



## Etkili Araştırma Sentezleri Yapabilmek için Bir Araştırma Yöntemi: Meta-Analiz \*

Ulaş Üstün <sup>1</sup>, Ali Eryılmaz <sup>2</sup>

### Öz

Son yıllarda, eğitim bilimlerindeki birincil çalışmaların sayısı artıkça kapsamlı ve sistematik araştırma sentezlerine olan ihtiyaç da artmaktadır. En etkili araştırma sentezi yollarından bir tanesi olan meta-analizin çeşitli uygulamalarının, sosyal bilimler ve eğitim bilimleri de dâhil olmak üzere birçok alanda teşvik edilmesinin temel sebebi budur. Bu makalenin temel amacı, meta-analizin diğer araştırma sentezi yöntemlerine kıyasla zayıf ve güçlü taraflarını sorgulayarak meta-analiz için kavramsal bir çerçeve oluşturmaktır. Bununla birlikte, sabit-etki ve rastgele-etkiler modellerinin karşılaştırılması, farklı etki büyüklüğü ölçüleri, analiz birimi, yayın yanlılığı ve birincil çalışmaların kalitesi gibi geçerlikle ilgili sorunlar ile heterojenlik, ara-değişken ve güç analizleri gibi bazı metodolojik ve istatistiksel hususlar detaylı şekilde tartışılmaktadır. Ayrıca bu makale kapsamında, meta-analizde kullanılan istatistiksel analizler için kullanılabilecek yazılımlar hakkında kısa ve öz bir bilgiye ve meta-analizlerin raporlanması için geliştirilen standartların özetine yer verilmiştir. Sadece bilimin birikimli doğası için değil aynı zamanda politika belirleyiciler ve uygulayıcılar için çok önemli bir yere sahip olması sebebiyle meta-analizlerin ne kadar iyi yapıldığı ve raporlandığı çok büyük bir önem taşımaktadır. Buradan hareketle, bu makaleyle meta-analiz uygulayıcılarına meta-analiz araştırmaları sırasında yararlanabilecekleri giriş seviyesinde bir rehber sağlamayı amaçlıyoruz.

### Anahtar Kelimeler

Meta-analiz  
Araştırma sentezi  
Etki büyüklüğü  
Ara-değişken analizi  
Sabit-etki modeli  
Rastgele-etkiler modeli  
Yayın yanlılığı

### Makale Hakkında

Gönderim Tarihi: 27.02.2014

Kabul Tarihi: 14.07.2014

Elektronik Yayın Tarihi: 06.08.2014

DOI: 10.15390/EB.2014.3379

### Giriş

Araştırma sentezleri sadece bilimsel bilginin birikimli doğası sebebiyle değil aynı zamanda güçlü bilimsel deliller ışığında politika belirleyicilere rehberlik sağlaması ve literatürdeki tutarsız sonuçları açıklamadaki potansiyeli nedeniyle bilimsel girişimde vazgeçilmez bir yere sahiptir (Üstün, 2012).

\* Bu çalışma büyük ölçüde Üstün (2012) doktora tezinin yöntem bölümüne dayanmaktadır.

<sup>1</sup> Artvin Çoruh Üniversitesi, Eğitim Fakültesi, İlköğretim Bölümü, Türkiye, [ulasustun@artvin.edu.tr](mailto:ulasustun@artvin.edu.tr)

<sup>2</sup> Orta Doğu Teknik Üniversitesi, Eğitim Fakültesi, Orta Öğretim Fen ve Matematik Alanları Eğitimi Bölümü, Türkiye, [eryilmaz@metu.edu.tr](mailto:eryilmaz@metu.edu.tr)

Bilimsel sürecin en önemli yönlerinden bir tanesi olan birikimli olma özelliği aynı zamanda bilimsel bilginin katlanarak gelişmesini de sağlamaktadır. Isaac Newton'ın 300 yıldan uzun zaman önce dile getirdiği "Eğer daha uzakları görebildiysem, omuzlarında yükseldiğim devler sayesinde" sözünün altında yatan mantık da budur. Bilim tarihi boyunca bu düşünce çok açık ve neredeyse tartışmasız olmasına rağmen bilim insanlarının yeni bilgileri entegre edebilmesi için eski bilgileri sentezleme sorumluluğu oldukça yakın zamanda kabul edilmiştir (Chalmers, Hedges, & Cooper, 2002). Bugün araştırma sentezlerinin, sadece var olan paradigmanın kapsamlı veya daha eksiksiz bir resmini sağlayarak eski ve yeni bilimsel bilgi arasında bağlantılar kurmakta değil aynı zamanda var olan bilginin kapsamının genişletilmesinde yardımcı olma konusunda anahtar bir role sahip olduğu yaygın bir şekilde kabul edilmiştir (Card, 2012; Chalmers vd., 2002; Chan & Arvey, 2012; Hunter & Schmidt, 2004; Mulrow, 1994).

Araştırma sentezlerinin bilimsel uğraşın birikimli doğasına katkısı vazgeçilmez olsa da, bu metodun artan akademik farkındalık ve popüleritesinin sebebi politika belirleyiciler ve uygulayıcılara sağladıklarıdır (Chalmers vd., 2002). Bu bağlamda, Petticrew ve Roberts (2006) tek bir birincil araştırma ile bir tarama çalışmasındaki tek bir katılımcı arasında bir analogi yapmaktadır. Bu analogi, bir tarama çalışmasında sonuca ulaşabilmek için çok sayıda katılımcıya ihtiyaç duyulması düşüncesine dayanmaktadır. Tek bir cevabın önemli olduğunu belirten araştırmacılar, buna rağmen diğer bir katılımcıdan tersi bir görüş elde etmenin her zaman olası olduğunu altını çizmektedir. Bu nedenle, araştırmaların sonuçları birçok katılımcıdan elde edilen birçok cevaba dayanmalıdır. Araştırmacılar, politika belirleyiciler ve uygulayıcıların vereceği kararların da birçok çalışma sonucunda ulaşılan fikir birliğine dayalı olarak yapılandırılması gerektiğini savunmaktadırlar. Benzer şekilde, Davies (2000) ne kadar iyi tasarlanmış ve uygulanmış olursa olsun tek bir deneyin, "zaman, örneklem ve bağlam" gibi kendine has özellikleri tarafından sınırlandırılacağını vurgulamaktadır. Ayrıca, Chalmers ve diğerleri (2002), araştırma sentezlerinin karar verme sürecindeki önemini vurgulamakta ve bu sentezlerin gelecekteki pozisyonunun büyük ihtimalle, birincil çalışmalar tarafından sağlanan bilgi kırıntılarının araştırma sonuçlarına dayalı karar almak isteyen insanlara yeterli yararı sağlamadığı gerçeğiyle yüzleşen akademik çevre dışından insanlar tarafından yaratılacağını iddia etmektedir.

Bilimsel bilginin birikimli doğasına katkısının ve politika belirleyicilere ve uygulayıcılara rehberlik sağlamasının yanında, araştırma sentezinin bilimsel uğraşın vazgeçilmez bir parçası olmasının bir başka sebebi de ilişkilerdeki tutarlılığı değerlendirme ve literatürdeki tutarsız sonuçları ve anlaşmazlıkları açıklama potansiyelidir (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hunt, 1997; Hunter & Schmidt, 2004; Mulrow, 1994; Petticrew & Roberts, 2006). Hangi bilimsel disiplin göz önünde bulundurulursa bulunsun, benzer bir konuda benzer şekilde tasarlanmış araştırma çalışmalardan çelişkili sonuçlar elde edilmesi seyrek rastlanan bir durum değildir (Rosenthal & DiMatteo, 2001). Ancak, sosyal bilimlerde ve eğitim bilimlerinde durum çok daha karmaşık hale gelmektedir çünkü insan davranışları çok daha karışık ve açıklanması zordur. Aynı zamanda, çalışmaların iç geçerliğini etkileyen ve tamamıyla yok edilmesi kolay olmayan çok sayıda tehdit bulunmaktadır. Bu bağlamda, Berliner (2002) "Bence, biz (eğitim araştırmacıları) bütün bilim dalları arasında en zor olana sahibiz! Biz doğa bilimcilerinin kabul edilemez bulduğu koşullarda bilim yapıyoruz" sözleriyle önemli bir noktaya dikkat çekmiştir. Ayrıca, eğitim araştırmalarında sınıftaki öğretim sırasındaki bağlamın, öğretmen davranışı, sosyo-ekonomik faktörler, öğrenme motivasyonu gibi birçok değişken arasındaki 10. veya 15. düzeyde ilişkiler içerdiğini ve bu ilişkilerin eğitim araştırmalarında birçok çelişkili bulguya sebep olduğunu iddia etmektedir. Bu yüzden, son yıllarda eğitim çalışmaları, politika belirleyicilere ve uygulayıcılara neyin çalışıp neyin çalışmadığı konusunda yardımcı olamaması sebebiyle ciddi şekilde eleştirilmektedir (J. Bennett, 2005). Bu bakış açısıyla, araştırma sentezleri, genel sonuçları özetleme ve bu sonuçlardaki her hangi bir heterojenliğin veya çelişkinin sebeplerini açıklama konusundaki işlevselliği sebebiyle ciddi şekilde teşvik edilmelidir.

Bu makalenin temel amacı, meta-analizin diğer araştırma sentezi yöntemlerine kıyasla zayıf ve güçlü taraflarını içeren kavramsal bir çerçeve oluşturmak ve araştırma sentezi yapmanın etkili bir yolu olarak meta-analizi tanıtmaktır. Ayrıca bu makalede, istatistiksel ve metodolojik hususlarla birlikte meta-analizin nasıl yapılacağı detaylı şekilde açıklanmaktadır.

Bu bağlamda, ilk olarak meta-analizin kendine has özellikleri ve kısaca tarihsel gelişimi hemen sonraki bölümlerde açıklanmaktadır. Daha sonra, meta-analizin diğer metotlara kıyasla güçlü tarafları tartışılmakta ve hemen ardından meta-analizle ilgili eleştiriler sorgulanmaktadır. Meta-analizde kullanılan başlıca iki model karşılaştırıldıktan sonra yayın yanlılığını tespit etme, büyüklüğünü belirleme ve düzeltme yolları açıklanarak meta-analizde geçerlikle ilgili sorunların nasıl çözüldüğü açıklığa kavuşturulmaktadır. Ardından, kodlama güvenilirliği, farklı etki büyüklüğü ölçüleri ve heterojenlik, ara değişken ve güç analizlerinin yapılma yolları analiz edilmektedir. Daha sonra, yazılım kullanılarak meta-analizin yapılışı tartışılmaktadır. Son olarak bu makale kapsamında, meta-analiz çalışmalarının raporlaştırılması için geliştirilen standartlar özetlenmektedir.

### Meta-Analiz Nedir?

Literatürde açıkça görüldüğü üzere, meta-analizin araştırma sentezi yapmanın en yaygın yollarından bir tanesi olduğu konusunda neredeyse bir görüş birliği vardır (Lipsey & Wilson, 2001; Schulze, 2007). Bununla birlikte, literatürde “meta-analiz” ile ne kastedildiğine dair bir uzlaşma bulunmamaktadır. Bazı araştırmacılar “meta-analizi” bir araştırma yöntemi olarak tanımlarken bazıları ise “meta-analizi” araştırma sentezi içerisinde kullanılan bir analiz tekniği olarak kabul etmektedir (Shelby & Vaske, 2008). Cooper ve Hedges (2009) meta-analiz teriminin genellikle araştırma sentezinin eş anlamlısı, yani bir araştırma yöntemi olarak kullanıldığını iddia etmektedir. Ancak, kendileri bu terimi, tüm araştırma sentezi girişiminden ziyade araştırma sentezi içinde kullanılan istatistiksel bir analiz olarak kullanmayı tercih etmektedir. Benzer şekilde, “meta-analiz” teriminin isim babası olan Glass bu terimi “bulguların entegre edilmesi amacıyla bireysel çalışmalardan ortaya çıkan çok sayıda analizin istatistiksel analizi” anlamında kullanmıştır (1976, s. 3). Bununla birlikte, Glass “meta-analizin olmazsa olmazının araştırma raporlarının bulgu ve karakteristiklerine araştırma yöntemlerinin uygulanması” olduğunu vurgulamıştır (1982, s. 93). Ayrıca, Glass, McGaw ve Smith (1981) meta-analizin “... bir teknik değil, aksine birçok ölçme ve istatistiksel analiz tekniklerini kullanan bir perspektif” (s. 21) olduğuna işaret etmektedir. Shelby ve Vaske (2008) “Gerçek bir meta-analizi neyin oluşturduğu tartışmalıdır” (s. 97) sözleriyle meta-analizin tanımıyla ilgili görüş ayrılıklarına dikkat çekmiştir. Diğer taraftan, Rosenthal ve DiMatteo (2001)’e göre, Smith ve Glass (1977)’in çalışmasıyla birlikte meta-analizin ne olduğu netleşmiş ve anlaşılmıştır ki: “meta-analiz istatistiksel bir teknikten fazlasıdır; (meta-analiz) bir grup araştırmayı sistematik olarak incelemek için kullanılan bir yöntemdir...”

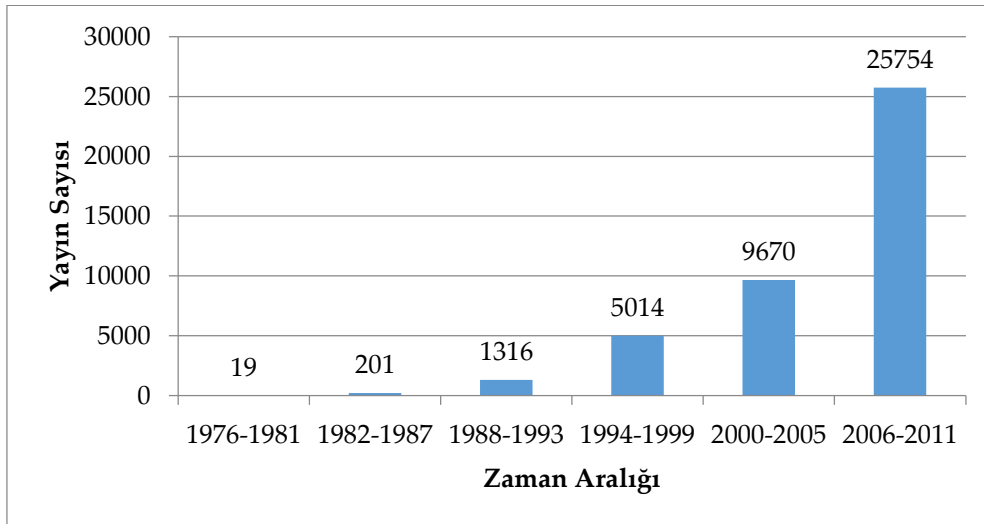
Buna ek olarak, Glass (1976) çalışmasında birincil analiz, ikincil analiz ve meta-analiz arasındaki ilişkiyi tanımlamaktadır. Glass birincil çalışmayı “araştırmadaki verilerin orijinal analizi” olarak, ikincil analizi ise “orijinal bir araştırma sorusuna daha iyi istatistiksel tekniklerle cevap vermek veya yeni soruları eski verilerle cevaplamak amacıyla verilerin yeniden analizi” olarak tanımlamıştır (s. 3). Meta-analizin ise “analizlerin analizi” anlamına geldiğini ve ikincil analiz uygulamalarını geliştirmeyi hedeflediğini iddia etmiştir.

Bu makalede, meta-analiz tüm araştırma sentezi sürecine karşılık gelecek şekilde kullanılmaktadır. Bir başka ifadeyle, “meta-analiz” terimi makale boyunca bir araştırma yöntemi olarak kullanılmaktadır. Bunun temel sebebi, meta-analizin bazı araştırma basamaklarında, olası ara değişkenlerin analizi gibi, kendine has özelliklerinin olmasıdır. Dolayısıyla, meta-analizin sadece istatistiksel bir teknik olarak tanımlanması bu karakteristik özelliklerinin göz ardı edilmesine sebep olacaktır. Literatürde bazı araştırmacıların meta-analizi benzer şekilde tanımladıkları görülmektedir (Fitz-Gibbon, 1985; Gliner, Morgan, & Harmon, 2003; Lundahl & Yaffe, 2007; Normand, 1999; Rosenthal & DiMatteo, 2001; Sánchez-Meca & Marín-Martínez, 2010a). Bu bağlamda, meta-analiz “belirli bir konu hakkındaki bir grup birincil çalışmanın sonuçlarını, o konudaki son gelişmelere karar verebilmek için nicel olarak entegre etmeyi amaçlayan bir araştırma yöntemi” (Sánchez-Meca & Marín-Martínez, 2010a, s. 274) olarak tanımlanabilir.

