggin ve diğerlerine (2011) göre farklı çalışmalardan elde edilen etki büyüklüklerinin başarılı bir şekilde birleştirilmesi (özetlenmesi) genellenebilir sonuçlar elde etmek ve diğer araştırmacılara hangi durumlarda uygulama düzeyindeki etkinin miktarının fazla ya da az olacağını göstermek adına faydalı olmaktadır. Farklı çalışmalardan elde edilen etki büyüklüğü değerlerini özetleme işi grup desenli çalışmalarda geleneksel meta-analiz modelleri vasıtasıyla ağırlıklı ortalama elde ederek gerçekleştirilmektedir. Geleneksel meta-analiz için önerilen aşamalar (alanyazın taraması, çalışmaların seçimi, etki büyüklüğü hesaplanması ve özetlenmesi) tek denekli deneysel çalışmaların meta-analizi için de geçerli olmaktadır. Tek denekli deneysel çalışmalarda en çok farklılık gösteren kısım ise farklı çalışmalardan elde edilen etki büyüklüğü değerlerinin nasıl özetleneceğidir.

Maggin ve diğerlerinin (2011) yaptığı araştırmaya göre tek denekli deneysel araştırma alanlarında araştırmacılar beş farklı yöntemi uygulamaktadır. Bu yöntemlerden en çok tercih edileni, tüm çalışmalardan elde edilen etki büyüklüğü değerlerinin aritmetik ortalama yoluyla özetlenmesidir. İkinci en çok kullanılan özetleme yöntemi ise ağırlıklı ortalama kullanılarak tüm çalışmalardan bir ortalama değer elde etme işidir. Ağırlıklı ortalamayı hesaplayabilmek için geleneksel meta-analizdeki gibi ters varyans ağırlıkları, güven aralıkları ya da standart hata gibi değerlerin elde ediliyor olması gerekmektedir. Ayrıca çok düzeyli modeller ile de ağırlıklı ortalama elde edebilmek mümkündür. Etki büyüklüklerinin özetlenmesinde kullanılan diğer yöntemler tüm çalışmaların ortanca değerini rapor etmek, betimleyici açıklamalar sunmak ya da elde edilen çalışmaların yüzde kaçının yüksek etkili, yüzde kaçının orta etkili ve yüzde kaçının düşük etkili olduğunu rapor etmek gibi uygulamaları içermektedir.

Tek denekli deneysel çalışmalarda yukarıda açıklanan yöntemlerden birini kullanarak ortak etki büyüklüğü değeri elde etme işinde araştırmacının birçok noktaya dikkat etmesi gerekmektedir (Vannest ve Davis, 2013, s.107–108). Öncelikle meta-analiz yapacak araştırmacının tek denekli deneysel araştırmaların doğasına hâkim olması gerekir. Araştırmacı incelediği tek denekli deneysel araştırma durumunda hangi düzeyler arasında karşılaştırma yapacağını iyi bilmelidir (Vannest ve Davis, 2013). Yukarıda açıklamaları sunulan indekslerin hesaplanması AB deseni durumları için gösterilmiştir. Bu indekslerin hemen hepsi karmaşık tek denekli deneysel desenlere de uygulanabilmektedir. Bu karmaşık desenlerde bile çoğu araştırmacı bir A düzeyi ve bir B düzeyi seçerek yukarıda açıklanan indeks değerlerini elde etmektedir. Önemli olan araştırmacının neden böyle bir karşılaştırma yaptığını gerekçelendirmesidir. Ayrıca tek denekli deneysel araştırmalarda meta-analiz yapmayı planlayan araştırmacı seçtiği etki büyüklüğü indeksinin hesaplanma yöntemini de seçme nedeni ile açıklamalı ve nasıl bir hesaplama yaptığını ayrıntılı bir şekilde sunmalıdır. WWC (Kratowill vd., 2010) ve birçok araştırmacı, tek denekli deneysel araştırmalardan elde edilen veri örüntülerine özgü tüm durumlara karşı hassas olan bir etki büyüklüğü indeksi geliştirilene dek, birden fazla indeks kullanılarak etki büyüklüğü hesaplanmasını ve rapor edilmesini önermektedir. Meta-analize katılan her bir çalışma için etki büyüklüğü değerini raporlamanın yanında araştırmacı çalışmalara ait katılımcı sayısı, incelenen davranış türü ve sayısı, çalışma sayısı ve kullanılan ağırlıklandırma türü gibi ayrıntıları da rapor etmelidir (Vannest ve Davis, 2013). Etki büyüklüğü indeksi değerlerini rapor ederken güven aralıklarının da sunulması gerekmektedir.

