



Güvenirlilik Genelleme Meta-Analizlerine Bir Rehber: REGEMA Kontrol Listesi ve Akış Şemasının Uyarılma Çalışması

Özen Yıldırım¹Endam Düzyol-Türk²

MAKALE BİLGİLERİ

DOI: 10.29299/kefad.1658099

Yükleme: 14.03.2025

Düzeltilme: 11.11.2025

Kabul: 27.11.2025

Anahtar Kelimeler:

Güvenirlilik,
Güvenirlilik Genellemesi,
Meta-Analiz,
REGEMA

ÖZ

Güvenirlilik genelleme [Reliability Generalization (RG)], bir testin farklı örneklemelere uygulandığında test puanlarına dayalı güvenirlilik tahminlerinin nasıl değiştiğini araştırmayı amaçlayan özel bir meta-analiz türüdür. Meta-analiz raporlamasına yardımcı olmak amacıyla PRISMA, AMSTAR, MOOSE, MARS gibi çeşitli raporlama yönergeleri müdahale etkileri veya değişkenler arasındaki ilişkileri inceleyen meta-analizler için geliştirilmiş olmasına rağmen güvenirlilik genelleme meta-analizleri için de sıklıkla raporlamada tercih edilmektedir. Güvenirlilik Genelleme Meta-Analizi [Reliability Generalization Meta-Analysis (REGEMA)] kontrol listesi ve akış şeması, güvenirlilik genelleme meta-analizlerinin doğru bir şekilde geliştirilmesi ve raporlanması için tasarlanmış bir araçtır. Güvenirlilik genelleme meta-analizlerinin nasıl raporlanacağına ilişkin özel olarak odaklanan bir kontrol listesi sadece REGEMA'ya rastlanmıştır. Bu bağlamda temel araştırma türünde oluşturulan çalışmada REGEMA Kontrol Listesi ve Akış Şeması'nın Türkiye'deki araştırmacılara sunulması ve içeriğinin tartışılması amaçlanmıştır. Çalışmada REGEMA Kontrol Listesi'ne ve açıklamalarına yer verilmiş olup, Akış Şeması sunulup örnekler ile açıklanmıştır. Bu çalışma aracılığıyla Türkiye'deki araştırmacıların REGEMA'ya kolay bir şekilde erişerek çalışmalarında yararlanabilecekleri düşünülmektedir. REGEMA, meta-analiz uygulayan araştırmacılar tarafından süreci takip etmeleri ve raporlamalarında oluşacak herhangi bir eksikliğin önüne geçmeleri için kullanılabilir.

1. Giriş

Günümüzde bilgiye erişim, geçmişe göre pratik ve kolay bir şekilde gerçekleşmesine rağmen doğru bilgiye erişimin sağlanamaması insan ve toplum üzerinde birçok etkiye sebep olmaktadır. Bu nedenle toplum için doğru bilgi önemli bir gereksinimdir. Nicel araştırmalarda çalışılan örneklemelerde kontrollü ve nesnel bilgiye ulaşması, verilerin geçerli ve güvenilir bir şekilde değerlendirilmesi temel alınır. Glass (1976), bilgiye kolay bir şekilde erişilebildiğini ve aynı konuda yapılan bağımsız çalışmalarda farklı bulgulara ulaşılmasının bilginin güvenirliliğine ilişkin şüphe yarattığını ve tekrar gözden geçirme ihtiyacı oluşturduğunu belirtmiştir. Bu ihtiyaçların karşılanması için "meta-analiz" ile araştırma sonuçlarının tek bir araştırma çatısı altında birleştirilmesini önermiştir.

1.1. Meta-Analiz

Meta-analiz, bir araştırma konusu ile ilgili önceden yapılan araştırmaların bulgularını sistematik olarak toplayıp tekrar analizini yapan bir yaklaşımdır (Paul ve Criado, 2020). Greenland (2008) benzer bir şekilde meta-analizi,

birbirinden bağımsız olarak belirli bir konuda gerçekleştirilmiş birden çok araştırmaların sonuçlarını bir araya getirme ve bulgularının istatistiksel analizini yapma yöntemi olarak belirtmiştir. Meta-analiz, belirli bir konuda gerçekleştirilen nicel araştırmalarda belirlenen sayısal verilerin istatistiksel olarak bütünleştirilerek ilgili konuya bütüncül bir bakış, yeni bir model veya kuramsal yapının oluşmasına yol açmaktadır (McMillan ve Schumacher, 2001). Çeşitli çalışmalarda belirlenen bilgilerin nicel sentezi olarak belirtilen meta-analiz (Lau vd., 1997), farklı çalışmalardan gelen bilgilerin bütünleştirilmesini sağlayarak araştırma bulgusunun güvenirliliğini değerlendirmede önemlidir (Ioannidis, 2003). Meta-analiz, aynı türdeki araştırma bulgularının istatistiksel olarak incelenerek, örneklem büyüklüğünü yükselterek anlamlılığı güçlendirmek, etki büyüklüğünü kestirmek ve çalışmaya başlarken tahmin edilemeyen sorulara cevap bulmak, sonuçlar birbiri ile uyumlu olmadığı zaman belirsizliğe ilişkin karar vermek için önemli bir yöntemdir (Whitehead, 2002). Meta-analitik yöntemler çoğunlukla sosyal ve sağlık bilimleri ile ilgili araştırmacılara ve uygulayıcılara bilgi vermek amacıyla kullanılmaktadır

(Hunter ve Schmidt, 2004; Wiernik ve Dahlke, 2020). Çalışmaların çoğunlukla sınırlı örneklem grupları ile yapılması sebebiyle tek başına güvenilir bulgular etmek kolay değildir. Meta-analiz sayesinde farklı araştırmalarda saptanan bulgular bir araya getirilerek istatistiksel gücü artar ve güçlü kanıtlar elde edilir. Çeşitli araştırmaların sonuçlarının sentezlenmesi ile eğitimciler, akademisyenler, politika yapımcılar ve toplum için genellenebilir ve güçlü veriler sunar (Borenstein vd., 2009). Bu çalışmanın amacı Sánchez-Meca vd. (2021) tarafından tasarlanan Güvenirlik Genelmesi Meta-Analizi [Reliability Generalization Meta-Analysis (REGEMA)] Kontrol Listesi ve Akış Şeması'nın Türkiye'deki araştırmacılara sunulması ve içeriğinin tartışılması olduğu için REGEMA adımlarını anlamlandırabilmek amacıyla öncesinde meta-analize adımlarına dayalı bilgilere kısaca yer verilmiştir.