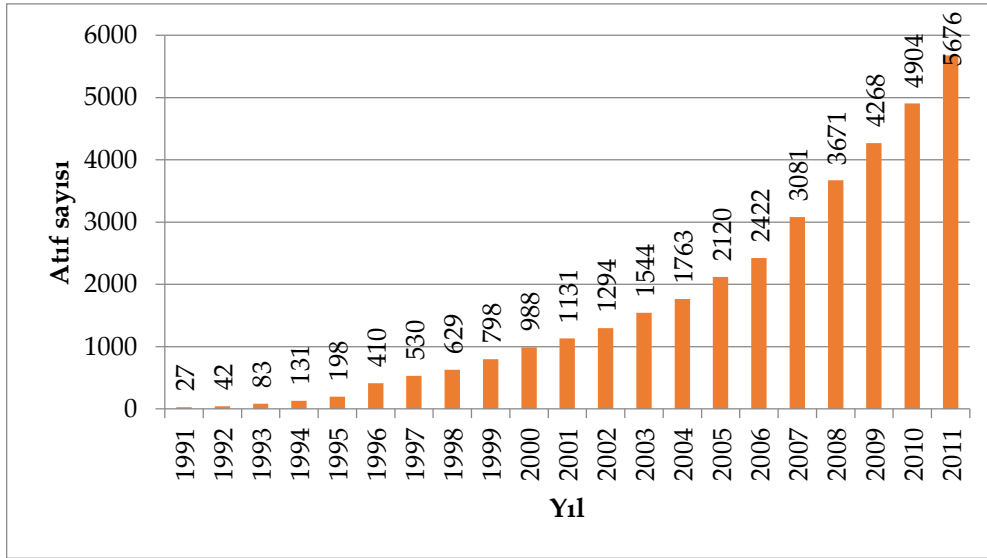
## Meta-Analizin Kısa Tarihçesi

Karl Pearson (1904) tarafından ortalama korelasyon katsayıları kullanılarak farklı çalışmaların bulgularının sentezlenmesi için gerçekleştirilen çalışma, bugün bildiğimiz anlamdaki araştırma sentezlerinin başlangıç noktası olarak kabul edilebilir (Chalmers vd., 2002; Lipsey & Wilson, 2001; O'Rourke, 2007). Ancak, Lipsey ve Wilson (2001) meta-analizin modern çağının Glass (1976), Rosenthal ve Rubin (1978), Schmidt ve Hunter (1977), Smith ve Glass (1977), Rosenthal ve Rubin (1978) ve Smith, Glass ve Miller (1980) gibi çalışmalarla başladığını ileri sürmektedir. Bir araştırma sentezi yöntemi olarak meta-analizin kullanılmasıyla ilgili bazı eleştiriler bulunsa da (Eysenck, 1978, 1984, 1994; Feinstein, 1995; Shapiro, 1994), farklı alanlardaki meta-analiz çalışmalarının sayısı giderek artmış ve Glass'ın "meta-analiz" terimini ilk kez kullandığı 1976 yılından bu yana meta-analiz nicel bir araştırma sentezi yöntemi olarak artan bir şekilde daha popüler olmuştur (Berman & Parker, 2002; Dalton & Dalton, 2008; Fitzgerald & Rumrill, 2003, 2005; Hedges, 1992; Hunter & Schmidt, 2004; Marin-Martinez & Sanchez-Meca, 1999; Sánchez-Meca & Marín-Martínez, 1998; Shelby & Vaske, 2008).

"Konu (topic)" olarak "Meta-analysis" anahtar kelimesiyle, Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (A&HCI) ile Conference Proceedings Citation Index in Science (CPCI-S) veritabanlarını kapsayan Web of Science ve Social Sciences and Humanities (CPCI-SSH) veritabanlarında yapılan arama 1976 ile 2012 yılları arasında yayınlanmış toplamda 45.519 sonuç vermektedir. Şekil 1, meta-analizin modern çağlarının başlangıcından 2011 yılına kadar beş yıllık zaman aralıklarında meta-analiz çalışmalarının nasıl arttığını göstermektedir. Ek olarak, Web of Science aracılığıyla "meta-analysis" ve "education" anahtar kelimeleri kullanılarak atıf alan referans taraması yapılmıştır. Bir önceki taramada kullanılan aynı zaman aralığı için ulaşılan 38.806 atıf sayısı meta-analizlerin eğitim çalışmalarındaki etkisiyle ilgili fikir vermektedir. Daha da önemlisi, Şekil 2'de görüldüğü gibi, atıfların sayısı özellikle son 20 yılda katlanarak artmıştır. 1991 ve 2000 yılları arasında 384 olan her yıl için ortalama atıf sayısı, 2001'den 2011'e kadar olan sonraki 11 yıllık sürede 2898'e ulaşmıştır. Son olarak, bu taramada elde edilen atıf raporuna göre, her bir çalışma için ortalama atıf sayısı 33,66 ve h-indeksi ise 94' tür. Bu sayı, Web of Science kapsamında 94 veya daha fazla atıf almış olan eğitimle ilgili 94 meta-analiz çalışması olduğu anlamına gelmektedir. Bu sonuç ise meta-analiz çalışmalarının eğitim araştırmaları için ne kadar önemli olduğunu göstermektedir.



Şekil 1. 1976-2011 Zaman Aralığında 'Meta-Analysis' Anahtar Kelimesi için Yapılan Aramanın Sonuçları



Şekil 2. 1991-2011 Zaman Aralığında 'Meta-Analysis' ve 'Education' Anahtar Kelimeleri için Yapılan Atıf Almış Referans Taramasının Sonuçları

### Neden Diğer Araştırma Sentezi Yöntemleri Yerine Meta-Analiz?

"Ampirik çalışmaları genellemeler yapabilmek amacıyla entegre etmeyi" (Cooper & Hedges, 2009, s. 6) amaçlayan araştırma sentezleri klasik (geleneksel, veya anlatı) derlemeler, geleneksel oy-sayımı yöntemi, birleştirilmiş anlamlılık testi ve meta-analiz gibi nitel, nicel ve karma yöntemlerle yapılabilir. Sonraki bölümlerde, meta-analiz diğer araştırma sentezleri yöntemleriyle kıyaslanmaktadır.

#### *Birleştirilmiş Anlamlılık Testi ve Oy-Sayımı Yöntemine Kıyasla Meta-Analiz*

Geleneksel oy-sayımı yöntemi ve birleştirilmiş anlamlılık testi araştırma sentezleri kapsamında kullanılacak iki nicel yöntemdir. Geleneksel oy-sayımı yöntemi basitçe istatistiksel olarak anlamlı ve anlamsız sonuçların sayımına dayanır ve son karar her bir kategorideki oylar sayılarak verilir (Borenstein vd., 2009; Bushman & Wang, 2009; Davies, 2000). Diğer taraftan, birleştirilmiş anlamlılık testi, derlenen çalışmalara ait sonuçların birleştirilmiş olasılıklarının anlamlı olup olmadığını istatistiksel olarak test etmeyi amaçlar (Bligh, 2000; Fitzgerald & Rumrill, 2003, 2005). Bu metodlar, öznel yargıyı en aza indirmeleri sebebiyle, anlatı derlemelerine kıyasla çok daha objektif olma gibi ortak bir avantaja sahip olsalar da, her ikisi de istatistiksel anlamlılık testinden kaynaklanan problemlerden dolayı sıkıntı yaşamaktadır (Fitzgerald & Rumrill, 2003, 2005). Ayrıca, Hedges ve Olkin (1980) gerçek bir etki varsa, istatistiksel gücü 0,50'den az olan çalışmaların sayısı artıça oy-sayımı yöntemiyle hatalı kararlar verme olasılığının arttığını göstermiştir. Bu nedenle, Hunter ve Schmidt (2004) bu konudaki görüşlerini "geleneksel oy-sayımı yöntemi kaçınılmaz şekilde istatistiksel ve mantıksal olarak hatalıdır" (s. 447) sözleriyle dile getirmiştir. Dahası, geleneksel derleme yöntemleri olarak hem oy-sayımı yöntemi hem de birleştirilmiş anlamlılık testi, araştırmacılara çalışma karakteristiklerinin etkilerini inceleme şansı vermediği için eleştirilmektedir (Fitzgerald & Rumrill, 2003, 2005).

Literatürde açıkça delillendirildiği gibi, sonuçların şans eseri olması beklenenden ne kadar farklı olduğunu gösteren istatistiksel anlamlılık, yanlış kullanıldığında, hatalı ve çelişkili sonuçlara sebep olmaktadır (Ellis, 2010; Fan, 2001; Hunter & Schmidt, 2004; Kirk, 1996, 2001; Olejnik & Algina, 2000; Schmidt, 1992, 1996; Vacha-Haase, 2001). Bu durum temelde, araştırmacıların istatistiksel anlamlılık ile sonuçların gerçek hayatta ne kadar kullanışlı olduğu hakkında fikir sağlayan pratiksel anlamlılığı nadiren ayırt edebilmesinden kaynaklanmaktadır (Ellis, 2010; Kirk, 1996). İstatistiksel olarak anlamlı olduğu gösterilen sonuçlar doğrudan pratik olarak da anlamlıymış gibi yorumlandığında daha da problemleri bir durum ortaya çıkmaktadır çünkü literatürde istatistiksel



olarak anlamlı ancak aynı zamanda önemsiz olan bir sonuca ulaşmak çok seyrek rastlanan bir durum değildir (Ellis, 2010; Olejnik & Algina, 2000). Bu nedenle, bazı araştırmacılar istatistiksel testten vazgeçilmesi gerektiğini savunurken (Hunter & Schmidt, 2004; Schmidt, 1996) diğer bazı araştırmacılar ise bu testlerin de kullanılması ancak etki büyüklüğünün daha fazla vurgulanması gerektiğini öne sürmektedir (Cohen, 1990; Kirk, 1996, 2001; Vacha-Haase, 2001). İstatistiksel testten nasıl yararlanılması gerektiği tartışmalı bir mesele olmasına rağmen istatistiksel anlamlılığın her zaman pratiksel anlamlılığı garanti etmediği hakkında literatürde görüş birliği oluşmuş durumdadır (Borenstein vd., 2009; Cohen, 1990; Ellis, 2010; Gravetter & Walnau, 2007; Hunter & Schmidt, 2004; Kirk, 1996, 2001; Schmidt, 1996; Vacha-Haase, 2001). Bu nedenle, Cohen "Bir araştırmanın birincil ürünün p değerleri değil bir veya daha fazla etki büyüklüğü ölçüsü olduğunu öğrendim ve öğrettim" sözleriyle bu durumun altını çizmiştir (1990, p. 1310). Cohen aynı çalışmasında başka bir noktayı şöyle vurgulamıştır:

Söylemekten mutluluk duyuyorum ki etki büyüklüğüne karşı uzun zamandır süren ilgi eksikliği sona yaklaşmış gibi görünüyor. Literatür derlemenin, p değerlerine dayalı kullanışsız ve temelde geçersiz olan kutu-skoru (box-score) yöntemi, Gene Glass (1977) tarafından formüle edilen etki büyüklüğüne dayalı meta-analiz ile yer değiştiriyor... Meta-analiz beni çok mutlu etti (1990, s.1309-1310).

Cohen tarafından ifade edildiği gibi, meta-analizin, araştırma sentezinin diğer nicel yöntemlerine kıyasla daha güçlü olmasını sağlayan istatistiksel anlamlılığa dayanmaması, bunun yerine sonuçların etki büyüklüğü ölçülerini kullanmasıdır (Borenstein vd., 2009; Shelby & Vaske, 2008). Bu sebeple, Hunter ve Schmidt (2004) istatistiksel anlamlılık testine iki alternatif önermektedir. Bunlar, birincil çalışmalar için güven aralığı, ikincil çalışmalar düzeyinde ise meta-analizdir.

p değerlerinden ziyade pratiksel anlamlılığa dayanmasından kaynaklanan gücüne ek olarak, meta-analizin diğer bir avantajı, araştırmacılara çalışmanın karakteristik özellikleri gibi ara değişkenlerin etkilerini inceleme olanağı sunmasıdır ve bunun diğer nitel ve nicel araştırma sentezi yöntemleriyle yapılması neredeyse olanaksızdır (Borenstein vd., 2009; Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001). Birincil çalışmalardan elde edilen çok miktarda veriyi kullanabilme olanağı, artan güç ve gelişmiş hassasiyet, meta-analizin araştırma sentezi yapmanın en kullanışlı yöntemlerinden bir tanesi olarak nitelendirilmesinin diğer sebeplerinden sadece bazılarıdır (Borenstein vd., 2009; Cohn & Becker, 2003; Gliner vd., 2003; Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001).

#### ***Klasik (Geleneksel) Derlemeye Kıyasla Meta-Analiz***

Araştırma sentezinin geleneksel, sistematik olmayan bir alternatifi olan klasik derleme, çok ciddi dezavantaj ve sınırlılıklara sahiptir (Borenstein vd., 2009; Bushman & Wells, 2001; Carlton & Strawderman, 1996; Cooper & Rosenthal, 1980; Fitzgerald & Rumrill, 2003, 2005; Littell, Corcoran, & Pillai, 2008; Petticrew & Roberts, 2006; Torgerson, 2003). Aynı zamanda geleneksel derleme veya anlatı derlemesi olarak da bilinen klasik derleme, üzerinde çalışılacak spesifik konudaki uzmanlar tarafından yapılır. Ancak, maalesef bu durum birincil çalışmaların yansız ve güvenilir bir özetinin yapılacağını garanti etmemektedir (Petticrew & Roberts, 2006). Düzeyi neredeyse hiç açıklanmayan subjektif yargılama, birincil çalışmaların sistematik olmayan yollarla dâhil edilmesi sonucu oluşan yanlı ve geneli iyi temsil etmeyen örneklem ve ağırlıklandırma prosedürü için açıkça belirtilen bir mantığın olmayışı klasik derlemenin problemlerinden bazıları olarak gösterilmektedir (Bushman & Wells, 2001; Carlton & Strawderman, 1996; Cooper & Rosenthal, 1980; Fitzgerald & Rumrill, 2003, 2005; Littell vd., 2008; Oakley, 2002; Petticrew & Roberts, 2006; Torgerson, 2003). Çalışma karakteristiklerinin etkilerini inceleme ve genel bir etki büyüklüğü oluşturma olanağının olmaması klasik derlemelerin diğer sınırlılıklarını oluşturmaktadır (Bushman & Wells, 2001; Fitzgerald & Rumrill, 2003, 2005). Son olarak, geleneksel anlatı derlemeleri, artan birincil çalışma sayısı sonucu sentezlenecek devasa büyüklükte bilgi oluştuğunda daha az kullanışlı olmaktadır (Borenstein vd., 2009; Glass, 2006; Hunter & Schmidt, 2004). Bu zayıflıklar sebebiyle, Cooper ve Rosenthal (1980),

Oakley (2002) ve Bushman ve Wells (2001)'in delillendirdiği gibi, aynı araştırma sorusu üzerinde klasik derleme çalışması yapan farklı araştırmacıların birbirinden farklı ve yanıltıcı sonuçlara ulaşması istisnai bir durum değildir (Fitzgerald & Rumrill, 2005).

Öncelikle, Cooper ve Rosenthal (1980) istatistiksel birleştirme prosedürleriyle geleneksel anlatı derlemelerini karşılaştırabilmek için deneysel bir çalışma tasarlamıştır. Bu çalışmada 41 araştırmacı, "kararlılık" psikolojik özelliğinde cinsiyet farkını araştıran aynı yedi çalışmanın derlemesini yapmak üzere istatistiksel birleştirme ve anlatı grubuna rastgele atanmıştır. Çalışmanın sonucunda, istatistiksel birleştirme yöntemlerini kullanan araştırmacılar kadınların daha kararlı olduğunu iddia eden hipotez için daha fazla destek tespit etmişlerdir. Ayrıca, bu gruptaki araştırmacılar geleneksel derleme yapanlara kıyasla daha büyük bir etki büyüklüğü rapor etmişlerdir. Bu sonuç, istatistiksel birleştirme prosedürlerinin çalışmanın gücünü artırmasından kaynaklanmış olabilir çünkü gücün artması araştırmacılara küçük etkileri bile saptayabilme olanağı ve daha hassas sonuçlar sağlamaktadır (Petticrew & Roberts, 2006).

Benzer şekilde, Oakley (2002) yaşlı insanlar ve kazaların önlenmesi hakkındaki 137 çalışmayı dâhil eden altı geleneksel derlemeyi incelemiş ve kaç tane birincil çalışmanın her altı derlemede ortak olarak bulunduğunu kontrol etmiştir. Sonuçlar şaşırtıcıdır: en azından iki çalışmada ortak olarak kullanılan yalnızca 33 çalışma vardır. Diğer taraftan, her altı derlemede birden bulunan yalnızca iki çalışma bulunmaktadır ve bunlardan yalnızca bir tanesi tüm derlemelerde tutarlı bir şekilde değerlendirilmiştir. Oakley ayrıca genç insanlar için sigara karşıtı eğitimle ilgili toplamda 27 çalışmayı dâhil eden iki derlemeyi karşılaştırmış ve her iki derlemede ortak yer alan yalnızca üç çalışma belirleyebilmiştir. Dahası, literatürde bu derlemelerin dâhil etme kriterlerine uyan en az 70 birincil çalışma daha olduğunu iddia etmiştir.

Son olarak, Bushman ve Wells (2001) 280 katılımcı ile gerçekleştirdikleri başka bir çalışmada anlatı derlemelerinin yanlı ve subjektif kararlarına karşı meta-analizin düzeltici özelliklerini örneklendirmişlerdir. Öncelikle katılımcılar tarafından derlenmek üzere benzerlik ve çekicilik arasındaki ilişkiyi inceleyen ve ortalamada  $d=0,2$  büyüklüğünde bir ilişkiye sahip olan 20 hayali araştırma sonucu yaratmışlardır. Daha sonra, çalışmaların dikkat çekici olma durumlarını ve sıralarını manipüle etmişlerdir. Sonuç olarak, anlatı derlemesi yapan gruptaki katılımcıların yargıları dikkat çekici başlıklar tarafından istatistiksel olarak anlamlı şekilde etkilenirken ( $p < .007$ ,  $d=0.50$ ) başlıkların dikkat çekici olma durumu meta-analiz grubunun sonuçlarını etkilememiştir ( $p=.71$ ,  $d=-0.07$ ). Altı çizilmesi gereken ilginç bir nokta da, başlıkların dikkat çekiciliğinin hem anlatı derlemesi hem de meta-analiz katılımcılarının hafızalarını önemli derecede etkilemiş olmasıdır. Ancak, meta-analiz katılımcılarının etki büyüklüğü tahmini başlıkların manipülasyonundan etkilenmezken anlatı derlemesi yapanlar için durum böyle olmamıştır. Dahası, Bushman ve Wells meta-analizin çok yakın bir etki büyüklüğü tahmini yaptığını ancak anlatı derlemesi yapanların etki büyüklüğünü olduğundan küçük tahmin ettiklerini belirtmişlerdir.