TARTIŞMA ve SONUÇ

Bu çalışmada tek denekli deneysel araştırmalar için geliştirilmiş olan ve etki büyüklüğü değeri olarak kullanılan 10 indeks incelenmiştir. Farklı araştırmacılar tarafından geliştirilen bu indekslerin nasıl hesaplanacağı örnek veriler üzerinde gösterilip açıklanmaya çalışılmıştır. Bu çalışmanın tek denekli deneysel çalışmalar yapan araştırmacılara fayda sağlayacağı beklenmektedir. İlk olarak yurt içi ve yurt dışı alanyazında en çok tercih edilen indeksler olan ÖVY ve SVY indeksleri açıklanmış ve bu indekslere alternatif olarak önerilen indekslerin özellikleri sunulmuştur. Bu çalışmada karşılaştırılan indekslerin arasında ÖVY'nin kolay hesaplanabilir olduğu için araştırmacılar tarafından çok fazla tercih edildiği fakat sınırlılıklarının göz ardı edildiği alanyazın taramamızda gözlenmiştir. ÖVY'ye benzer diğer indekslerden bazılarının (ÖTVY, TÇÖ, BOAVY ve Tau) ÖVY'nin sınırlılıklarına çözüm bulmasına karşın geleneksel anlamdaki etki büyüklüğü değerlerinin sağladığı avantajları sunmadığının altı çizilmiştir.

Etki büyüklüğü indekslerinin farklı davranış durumlarında kullanılıyor olması bu indekslerin ayırt edici özellikleri arasında yer almaktadır. Çalışmamızda incelenen indekslerin çoğu hedef davranışın artmasının ya da azalmasının beklendiği durumlarda kullanılırken SVY indeksi sadece hedeflenen davranışın ortadan kalkması beklenen durumlarda kullanılmaktadır. Eğer araştırmacı katılımcıdaki davranışın ortadan kaldırılması üzerine bir uygulama yapıyorsa tercih edeceği indeks SVY indeksi olabilir. Başka bir farklılık da indeksler tarafından kullanılan veri sayısıdır. Bazı indeksler (örneğin SVY) sadece uygulama düzeyindeki veri değerlerini hesaba katmakta iken bazı indeksler (ÖVY, BOAVY, ÖTVY vb.) başlama düzeyindeki verilerden birini ya da çok azını hesaba katmaktadır. Bu nedenle hem başlama hem de uygulama düzeylerindeki bütün verileri hesaba katan indekslerin (TÇÖ, Tau, Tau-U, SOF, DVY) kullanılması daha doğru bir uygulama olacaktır. Verideki değerlerin hepsini hesaba katan istatistikler aykırı değerlerden daha az etkilenmektedir. Özellikle TÇÖ, Tau ve Tau-U indeksleri diğer nonparametrik indekslerden daha fazla istatistiksel güce sahip oldukları için daha kesin sonuçlar üretme eğilimindedir (Parker vd., 2011). Çünkü bazı veri noktalarının göz ardı edilmesi uygulama düzeyinin etkililiği hakkında yanıltıcı kararlar verilmesine yol açabilir. Tüm veri noktalarını hesaba katmanın yanında SOF ve DVY gibi indeksler geleneksel etki büyüklüğü değerlerine dönüştürülebilme ya da aynı şekilde yorumlanabilme özelliğine sahip olmasıyla öne çıkmaktadır. Bu bağlamda özellikle SOF hesaplamalarında elde edilen değerlerin grup desenli çalışmalara oranla daha büyük çıktığı gözlenmektedir. Bu özelliğe sahip bir başka indeks de Phi değerini elde ederek *d* değerine dönüştürülebilme ÖTVY indeksidir. Bu çalışmada sunulan yüzdelik indekslerinin çoğu grup desenli çalışmalardan elde edilen etki büyüklüğü değerleri şeklinde yorumlanabilen değerler üretmemektedir. Bu çalışmada ele alınan indekslerin sahip olduğu başka bir sınırlılıkta çoğu indeksin tek denekli deneysel çalışmalardaki (zaman serilerinde olduğu gibi) otokorelasyon ve eğilim etkisini hesaba katamamasıdır. Otokorelasyon etkisi aynı bireyden toplanan tekrarlı verilerin birbirleriyle pozitif ya da negatif korelasyon göstermesi olarak tanımlanmaktadır. Bu etki standart sapma değerlerinin yanlış hesaplanmasına yol açmaktadır. Bu olumsuz etkiyi hesaba katabilen indeks (Tau-U) değerleri araştırmacılar tarafından daha ideal olarak adlandırılmaktadır. Otokorelasyon ve eğilim olmayan veri durumlarında ÖVY ve BOEAVY'nin kullanımını kolaylık olması bakımından bir müdahalenin etkili olup olmadığını belirlemek için uygulamacılara önerilebilir. Ancak meta-analizler kapsamında daha gelişmiş yöntemlerin (ör., Tau-U) kullanılması daha doğru olacaktır.