1.2. Meta-Analizin Adımları

Bir meta-analizin gerçekleşmesi için birbiri ardına izlenmesi ve uyulması gereken adımlar vardır. Literatürde meta-analizin adımlarına yönelik çeşitli kaynaklar bulunmaktadır (Borenstein vd., 2009, 2010; Hansen vd., 2022; Hoon, 2013; Page, 2021; Tuttle vd., 2009). İlgili kaynaklardan yararlanılarak meta-analizin adımları aşağıdaki şekilde özetlenmiştir:

1. Adım: Araştırma Sorusunun Tanımlanması

Meta-analiz çalışmalarında ilk adım araştırma sorusunun belirlenmesidir. Araştırma sorusu etkileri analiz edilecek müdahalelerin türünü veya dikkate alınacak yapıların alanını belirler (Hansen vd., 2022). Örneğin "Açık alanda gerçekleştirilen etkinliklerin 60-72 aylık çocukların bilimsel süreç becerileri üzerindeki genel etki büyüklüğü nedir?" şeklinde bir soru hazırlanabilir.

2. Adım: Literatür Taraması

Web of Science, ScienceDirect, Scopus, ERIC, EBSCO, PsychInfo, ProQuest, Pubmed/Medline, Google Akademik, Ulakbim Keşif ve YÖK Ulusal Tez Merkezi gibi bazı çevrim içi veri tabanı ve arama motorları kullanılarak uygun anahtar kelimeler ile veya Boolean mantığından yararlanarak literatür taraması yapılabilir (Tuttle vd., 2009). Boolean mantığında, "VE", "VEYA" ve "DEĞİL" bağlaçları kullanılarak elde edilen bulguların genişlemesini veya daralmasını sağlar (Basu, 2017). Literatür taraması sürecinde meta-analiz çalışması sistematik, tekrarlanabilir ve şeffaf olmalıdır. Konu ile ilgili bütün araştırmaları kapsayan bir örneklemle sonuçlanmalıdır (Hansen vd., 2022). Örneğin küçük çocuklarda bilimsel süreç becerilerine yönelik ölçüklerin güvenilirlik genellemesine incelemek isteyen bir çalışmada ("okul öncesi" VEYA "erken çocukluk" VEYA "anaokulu") VE ("bilimsel süreç becerileri") VE ("cronbach alpha") kodu ile tarama yapılabilir.

3. Adım: Kodlama, Etki Büyüklüğü Ölçüsünün ve Analitik Yöntemin Seçimi

Kodlama sürecinde kodlama sayfasının tasarımı ilk adımdır. Kodlama sayfasının tasarımı kullanılan yöntemlere, yazılıma ve araştırma tasarımına bağlı olduğu için evrensel bir şablon yoktur (Hansen vd., 2022). Veri setini kodlamadan önce, amaçlanan araştırma modelini ve ilgili hedef olmayan değişkenleri dikkate almak çoğunlukla önemlidir. İlgili moderatör veya kontrol değişkenlerini dahil etme, bu değişkenlerin araştırılan etkiyi nasıl etkileyebileceği hakkında teorik gerekçelere dayanmalıdır (Berneth ve Aguinis, 2016). Meta-analizler yürütürken uzun süredir tartışılan bir konu, tek bir birincil çalışma içinde aynı yapı için yalnızca bir veya tüm mevcut etki boyutlarının kullanılıp kullanılmayacağıdır. Hansen vd. (2022) tüm mevcut etki büyüklüğü gözlemlerini kullanmanın avantajlarından yararlanılmasını fakat çok düzeyli modeller, panel regresyon modelleri veya sağlam varyans tahmini (robust variance estimation) gibi uygun yöntemler kullanarak bağımlılıklarının dikkatlice değerlendirilmesini belirtmiştir.

Elde edilecek olan istatistiksel sonuçlardan hesaplanan standart, ortak bir değer olan etki büyüklüğü, meta-analiz kodlamasında kullanılır. Standartlaştırılmış ortalama farkı (Cohen d, Hedges g ve Glass delta), korelasyon katsayısı (Pearson korelasyon katsayısı) ve ikili veriye dayalı (olasılık oranı, risk oranı ve risk farkı) değerleri meta-analiz çalışmalarında kullanılan etki büyüklüğü değerleridir (Page, 2021). Meta-analizde birincil etki büyüklüğü ölçüsünü tanımladıktan sonra, daha sonraki kodlama sürecinde seçilen birincil etki büyüklüğünden farklı etki büyüklüklerinde raporlanan çalışma bulgularını dönüştürmek gerekebilir (Hansen vd., 2022). Meta-analiz yapmak için yaygın olarak kullanılan sabit etki modeli ve rastgele etkiler modeli üzere iki istatistiksel model vardır (Borenstein vd., 2009).

Doğrudan meta-analizin araştırma sorusu ile bağlantılı bir şekilde hangi meta-analitik yöntemin ve yazılımın kullanılacağı seçilir. Bu araştırma soruları, yapılar arasındaki bir ilişkiye, bir müdahalenin etkisine veya etkileri düzenlemeye veya aracılık etmeye odaklanabilir. Geleneksel tek değişkenli meta-analiz, meta-regresyon, meta-analitik yapısal eşitlik modellemesi ve nitel meta-analiz uygulanabilir (Hoon, 2013).