Özetle, meta-analizin en önemli güçlü yönlerinden bir tanesi, geleneksel anlatı derlemelerinin sorun yaşadığı, yanlı ve subjektif kararlar ve temsil gücü zayıf örneklemeler gibi sınırlılıklara karşı dirençli olmasıdır. Ayrıca, meta-analizlerde sentezlenecek birincil çalışmaların sayısının artması hem istatistiksel gücün ve hassasiyetin artmasına sebep olmakta hem de sonuçlarda tutarsızlıklar varsa sebeplerini araştırma esnekliği kazandırmaktadır. Diğer taraftan, aynı durum, insanın çok büyük miktarda veriyi aynı anda güvenilir ve geçerli şekilde işleme kabiliyetinin olmaması sebebiyle, anlatı derlemelerinde kaotik ve kullanışsız olabilmektedir (Borenstein vd., 2009; Glass, 2006; Glass vd., 1981; Hunter & Schmidt, 2004; Petticrew, 2003; Petticrew & Roberts, 2006; Wolf, 1986).

#### ***Meta-Analizin Geleneksel Araştırma Sentezi Yöntemlerine Kıyasla Avantajlarının Özeti***

Lipsev ve Wilson (2001) bir grup araştırmayı analiz etmek ve özetlemek için neden klasik derleme yöntemlerinden ziyade meta-analiz kullanmamız gerektiğini açıklayan dört sebebe işaret etmektedir. Bu sebepler aynı zamanda meta-analizin birincil avantajlarını oluşturmaktadır. İlk olarak, araştırma bulgularının sentezlenmesi sürecinde meta-analiz prosedürleri araştırmacıyı kullanışlı bir

disiplinliliğe zorlar. Meta-analizin, birincil çalışmalara benzer şekilde, önceden belirlenmiş basamakları vardır ve sürecin inceleme ve tekrara açık olabilmesi için, meta-analizcilerden araştırma sentezi sırasında takip ettikleri her basamağı açıkça raporlamaları beklenir. İkinci sebep ise meta-analizlerin klasik anlatı derlemelerine kıyasla çalışmanın ana bulgularını çok daha etkili ve sofistike şekilde özetleyebilmesidir çünkü klasik derlemeler nitel özetlere veya oy-sayımı yöntemine dayanır ki, bu yöntem temelde örneklem büyüklüğü tarafından şekillenen örneklem hatasına karşı son derece duyarlı olması sebebiyle eleştirilen istatistiksel anlamlılığı temel alır. Diğer derlemelerden ziyade meta-analizin tercih edilmesi için üçüncü önemli sebep ise meta-analizin bize, araştırmaları özetlemekte kullanılan diğer yöntemlerde belirsiz olan etki ve büyüklükleri bulabilme kabiliyeti sağlamasıdır. Son olarak, meta-analiz araştırmacılara, incelenen çok miktarda araştırma bulgusunun son derece organize bir şekilde üstesinden gelme becerisi kazandırır.

Ek olarak, “meta-analizin” isim babası olan Glass (1982), meta-analizi “etki büyüklüklerinin ortalamasının alınması” şeklinde etiketlenen, varyans analizini “toplama ve çarpma” olarak tanımlamaktan daha az hatalı olmayan bir yanlış yorumlama olduğunu ileri sürmektedir. Bununla birlikte, Glass meta-analizin üç çok önemli karakteristik özelliği olduğunu belirtmektedir. İlk olarak, meta-analiz, çok büyük miktarda verinin sentezlenmesi için bir dizi istatistiksel yöntemin kullanıldığı nicel bir süreçtir. Ayrıca, meta-analiz araştırma bulgularını araştırma kalitesi açısından önceden yargılamaz ve bu özelliği meta-analizi araştırma sentezinde kullanılan diğer yaklaşımlardan ayırmaktadır. Son olarak, meta-analiz genel sonuçlara ulaşmaya çalışır; yani, anlamlı genellemeler üretmeyi amaçlar.

Ayrıca, Rosenthal ve DiMatteo (2001) meta-analizin araştırmacılara, her hangi bir birincil çalışmanın veya anlatı derlemesinin sağlayabileceğinden çok daha doğru ve güvenilir sonuçlar sağlayacağını vurgulamaktadır. Ardından, meta-analizin avantajlarını, araştırma girişimin genel görüntüsünü algılayabilme, istatistiksel anlamlılığı bir bütünlük içerisinde değerlendirme, hiçbir veriyi boşa harcamama, veriyle yaklaşma, bir araştırma hipotezine odaklanma ve ara değişkenlerin belirlenmesi olarak özetlemektedirler.

### Meta-Analize Yönelik Eleştiriler

Bir önceki bölümde, meta-analizin, farklı araştırmacılar tarafından ifade edilen güçlü yanları özetlenerek neden diğer nitel ve nicel yöntemlerden ziyade meta-analizin bir araştırma sentezi yöntemi olarak kullanılmasının teşvik edilmesi gerektiğinin sebepleri açıklanmaktadır. Ancak, literatürde meta-analize yönelik eleştiriler de yer almaktadır. Glass (1982) bu eleştirileri dört grupta sınıflandırmıştır. Birinci grup “elma ve armut problemi” olarak isimlendirilen eleştirileri temsil etmektedir. Bu eleştiri, araştırma sentezinde meta-analizin elma ve armutları birbirine karıştırdığı görüşüne dayanmaktadır. Sonuçları farklı ölçüm teknikleri, değişken tanımları ve katılımcılara dayanan çalışmaların karşılaştırılmasıyla anlamlı genellemelerin yapılamayacağı çünkü bu çalışmaların birbirinden çok farklı olduğu iddia edilmektedir. Diğer taraftan, Glass her yönüyle aynı olan çalışmaları karşılaştırmaya ihtiyaç olmadığını çünkü bu çalışmaların açık bir şekilde, istatistiksel hatalar dâhilinde, çok benzer sonuçlar sağlayacaklarını savunmakta ve bu konuyu şöyle vurgulamaktadır: “karşılaştırılmasına veya entegre edilmesine ihtiyaç duyulan çalışmalar yalnızca farklı olanlardır” (s. 102). Ayrıca, Glass bu durumun birincil bir çalışmada farklı kişilerden veri toplamaktan çok farklı olmadığını çünkü bu kişilerin de en az elma ve armut kadar farklı olduğunu ileri sürmektedir.

İkinci eleştiri ise meta-analizin birincil çalışmaların kalitesiyle ilgili “muhakemelerde düşük standartları savunduğu” iddiasıdır. Diğer bir ifadeyle, iyi tasarlanmış çalışmaların sonuçlarıyla birlikte zayıf şekilde tasarlanmış çalışmalardan elde edilen sonuçlar da sentezlenmek üzere meta-analize dâhil edilmektedir. Glass ise subjektif yargılara dayanan kriterleri karşılamayan araştırmaların elenmesinin sağlıklı olmayan sonuçlara sebep olabileceğini savunmaktadır. Bu problemi aşmak için alternatif yollar önermektedir. Örneğin, birincil araştırmaların tasarım ve analiz özelliklerinin tanımlanarak araştırma bulguları ile kovaryanslarının çalışılması bu eleştirinin zayıflatılması için bir



yol olabilir. Aynı zamanda, bu yol bize tasarımla ilgili meselelerde farklı özelliklere sahip olan çalışmaların deneysel etkileri arasında fark olup olmadığını inceleme olanağı tanımaktadır. Ayrıca, Glass, tasarım kalitesiyle araştırmaların bulguları arasında bir ilişki olup olmadığını kontrol edebilmek için 12 meta-analiz çalışmasının bulgularını incelemiş ve sonuçları şu şekilde dile getirmiştir: “Geçerliği yüksek olan deneylerle düşük olanlar arasındaki farkın standart sapmanın onda birinden çok daha fazla olduğu durumlar oldukça nadirdir” (s. 104). Diğer taraftan, literatürde de belirtildiği gibi, ara değişken analizi meta-analizcilere iyi ve zayıf tasarlanmış çalışmaların etki büyüklüğü ölçüleri bakımından birbirlerinden ne derece farklı olduklarını inceleme şansı verir (Borenstein vd., 2009; Card, 2012; Wolf, 1986).

Üçüncü eleştiri ise “basılı literatürde yer alan araştırmalar, tamamlanmış olan çalışmalar popülasyonunu sistematik olarak yanlış temsil ettiğinde ortaya çıkan durum için kullanılan terim” (Rothstein, Sutton, & Borenstein, 2005, s. 1) olan yayın yanlılığıdır. İstatistiksel olarak anlamlı olmayan sonuçlar yayınlanmaya nadiren kabul edildiği için, basılı araştırmaların istatistiksel olarak anlamlı sonuçlar lehine yanlı olduğu öne sürülmektedir. Sonuçta bu durum meta-analiz sonuçlarının da yanlı olmasına sebep olmaktadır. Rosenthal (1979) bu olguyu, “dosya çekmecesi problemi (file drawer problem)” olarak adlandırmaktadır çünkü bu sorun anlamlı sonuçlar yayınlanmak üzere gönderilirken anlamlı olmayan sonuçların dosya çekmecelerine sürgün edilmesinden kaynaklanmaktadır (Rosenthal & DiMatteo, 2001). Glass, bir önceki eleştiride olduğu gibi, birçok meta-analiz incelemiş ve “...dergilerde raporlanan bulguların doktora ve yüksek lisans tezlerinde raporlananlara kıyasla, ortalama olarak, standart sapmanın üçte biri oranında araştırmacıların hipotezlerini desteklemeye daha eğilimli” (s. 106) olduğu sonucuna ulaşmıştır. Ayrıca, Rothstein ve diğerleri (2005) yayın yanlılığının, araştırma sentezlerinin geçerliğini etkileyen, muhtemelen en dikkat edilmesi gereken tehdidi oluşturduğunu öne sürmektedir. Diğer taraftan, bu olguyla ilgili iki önemli noktaya dikkat çekmişlerdir: öncelikle, diğer birçok araştırmacının birçok kez belirttiği gibi (Borenstein vd., 2009; Card, 2012; Rosenthal & DiMatteo, 2001; Sutton, 2009), bu problem meta-analize özgü değildir aksine sentez veya derlemelerin tüm çeşitleri için ortak bir problemdir. Ayrıca, yayın yanlılığı meta-analiz veya diğer herhangi bir araştırma sentezi yönteminin sebep olduğu bir problem değildir. Tam tersine, sonuçları özetlemek için araştırma sentezlerinin yapılıp yapılmamasından bağımsız olarak literatürde var olan bir olgudur. Bu nedenle, literatürdeki yayın yanlılığının varlığı araştırma sentezlerine karşı bir argüman olmamalı ve yayın yanlılığının literatürden sonuç çıkaran birincil çalışmaları da etkilediği unutulmamalıdır (Rothstein vd., 2005; Sutton, 2009).

Aslında, meta-analiz bu konuda problemin bir kaynağı değil çözümün bir parçasıdır çünkü meta-analiz yayın yanlılığının tespit edilebilmesi ve sonuçları ne derece etkilediğinin tahmin edilebilmesi için çeşitli yaklaşımlar önermektedir (Glass, 1982; Sutton, 2009). Sonuçların yayın türüne göre ayrı ayrı incelenmesi, ara değişken analizinin yapılması veya tanılama amacıyla huni grafiğinin kullanılması, bir meta-analiz çalışmasında yayın yanlılığının incelenmesi için kullanılacak yollardan sadece bazılarıdır. Yayın yanlılığının sadece tanınması için değil aynı zamanda düzeltilmesi amacıyla kullanılacak birçok yöntem bu makale kapsamında “yayın yanlılığı” bölümünde detaylı şekilde açıklanmaktadır.

Son olarak, dördüncü eleştiri ise “yumrulanma (bağımsız olmayan veri)” olarak adlandırılmaktadır. Sıklıkla aynı çalışmadan gelen birden fazla sonuç kullanılmaktadır ve bu durum yanlılığa veya meta-analizin geçerliğinin olumsuz etkilenmesine sebep olabilir. Ayrıca, sonuçlar birbirinden bağımsız olmadığı için gerçekte olduğundan daha güvenilir görünür. Örneğin, etki büyüklükleri 0,3, 0,3 ve 0,3 olan bir çalışma ile 0,5, 0,5 ve 0,5 olan ikinci bir çalışma aynı meta-analiz kapsamında yer alıyorsa, bu meta-analizde dikkate alınması gereken gerçek bağımsızlık derecesi etki büyüklüklerinin sayısı olan altı değil birincil çalışma sayısı, yani iki olmalıdır. Glass (1982) bu probleme basit bir çözüm olarak bir çalışma dâhilindeki tüm etki büyüklüklerinin ortalamasının alınmasını önermektedir. Ayrıca, yüksek lisans ve doktora tezlerine dayanan makaleler konusunda dikkatli olmalıyız çünkü hiçbir çalışma meta-analize birden fazla kez dâhil edilmemelidir. Bununla birlikte, bağımlı etki büyüklüklerinin ortalamasını almanın daha sofistike yollarını kullanmak da mümkündür (Gleser & Olkin, 2009; Hedges & Olkin, 1985; Marin-Martinez & Sanchez-Meca, 1999; Rosenthal & Rubin, 1986).

Benzer şekilde, Rosenthal ve DiMatteo (2001) meta-analizle ilgili eleştirileri beş grupta sınıflandırarak açıklamıştır. Bu gruplar, bulguların örneklemindeki yanlılık, çöp giren çöp çıkar (garbage in garbage out), etkilerin bağımlılığı ve tekliği, her bir etkinin aşırı vurgulanması ve elmayla armudun birleştirilmesinden oluşmaktadır. Meta-analizin eleştirilerinin özetlendiği bu gruplar, Glass (1982) tarafından oluşturulan gruplara oldukça benzerdir. Rosenthal ve DiMatteo, ek olarak, meta-analizin, sadece bağımlı ve bağımsız değişkenler arasındaki bireysel etkileri incelediği için eleştirildiğinden bahsetmiştir. Bu eleştiriyle ilgili olarak, farklı değişkenler arasındaki etkileşimler incelenmeden önce, meta-analizin bize her bir bireysel bileşenin basit etkilerinin anlaşılır bir resmini sağladığını savunmuşlardır. Son olarak, meta-analizle ilgili yapılan eleştirilerin çoğunluğunun meta-analizin nasıl yapıldığı ile ilgili yanlış anlamalara dayandığına dikkat çekmişlerdir.

### Meta-Analiz Nasıl Yapılır?

Rosenthal ve DiMatteo (2001) meta-analizin sadece bir istatistiksel teknik olmaktan ziyade birincil araştırmalardakilere benzer basamakları dikkatlice takip ederek araştırma sentezi yapmanın bir yolu olduğunu altını çizmişlerdir. Ardından, meta-analiz yapmanın temel basamaklarını şöyle açıklamışlardır:

- İlgili bağımlı ve bağımsız değişkenleri tanımlayın, örneğin; probleme dayalı öğrenmenin öğrencilerin başarısı, fenedeki motivasyonları ve fene karşı tutumları üzerindeki etkisi.
- Birincil çalışmaları sistematik bir yolla toplayıp seçtikten sonra her bir makaleyi dikkatlice okuyun.
- Elde edilen etki büyüklükleri arasındaki heterojenliği, grafikler ve tablolar veya ki-kare anlamlılık testi ile inceleyin. Ancak, ki-kare testi ihtiyatlı bir şekilde yorumlanmalıdır çünkü diğer anlamlılık testleri gibi bu test de örneklem büyüklüğüne, yani meta-analize dâhil edilen birincil çalışma sayısına bağımlıdır. Ek olarak, ilgili ara değişkenlerin etki büyüklükleri arasındaki değişkenlik üzerindeki etkisi araştırılmalıdır.
- Ağırlıklandırılmış ortalama gibi merkezi eğilim ölçülerini kullanarak birincil çalışmalardan elde edilen etki büyüklüklerini birleştirin.
- Merkezi eğilim indekslerinin anlamlılık seviyesini inceleyin.
- Elde edilen ortalama etki büyüklüğünün önemini değerlendirin.

Benzer şekilde, Glass (2006) meta-analizin temel basamaklarını, problemin tanımlanması, literatürün elde edilmesi, çalışmaların kodlanması, bulguların ortak bir ölçeğe dönüştürülmesi ve istatistiksel olarak analiz edilmesi olarak özetlemiştir.

İstatistiksel modeller bakımından, meta-analiz sürecinde kullanılacak farklı sayıtlara sahip iki ana yaklaşım bulunmaktadır. Bunlar, bir grup çalışmadan elde edilen ortalama etki büyüklüğü hakkında çıkarımda bulunmak için geliştirilen sabit-etki ve rastgele-etkiler modelleridir (Borenstein vd., 2009; Hedges & Vevea, 1998; Hunter & Schmidt, 2000, 2004; Tweedie, Smelser, & Baltés, 2004). Sonraki bölümde, bu modeller farklı açılardan detaylı şekilde karşılaştırılmaktadır.