KAYNAKÇA

- Alresheed, F., Hott, B. L., & Bano, C. (2013). Single subject research: A synthesis of analytic methods. *The Journal of Special Education Apprenticeship*, 2(1), 1–18.
- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case*. *Behaviour Research and Therapy*, 31(6), 621–631.
- Allison, D. B., & Gorman, B. S. (1994). "Make things as simple as possible, but no simpler." A rejoinder to Scruggs and Mastropieri. *Behaviour Research and Therapy*, 32(8), 885–890.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

- Aslan, C., Yalçın, G. & Özdemir, S. (Mayıs, 2016). *Sosyal öykü tekniğinin etkililiği: Betimsel değerlendirme ve meta-analiz çalışması*. Teacher Education in Special Education, Vocational Training and Sports Konferansı (ELMIS), Konya, Türkiye.
- Aydın, O. (2017). *Otizm spektrum bozukluğu olan bireylere matematik becerilerinin öğretimi: tek-denekli araştırmalarda betimsel ve meta analiz* (Yayımlanmamış yüksek lisans tezi). Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Eskişehir.
- Bell, R. J., Skinner, C. H., & Fisher, L. A. (2009). Decreasing putting yips in accomplished golfers via solution-focused guided imagery: A single-subject research design. *Journal of Applied Sport Psychology*, 21(1), 1–14.
- Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, 2, 129–141.
- Bozkuş-Genç., G. (2017). *Otizm spektrum bozukluğu olan çocuklara soru sorarak iletişim başlatmanın kazandırılmasında temel tepki öğretiminin etkileri* (Yayımlanmamış doktora tezi). Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Eskişehir.
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp.187–212). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- 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, 120–138.
- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28(2), 234–246.
- Carr, J. E., Severtson, J. M., & Lepper, T. L. (2009). Noncontingent reinforcement is an empirically supported treatment for problem behavior exhibited by individuals with developmental disabilities. *Research in Developmental Disabilities*, 30(1), 44–57.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3–8.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Heyvaert, M., Saenen, L., Campbell, J. M., Maes, B., & Onghena, P. (2014). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: An updated quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 35(10), 2463–2476.
- Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment*, 7, 107–118.
- Huitema, B. E., & Mckean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, 60(1), 38–58.
- Karasu, N. (2009a). Özel eğitimde delile dayalı yöntemlerin belirlenmesi: Tek denekli çalışma analizleri ve karşılaştırmaları. *Türk Eğitim Bilimleri Dergisi*, 7(1), 143–163.
- Karasu, N. (2009b). Otizmden etkilenmiş bireylerde sosyal ve iletişim becerilerini arttıran yöntemlerin delile dayalı yöntem olarak belirlenmesi: Bir meta-analiz örneği. *Türk Eğitim Bilimleri Dergisi*, 7(3), 713–739.
- Karasu, N. (2011). Otizimli bireylerin eğitiminde video ile model olma uygulamalarının değerlendirilmesi: Bir alanyazın derlemesi ve meta-analiz örneği. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 12(2), 1–12.
- Kavale, K. A. (2001). Decision making in special education: The function of meta-analysis. *Exceptionality*, 9, 245–268.
- Kaya, F. (2015). *Otizm spektrum bozukluğu olan öğrencilere yiyecek-içecek hazırlama becerilerinin öğretiminde sesli anlatım içeren ve içermeyen video ipucunun karşılaştırılması* (Yayımlanmamış yüksek lisans tezi). Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Eskişehir.
- Kırcaali-İftar, G., & Tekin, E. (1997). *Tek denekli araştırma yöntemleri*. Ankara: Türk Psikologlar Derneği Yayınları.
- Korkmaz, Ö. T., & Diken, İ. H. (2010). Stereotipik davranışların azaltılmasında kullanılan yöntemlerin etkililiği: Betimsel ve meta analizi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 11(2), 1–12.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse*. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf adresinden alınmıştır.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