4. Adım: Analiz

Birincil analizi yürütmeden önce bulguların sağlamlığını garanti altına alması gereken bazı ön duyarlılık analizleri gerekli olabilir (Rudolph vd., 2020). Etkili aykırı değer gözlemleri, özellikle toplam etki büyüklüğü sayısı küçükse, gözlenen sonuçları potansiyel olarak yanlı hale getirebilir. Meta-analitik veri kümelerindeki aykırı değerleri saptamak amacıyla farklı istatistiksel yöntemlerden yararlanılabilir. Orman grafiği (forest plot) ve huni grafiği (funnel plot) meta analizde aykırı değerleri belirlemede kullanılan yöntemler arasındadır (Aguinis vd., 2013). İkinci olarak

literatür taraması bağlamında belirtildiği gibi yanlılık bir sorun olabilir. Bu nedenle yanlılık çeşitli yollarla incelenebilir. Huni grafiği ve bu grafikte tercih edilen Trim ve Fill yöntemi, Begg testi veya Egger Testi yanlılık belirleme yöntemleri arasındadır (Rothstein vd., 2005).

Etkili aykırı değerler veya yanlılık kontrol edip düzelttikten sonra, meta-analizdeki bir sonraki adım, meta-analistlerin farklı varsayımlara dayanan sabit etkiler ve rastgele etkiler olmak üzere iki farklı model türü arasında karar verilmesi gereken analizdir (Borenstein vd., 2010).

5. Adım: Sonuçların Raporlanması

Meta-analizde sonuçların raporlanması son adımdır. Tüm adımlar ve metodolojik kararlar okuyucular tarafından anlaşılır olmalıdır. Meta-analizlerin raporlanmasını standardize etmek için bazı protokoller geliştirilmiştir. Bunlardan günümüzde en yaygın kullanılanı Preferred Reporting Items For Systematic Review and Meta-Analysis Protocols (PRISMA-P) olarak bilinmektedir (Moher vd., 2015).

Yukarıda yer alan meta-analiz aşamaları incelendiğinde; araştırma sorusu tanımlama, literatür taraması gerçekleştirme, kodlama, etki büyüklüğü ölçüsünün ve analitik yöntemin seçimi, analiz ve sonuçların raporlanması adımları ile bir meta-analiz çalışması gerçekleştirilebilir.

1.3. Güvenirlilik

Meta-analiz yönteminin ayırt edici yönü, araştırma sonuçlarını değerlendirirken yalnızca öznel yargılara güvenmek yerine, nicel yöntemleri kullanmasıdır. Araştırmacılar bu analiz ile araştırdıkları konuya yönelik minimum varyanslı, güvenilir ve geçerli parametreleri öngörürler (Mosteller ve Colditz, 1996). Bir ölçeğin standardize olabilmesi için birçok standardın sağlanmasıyla birlikte "güvenirlilik ve geçerlik" olmazsa olmaz iki şarttır. . Güvenirlilik, aynı şartlarda bir ölçme aracının tekrarlanan ölçümlerdeki ölçüm değerlerinin kararlılığını; geçerlik ise bir ölçme aracının ölçmeyi hedeflediği özelliği, başka bir özellik ile karıştırmadan ölçmesidir (Carmines ve Zeller, 1979; Crocker ve Algina, 1986). Ölçme aracının tekrarlanan ölçümlerde tutarlılığı güvenirliliği temsil eder (Ho, 2014). Güvenirlilik, geçerliliği de etkilemekte ve ölçme araçlarındaki puanların doğru yorumlanmasında önemlidir. Aynı ölçme aracı ile gerçekleştirilen farklı araştırmalarda güvenirlilik sonuçları farklılık gösterebilir. Bu nedenle güvenirlilik tahminleri her araştırma için tekrar kestirilmelidir. Güvenirlilik belirlemede birden çok uygulama ve tek uygulamaya dayalı farklı güvenirlilik kestirim yöntemleri vardır. Örneğin paralel testler yöntemi, test-tekrar-test yöntemi birden çok uygulamaya dayalı yöntemlerken yarıya bölme yöntemi ve Cronbach alfa yöntemi tek uygulamaya dayalı yöntemler arasındadır. Bu yöntemlerde güvenirlilik katsayısı bir korelasyon değeri olarak kestirilir. Test-tekrar test yönteminde güvenirlilik kararlılık katsayısı olarak

kestirilirken, Cronbach Alfa yönteminde iç tutarlılık katsayısı elde edilir (Mor, 2021). Örneğin, "Cronbach alfa güvenirlilik katsayısı" psikolojik özelliklerin ölçülmesinde yararlanılan ölçeklerin güvenirliliğinin saptanmasında sıklıkla kullanılmaktadır. Cronbach alfa katsayısı, ölçekte bulunan k maddenin varyansları toplamının genel varyansa oranlanması ile belirlenen ağırlıklı standart değişim ortalamasıdır. Her bir madde için tek bir α değeri tespit edilebileceği gibi, tüm maddelere ilişkin ortalama bir α değeri de bulunabilir. Belirlenen α değeri ölçeğin toplam güvenirliliğini yansıtmaktadır ve .70 ve büyük olması ölçekte maddeler arası tutarlılığın yüksek olduğunu diğer bir deyişle ölçekten elde edilen puanların güvenilir olduğunu göstermektedir (Cronbach, 1951).