#### *Sabit-Etki ve Rastgele-Etkiler Modellerinin Karşılaştırması*

Sabit-etki modelinin en önemli sayıltısı meta-analizdeki tüm çalışmalar için sadece bir tane gerçek etki büyüklüğünün olduğudur. Bu sayıltı aynı zamanda gözlenen etkilerdeki tüm farklılıkların sadece örnekleme hatasından kaynaklandığı anlamına gelmektedir. Diğer taraftan, rastgele-etkiler modeli katılımcıların yaşları, eğitim seviyesi veya sınıf büyüklüğü gibi bazı ara değişkenler nedeniyle gerçek etki büyüklüğünün çalışmadan çalışmaya değişebileceği düşüncesine dayanmaktadır. Bu nedenle, gerçek etki büyüklüğü bazı ortalamalar etrafında dağılmaktadır. Meta-analizlerdeki çalışmalardan elde edilen etki büyüklüklerinin bu dağılımın rastgele bir örnekleme olduğu varsayılmaktadır.

Sabit-etki modelinde, etki büyüklüğünü etkileyebilecek tüm faktörlerin sabit olduğu varsayıldığı için her bir çalışmada gözlenen etki büyüklüğü popülasyon ortalaması ve örnekleme

hatası ile hesaplanabilmektedir. Diğer taraftan, rastgele-etkiler modelinde gerçek etki büyüklüğünün dağılım gösterdiği varsayıldığı için çalışmalar arası varyans ortaya çıkmaktadır. Bundan dolayı, rastgele-etkiler modelinde, her bir çalışmada gözlenen etki büyüklüğü, çalışmalar arası varyansın kaynaklanan başka bir hata daha eklenerek hesaplanabilmektedir (Borenstein vd., 2009).

Her iki modelde de, özet etkinin (sabit-etki modeli için popülasyon ortalaması ve rastgele-etkiler modeli için genel ortalama) daha hassas tahmininin elde edilmesi için, yani varyansın en aza indirilmesi için, kesinliği daha yüksek olan çalışmalara daha fazla ağırlık verilerek ağırlıklandırılmış bir ortalama hesaplanmaktadır. Hangi çalışmanın kesinliğinin daha yüksek olduğuna karar verebilmek için, çalışmanın varyansı dikkate alınır. Bir başka ifadeyle, her iki modelde de daha az varyansa sahip olan çalışmaya daha fazla ağırlık atanır.

Ayrıca, sabit-etki ve rastgele-etkiler modelleri arasında özet etkinin tahmini bakımından önemli bir fark bulunmaktadır. Sabit-etki modelindeki temel amaç tek bir gerçek etki büyüklüğünü tahmin etmek olduğu için küçük örnekleme sahip olan çalışmalardan elde edilen bilgilerin önemi azımsanırken, daha büyük örnekleme sahip olan çalışmalara çok daha fazla ağırlık verilir. Diğer taraftan, rastgele-etkiler modelinde temel amaç etki büyüklüklerinden oluşan dağılımının ortalamasının tahmin edilmesidir. Bu nedenle, ister küçük ister büyük örnekleme sahip olsun, her bir çalışma özet etki içerisinde temsil edilmelidir. Sonuç olarak, rastgele-etkiler modelinde atanan göreceli ağırlıklar daha dengeli dağılmaktadır (Borenstein vd., 2009).

Standart hata ve güven aralığının büyüklüğü iki model arasındaki diğer bir farkı oluşturmaktadır. Rastgele-etkiler modeli, çalışma-içi varyansa ek olarak çalışmalar-arası varyansın olduğunu varsaydığı için aynı birincil çalışmaların meta-analizi için özet etkiye ait standart hata ve güven aralığının rastgele-etkiler modelinde sabit-etki modeline kıyasla her zaman daha büyük olması beklenir.

Sabit-etki modeli, örnekleme hatasını düşük tahmin ettiği için ciddi şekilde eleştirilmektedir. Bu durum aynı zamanda ortalama etki büyüklüğü için güven aralığının gerçekte olduğundan daha dar olmasına ve modelin birçok durum için gerçekçi görünmeyen sayıltıları sağlanmadığında, kesinliğin abartılmasına sebep olmaktadır (Borenstein vd., 2009; Erez, Bloom, & Wells, 1996; Hunter & Schmidt, 2000; Overton, 1998; Schmidt, Oh, & Hayes, 2009). Buna rağmen, yakın zamana kadar yapılan meta-analizlerde sabit-etki modelinin çok daha yaygın şekilde kullanılmış olduğu literatürde açıkça görülmektedir (Cooper, 1997; Hunter & Schmidt, 2000; National Research Council, 1992; Overton, 1998; Schmidt vd., 2009). Meta-analiz yapan araştırmacıların rastgele-etkiler modelinin yerine sabit-etki modelini tercih etmesinin sebebi, sabit-etki modelinin daha kolay yürütülmesi (Cooper, 1997) ve kavramsal altyapı ve sayısal analizleri açısından çok daha kolay olmasıdır (National Research Council, 1992).

Diğer taraftan, yürütülmesi kolay olmasına rağmen, birçok araştırmacı (Field, 2003; Hedges & Vevea, 1998; Hunter & Schmidt, 2000; Overton, 1998; Schmidt vd., 2009) sabit-etki modelinin, homojenlik sayıltısı karşılanmadığında, istatistiksel testler için Tip 1 hata oranlarının artmasına sebep olduğuna dikkat çekmektedir. Hunter ve Schmidt (2000), Tip 1 hata oranının heterojenlik ve meta-analize dâhil edilen çalışmaların örneklem büyüklüğünden nasıl etkilendiğini açıklamış ve homojenlik sayıltısı daha ciddi şekilde ihlal edildikçe Tip 1 hata oranının arttığını vurgulamıştır. Sezgilerimize aykırı şekilde, meta-analize dâhil edilen çalışmaların ortalama örneklem büyüklüğü arttıkça Tip 1 hata yapma olasılığı da dramatik şekilde artmaktadır. Sonuç olarak, ortalama örneklem büyüklüğü 100 ve standart hata 0,25 olduğunda, alfa değeri 0,46'ya yükselmektedir. Bu değer, neredeyse her iki meta-analizden bir tanesinin hatalı şekilde anlamlı sonuçlara ulaştığı anlamına gelmektedir. Dahası, meta-analize dâhil edilen birincil çalışma sayısının artması bu yükselmiş oranı düşürmemektedir. Bu nedenle, Hunter ve Schmidt "Meta-analizlerdeki gerçek veriler için, SE (sabit-etki) model ve prosedürleri, eğer hiç değilse, nadiren uygundur ve bu nedenle RE (rastgele-etkiler) model ve prosedürleri, SE model ve prosedürlerine tercih edilmelidir" (s. 284) sonucuna ulaşmışlardır ve bu sonuç National Research Council (1992)'in tavsiyeleriyle benzerdir.

Diğer taraftan, Hedges ve Vevea (1998) modeller arasındaki kavramsal farkı açıklamayı amaçlamakta ve uygun modele karar verilirken en önemli meselenin yapılmak istenen çıkarımın

doğası olması gerektiğini savunmaktadırlar. Sabit-etki modelinin sadece meta-analize dahil edilen çalışmalardaki parametrelerle ilgili çıkarımlar yapmak için kullanılmasını tavsiye ederken, meta-analize dahil edilen çalışmaların örneklem olarak seçildiği popülasyon ile ilgili çıkarımlar, yani koşulsuz çıkarımlar için sabit-etki modelinin uygun olmadığını dile getirmişlerdir. Bu tür çıkarımlar için ise rastgele-etkiler modelinin kullanılmasını önermişlerdir. Ancak, Borenstein ve diğerleri (2009) ve Erez, Bloom ve Wells (1996) meta-analizdeki tüm çalışmalar için sadece bir tane gerçek etki büyüklüğü öngören sabit-etki modelinin temel sayılısının birçok durum için gerçekçi olmadığını savunmuşlardır. Benzer şekilde, Schmidt ve diğerleri (2009) sabit-etki modelinin uygun olacağı koşulların çok sınırlı olduğunu belirtmişlerdir. Bu nedenle, birçok araştırmacı meta-analiz çalışmalarını için sabit-etki modelinden ziyade rastgele-etkiler modelinin kullanılmasını tavsiye etmektedir (Borenstein vd., 2009; Field, 2003; Hunter & Schmidt, 2000; National Research Council, 1992).

### Etki Büyüklüğü İndeksleri

“Örneklem büyüklüğü ve istatistiksel testlerin sonuçlarından bağımsız olarak, araştırılan olgunun çalışma sonuçlarında ne ölçüde var olduğu” (Sánchez-Meca & Marín-Martínez, 2010b, s. 274) şeklinde tanımlanabilecek olan etki büyüklüğünün birçok indeksi bulunmaktadır. Tablo 1, detayları literatürde birçok kaynakta sunulan (Borenstein, 2009; Borenstein vd., 2009; Ellis, 2010; Fleiss & Berlin, 2009; Olejnik & Algina, 2000) yaygın etki büyüklüğü indekslerinden bazılarını göstermektedir. Ayrıca, Huberty (2002) etki büyüklüğü indekslerinin tarihi hakkında detaylı bilgi sağlamaktadır.

Bütün etki büyüklüğü indeksleri hakkında kapsamlı açıklamalar sunulması bu makalenin kapsamı dışındadır. Ancak, kavramın özü bazı örnekler yardımıyla açıklanabilir. Eğitim araştırmalarında, sürekli bir bağımlı değişken üzerinden grupların karşılaştırılması oldukça yaygındır. Bu nedenle, sürekli çıktılarda karşılaştırılan grupları temsil eden etki büyüklüklerinden en tanıdık olanı, Cohen d, etki büyüklüğü kavramının örneklendirilmesi için iyi bir başlangıç noktası olabilir.

Cohen d, iki grup arasındaki düzeltilmemiş ve standardize edilmiş ortalama farkıdır. Birleştirilmiş standart sapmaya dayanan Cohen d şöyle gösterilebilir:

$$\text{Cohen } d = \frac{X_e - X_c}{S_p}$$

burada  $X_e$  deneysel grubun ortalamasını,  $X_c$  kontrol grubunun ortalamasını ve  $S_p$  iki grubun birleştirilmiş standart sapmasını temsil etmektedir. Birleştirilmiş standart sapma ise şöyle hesaplanabilir:

$$S_p^2 = \frac{(N_e - 1)S_e^2 + (N_c - 1)S_c^2}{(N_e + N_c - 2)}$$

burada  $N_e$  deneysel gruptaki katılımcı sayısını,  $N_c$  kontrol grubundaki katılımcı sayısını,  $S_e^2$  deneysel grubun varyansını ve  $S_c^2$  kontrol grubunun varyansını temsil etmektedir (Borenstein, 2009).

Son olarak Cohen d'nin varyansı ( $v_d$ ) ise şöyle hesaplanabilir:

$$v_d = \frac{N_e + N_c}{N_e N_c} + \frac{d^2}{2(N_e + N_c)}$$

Glass  $\Delta$ , başka bir iki grup arasındaki düzeltilmemiş ve standardize edilmiş ortalama farkıdır, ancak kontrol grubunun standart sapmasına dayanmaktadır ve şöyle hesaplanabilir:

$$\text{Glass } \Delta = \frac{X_e - X_c}{S_c}$$

burada  $S_c$  kontrol grubunun standart sapmasıdır.

**Tablo 1.** Yaygın Etki Büyüklüğü İndekslerinden Bazıları

		Etki büyüklüklerinin yorumlanması *			
		Küçük	Orta	Büyük	
d ailesi	İkili çıktılarda karşılaştırılan gruplar	Risk farkları (RD)			
		Görel risk (RR)			
	Sürekli çıktılarda karşılaştırılan gruplar	Olasılık oranları (OR)			
		Cohen d	0.20	0.50	0.80
r ailesi	Korelasyon indeksleri	Glass delta ( $\Delta$ )	0.20	0.50	0.80
		Hedge g	0.20	0.50	0.80
	Varyans oranları indeksleri	Response ratios I			
		Pearson korelasyon katsayısı r	0.10	0.30	0.50
r ailesi	Korelasyon indeksleri	Kendall tau ( $\tau$ )			
		Phi coefficient ( $\phi$ )	0.10	0.30	0.50
	Varyans oranları indeksleri	Kruskal lambda ( $\lambda$ )			
		Belirleme katsayısı ( $r^2$ )	0.01	0.09	0.25
	Varyans oranları indeksleri	R kare ( $R^2$ )	0.02	0.13	0.26
		Cohen's f	0.10	0.25	0.40
Varyans oranları indeksleri	Eta kare ( $\eta^2$ )	0.01	0.06	0.14	
	Epsilon kare ( $\epsilon^2$ )				
Varyans oranları indeksleri	Omega kare ( $\omega^2$ )				

\*Etki büyüklüklerinin yorumlanması Cohen (1988)'e dayanmaktadır.

Başka bir deyişle, hem Cohen d hem de Glass  $\Delta$  popülasyon etki büyüklüğünün düzeltilmemiş, yani yanlış, bir kestirimidir. Aralarındaki tek fark ise ortalamalar arasındaki farkın standardize edilmesi için hangi standart sapmanın kullanılacağına yatmaktadır. Glass  $\Delta$ 'da, birleştirilmiş standart sapma değil kontrol grubun standart sapması kullanılmaktadır. Bunun sebebi, kontrol grubu uygulamanın etkileriyle bozulmamış olduğu için bu grubunun standart sapmasının popülasyonun standart sapmasını daha iyi temsil edeceği düşüncesidir (Ellis, 2010).

Hem Cohen d hem de Glass  $\Delta$  popülasyonun etki büyüklüğünün kestirilmesinde, özellikle küçük örneklerde, hafif bir yanlılığa sahiptir. Parametreyi olduğundan biraz daha büyük tahmin ederler. Bu durum, Hedge g'de J ile gösterilen bir düzeltme faktörü kullanılarak düzeltilmektedir. Bu faktör şöyle hesaplanabilir:

$$J = 1 - \frac{3}{4d_f - 1}$$

burada  $d_f$  çalışma-içi standart sapmanın ( $S_{within}$ ) kestirilmesi için kullanılan bağımsızlık derecesidir.



Ardından,  $g$ , karşılık gelen varyans ( $v_g$ ) ve standart hata ( $SE_g$ ) şöyle hesaplanabilir:

$$g=J.d$$

$$v_g=J^2.v_d$$

$$SE_g=\sqrt{v_g}$$

$J$  her zaman birden küçüktür, bu nedenle, Hedge  $g$  her zaman Cohen  $d$ 'den biraz daha küçüktür, aynı durum Cohen  $d$ 'ye kıyasla Hedge  $g$ 'nin varyansı için de doğrudur. Aradaki fark, azalan örneklem büyüklüğü ile artmaktadır (Borenstein vd., 2009).

Bir araştırma çalışmasında elde edilen etki büyüklüğünün yorumlanması kolay bir iş değildir ve aslında uygulama etkisinin değerlendirileceği bağlama bağlıdır (Ellis, 2010). Diğer taraftan, etki büyüklüğü değerlerini daha kolay yorumlayabilmek için Cohen (1988) tarafından bazı sınır değerleri önerilmiştir. Cohen, küçük, orta ve büyük etki büyüklükleri için sırasıyla 0,20, 0,50 ve 0,80 sınır değerlerini belirlemiştir ve bu değerler Hedge  $g$  de dahil olmak üzere yukarıda bahsedilen her üç etki büyüklüğü indeksi için de geçerlidir. Sınır değerlerini kullanmanın kolay olmasına ve Cohen bu değerlerin yeterince mantığa dayandırıldığını iddia etmesine rağmen, etki büyüklüğünün yorumlanmasında Cohen'in kriterlerinin kullanılması halen tartışmalı bir konudur. Glass ve diğerleri (1981) bu konudaki görüşlerini "Hangi faydaların neye mal olarak kazanılacağına bağlı olarak 2,0 etki büyüklüğü 'zayıf' ve 0,1 etki büyüklüğü 'iyi' olabilir" sözleriyle açıklamıştır. Buna rağmen, Cohen tarafından belirtilen sınır değerleri, literatürde etki büyüklüklerinin yorumlanması için en yaygın kullanılan kriterleri oluşturmaktadır ve sonuçların yorumlanmasında başvurulması tavsiye edilmektedir. Ancak, bağlamın önemi dikkate alınmalı ve etki büyüklüğü, bilgiye katkısı açısından da değerlendirilmelidir (Ellis, 2010).

### Analiz Birimi

Bir meta-analiz çalışmasında, meta-analize dâhil edilen her bir birincil çalışma veya bu çalışmalardan elde edilen her bir etki büyüklüğü analiz birimi olarak kabul edilebilir. Her iki durum için de, bağımlı veriler sonucunda oluşan yumrulanmayı önlemek için bazı tedbirlerin alınması gerekmektedir. İster birincil çalışmalar ister etki büyüklüğü analiz birimi olarak kabul edilsin, her bir birincil çalışmanın birbirinden bağımsız olup olmadığı kontrol edilmelidir. Başka bir deyişle, aynı meta-analiz kapsamında, dergilerde basılan bazı makalelerin, birincil çalışma örneğine doktora veya yüksek lisans tezi olarak da eklenmiş olmasından dolayı aynı örnekleme paylaşan hiçbir çalışma olmamalıdır. Bir meta-analizde etki büyüklüğü analiz birimi olarak kullanıldığında dikkatli olmamız gereken diğer bir nokta ise, bazı çalışmaların aynı yapıyı ölçmek için farklı ölçüm araçları kullanması sonucunda aynı çıktıyla ilgili birden fazla etki büyüklüğü sağlamasıdır.