- Lydon, S., Healy, O., O'Reilly, M., & McCoy, A. (2013). A systematic review and evaluation of response redirection as a treatment for challenging behavior in individuals with developmental disabilities. *Research in Developmental Disabilities, 34*(10), 3148–3158.
- Ma, H. (2006). An alternative method for quantifying synthesis of single-subject research: Percent of data points exceeding the median. *Behavior Modification, 30*, 598–617.
- Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985–2009. *Exceptionality, 19*(2), 109–135.
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O'keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*(3), 301–321.
- Manolov, R., Sierra, V., Solanas, A., & Botella, J. (2014). Assessing functional relations in single-case designs: Quantitative proposals in the context of the evidence-based movement. *Behavior modification, 38*(6), 878–913.
- Manolov, R., & Solanas, A. (2008). Comparing N=1 effect size indices in presence of autocorrelation. *Behavior Modification, 32*, 860–875.
- Manolov, R., Solanas, A., & Leiva, D. (2010). Comparing “visual” effect size indices for single-case designs. *Methodology, 6*, 49–58.
- Manolov, R., Solanas, A., Sierra, V., & Evans, J. J. (2011). Choosing among techniques for quantifying single-case intervention effectiveness. *Behavior Therapy, 42*(3), 533–545.
- O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons with Severe Handicaps, 15*(3), 148–159.
- Olive, M. L., & Franco, J. H. (2008). (Effect) size matters: And so does the calculation. *The Behavior Analyst Today, 9*(1), 5–10.
- Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology, 25*(2-3), 313–324.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40*, 194–204.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*(4), 357–367.
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75*(2), 135–150.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*(4), 303–322.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*(2), 284–299.
- 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.
- Reichle, J. (2007). Amongst methodologies of functional behavioral assessment, functional analysis yields more effective suppression outcomes. *Evidence-Based Communication Assessment and Intervention, 1*(4), 153–155.
- Schlosser, R. W., & Koul, R. K. (2015). Speech output technologies in interventions for individuals with autism spectrum disorders: a scoping review. *Augmentative and Alternative Communication, 31*(4), 285–309.
- Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96*, 233–256.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*(3), 221–242.
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*(1), 9–19.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*, 24–33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11*, 260–71.
- Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123–147.

- Sönmez, M., & 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.
- Strain, P. S., Kohler, F. W., & Gresham, F. (1998). Problems in logic and interpretation with quantitative syntheses of single-case research: Mathur and colleagues (1998) as a case in point. *Behavioral Disorders*, 24, 74–85.
- Swaminathan, H., Rogers, H. J., Horner, R. H., Sugai, G., & Smolkowski, K. (2014). Regression models and effect size measures for single case designs. *Neuropsychological Rehabilitation*, 24(3-4), 554–571.
- Tavil, Y. Z. ve Karasu, N. (2013). Aile eğitim çalışmaları: Bir gözden geçirme ve meta-analiz örneği. *Eğitim ve Bilim*, 38(168), 85–95.
- Uysal, H. (2017). *Zihin yetersizliği olan öğrencilere temel toplama işlemlerinde akıcılık kazandırmada iki farklı uygulamanın karşılaştırılması* (Yayımlanmamış yüksek lisans tezi). Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Eskişehir.
- Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments & Computers*, 35, 1–10.
- Vannest, K.J., & Davis, H. S. (2013). Synthesizing single case research to identify evidence based treatments. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *Evidence-Based practices* (pp. 93–119). UK: Emerald Group Publishing.
- Wehmeyer, M. L., Palmer, S. B., Smith, S. J., Parent, W., Davies, D. K., & Stock, S. (2006). Technology use by people with intellectual and developmental disabilities to support employment activities: A single-subject design meta analysis. *Journal of Vocational Rehabilitation*, 24(2), 81–86.
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Merrill.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44(1), 18–28.