1.4. Güvenirlilik Genellemesi

Güvenirlilik Genellemesi [Reliability generalization (RG)], meta-analiz türlerinden biridir. Vacha Haase (1998) bir ölçeğin farklı çalışmalarda güvenirlilik değerlerini birleştirerek genellemesi için meta-analizi önermiştir. Güvenirlilik genellemesi, bir testin farklı örneklemelere uygulandığında test puanlarına dayalı güvenirlilik tahminlerinin nasıl değiştiğini araştırmayı amaçlayan özel bir meta-analiz türüdür. RG güvenirlilik katsayılarının farklı araştırmalarda sistematik olarak analiz edilmesini sağlayan önemli bir metottur (Thompson, 2003). Bir RG meta-analizinin temel amacı, araştırmacıları ve uygulayıcıları bir testten/ölçekten elde edilen puanların beklenen güvenirliliğine yönelik bilgilendirmek, test formatının ve uygulama şartlarının hangilerinin güvenirlilik tahminlerini etkileyebileceğini göstermektir (Vacha Haase, 1998). Meta-analiz, ölçme araçlarının genellikle ölçmek istediği özellikleri ölçmede etkili olup olmadığı ya da geçerliliğine odaklanırken, RG aracılığıyla farklı araştırmalarda ölçme aracının tutarlılık gösterme durumu incelenir. Araştırmacıların sonraki çalışmalarında da ölçek seçimini daha bilinçli yapmasını sağlar (Osborne ve Fitzpatrick, 2012).

Araştırmacılar ve uygulayıcılar arasında güvenirliliğin teste özgü bir özellik olduğu ve testin uygulandığı yerden bağımsız olarak sabit kaldığına dair yaygın bir görüş vardır. Ancak psikometrik teori, güvenirliliğin testin değişmez bir özelliği değil, test puanlarına özgü bir özellik olduğunu ve bu nedenle her uygulamada değiştiğini göstermektedir. Meta-analize dayalı güvenirlilik genellemesi, ortak etki büyüklüğü indeksi belirli bir testi uygulayan birincil çalışmalarda bildirilen güvenirlilik katsayılarının (örneğin, iç tutarlılık katsayıları, test-tekrar test katsayıları, değerlendiriciler arası korelasyonlar, vb.) bir araya getirildiği tek meta-analiz türüdür (Badenes-Ribera, 2023; Traub, 1994). RG farklı bağlamlarda ve farklı örneklem gruplarında daha önceden güvenirliliği belirlenen araştırmalarda elde edilen güvenirlilik tahminlerinin istatistiksel bir şekilde birleştirilmesi süreci olarak da belirtilmektedir (Badenes Ribera vd., 2023). Vacha Haase (1998) güvenirlilik genellemesi meta-analizinin amaçlarını;

- testin farklı örneklem ve bağlamlarda birçok kez uygulanması yoluyla test puanlarının ortalama güvenilirliğini tahmin etmek,
- güvenirliğin bir test uygulamasından diğerine genellenebilirliğini değerlendirmek,
- güvenirlik test uygulamaları arasında değişiyorsa, güvenilirlik katsayılarıyla istatistiksel olarak ilişkili olabilecek potansiyel çalışma ve örneklem özelliklerini belirlemek şeklinde ifade etmiştir.

Meta-analiz çalışmalarının raporlamasına yardımcı olmak ve yol göstermesi amacıyla alanda PRISMA (Moher vd., 2015), AMSTAR (Shea vd., 2009), MOOSE (Stroup vd., 2000), MARS (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) gibi çeşitli raporlama yönergeleri bulunmaktadır. PRISMA sistematik derlemeler ve meta-analizlerin planlanması, yürütülmesi ve raporlanması için standart bir raporlama çerçevesi sunarken (Page vd., 2021) meta analizle en sık kullanılan kontrol listesi PRISMA 2020 Türkçe (Hür, 2021) ve diğer 15 dile uyarlaması yapılarak literatüre kazandırılmıştır. AMSTAR sistematik derlemelerin metodolojik kalitesini, çalışmaların tarafsızlığını, geçerliğini değerlendirmek için kullanılan bir kontrol listesi olup (Shea vd., 2007), MOOSE özellikle hasta kayıtları, vaka-kontrol çalışmaları ve kohort çalışmalarının analizinde kullanılan bir dizi öneri ve kontrol listesi (Stroup vd., 2000), MARS ise Amerikan Psikoloji Derneği (APA) tarafından geliştirilmiş standartlar kümesi olmak ile beraber hangi istatistiksel yöntemlerin analiz nasıl raporlanması gerektiğini ifade eder (Appelbaum vd., 2018). Ancak bu yönergeler RG meta-analizlerinin raporlanması için değil, müdahale etkileri veya değişkenler arasındaki ilişkileri inceleyen meta-analizlerin raporlanması için geliştirilmiştir. Güvenirlik genelleme meta-analizlerinin kendine has özel karakteristikleri, literatürde şu ana kadar önerilen kontrol listelerinin verimli bir şekilde uygulanarak RG'e uyarlanması zorlaştırmaktadır (Sánchez-Meca vd., 2021). Güvenirlik genellemesi meta-analizlerinin nasıl raporlanacağına ilişkin özel olarak odaklanan kontrol listesi sadece REGEMA'ya (Sánchez-Meca vd., 2021) rastlanmıştır. Sánchez-Meca vd. (2021) tarafından geliştirilen Güvenirlik Genellemesi Meta-Analizi [Reliability Generalization Meta-Analysis (REGEMA)] kontrol listesi ve akış şeması, güvenilirlik genelleme meta-analizlerinin doğru bir şekilde geliştirilmesi ve raporlanması için tasarlanmış bir araçtır. Sánchez-Meca vd. (2021), araştırmacıların güvenilirlik genellemesi meta-analizinin seçim ölçütlerini karşılayan birincil araştırmaları arama ve seçme sürecini gerçekleştirmelerine destek olmak amacıyla PRISMA'da önerilene yakın fakat güvenilirlik genellemesi meta-analizlerinin özel durumlarına göre uyarlanan bir akış şeması geliştirmişlerdir. Bu çalışmada güvenilirlik genellemesi meta-analizi için kavramın uluslararası tanınırlığını ve ortak kullanımını artırmak amacıyla REGEMA kısaltması kullanılmıştır.