### Meta-Analizdeki Geçerlik Sorunları

Yayın yanlılığı ve birincil çalışmaların kalitesi bir meta-analiz çalışmasının geçerliği ile ilgili temel sorunları oluşturmaktadır (Borenstein vd., 2009; Lipsey & Wilson, 2001; Rendina-Gobioff, 2006). Sonraki bölümlerde "yayın yanlılığı" ve "çalışmaların kalitesi" terimlerinin ne anlama geldiği ve neden meta-analizin geçerliği için potansiyel bir tehdit olduklarıyla ilgili detaylı açıklamalar bulunmaktadır.

#### Yayın Yanlılığı

Literatürde açıkça belirtildiği gibi, yayın yanlılığı, veya "dosya çekmecesini problemi" ilgili çalışmaların elde edilmesindeki en ciddi problemlerden bir tanesidir (D. A. Bennett, Latham, Stretton, & Anderson, 2004; Borenstein vd., 2009; Rendina-Gobioff, 2006; Rothstein vd., 2005; Song, Khan, Dinnes, & Sutton, 2002; Thornton & Lee, 2000; Tweedie vd., 2004). Rothstein ve diğerleri (2005) diğer metodolojik konularda ne kadar sorunsuz olursa olsun meta-analize dâhil edilen çalışmalar yanlı ise, meta-analizden elde edilen sonuçların geçerliğinin tehdit altında olduğunu altını çizmektedir. Bu konudaki spesifik kaygı, dergilerin negatif (istatistiksel olarak anlamlı olmayan) sonuçlara sahip olan

çalışmaları reddetme eğiliminde olmalarıdır. Başka bir ifadeyle, istatistiksel olarak anlamlı sonuçlara sahip olan çalışmaların yayınlanma olasılığının daha yüksek olması basılı literatürde bir yanlılığa sebep olmaktadır ve ardından bu yanlılık literatüre dayalı olan meta-analizlere taşınmaktadır (Borenstein vd., 2009). Tablo 2, Rendina-Gobioff (2006)'un çalışmada gözlenen etki büyüklüğü ve varyansın yayınlanma olasılığı üzerindeki etkisini nasıl açıkladığını göstermektedir. Tabloda açık şekilde görüldüğü gibi, istatistiksel anlamlılık sadece uygulamanın etki büyüklüğüne değil aynı zamanda çalışmanın örneklem büyüklüğü ile ters orantılı olan varyansına da bağlıdır. Birçok araştırmacı örneklem büyüklüğüne bağlılığının istatistiksel testlerin zayıflıklarından bir tanesi olduğunu kabul etmektedir ve bu durum pratikte hiç anlamlı olmamasına rağmen istatistiksel olarak anlamlı olan sonuçların oluşmasına sebep olabilmektedir (Borenstein vd., 2009; Cohen, 1990; Ellis, 2010; Gravetter & Walnau, 2007; Hunter & Schmidt, 2004; Kirk, 1996, 2001; Schmidt, 1996; Vacha-Haase, 2001). Ancak, yayın yanlılığına sebep olan durum bu değildir. Yanlı sonuçlara sebep olan şey, istatistiksel testlerin etki büyüklüğüne de bağlı olması sebebiyle, istatistiksel olarak anlamlı olmayan sonuçların küçük etki büyüklüğüne sahip olma eğiliminde olmalarıdır. Yani, küçük etki büyüklüğüne sahip olmaları daha muhtemel olan istatistiksel olarak anlamlı olmayan sonuçlara sahip olan çalışmaların yayınlanma olasılığı daha düşük olduğu için, sadece yayınlanmış çalışmaları içeren her hangi bir meta-analiz, muhtemelen abartılmış bir ortalama etki büyüklüğü değerine işaret edecektir.

**Tablo 2.** Bir Çalışmada Gözlenen Varyans ve Etki Büyüklüğünün Yayınlanma Olasılığına Etkisi (Rendina-Gobioff, 2006)

		Etki Büyüklüğü	
		Küçük	Büyük
Varyans	Küçük (N=büyük)	Yayınlanmış (İstatistiksel olarak anlamlı)	Yayınlanmış (İstatistiksel olarak anlamlı)
	Büyük (N= küçük)	Yayınlanmamış (İstatistiksel olarak anlamlı değil)	Yayınlanmış (İstatistiksel olarak anlamlı)

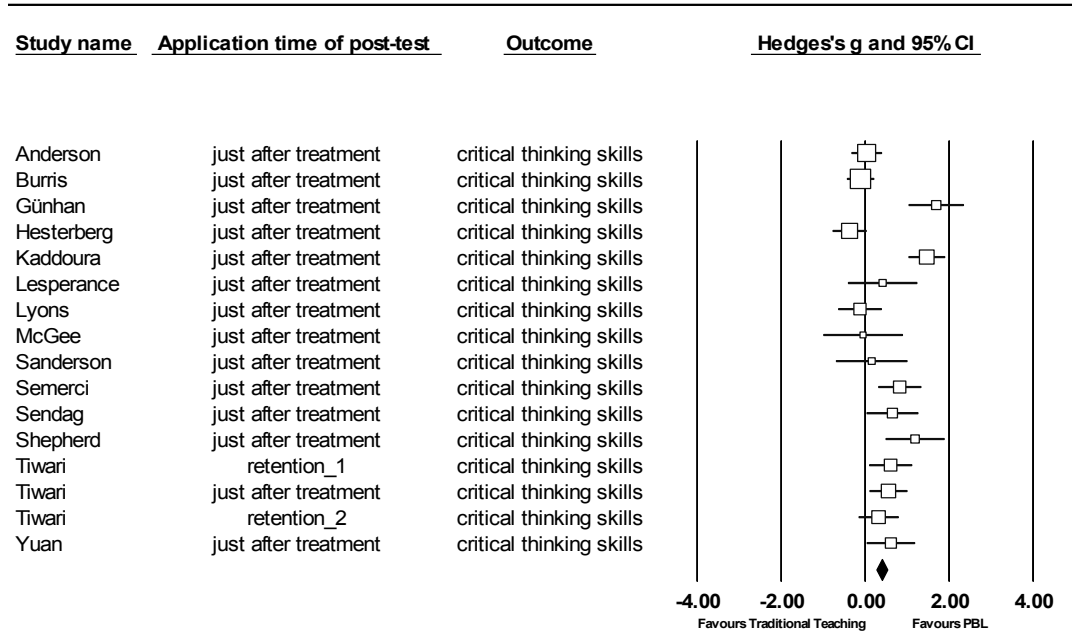
\* N çalışmadaki katılımcı sayısını temsil etmektedir.

Yayın yanlılığı meta-analiz yöntemine özgü bir tehdit değildir, aynı zamanda anlatı derlemeleri ve her hangi bir çeşit literatür derleme yöntemi için de bir problem oluşturmaktadır (Borenstein vd., 2009; Rosenthal & DiMatteo, 2001). Aslında, daha önce de bahsedildiği gibi, meta-analiz problemin bir kaynağı değil çözümün bir parçasıdır çünkü meta-analizcilere yayın yanlılığını tespit etmek ve olası etkilerini kontrol edebilmek için çeşitli yöntemleri kullanma olanağı sunmaktadır. Orman grafiği, huni grafiği, Rosenthal'ın güvenli N (fail-safe N-FSN) ve Duval ve Tweedie'nin Kes ve Ekle (Trim and Fill) yöntemleri literatürde sıkça atıf alan yöntemlerden bazılarıdır (Duval & Tweedie, 2000a, 2000b; Egger, Smith, Schneider, & Minder, 1997; Lewis & Clarke, 2001; J. A. C. Sterne & Egger, 2001; J. A. C. Sterne & Harbord, 2004; Thornton & Lee, 2000; Tweedie vd., 2004; Yeh & D'Amico, 2004). Ancak, yayın yanlılığının zararlı etkilerinden korunmanın en etkili yolunun, hem yayınlanmış hem de yayınlanmamış çalışmaları meta-analize dâhil ederek, yayın yanlılığının önlenmesi olduğunun vurgulanması son derece önemlidir. Yine de, yayınlanmamış çalışmaların da dâhil edilmesi yayın yanlılığının olmayacağını garantilemez. Bu nedenle, meta-analizin sonuçlarının negatif sonuçlara sahip yeni çalışmaların eklenmesine karşı yeterli derecede dayanıklı olduğuyla ilgili delil sağlanabilmesi için, yanlılığın tanınması ve giderilmesine yönelik geliştirilen yöntemler kullanılmalıdır.

Her yöntem kendine has zayıf ve güçlü yönleri sahip olduğu için, meta-analiz çalışmaları kapsamında yayın yanlılığının tanınması ve sonuçları ne derece etkilediğinin kestirilmesi için çeşitli yöntemlerden yararlanılmalıdır. Bu bağlamda, yayın yanlılığının tanınması ve düzeltilmesi için geliştirilen orman grafiği, huni grafiği, Egger'ın doğrusal regresyon yöntemi, Rosenthal'ın FSN, Orwin'in FSN ve Duval ve Tweedie'nin Kes ve Ekle yöntemleri sonraki bölümlerde açıklanmaktadır.

### Orman grafiği

Borenstein (2005) verilerin görsel temsili olan orman grafiğinin her hangi bir meta-analizdeki anahtar unsurlardan bir tanesi olduğunu öne sürmektedir. Şekil 3, orman grafiğinin bir örneğini göstermektedir. Bu örnekte orman grafiği, probleme dayalı öğrenmenin kritik düşünme becerileri üzerindeki etkisini inceleyen 16 çalışmadan elde edilen Hedge g, yani düzeltilmiş ve standardize edilmiş ortalama farkı tahmini, verilerini içermektedir (Üstün, 2012). Bu grafikte, bireysel kareler her bir çalışmanın etki büyüklüğü tahminini ve karelerin iki yanından uzanan çizgiler de tahmin için %95 güven aralığını göstermektedir. Her bir karenin alanı bireysel çalışmaların meta-analizdeki ağırlığına karşılık gelmektedir. Örneklem büyüklüğü ve kesinlik arttıkça, meta-analizde çalışmaya atanan ağırlık da arttığı için büyük kareler aynı zamanda büyük örnekleme sahip çalışmaları işaret etmektedir. Son olarak, meta-analizden elde edilen genel etki büyüklüğü tahmini ve ona ait güven aralığı en aşağıda bulunan bir elmasla temsil edilmektedir.

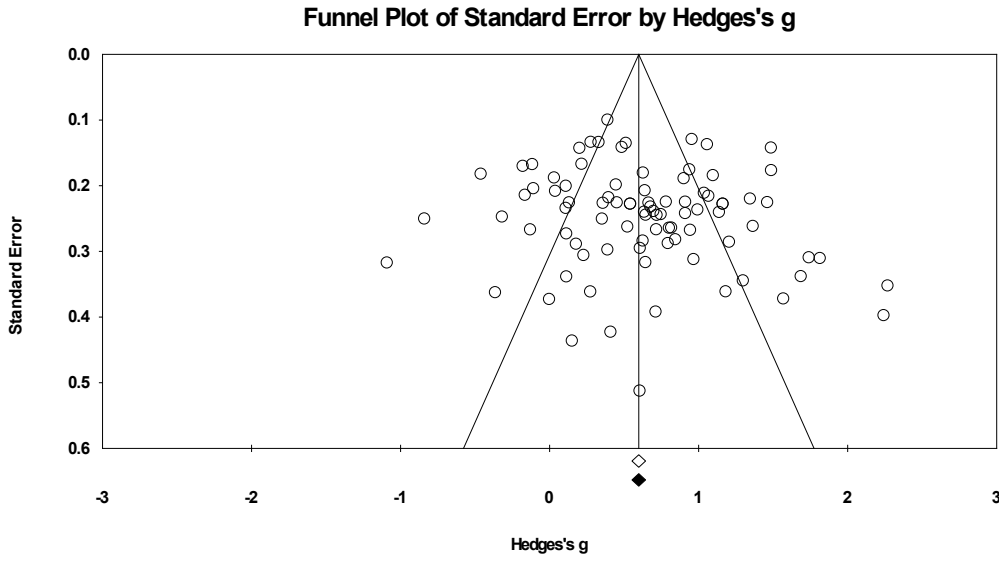


Şekil 3. PBL'nin Kritik Düşünme Becerileri Üzerindeki Etkisini İnceleyen 16 Çalışma için Güven Aralıklarıyla Birlikte Hedge g Değerlerini Gösteren Bir Orman Grafiği Örneği

Orman grafiği yayın yanlılığından ziyade meta-analizin özüyle daha fazla ilgiliymiş gibi görünse de, bu grafiğin incelenmesi her hangi bir analize başlamak için mantıklı bir ilk adım olacaktır (Borenstein, 2005) çünkü orman grafiği sadece okuyuculara bir bakışta meta-analizdeki bireysel çalışmalarla ilgili bilgi sağlamaz aynı zamanda birleştirilmiş sonuçla genel etkiyi özetler. Dahası, çalışmalar arasında ne kadar varyasyon olduğu orman grafiği yardımıyla kolayca görülebilmektedir (Yeh & D'Amico, 2004).

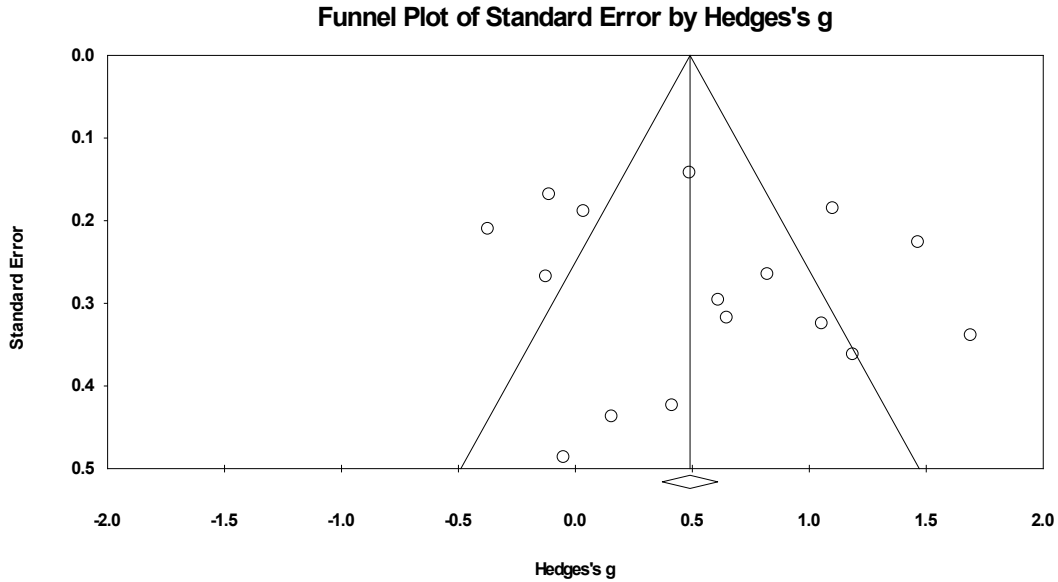
### Huni grafiği

Huni grafikleri, çalışmaların örneklem büyüklüğünün bir ölçüsüne karşı her bir çalışmada tahmin edilen etki büyüklüğünün basit bir dağılım grafiğidir. Geleneksel olarak, huni grafikleri, X eksenini etki büyüklükleri değerlerini gösterirken Y eksenini örneklem büyüklüğü, varyans veya standart hatayı gösterecek şekilde oluşturulur. "Huni grafiği" ismi çalışmaların örneklem büyüklüğü arttıkça uygulamanın etki büyüklüğü tahminindeki kesinliğin de artacağı düşüncesinden gelmektedir (Sterne & Harbord, 2004). Küçük örnekleme sahip çalışmalardan elde edilen sonuçlar grafiğin tabanında geniş bir alana saçılırken büyük örneklemler yukarıda daha küçük bir saçılım gösterirler. Bu nedenle, yanlılığın olmadığı durumlarda grafiğin Şekil 4'de görüldüğü gibi ters dönmüş simetrik bir huniye benzemesi beklenir (Üstün, 2012).



Şekil 4. Yanlılığın Olmadığı Simetrik Bir Huni Grafiği

Diğer taraftan, eğer bir yayın yanlılığı varsa, genellikle Şekil 5'te görüldüğü gibi huni grafiğinde çarpık ve asimetrik bir dağılım olması beklenir. Şekil 5'teki örnekte, meta-analizdeki genel etki uygulamanın etkisini olması gerekenden daha büyük tahmin ederek, 0,38 etki büyüklüğüne ulaşmıştır. Kes ve Ekle yöntemine göre, yanlılığın olmadığı durumda, bu değer 0,09 olması beklenmektedir (Üstün, 2012).



Şekil 5. Olası Bir Yanlılık İçeren Asimetrik Bir Huni Grafiği

Ancak, literatürde yayın yanlılığının huni grafiğindeki asimetrinin sebeplerinden yalnızca bir tanesi olduğu, bu nedenle grafiğin şeklinin yanıltıcı olabileceği ve huni grafiğinin ihtiyatlı bir şekilde yorumlanması gerektiği ciddi şekilde vurgulanmaktadır (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006; J. A. C. Sterne & Harbord, 2004; Terrin, Schmid, & Lau, 2005). Ek olarak, Tang ve Liu (2000) kesinlik ve/veya etki büyüklüğü ile ilgili farklı bir tanım kullanıldığında, huni grafiğinin şeklinin önemli derecede değişebileceğini savunmaktadırlar. Ayrıca, huni grafiğindeki her hangi bir asimetrinin gerçek bir heterojenlikten kaynaklanabileceğini belirtmektedirler.