Türkiye'de RG kullanımı sınırlı olmak ile beraber Sánchez-Meca vd. (2021) tarafından geliştirilen REGEMA Kontrol Listesinin ve Akış Şemasının, RG çalışmalarının sistematik yürütülmesine ilişkin önemli bir rehber olduğu düşünülmektedir. Güvenirlik genellemesi çalışmaları için standartlaştırılmış bir araç olan REGEMA'nın Türkiye'deki araştırmacılar tarafından kullanılması, güvenilirlik genellemesi çalışmalarının sistematik biçimde değerlendirilmesine, ortak bir dil birliği oluşmasına katkıda bulunacaktır. Özellikle her bilim alanının kendine has bir dili ve kavramları olduğu düşünüldüğünde, uzmanlar tarafından ilgili akış şeması ve kontrol listesinin Türkçeye uyarlamasının ve alana kazandırılmasının önemli olduğu düşünülmektedir. Bu bağlamda çalışmada güvenilirlik genellemesine dayalı araştırmaların sistemli yürütülmesi ve raporlanmasında yol gösterecek REGEMA tanıtılıp tartışılmıştır. Bu sayede Türkiye'deki araştırmacıların REGEMA'ya kolay bir şekilde erişerek çalışmalarında yararlanabilecekleri düşünülmektedir. Bilimsel araştırmalarında REGEMA kullanarak güvenilirlik genellemesinin kullanılması, meta-analiz uygulayan araştırmacılar tarafından süreci takip etmeleri ve raporlamalarında oluşacak herhangi bir eksikliğin önüne geçmelerinde veya dergi editörlerinin ve hakemlerinin RG meta-analizini incelemelerinde yardımcı olabilir.

2. Yöntem

2.1. Desen

Bu çalışmada REGEMA Kontrol Listesi ve Akış Şeması'nın İngilizceden Türkçeye uyarlaması yapılmıştır. Araştırma bir uyarlama çalışması olduğundan dolayı temel araştırma türündedir. Araştırmada yeni bir bilginin standartlara uygunluğu tartışılmıştır.

2.2. İşlem

REGEMA Kontrol Listesi ve Akış Şeması ayrıntılı biçimde incelenmiştir. Çeviride meta-analize dayalı çalışmaları olan ölçme ve değerlendirme, İngilizce dil eğitimi alanında uzman dört akademisyenden destek alınmıştır. Akış şeması ve kontrol listeleri Türkçeye çevrildikten sonra bir ölçme ve değerlendirme uzmanı tekrar İngilizce ve Türkçe çeviriyi karşılaştırılmalı kontrol etmiştir. Ayrıca yazarlar tarafından akış şeması ve kontrol listelerindeki kavramlar literatüre göre gözden geçirilmiş ve anlamsal olarak en doğru çeviriyi yansıtan kavramlar seçilmeye çalışılmıştır. Bu amaçla makalelerinde REGEMA'ya yer veren uluslararası literatürdeki farklı çalışmalar incelenmiş ve tanımlamalarına bakılmıştır. Taslak Türkçe form, ölçme ve değerlendirme alanında deneyimli üç uzmana tekrar gönderilerek kavramsal ve dilsel uygunluk, anlaşılabilirlik, terminolojik doğruluk gibi açılardan dönüt alınmıştır.

Uzman görüşleri doğrultusunda metin revize edilerek nihai hali oluşturulmuştur.

2.3. Veri Toplama Araçları

2.3.1. REGEMA kontrol listesi ve akış şeması

Sánchez-Meca vd. (2021) tarafından REGEMA Kontrol listesi ve akış şeması geliştirilmiştir. REGEMA Kontrol Listesi güvenilirlik genellemesi meta-analizi çalışmalarını standartlaştırmak amacıyla geliştirilmiştir. Toplam 30 maddeden oluşan bu liste, araştırmacıların hangi bilgileri, hangi aşamalarda ve nasıl raporlamaları gerektiğini adım adım açıklamaktadır. REGEMA Akış Şeması ise, bir RG meta-analizinde takip edilmesi gereken temel adımları araştırmacılara görsel olarak sunmaktadır.

2.4. Veri Analizi

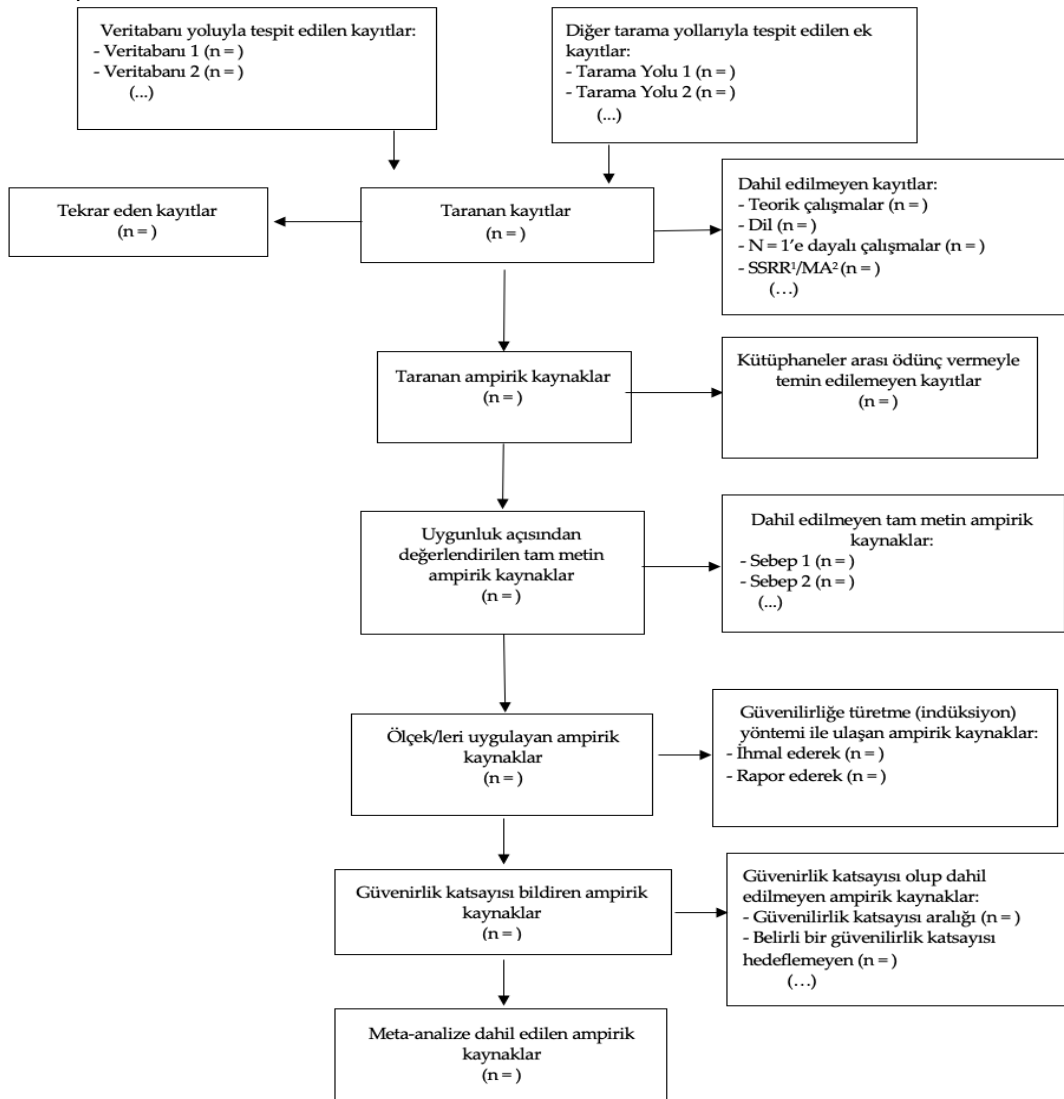
Çeviri, uzman görüşü ve geri çeviri aşamalarında elde edilen dönütler sistematik olarak incelenerek nihai hali oluşturulmuştur.

3. Bulgular

Aşağıda RG araştırma sürecinde verilerin toplanmasında izlenecek yol hakkında bilgi veren REGEMA akış şeması ve RG analizlerinin raporlanmasında dikkat edilecek ölçütleri içeren REGEMA kontrol listesi açıklanmıştır.

3.1. REGEMA Akış Şeması

REGEMA akış şeması literatür taraması, tarama süreci ve incelenecek araştırmaların seçim sürecini göstermek amacıyla tasarlanmıştır. Şekil 1'de Sánchez-Meca vd. (2021) tarafından oluşturulan REGEMA Akış Şeması'na yer verilmiştir.



Not. ¹Sistemantik İncelemeler. ²Meta-analiz

Şekil 1. REGEMA akış şeması 2019

REGEMA Akış Şemasında yer alması gereken adımların açıklamaları ve bu adımlara ilişkin örnekler aşağıda sunulmuştur.

Veritabanı yoluyla tespit edilen kayıtlar: Belirlenen konu doğrultusunda literatür incelenir. Belirlenen anahtar kelimeler ile belirlenen veri tabanlarında aramalar gerçekleştirilir. WoS, Scopus, ScienceDirect, Eric gibi veri tabanları burada yer almaktadır. Anahtar kelimeler kullanılırken "AND" veya "OR" kullanılabilir. Örneğin WoS'ta gelişmiş aramada başlık bölümünde TI=("preschool" OR "pre-school" OR "early child" OR "early childhood" OR "kindergarten") AND ("scientific process skills") AND ("cronbach alpha") kodu ile tarama yapılabilir.

Diğer tarama yollarıyla tespit edilen ek kayıtlar: Belirlenen anahtar kelimeler ile diğer literatür tarama yolları aracılığıyla aramalar gerçekleştirilir. Kayıtlara ulaşılabilecek diğer tarama yollarına örnek; ResearchGate, OpenGrey, Google Scholar, Yazarlarla İletişim vb. verilebilir.

Taranan kayıtlar: Veri tabanı yoluyla ve diğer kaynaklar aracılığıyla tespit edilen ek kayıtların toplamı taranan kayıtlar bölümüne yazılır. Bu ilk üç adımda dikkat edilmesi gereken en önemli kriterlerden biri kaynaklara göre elde edilen kayıtların sayılarının ve araştırıldığı tarihlerin not edilerek kaydedilmesidir.

Tekrar eden kayıtlar: Diğer kaynaklardan ve veri tabanlarından elde edilen kayıtlar bir araya getirilerek yinelenen çalışmalar tespit edilir. Bu çalışmalardan sadece bir tanesi araştırmaya dahil edilir.

Dahil edilmeyen kayıtlar: Kayıtlar incelenirken belirli dışarda tutma kriterleri dikkate alınabilir. Örneğin, teorik çalışmalar, belirli bir dil kriterini sağlamayan (sadece İngilizce) çalışmalar, belirli bir örneklem sayısının altındaki çalışmalar, sistematik derlemeler, meta-analiz çalışmaları araştırmanın amacına göre dışarda tutma kriterleri içerisinde ele alınabilir. Araştırmacı araştırmaya dahil etmeyeceği kayıtları hangi kriterlere göre ve ne kadarını dışarda tuttuğunu not almalıdır.

Taranan ampirik kaynaklar: Tekrar eden ve dahil edilmeyen kayıtların toplamı taranan kayıtlardan çıkarılarak kalan sayı bu bölüme yazılır.