Egger ve diğerleri (1997) ve Sterne ve Harbord (2004) huni grafiğindeki asimetrinin olası sebeplerini şöyle özetlemişlerdir: seçim yanlılığı (yayın yanlılığı, lokasyon yanlılığı), gerçek heterojenlik, veri düzensizlikleri, insan eseri (artifact), yani etki büyüklüğü ölçümünün yanlış seçiminden kaynaklanan heterojenlik ve tek başına şans. Şans faktörüyle huni grafiğindeki asimetrinin yanlılıktan kaynaklanmak zorunda olmadığı vurgulanmaktadır.

#### *Egger'in doğrusal regresyon yöntemi*

Huni grafikleri yayın yanlılığı hakkında veriyi anlamlandırmak için kullanılabilir yararlı görsellerdir. Ancak, yanlı sonuçları tespit etmek için nicel bir yol sağlamazlar. Diğer taraftan, Egger ve diğerleri (1997) meta-analizdeki verilerde her hangi bir yanlılık olup olmadığını istatistiksel olarak test edebilmek için bir doğrusal regresyon yaklaşımı önermektedir. İstatistiksel test, standart normal sapmanın ( $z$ =etki büyüklüğü tahmini/standart hata) bu değerlerin kesinliğine ( $prec=1/\text{standart hata}$ ) karşı regresyonunu içeren bir modele dayanmaktadır (Sterne & Egger, 2005).

$$E[z] = \beta_0 + \beta_1 prec$$

Simetrik bir huni grafiği için, regresyon çizgisinin orijinden geçerek  $\beta_0$  değerini 0 yapması beklenir. Diğer taraftan, eğer bir asimetri varsa, kesim noktası olan  $\beta_0$  asimetrinin bir ölçüsünü verir. Bu nedenle, istatistiksel test " $\beta_0 = 0$ " sıfır hipotezini kontrol etmek için kullanılır.

Egger'in doğrusal regresyon testinin de istatistiksel anlamlılık testlerinin zayıflıklarını paylaştığının gözden kaçırılmaması önemlidir. Dahası, Borenstein (2005) Egger testinin, farklı örneklem büyüklüğüne sahip çalışmaları ve en azından bir tane orta büyüklükte etki büyüklüğünü içeren veriler için uygun olduğunun altını çizmektedir.

#### *Rosenthal'in güvenli N yöntemi*

Rosenthal (1979) tarafından önerilen güvenli N (fail-safe N-FSN) veya "dosya-çekmecesesi sayısı" (file-drawer number), sosyal bilimlerde yayın yanlılığı problemiyle ilgilenmek için kullanılan en eski ve hala en popüler yaklaşımlardan bir tanesidir (Becker, 2005). Rosenthal FSN, bir meta-analizde elde edilen etkiyi "sınırlamak" için meta-analize eklememiz gereken yeni çalışma sayısı olarak tanımlanabilir (Borenstein vd., 2009). Bir başka değişle, birleştirilmiş testten elde edilen genel olasılık değerini, istatistiksel anlamlılık için genellikle 0,05 veya 0,01 olarak belirlenen kritik değerden daha büyük bir değere dönüştürmek için gerekli yeni çalışma sayısıdır (Rosenthal, 1991). Rosenthal eğer FSN değeri, gözlenen çalışmaların sayısına kıyasla oldukça büyükse, sonuçların yayın yanlılığına karşı dirençli olduğunun varsayılabilirliğini savunmaktadır. Yayın yanlılığından yeterince uzaklaşabilmek için FSN'nin ne kadar büyük olması gerektiğiyle ilgili kesin bir kural bulunmamasına rağmen, Mullen, Muellerleile ve Bryant (2001), Rosenthal'ın önerisini temel alarak,  $N/(5k+10)$  (burada  $k$ , meta-analize dâhil edilen çalışma sayısıdır) değerinin 1'i geçmesi durumunda, meta-analizin sonuçlarının gelecekteki çalışmalar için yeterince dirençli görüldüğü sonucuna ulaşılabileceğini önermişlerdir.

Tablo 3, probleme dayalı öğrenmenin yaratıcılık üzerindeki etkisini inceleyen altı çalışma için gerçekleştirilen Rosenthal FSN hesaplarını göstermektedir (Üstün, 2012). Buna göre,  $N/(5k+10)$  oranı 1,95 olarak hesaplanmaktadır ve bu değer meta-analizdeki çalışma sayısı oldukça az olmasına rağmen, meta-analizin sonuçlarının sonraki çalışmalar için yeterince toleranslı olduğunu göstermektedir.

**Tablo 3.** Probleme Dayalı Öğrenmenin Yaratıcılık Üzerindeki Etkisini İnceleyen Altı Çalışma için Gerçekleştirilen Rosenthal FSN Hesaplamaları için Bir Örnek Çıktı

Gözlenen çalışmalar için Z-değeri	7.29293
Gözlenen çalışmalar için p-değeri	0.00000
Alfa	0.05
Yön	2
Alfa için Z değeri	1.95996
Gözlenen çalışma sayısı	6
FSN	78



### Orwin'in güvenli N yöntemi

Rosenthal FSN bize yayın yanlılığını tespit etmenin açık ve nicel bir yolunu sağlıyor olmasına rağmen istatistiksel anlamlılığa bağlı olması ve kayıp çalışmaların ortalama etki büyüklüğünün sıfır varsayılması sebebiyle eleştirilmektedir (Borenstein, 2005). Alternatif olarak, Orwin FSN pratiksel anlamlılığa bağlı olarak hesaplanmakta ve araştırmacılara sadece kayıp çalışmaların etki büyüklüğünü değil aynı zamanda kayıp çalışmaların eklenmesiyle genel etki büyüklüğünün azalacağı spesifik etki büyüklüğü değerini de belirleme şansı vermektedir. Böylece, kayıp çalışmaların bir dizi dağılımlarının modellenmesine olanak sağlamaktadır (Becker, 2005; Borenstein vd., 2009). Tablo 4, bir önceki Rosenthal FSN örneğindeki çalışmalar için Orwin FSN hesaplamalarını göstermektedir (Üstün, 2012). Sonuçlar genel etki büyüklüğünün "önemsiz" olarak seçilen 0,1 değerine inmesi için sıfır etkiye sahip 370 ek çalışmaya ihtiyaç olduğunu göstermektedir. Eğer eklenecek çalışmaların etki büyüklüğü değeri sıfırdan 0,005'e yükseltirse, eklenmesi gereken çalışma sayısı 740'a yükselmektedir. Önceden belirlenen farklı değerler için farklı FSN değerleri elde etmek mümkündür.

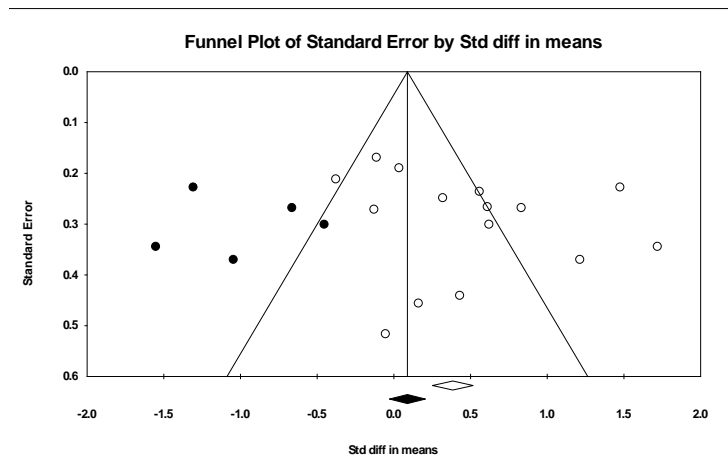
**Tablo 4.** Probleme Dayalı Öğrenmenin Yaratıcılık Üzerindeki Etkisini İnceleyen Altı Çalışma için Gerçekleştirilen Orwin FSN Hesaplamaları için Bir Örnek Çıktı

Gözlenen çalışmalara ait Hedge's g	0.62592
"Önemsiz" Hedge g için ölçüt	0.10000
Kayıp çalışmalar için ortalama Hedge g	0.00000
FSN	370

### Duval ve Tweedie'nin kes ve ekle yöntemi

Kes ve Ekle yöntemi meta-analizdeki olası kayıp çalışma sayısını ve bu kayıp çalışmaların genel bulgu üzerindeki etkisini tahmin etmek için Duval ve Tweedie (2000a, 2000b) tarafından geliştirilmiştir. Bu yöntem tekrarlı bir süreç içermektedir. Bu süreçte öncelikle teorik olarak etki büyüklüğünün yansız bir tahmini olan "düzeltilmiş etki büyüklüğünü" hesaplayabilmek için huni grafiğinin asimetric parçası kesip çıkarılır. Ancak, bu prosedür etki büyüklüğünün varyansını da etkileyerek çok dar bir güven aralığının oluşmasına sebep olur. Bu nedenle, çıkarılan çalışmalar analize yeniden dâhil edilir fakat yansız bir örneklem yaratabilmek için sanal simetrik çalışmalar da eklenir. Eklenmiş olan bu sanal simetrik çalışmalar genel etki büyüklüğünün düzeltilmiş tahminini değiştirmez (Borenstein vd., 2009; Duval, Rothstein, Sutton, & Borenstein, 2005; Duval & Tweedie, 2000a, 2000b).

Şekil 6, daha önce Şekil 5'de yer alan veri için Kes ve Ekle düzeltmesinin dikkate alındığı huni grafiğini göstermektedir. Sonradan eklenen beş çalışma içi dolu çemberle gösterilmektedir. İçi dolu elmas ise düzeltilmiş genel tahmini belirtmektedir. Bu örnekte, düzeltilmiş tahmin sıfır etkiye oldukça yakındır (Üstün, 2012).



**Şekil 6.** Kes ve Ekle Yöntemi Tarafından Eklenen Çalışmaları ve Düzeltilmiş Etki Büyüklüğünü Gösteren Bir Huni Grafiği Örneği

### ***Birincil Çalışmaların Kalitesi***

Meta-analiz sonuçlarının geçerliği ile ilgili diğer bir önemli kaygı birincil çalışmaların kalitesidir (Lipsey & Wilson, 2001; Rendina-Gobioff, 2006). Fakat hem çalışma kalitesinin değerlendirilmesi hem de bu değerlendirmenin meta-analize nasıl dâhil edileceği bazı gerginliklere sebep olmaktadır. Öncelikle, “kalite” teriminin tanımlanması kolay değildir çünkü bir çalışmayı neyin kaliteli yaptığı “değerlendirmenin neden yapıldığına” bağlı olarak değişmektedir ve bu durum kavramın çok boyutlu olmasına sebep olmaktadır (Jüni, Altman, & Egger, 2001; Valentine, 2009). Çok boyutlu yapısı nedeniyle çalışma kalitesinin değerlendirilmesindeki zorluk diğer bir gerginliğe sebep olmakta ve standardize olmuş farklı kalite ölçeklerini kullanan araştırmacıların aynı çalışma için farklı skorlar elde etmesi sonucunu doğurmaktadır (Herbison, Hay-Smith, & Gillespie, 2006). Çalışma kalitesinin değerlendirilmesiyle ilgili diğer bir zorluk da çalışma kalitesiyle raporlama kalitesinin girişiminden kaynaklanmaktadır (Wells & Littell, 2009). Birçok durumda, meta-analiz yapan araştırmacıların çalışma kalitesinin unsurlarını kodlayabilmesi için gerekli bilgi çalışmalarda yer almadığı gibi araştırmacının bu tip durumlarda izleyebileceği açık bir prosedür de bulunmamaktadır (Valentine, 2009). Son gerilim ise çalışma kalitesinin meta-analizlerde nasıl kullanılacağından doğmaktadır. Bir meta-analiz çalışmasında çalışma kalitesinin kullanılmasının sıradan yollarından bir tanesi basitçe düşük standartlara sahip çalışmaların dışlanması, yani meta-analize dâhil edilmemesidir (Lipsey & Wilson, 2001; Valentine, 2009). Ancak, Glass (1982, 2006) çalışma kalitesinin dışlama ölçütlerinden bir tanesi olarak kullanılması fikrine katılmamaktadır çünkü her hangi bir birincil çalışmayı kalite kaygılarıyla dışlamak subjektif bir yargılamaya dayanır ve bu durum sağlıklı sonuçlar doğurabilir. Çalışma kalitesinin meta-analizlerde kullanılmasıyla ilgili diğer bir yaklaşım ise kalite kaygılarından bağımsız olarak tüm birincil çalışmaların meta-analize dâhil edilmesi ve daha sonra çalışma kalite göstergeleri için ara değişken veya alt-grup analizi yapılmasıdır (Littell vd., 2008).

Literatürde, araştırma sentezleri yapılırken birincil çalışmaların kalitesinin değerlendirilmesi için hazırlanmış bir dizi standart ve kriter listesini içeren çok sayıda ölçüm aracı bulunmaktadır (Herbison vd., 2006; Littell vd., 2008; Valentine, 2009). Ayrıca, çalışma kalitesi için geliştirilmiş olan bu ölçüm araçlarını değerlendiren araştırmalar da literatürde yer almaktadır (Deeks vd., 2003; Herbison vd., 2006; Jüni vd., 2001; Jüni, Witschi, Bloch, & Egger, 1999; Wells & Littell, 2009). Örneğin, Deeks ve diğerleri (2003) rastgele atama içermeyen çalışmaların kalitesini değerlendirmek için geliştirilen 194 ölçüm aracını incelemiş ve çalışmaların hiçbirinin revize edilmeden bu amaçla kullanmaya tamamen uygun olmadığı sonucuna varmıştır. Benzer şekilde, Herbison ve diğerleri (2006) çalışma kalitesi skoru elde etmek amacıyla geliştirilen 45 ölçeğin geçerliğini ampirik olarak araştırmış ve “güncel kalite skorları meta-analizlerin yararlılığını geliştirmede hiçbir değere sahip değildir veya çok az değere sahiptir. Aslında, yanlılığa sebep olabilirler çünkü hangi kalite skorunu kullandığınıza bağlı olarak farklı cevaplar alırsınız” (s. 1251) sonucuna ulaşmışlardır. Çalışma kalitesinin açık bir şekilde önemli olduğunu kabul etmekle birlikte kalite skorlarının bu durum için bir çözüm sunmadığının altını çizmektedirler.

Sonuç olarak, meta-analizlerde çalışma kalitesi ölçeklerine dayalı toplam bir puan atanmasından vazgeçilmesi gerektiği literatürde yaygın şekilde kabul görmektedir (Herbison vd., 2006; Jüni vd., 1999; Littell vd., 2008; Wells & Littell, 2009). Bunun yerine, meta-analiz çalışmalarında çalışma kalitesinin spesifik boyutlarının ara değişken analiziyle incelenmesi tavsiye edilmektedir (Herbison vd., 2006; Jüni vd., 2001; Littell vd., 2008). Diğer taraftan, Wells ve Littell (2009) yayın statüsünün çalışma kalitesinin iyi bir göstergesi olmadığını savunmakta ve raporlama kalitesinin çalışma kalitesiyle karıştırılmaması gerektiğini vurgulamaktadırlar.

## Kodlama Güvenirliği

Meta-analizlerde kodlama güvenirliliğinin sağlanması çok önemlidir çünkü kodlama sayfasındaki maddeleri nasıl kodladığınız, birincil çalışmalarını kodlarken kodlayıcıların kaçınılmaz şekilde kullandıkları yargılama sürecinin sonucu olarak bir miktar değişkenlik gösterebilir. Kodlama güvenirliliğinin iki boyutu vardır. Bunlardan bir tanesi “kodlayıcı güvenirliliği”, yani tek bir kodlayıcının çalışmadan çalışmaya kodlama tutarlılığıdır. Diğeri ise “kodlayıcılar-arası güvenirlilik”, yani farklı kodlayıcılar arasındaki tutarlılıktır (Lipsey & Wilson, 2001).

Kodlayıcı ve kodlayıcılar-arası güvenirliliğin bir ölçüsü olarak “uzlaşma oranı (agreement rate)” (AR) kullanılabilir. AR aşağıdaki formül kullanılarak hesaplanabilir (Orwin & Vevea, 2009):

$$AR = \frac{\text{üzerinde uzlaşılan görüş sayısı}}{\text{toplam görüş sayısı}}$$

Orwin ve Vevea (2009) kodlama kararlarının nasıl değerlendirilebileceği ile ilgili daha detaylı bilgi sağlamaktadır.

## Heterojenlik Analizi

Huedo-Medina, Sanchez-Meca, Marin-Martinez ve Botella (2006) meta-analizin üç önemli amacı olduğunu ileri sürmektedirler. Bunlardan birincisi çalışılan ilişkinin etki büyüklüğüyle ilgili güven aralığı ve istatistiksel anlamlılığı ile birlikte genel bir indeks elde edilmesi, ikincisi etki büyüklüklerinin heterojenliklerinin test edilmesi ve sonuncusu ise birincil çalışmalardan elde edilen etki büyüklükleri arasında heterojenlik varsa sonucu etkileyebilecek olası ara değişkenlerin tespit edilmesidir. Yani, heterojenliğin test edilmesi meta-analizin temel amaçlarından bir tanesidir çünkü heterojenlik sadece ara değişkenlerin varlığına işaret etmez aynı zamanda rastgele-etkiler modeli fikrinin altında yatan sayıtlılardan bir tanesidir.

Gerçek etki büyüklükleri, yani çalışmaların temsil ettiği popülasyonların etki büyüklükleri, arasındaki heterojenliği tespit etmekteki güçlük, gözlenen varyansı kullanarak gerçek heterojenliği tahmin etmeye çalışmamızdan ve bu varyansın aynı zamanda rastgele hata içermesinden kaynaklanmaktadır (Borenstein vd., 2009). Diğeri bir değişle, değişkenliğin iki kaynağı vardır. Bunlar, aynı zamanda çalışma-içi değişkenlik olarak da bilinen örnekleme hatası ve çalışmalar-arası değişkenliktir. Örnekleme hatası meta-analizlerde her zaman bulunurken çalışmalar-arası değişkenlik sadece, gözlenen etki büyüklükleriyle tahmin edilen, popülasyonların etki büyüklükleri arasında heterojenlik olduğunda var olmaktadır (Huedo-Medina vd., 2006). Bu durumda, miktarını belirlemeye çalıştığımız şey, örnekleme hatasından arındırılmış olan çalışmalar-arası heterojenliktir.

Meta-analizlerde heterojenliği tanımlama ve miktarını belirlemenin farklı yolları vardır. Heterojenliği ölçmenin alışılmış yolu olan Q istatistiği ve karşılık gelen ki-kare anlamlılık testinin avantaj ve eksiklikleri önümüzdeki bölümde tanıtılırken I<sup>2</sup> ve τ<sup>2</sup> gibi alternatifleri ilerleyen bölümlerde kısaca açıklanacaktır.

### *Q İstatistiği ve İlişkili Ki-kare Anlamlılık Testi*

Hesaplanırken, ortalama etki büyüklüğünden sapmaların, varyansın tersiyle ağırlıklandırıldığı Q istatistiği, basitçe ağırlıklandırılmış kareler toplamıdır. Böylece çalışma-içi varyansı da içeren toplam varyans hakkında bir ölçüm sağlar. Gerçek heterojenlik ise Q istatistiğinden birincil çalışma sayısının bir eksiği, yani bağımsızlık derecesi (df) çıkarılarak kestirilmeye çalışılır. Ancak, bu değerin bir ortalama değil sapmalar toplamı olduğuna dikkat edilmelidir, yani Q istatistiği sezgisel bir ölçüm değildir. Bu nedenle, Q istatistiği tüm çalışmaların ortak bir etki büyüklüğünü paylaştığını iddia eden sıfır hipotezini, ki-kare dağılımıyla test etmek için kullanılır (Borenstein vd., 2009).

Diğer taraftan, anlamlılık testi halen diğer istatistiksel anlamlılık testlerinin sınırlıklarını paylaşmaktadır çünkü örneklem büyüklüğüne, yani meta-analize dâhil edilen birincil çalışmaların sayısına, son derece bağlıdır. Huedo-Medina ve diğerleri (2006) meta-analizdeki çalışma sayısı ve/veya ortalama örneklem büyüklüğü düşük olduğunda, Q istatistiğini kullanan istatistiksel test olan Q testinin düşük istatistiksel güçten sıkıntı yaşadığını iddia etmişlerdir. Ayrıca, Q testinin sadece heterojenliğin varlığına veya yokluğuna işaret ettiğini ancak heterojenliğin miktarını belirleyemediğini vurgulamışlardır.

#### *$\tau^2$ 'nin Tahmin Edilmesi*

$\tau^2$  gerçek etki büyüklüğünün varyansına karşılık gelen bir parametredir. Meta-analizlerde  $\tau^2$  gözlenen etki büyüklüğünün varyansı ( $T^2$ ) kullanılarak tahmin edilmeye çalışılır. Bu tahmin (Q-df) değerine dayanmaktadır ancak farklı olarak etki büyüklüğüyle aynı ölçekte mutlak bir değer sağlayarak gerçek değişkenliğin miktarını belirtir (Borenstein vd., 2009). Dahası, bu değer karekökü, popülasyonun standart sapması olan tau'nun ( $\tau$ ) bir tahminini vermektedir.

Hem  $\tau^2$  hem de  $\tau$  etki büyüklüğünün ne derece heterojen olduğu hakkında bilgi verirler ve bu değer Q istatistiğinden doğrudan çıkarılamaz. Ancak,  $\tau^2$ 'nin büyüklüğü ile ilgili tartışma etki büyüklüğü indeksine bağlıdır çünkü  $\tau^2$  çalışmalar-arası varyansının büyüklüğünü etki büyüklüğü indeksi ile aynı ölçekte belirler.  $\tau^2$ 'in gerçek etki büyüklüğünün heterojenliğini belirlemede nasıl kullanıldığını örneklendirmek için 0,566 Hedge g değerine ait bir dağılım olduğunu ve bu dağılıma ait  $\tau^2$  değerinin 0,212 olduğunu varsayalım. Bu  $\tau$  değerinin 0,461 olduğu anlamına gelir. Bu örnekte, uygun hesaplamalarla tüm durumların %95'inde, yeni bir çalışmanın etki büyüklüğünün -0,416 ile 1,548 aralığına düşeceği sonucuna kolaylıkla ulaşılabilir. Hesaplamaların detaylarına Borenstein ve diğerleri (2009)'da ulaşılabilir.

#### *$I^2$ İstatistiği*

Heterojenliğin miktarını belirlemenin bir diğer yolu ise gerçek varyansın toplam varyansa oranına karşılık gelen  $I^2$  istatistiğinin oluşturulmasıdır.  $I^2$  istatistiği de Q istatistiğine dayalı olmasına rağmen Q istatistiğine kıyasla heterojenliğin ölçüsünü bize çok daha sezgisel bir ölçekte sağlar. Etki büyüklüğü indeksi ile aynı ölçekte mutlak değerler sunan  $\tau^2$  ve  $\tau$ 'dan farklı olarak,  $I^2$  istatistiği etki büyüklüğü ölçeğine bağlı olmayan göreceli bir ölçekte oran sağlamaktadır.

Higgins, Thompson, Deeks, ve Altman (2003), heterojenliğin bir ölçüsü olarak  $I^2$  kullanmanın bazı avantajlarını şöyle özetlemektedirler:

- Bir oran sağladığı için yorumu sezgiseldir.
- Hesaplanması kolaydır.
- Doğası gereği örneklem büyüklüğüne bağlı değildir.
- Etki büyüklüğü ölçeğinden bağımsız olarak yorumlamak mümkündür.

Ayrıca, düşük, orta ve yüksek seviye heterojenlik için sırasıyla %25, %50 ve %75 sınır değerlerini önermişler ve  $I^2$  istatistiğinin heterojenliğin ölçüsü olarak tercih edilebilir olduğunu iddia etmişlerdir.

Diğer taraftan, Borenstein ve diğerleri (2009),  $I^2$ 'nin sadece çalışmalar-arası varyansın toplam varyansa oranını yansıttığını ve gerçek varyansın mutlak değerini sağlamadığını vurgulamışlardır. Bu nedenle, büyük miktarda bir gerçek varyans, zayıf kesinliğin, yani geniş güven aralıklarının sonucu olarak ortaya çıkan yüksek miktarda rastgele hata ile kolaylıkla maskelenebilir. Ayrıca, gerçek heterojenliğin bilgilendirici bir tanıtımı için hem heterojenliğin büyüklüğü ile ilgili bir ölçünün hem de belirsizliğin bir ölçüsünün raporlanması gerektiğini dile getirmişlerdir. Heterojenliğin büyüklüğü  $\tau^2$ 'nin bir tahmini olarak  $T^2$  veya  $I^2$  ile belirsizliğin ölçümü ise Q-test veya  $T^2$  veya  $I^2$  için güven aralıkları ile gösterilebilir.

## Ara Değişken Analizi

Meta-analizin temel amaçlarından bir tanesi, çalışmaların farklı alt-grupları için ortalama etkilerin karşılaştırılarak, meta-analize dâhil edilen birincil çalışmalardan elde edilen etki büyüklükleri arasındaki değişkenliğin analiz edilmesidir (Borenstein vd., 2009; Huedo-Medina vd., 2006). Ancak, birincil çalışmalardaki alt-grupların karşılaştırılması için kullanılan varyans analizi (ANOVA) doğrudan meta-analizlerde uygulanamaz çünkü birincil çalışmadaki katılımcıların bireysel skorlarının yerini her bir birincil çalışmadan elde edilen etki büyüklükleri almaktadır. Bu nedenle, meta-analizlerde alt-grupları karşılaştırmak için istatistiksel test olarak Q-testine dayanan “analog ANOVA” kullanılır.

Analog ANOVA testi, sabit-etki, rastgele-etkiler (aynı zamanda tamamen rastgele-etkiler olarak da adlandırılır) ve karma-etkiler modeli gibi farklı modeller temel alınarak uygulanabilir. Bu modellerden her birinin, alt-gruplar içinde etki büyüklüklerinin değişkenliği ve alt-grupların değişkenliği ile ilgili farklı sayıtları vardır. “Alt-gruplar içi” seviyesinde, sabit-etki ile rastgele etkiler modelleri arasındaki fark genel etki büyüklüğünün hesaplanması sırasında kullanılanlarla aynıdır. Yani, sabit-etki modeli bir tane gerçek popülasyonu temsil eden sadece bir tane gerçek etki büyüklüğü olduğunu ve alt-gruplar içerisinde etki büyüklüklerindeki değişkenliğin sadece örnekleme hatasından kaynaklandığını varsaymaktadır. Rastgele etkiler modeli ise farklı popülasyonları temsil eden farklı etki büyüklüklerine izin vermekte ve toplam varyansı “alt-gruplar arası” ve “alt-gruplar içi” varyansları olarak ikiye ayırmaktadır. Diğer taraftan, “alt-gruplar arası” seviyesinde, sabit ve rastgele farklı anlamlara gelmektedir. “Sabit”, benzer analizleri yapacak olan her araştırmacı için alt-grupların sabit veya aynı olduğu anlamına gelmektedir. Örneğin, cinsiyet ara değişkeninin alt grupları sabit olarak atanabilir. Ancak, ülke değişkeni, belirli bir meta-analiz çalışmasında yer almayan ülkelere de genelleme yapılabilmesi için, alt-gruplar arası seviyesinde rastgele olarak atanabilir (Borenstein vd., 2009).

Ara değişken analizi için kullanılan sabit-etki modeli alt-grup içerisinde sadece bir tane gerçek etki büyüklüğü olduğunu ve alt-gruplar arası seviyesinde alt-grup kategorilerinin sabit olduğunu varsaymaktadır. Rastgele-etkiler modeli ise her iki seviyede de rastgele değişkenliği kullanmaktadır. Ayrıca, karma-etkiler modeli adında bir model daha vardır ve bu model “alt-grup içi” seviyesinde rastgele-etkiler modelini kullanmakta ama alt-grup kategorilerinin sabit olduğunu varsaymaktadır.

Son olarak, Lipsey ve Wilson (2001) analog ANOVA’nın ara değişkenlerle ilgili sınırlı sayıda öncül hipotezin test edilmesi için kullanılması gerektiğini iddia etmektedir. Çok sayıda kategorik değişkenin analog ANOVA ile test edilmesinin yaygın, ancak Tip 1 hata oranlarını artırması sebebiyle, yanlış bir uygulama olduğunun altını çizmektedirler.

Ayrıca, birincil çalışmalarda kullanılan regresyon veya çoklu regresyona benzer bir analiz olan meta-regresyonun meta-analizcilere meta-analiz kapsamında ara değişken analizi yapabilmek için diğer bir seçenek sunuyor oluşu hatırlatılması gereken diğer bir noktadır. Regresyon ile meta-regresyonun mantığı arasındaki tek fark meta-regresyonda, kovaryansların birincil çalışmalardaki katılımcılar seviyesinde değil birincil çalışmalar seviyesinde tanımlanmasıdır. Ancak, regresyondan farklı olarak, meta-regresyon yapılırken her bir birincil çalışmanın ağırlıklandırılması gerekmektedir. Bu nedenle, SPSS veya STATA gibi genel amaçlar için hazırlanmış olan istatistiksel paketlerin yapılandırdığı standart regresyon modellerini kullanmak mümkün değildir.

### *Açıklanan Varyansın Oranı*

Analog ANOVA testi istatistiksel anlamlılık testlerinin doğasında var olan zayıflıklara sahiptir. Ayrıca, literatürde belirtildiği gibi ara değişken analizleri için kullanılan anlamlılık testleri genellikle düşük istatistiksel güce sahiptir (Borenstein vd., 2009; Pigott, 2012). Bu nedenle, bu anlamlılık testlerinden elde edilen anlamlı olmayan sonuçlar ihtiyatlı bir şekilde yorumlanmalıdır. Son olarak, bu test sadece alt-grupların ortalama etki büyüklükleri arasında istatistiksel olarak anlamlı bir fark olup olmadığını kontrol etmektedir ancak bu farkın büyüklüğünü belirlememektedir.



Birincil çalışmalarda, açıklanan varyansın toplam varyansa oranı olarak tanımlanan bir indeks olan  $R^2$ , kovaryantın bağımlı değişken üzerindeki etkisinin büyüklüğünü belirlemek için kullanılmaktadır. Ancak, tamamen dışlanması mümkün olmayan çalışma-içi varyans sebebiyle bu indeksin doğrudan meta-analizlerde kullanılması mümkün değildir. Bu nedenle,  $R^2$  meta-analizlerde sadece gerçek varyans olan  $\tau^2$ 'e odaklanacak şekilde yeniden tanımlanır. Yani,  $R^2$  kovaryant tarafından açıklanan toplam varyansın değil gerçek varyansın oranı olarak yeniden tanımlanır (Borenstein vd., 2009). Bu indeks şu şekilde hesaplanabilir:

$$R^2 = 1 - \frac{T_{within}^2}{T_{total}^2}$$

burada  $T_{within}^2$  alt-gruplar boyunca birleştirilmiş varyanstır ve şöyle hesaplanabilir:

$$T_{within}^2 = \frac{Q_{total} - df}{C_{total}}$$

burada C bir ölçeklendirme faktörüdür ve Comprehensive Meta-analysis (CMA) tarafından hesaplanmaktadır veya şu formül kullanılabilir:

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

burada w her bir çalışmanın ağırlığıdır (Ellis, 2010).  $T_{within}^2$  örnekleme sorunları nedeniyle negatif bir değer olabilir. Bu durumda sıfır olarak kabul edilmelidir (Borenstein vd., 2009).

$R^2$  indeksi CMA tarafından sağlanmamaktadır fakat yukarıda verilen formüllerle ve CMA tarafından sağlanan C değeri kullanılarak hesaplanabilir. Cohen (1988)  $R^2$  indeksinin küçük, orta ve büyük kategorileri için sırasıyla 0,02, 0,13 ve 0,26 değerlerini sınır değerleri olarak önermektedir.

### Güç Analizi

İstatistiksel güç “bir testin gerçek bir etkiyi doğru şekilde belirleme olasılığını” tanımlar. “Teknik olarak, bir testin gücü testin yanlış bir sıfır hipotezi reddetme olasılığı olarak tanımlanabilir” (Ellis, 2010, s. 52). Birincil bir çalışmada istatistiksel gücü etkileyen dört faktör bulunmaktadır. Bunlar: etki büyüklüğünün değeri, araştırmacı tarafından belirlenen alfa değeri, testin tek yönlü veya çift yönlü oluşu ve son olarak örneklem sayısıdır (Gravetter & Walnau, 2007). Etkilerin yönü şöyle özetlenebilir: istatistiksel güç, artan uygulama etkisiyle ve çalışmanın artan kesinliğiyle artmaktadır. Bu durum, meta-analizlerin istatistiksel gücü için de aynen doğrudur (Borenstein vd., 2009). Bu nedenle, sabit-etki modeli kullanılarak yapılan bir meta-analizin istatistiksel gücünün meta-analize dâhil edilen her bir çalışmanın gücünden daha büyük olması şaşırtıcı değildir. Sabit-etki modeliyle yapılan meta-analizlerin istatistiksel olarak ne derece güçlü olduğu bu modelde ortalama etki büyüklüğünün güven aralığının birincil çalışmaların güven aralığından her zaman daha dar olmasından da kolayca tahmin edilebilir. Güven aralığının darlığı aynı zamanda örneklemin büyük olmasından dolayı artan kesinliğe işaret etmektedir.

Diğer taraftan, daha önce açıklandığı gibi iki farklı hata kaynağı içeren rastgele-etkiler modelinde durum oldukça farklıdır. Heterojenliğin işaretçisi olan çalışmalar-arası varyans istatistiksel gücü de etkiler; bu nedenle, rastgele-etkiler modeliyle yapılan bir meta-analizde meta-analizin, dâhil ettiği birincil çalışmalardan daha küçük istatistiksel gücünün olması mümkündür.

Ana etki için kullanılan istatistiksel testler için yapılan güç analizleri birincil çalışmalarda yapılanlara çok benzerdir. Tek fark, rastgele-etkiler modelinde artan heterojenlikle artan, ortalama etki büyüklüğünün varyansının hesaplanmasından kaynaklanmaktadır. Bir kere varyans hesaplandığında,  $\lambda$  parametresi şöyle hesaplanabilir:

$$\lambda = \frac{\delta}{\sqrt{V_\delta}}$$

burada  $\delta$  gerçek etki büyüklüğü,  $V_\delta$  ise bu etki büyüklüğünün varyansdır. Ardından, istatistiksel güç şöyle hesaplanabilir:

$$\text{Power} = 1 - \Phi(c_\alpha - \lambda) + \Phi(-c_\alpha - \lambda)$$

burada  $c_\alpha$ ,  $\alpha$  anlamlılık seviyesiyle ilişkili Z'nin kritik değeridir. Bu değer  $\alpha$ 'nın 0,05 değeri için 1,96'dır.  $\Phi(x)$  ise EXCEL'de NORMSDIST fonksiyonu kullanılarak hesaplanabilir (Pigott, 2012).