Kütüphaneler arası ödünç vermeyle temin edilemeyen kayıtlar: Araştırmacının kayıtlar arasında yer almasına rağmen kütüphaneler aracılığı ile elde edeceği ancak kütüphanenin paylaşmadığı kayıtların sayısı bu bölüme yazılır.

Uygunluk açısından değerlendirilen tam metin ampirik kaynaklar: Taranan ampirik kaynaklardan kütüphaneler arası ödünç vermeyle temin edilemeyen kayıtlar çıkarılarak kalan sayı bu bölüme yazılır.

Dahil edilmeyen tam metin ampirik kaynaklar: Ölçek/leri kullanmayan çalışmaların sayıları sebepleri açıklanarak bu

bölüme yazılır. Örneğin analizlerin yürütülmesi için yeterli veriye yer verilmeyen kaynaklar.

Ölçekleri uygulayan ampirik kaynaklar: Uygunluk açısından değerlendirilen tam metin ampirik kaynaklardan, dahil edilmeyen tam metin ampirik kaynaklar çıkarılarak kalan sayı bu bölüme yazılır.

Güvenilirliğe türetme (indüksiyon) yöntemi ile ulaşılan ampirik kaynaklar: Ölçek/leri kullanan çalışmalar detaylı bir şekilde incelenir. Bu çalışmalarda güvenilirliği kendi örnekleme üzerinden kestirmeyen araştırmalar ayırt edilir. Araştırmacılar, başka bir araştırmada ölçeğin güvenilirliği önceden kestirildiği için kendi araştırmasında ölçekten elde edilen puanların zaten güvenilir olduğunu düşünüp güvenilirlik katsayısını tekrar hesaplamayabilir ya da aynı ölçek için farklı bir araştırmacının hesapladığı bir güvenilirlik katsayısını kendi araştırmasının güvenilirlik değerimiş gibi raporlayabilir. Bu durum, güvenilirlik türetme (indüksiyonu) olarak adlandırılır. RG meta-analizlerine özgü özel bir yayın yanlılığı türüdür. Güvenirliğin türetilmesi, şüpheli ölçüm uygulamalarından biri olarak kabul edilebilir; bunlar, 'araştırmacıların ölçümlerin geçerliği hakkında şüpheler uyandıran ve nihayetinde çalışma sonuçlarının geçerliğini sorgulayan kararlar' olarak tanımlanır (Sánchez-Meca vd., 2021). Güvenirliğin iki tür türetilmesi söz konusudur. 'Raporlama yoluyla' güvenilirlik türetme, bir çalışmanın başka bir çalışmadan elde edilen test puanlarının güvenilirliğini raporlaması; 'ihmal yoluyla' güvenilirlik türetme, bir veya birden fazla psikometrik aracın kullanıldığı bir çalışmada güvenilirliğin hiç bahsedilmemesi durumudur. İhmal ve raporlama yöntemlerine göre güvenilirliği türeten kayıtların sayısı rapor edilmelidir. Hatta hangi tür çalışmalarda bu türetmelerin yapıldığının detaylı incelenerek kaydedilmesi gerekebilir.

Güvenirlik katsayısı bildiren ampirik kaynaklar: Ölçek/leri kullanan ampirik kaynaklardan, güvenilirliği türeten yaklaşımla ulaşılan kaynakların çıkarılması durumunda elde edilen sayı bu bölüme yazılır.

Güvenirlik katsayısı olup dahil edilmeyen ampirik kaynaklar: Bu bölümde araştırmanın amacına dayalı olarak güvenilirlik katsayıları incelenir. Örneğin sadece Cronbach Alfa katsayısını kullanacak bir araştırmacı için test-tekrar-test, puanlayıcılar arası güvenilirlik, yarıya bölme gibi güvenilirlik kestirim yöntemlerinden elde edilen katsayılar araştırmanın dışında tutulabilir. Araştırmalarda ilgili katsayı bir aralık değeri olarak verildiğinde veya belirli bir katsayı hedefi belirtilmemişse dahil edilmeyebilir.

Meta-analize dahil edilen ampirik kaynaklar: "Bazı güvenilirlik katsayısı bildiren ampirik kaynaklardan", "güvenirlik katsayısı olup dahil edilmeyen ampirik kaynaklar" sayısı çıkarılarak kalan sayı bu bölüme yazılır. Bu bölümde kalan çalışmalar meta-analize dahil edilir. Eldeki kaynaklarda güvenilirlik katsayısı birden fazla kez hesaplanması durumunda (Örn. farklı örneklemlerden

hesaplama, farklı uygulamalara dayalı (ön test- son test, boylamsal uygulama) hesaplama vb.). Her bir katsayı ayrı bir çalışma gibi değerlendirilmelidir. Örneğin deneysel modele dayalı bir çalışmada aynı örnekleme iki farklı zamanda testin uygulanmasından hesaplanan iki katsayı farklı çalışmalar olarak analize dahil edilmelidir. Bu durum veri ayıklamada dikkate alınmalı ve bu kısımda belirtilmelidir.

3.2. REGEMA Kontrol Listesi

RG araştırması yaparken araştırmacının raporlama sürecinde dikkat etmesi gereken kriterlere Sánchez-Meca vd. (2021) REGEMA Kontrol Listesi'nde yer vermiştir. Sánchez-Meca vd. (2021) REGEMA kontrol listesinin oluşturulması aşamasında detaylı bir literatür taraması yapmışlar ve literatürde meta-analiz raporlamada kullanılan PRISMA, MOOSE, AMSTAR, AMSTAR 2, MARS gibi kontrol listelerine ve 80 metodolojik yayına başvurmuşlardır. REGEMA kontrol listesi sekiz boyutta yapılandırılmıştır: Başlık (bir madde), Özet (bir madde), Giriş (iki madde), Yöntem (14 madde), Sonuçlar (altı madde), Tartışma (dört madde), Finansman (bir madde), ve Protokol (bir madde) olmak üzere sekiz boyut altında toplam 30 madde yer almaktadır. Her maddede, o maddenin amacını açıklayan "Evet", "Hayır", "Belirsiz" ve "Uygun Değil" gibi seçenekler sunulur. Ek materyalde REGEMA kontrol listesindeki bölümler, açıklamalar ve kontrol seçenekleri bulunmaktadır.