### İstatistiksel Analizler için Yazılımlar

SPSS, SAS, STATA ve R gibi genel amaçlar için tasarlanmış istatistiksel paketler, meta-analiz için içerisinde var olan bir desteğe sahip değildir. Bu yazılım paketlerinin her hangi birinde özellikle rastgele-etkiler modelinde ihtiyaç duyulan ağırlıkların atanması kolay değildir. Alt-grup analizinde (analog ANOVA) ve meta-regresyonda ise meta-analizlerdeki bağımsızlık derecelerinin atanmasındaki farklı kurallar nedeniyle genel amaçlı paketler yanlış p-değerleri oluşturmaktadırlar (Borenstein vd., 2009). Ancak, bu istatistiksel paketler için geliştirilen bazı makrolar yardımıyla bu yazılımlarla meta-analiz yapabilmek mümkündür ve her birinin kendine has güçlü yanları ve sınırlılıkları bulunmaktadır.

Diğer taraftan, Comprehensive Meta-Analysis (CMA), RevMan ve MIX gibi özellikle meta-analiz için geliştirilmiş olan bazı istatistiksel yazılımlar bulunmaktadır. Bax, Yu, Ikeda ve Moons (2007) meta-analize adanmış olan altı istatistiksel programı karşılaştırmış ve en uygun meta-analiz programının hangisi olduğunun, beklentilerine bağlı olarak her bir kullanıcı için değişebileceği sonucuna ulaşmışlardır. Bununla birlikte, CMA'nın en çok yönlü program ve MIX ve CMA'nın en kullanıcı-dostu meta-analiz programı olduğunu belirtmişlerdir. CMA, sabit-etki ve rastgele-etkiler modelinde ana etkilerin hesaplanması için ve farklı heterojenlik ve yayın yanlılığı analizlerinin yanında alt-grup analizi ve meta-regresyon yapabilmek için birçok istatistiksel analiz yürütülmesine olanak sağlayan ticari bir yazılımdır. Ayrıca, bu yazılımla orman ve huni grafikleri oluşturmak ve bu grafiklerde bazı değişiklikler yapmak mümkündür. CMA'nın diğer bir önemli avantajı ise araştırmacılara 100 farklı formatta veri girişi olanağı sağlamasıdır. Ancak, ne CMA ne de MIX araştırmacılara çoklu seviyeli veya Bayes modellerle çalışma veya çok değişkenli analizler yapabilmek olanağı vermektedir. Son olarak, Wallace, Schmid, Lau ve Trikalinos (2009), STATA ve R makroları ve MIX, CMA, RevMan ve Meta-Analyst gibi meta-analiz yazılımlarının detaylı karşılaştırmasını sunmaktadır.

### Meta-Analiz Nasıl Rapor Edilir?

Şu ana kadar, araştırma sentezi yapmak için kullanılan bir araştırma yöntemi olarak meta-analizin kavramsal çerçevesini oluşturmaya ve nasıl yapılacağı hakkında açıklamalar sağlamaya çalıştık. Ancak, bir araştırma çalışmasının nasıl yapılabileceği akademik işin sadece bir yönüdür, diğer yönünü ise raporlama meseleleri oluşturur. Bir araştırma çalışması hakkında diğer araştırmacıların ne bildiği, ilgili makale veya tezde ne rapor edildiğiyle sınırlı olduğu için raporlama kalitesi meta-analizler dâhil her türlü akademik iş için hayati öneme sahiptir (Clarke, 2009).

Bu bağlamda, Ahn, Ames ve Myers (2012) eğitim alanında yapılan meta-analizlerin özellikle veri analizi ve değerlendirmesinde birçok metodolojik zayıflıkları olduğuna dikkat çekmektedir. Araştırmacılar, bu çalışma kapsamında, son zamanlarda eğitim alanında yayınlanmış olan 56 meta-analizi incelemişler ve özellikle istatistiksel metotların kullanımında meta-analiz çalışmalarındaki araştırma ve raporlama kalitesinin artırılması gerektiğinin altını çizmişlerdir. Bu nedenle, bütün meta-analiz yapan araştırmacılara, meta-analizlerini olabildiğince iyi tasarlayabilmeleri için, çalışmalarına başlamadan önce kontrol listelerinden uygun olanları dikkatlice incelemeleri çok fazla tavsiye edilmektedir.

Bu bakış açısıyla, meta-analizler için raporlama standartları geliştirmek için literatürde kayda değer bir çaba görülmektedir. "The Quality of Reporting of Meta-analysis" (QUOROM) (Moher vd., 1999) ve güncellenmiş versiyonu, "Preferred Reporting Items for Systematic Reviews and Meta-

analyses" (PRISMA) (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009), ve "Meta-analysis of Observational Studies in Epidemiology" (MOOSE) (Stroup vd., 2000) geliştirilen standartlardan bazılarını oluşturmaktadır. Ayrıca, APA, "Meta-analysis Reporting Standards" (MARS) (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) adında meta-analizler için diğer bir raporlama standardı yayınlamıştır. İlgili referansta belirtilen web sitesinde çevrimiçi bir doküman olarak sunulan MARS, başlık, öz, giriş, dâhil etme ve dışlama kriterlerini, ara değişken analizini, arama stratejilerini, kodlama sürecini ve istatistiksel metotları içeren metot bölümü ve sonuçlar ve tartışma bölümleri gibi meta-analizin her bir bölümünde ne raporlanması beklendiğini detaylı şekilde açıklamaktadır. MARS eğitim alanındaki birincil çalışmaların meta-analizleri için en uygun kontrol listesi gibi görünmektedir.

### **Sonuç**

Bu makalede, araştırma sentezlerinin bilimsel girişimdeki çok önemli rolünü vurgulamaya ve meta-analizin araştırma sentezi içerisindeki yerini tartışarak meta-analiz kavramını açıklamaya çalıştık. Daha sonra, güçlü ve eleştirilen yanları ve kısmen metodolojik ve istatistiksel temelleri gibi farklı açılardan meta-analizi inceledik. Son olarak, meta-analizin nasıl raporlanması gerektiğine dair geliştirilen standartları özetleyerek daha iyi tasarlanmış meta-analizler yapabilmek ve bu meta-analizleri, tüm okuyucuların çalışmanın detaylarını anlayabileceği şekilde, raporlayabilmek için bu tip standartların meta-analiz yapan araştırmacılar tarafından takip edilmesinin ne derece önemli olduğunu vurguladık. Araştırma sentezinin etkili bir yöntemi olarak meta-analizlerin sayısı, son yıllarda eğitim araştırmalarının da dâhil olduğu birçok disiplinde hızla artmaktadır. Ancak, bu meta-analizlerin önemli bir kısmı metodolojik, istatistiksel ve raporlama zayıflıklarından zarar görmektedir (Ahn vd., 2012). Bu nedenle, meta-analiz çalışmalarında, özellikle istatistiksel yöntem bölümlerinde, hem araştırma hem de raporlama kalitesinin artırılması gerekmektedir. Bu bağlamda, bu makalenin meta-analizle uğraşan araştırmacılara meta-analizlerini yaparken ve raporlarken genel bir çerçeve ve rehberlik sağlamasını ümit ediyoruz.

### **Teşekkürler**

Bu makale hakkındaki değerli dönütleri için Yrd. Doç. Dr. Ertuğrul Özdemir ve Yrd. Doç. Dr. Süleyman Davut Göker'e teşekkür ederiz.

### Kaynakça

- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82*(4), 436-476. doi: 10.3102/0034654312458162
- APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*(9), 839-851. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2957094/>
- Bax, L., Yu, L. M., Ikeda, N., & Moons, K. G. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology, 7*(1), 40.
- Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex, England: John Wiley & Sons, Ltd.
- Bennett, D. A., Latham, N. K., Stretton, C., & Anderson, C. S. (2004). Capture-recapture is a potentially useful method for assessing publication bias. *Journal of Clinical Epidemiology, 57*(4), 349-357.
- Bennett, J. (2005). Systematic reviews of research in science education: Rigour or rigidity? *International Journal of Science Education, 27*(4), 387-406.
- Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18.
- Berman, N., & Parker, R. (2002). Meta-analysis: Neither quick nor easy. *BMC Medical Research Methodology, 2*(1), 10.
- Bligh, J. (2000). Problem-based learning: The story continues to unfold. *Medical Education, 34*(9), 688-689.
- Borenstein, M. (2005). Software for Publication Bias. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication Bias for Meta-Analysis: Prevention, Assessment and Adjustments*. West Sussex, England: John Wiley & Sons Ltd.
- Borenstein, M. (2009). Effect size for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons, Ltd.
- Bushman, B. J., & Wang, M. C. (2009). Vote-counting procedures in meta-analysis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 207-220). New York: Russell Sage Foundation
- Bushman, B. J., & Wells, G. L. (2001). Narrative impressions of literature: The availability bias and the corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin, 27*(9), 1123-1130.
- Card, N. A. (2012). *Applied meta-analysis for social science research*. New York: The Guilford Press.
- Carlton, P. L., & Strawderman, W. E. (1996). Evaluating cumulated research I: The inadequacy of traditional methods. *Biological Psychiatry, 39*(1), 65-72.
- Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & The Health Professions, 25*(1), 12-37.
- Chan, M. L. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science, 7*(1), 79-92.
- Clarke, M. (2009). Reporting format. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2 ed., pp. 521-534). New York: Russell Sage Foundation.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304.
- Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, 8(3), 243-253.
- Cooper, H. (1997). Some finer points in the meta-analysis. In M. Hunt (Ed.), *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.
- Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 3-16). New York: Russell Sage Foundation
- Cooper, H., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87(3), 442.
- Dalton, D. R., & Dalton, C. M. (2008). Meta-analyses. *Organizational Research Methods*, 11(1), 127-147.
- Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26(3-4), 365-378.
- Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovich, C., Song, F., . . . Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7(27).
- Duval, S., Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *The Trim and Fill Method Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. West Sussex, England: John Wiley & Sons, Ltd.
- Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98.
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
- Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49(2), 275-306.
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 33(5), 517.
- Eysenck, H. J. (1984). Meta-analysis: An abuse of research integration. *The Journal of Special Education*, 18(1), 41-59.
- Eysenck, H. J. (1994). Systematic reviews: Meta-analysis and its problems. *British Medical Journal*, 309, 789-792.
- Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research*, 94(5), 275-282.
- Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century *Journal of Clinical Epidemiology*, 48(1), 71-79.
- Field, A. P. (2003). The problem in using fixed-effects models of meta-analysis on real world data. *Understanding Statistics*, 2, 77-96.
- Fitz-Gibbon, C. T. (1985). The implications of meta-analysis for educational research. *British Educational Research Journal*, 11(1), 45-49.
- Fitzgerald, S. M., & Rumrill, P. D. (2003). Meta-analysis as a tool for understanding existing research literature. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 21(1), 97-103.
- Fitzgerald, S. M., & Rumrill, P. D. (2005). Quantitative alternatives to narrative reviews for understanding existing research literature. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 24(3), 317-323.



- Fleiss, J. L., & Berlin, J. A. (2009). Effect size for dichotomous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Glass, G. V. (1982). Meta-analysis: An approach to the synthesis of research results. *Journal of Research in Science Teaching*, 19(2), 93-112.
- Glass, G. V. (2006). Meta-analysis: The quantitative synthesis of research findings. In J. L. Green, P. B. Elmore & G. Camilli (Eds.), *Handbook of Complementary Methods in Education Research*. Mahwah: Lawrence Erlbaum Associates.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. CA: Sage Publications.
- Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Gliner, J. A., Morgan, G. A., & Harmon, R. J. (2003). Meta-analysis: Formulation and interpretation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(11), 1376.
- Gravetter, F. J., & Walnau, L. B. (2007). *Statistics for behavioral sciences*. Belmont, CA: Thomson Learning, Inc.
- Hedges, L. V. (1992). Meta-analysis. *Journal of Educational and Behavioral Statistics*, 17(4), 279-296.
- Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, 88(2), 359-369.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Herbison, P., Hay-Smith, J., & Gillespie, W. J. (2006). Adjustment of meta-analyses on the basis of quality scores should be abandoned. *Journal of Clinical Epidemiology*, 59, 1249-1256.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557-560.
- Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, 62(2), 227.
- Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analyses: Q statistic or I<sup>2</sup> index? *Psychological Methods*, 11(2), 193-206.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. NY: The Russell Sage Foundation.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. Random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8(4), 275-292.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2 ed.). California: Sage Publications.
- Jüni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Medical Journal*, 323, 42-46.
- Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, 282, 1054-1060.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.
- Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, 61(2), 213-218.

- Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). Evidence based medicine: The case of the misleading funnel plot. *BMJ: British Medical Journal*, 333(7568), 597.
- Lewis, S., & Clarke, M. (2001). Forest plots: trying to see the wood and the trees. *British Medical Journal*, 322(7300), 1479.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. California: Sage Publications.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford: Oxford University Press.
- Lundahl, B., & Yaffe, J. (2007). Use of meta-analysis in social work and allied disciplines. *Journal of Social Service Research*, 33(3), 1-11.
- Marin-Martinez, F., & Sanchez-Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *The Spanish Journal of Psychology*, 2(1), 32-38.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D., & The QUOROM group. (1999). Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. *Lancet*, 354(9193), 1896-1900.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151(4), 264-269.
- Mullen, B., Muellerleile, P., & Bryant, B. (2001). Cumulative meta-analysis: a consideration of indicators of sufficiency and stability. *Personality and Social Psychology Bulletin*, 27(11), 1450.
- Mulrow, C. D. (1994). Systematic reviews: Rationale for systematic reviews. *British Medical Journal*, 309, 597-599.
- National Research Council. (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy Press.
- Normand, S. L. T. (1999). Tutorial in biostatistics meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3), 321-359.
- O'Rourke, K. (2007). An historical perspective on meta-analysis: Dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, 100(12), 579-582.
- Oakley, A. (2002). Social science and evidence-based everything: The case of education. *Educational Review*, 54(3), 277-286.
- Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations and limitations. *Contemporary Educational Psychology*, 25(3), 241-286.
- Orwin, R. G., & Vevea, J. L. (2009). Evaluating Coding Decisions. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3(3), 354-379.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243-1246.
- Petticrew, M. (2003). Why certain systematic reviews reach uncertain conclusions. *British Medical Journal*, 326(7392), 756-758.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden: Blackwell Publishing.
- Pigott, T. D. (2012). *Advances in meta-analysis*. NY: Springer Verlag.
- Rendina-Gobioff, G. (2006). *Detecting publication bias in random-effects meta-analysis: An empirical comparison of statistical methods* Unpublished doctoral dissertation. University of South Florida. Florida.

- Rosenthal, R. (1979). The 'file drawer' problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. (Vol. 6). CA: Sage Publication.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52(1), 59-82.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 3, 377-386.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400-406.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex, England: John Wiley & Sons.
- Sánchez-Meca, J., & Marín-Martínez, F. (1998). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical and Statistical Psychology*, 51(2), 311-326.
- Sánchez-Meca, J., & Marín-Martínez, F. (2010a). Meta-analysis in psychological research. *International Journal of Psychological Research*, 3(1), 150-162.
- Sánchez-Meca, J., & Marín-Martínez, F. (2010b). Meta Analysis. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (Vol. 7, pp. 274-282). Oxford: Elsevier.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173-1181.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2), 115-129.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62(5), 529-540.
- Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed- versus random effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62, 97-128.
- Schulze, R. (2007). The state and the art of meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology*, 215(2), 87-89.
- Shapiro, S. (1994). Meta-analysis/Shmeta-analysis. *American Journal of Epidemiology*, 140(9), 771-778.
- Shelby, L. B., & Vaske, J. J. (2008). Understanding meta-analysis: A review of the methodological literature. *Leisure Sciences*, 30(2), 96-110.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752-760.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
- Song, F., Khan, K. S., Dinnes, J., & Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology*, 31(1), 88.
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046-1055.
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex, England: John Wiley & Sons, Ltd.
- Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, 4(2), 127-141.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., . . . Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *The Journal of the American Medical Association*, 283(15), 2008-2012.

- Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435-452). New York: Russell Sage Foundation.
- Tang, J. L., & Liu, J. L. Y. (2000). Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, 53(5), 477-484.
- Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, 58(9), 894-901.
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207-216.
- Torgerson, C. (2003). *Systematic reviews*. London: Continuum International Publishing Group.
- Tweedie, R. L., Smelser, N. J., & Baltés, P. B. (2004). Meta-analysis: Overview *International Encyclopedia of the Social & Behavioral Sciences*. (pp. 9717-9724): Elsevier Science Ltd.
- Üstün, U. (2012). *To what extent is problem-based learning effective as compared to traditional teaching in science education? A meta-analysis study*. Unpublished doctoral dissertation. METU. Ankara.
- Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. *Educational and Psychological Measurement*, 61(2), 219-224.
- Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.
- Wallace, B. C., Schmid, C. H., Lau, J., & Trikalinos, T. A. (2009). Meta-Analyst: Software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology*, 9(80). doi: 10.1186/1471-2288-9-80
- Wells, K., & Littell, J. H. (2009). Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice*, 19(1), 52-62.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. California: Sage Publications Inc.
- Yeh, J., & D'Amico, F. (2004). Forest plots: data summaries at a glance. *The Journal of Family Practice*, 53, 1007.