4. Tartışma

Bu çalışmada Sánchez-Meca vd. (2021) tarafından tasarlanan REGEMA Kontrol Listesi ve Akış Şeması'nın Türkiye'deki araştırmacılara sunulması ve içeriğinin tartışılması amaçlanmıştır. Çalışmada REGEMA Kontrol Listesi'ne ve açıklamalarına yer verilmiş olup, Akış Şeması örnekler ile sunulmuştur. Meta-analizler yararlılığı raporlama kalitesine bağlı olup, yetersiz raporlanan bir meta-analiz gelecek araştırmalarda sınırlı düzeyde yarar sağlar. Bu sınırlılıktan RG meta-analizleri de etkilenmektedir (Nosek vd., 2015). Sánchez-Meca vd. (2019) tarafından yapılan sistematik derleme çalışmasında psikolojik testler üzerine yapılmış 150 RG meta-analizin raporlama uygulamaları analiz edilmiştir. Çalışmada RG meta-analizin yürütülmesindeki temel metodolojik yönlerde önemli raporlama eksikliklerinin bulunduğu sonucuna ulaşılmıştır. Sánchez-Meca vd. (2021) zayıf bir şekilde raporlanan bir meta-analizin tekrarlanabilirliği engellediğini ve Açık Bilim Çerçevesi tarafından teşvik edilen şeffaflık ilkesine aykırı olduğunu belirtmişlerdir. Buradan yola çıkarak araştırmacılar REGEMA Kontrol Listesi ve Akış Şeması'nı geliştirerek bu sınırlılığa alternatif çözüm olacak bir çalışma gerçekleştirmişlerdir. Bu nedenle bu çalışmada REGEMA'nın araştırmacılara sunulması ve içeriğinin tartışılmasına karar verilmiştir.

Bir ölçeğin güvenilirlik katsayısının raporlanmasının, farklı çalışmalar için diğer araştırmacıların o ölçekten yararlanılıp yararlanamayacağı konusunda önemli olduğu

belirtilmektedir. Herhangi bir ölçeği araştırmalarında uygulayan araştırmacıların güvenilirlik katsayısını belirlemeleri durumunda, o ölçeği kullanan çalışmaların güvenilirlik katsayılarında bütünleştirme gerçekleştirilerek güvenilirlik genellemesi yapılmaktadır. Böylece bu ölçekler arasında karşılaştırma yaparak araştırmacılar bilgi sahibi olabilmektedirler (Deng vd., 2019). Güvenirlik genellemesi yaparken ölçekten yararlanan bütün çalışmalara ulaşmak önemlidir. Bu nedenle bir ölçeği kullanan araştırmacıların güvenilirlik katsayılarını hesaplamalarının önemli olduğu belirtilmektedir (Yörük ve Sen, 2023). REGEMA'da da güvenilirlik katsayılarının belirlenmesi üzerinden bir raporlama süreci bulunmaktadır. Kapsamlı bir kontrol listesi ve akış şeması olması nedeniyle REGEMA'daki süreçler dikkatli bir şekilde takip edilerek araştırma yapıldığında güvenilir sonuçların elde edileceği düşünülmektedir.

Etik Bildirim: Yapılan bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması belirtilen tüm kurallara uyulmuş; yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir. REGEMA Kontrol Listesi ve Akış Şeması'nı geliştiren kişi (Julio Sánchez-Meca) ile iletişime geçilip araştırma ile ilgili bilgi verilerek kullanımına ilişkin gerekli izin alınmıştır. Bu çalışmanın uyalama niteliğinde olması ve insan veya hayvan katılımcılar üzerinde deneysel uygulama ya da veri toplama süreci içermemesi sebebiyle etik kurul onayı alınmamıştır.

Yazar Katkıları: İki yazar eşit katkı sağlamıştır.

Finansman: Bu çalışma için herhangi bir finansal destek alınmamıştır.

Çıkar Çatışması: Çıkar çatışması bulunmamaktadır.

Veri Erişilebilirliği: Uygulamaya dayalı veri toplanmamıştır.

Kaynakça

- Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational research methods*, 16(2), 270-301. <https://doi.org/10.1177/1094428112470848>
- American Psychological Association. 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. <https://doi.org/10.1037/0003-066X.63.9.839>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25.
- Badenes-Ribera, L., Duro-García, C., López-Ibáñez, C., Martí-Vilar, M., & Sánchez-Meca, J. (2023). The adult prosocialness behavior scale: A reliability generalization meta-analysis. *International Journal of Behavioral Development*, 47(1), 59-71. <https://doi.org/10.1177/01650254221128280>
- Basu, A. (2017). How to conduct meta-analysis: A basic tutorial. *PeerJ Preprints*, 5, e2978v1. <https://doi.org/10.7287/peerj.preprints.2978v1>
- Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, 69(1), 229-283. <https://doi.org/10.1111/peps.12103>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111. <https://doi.org/10.1002/jrsm.12>
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage. <https://doi.org/10.4135/9781412985642>
- Cheung, M. W. L., & Vijayakumar, R. (2016). A guide to conducting a meta-analysis. *Neuropsychology Review*, 26, 121-128. <https://doi.org/10.1007/s11065-016-9319-z>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Deng, J., Wang, M. C., Zhang, X., Shou, Y., Gao, Y., & Luo, J. (2019). The inventory of callous unemotional traits: A reliability generalization meta-analysis. *Psychological Assessment*, 31(6), 765. <https://doi.org/10.1037/pas0000698>
- Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly*, 68, 103-106. <https://doi.org/10.1007/s11301-018-0142-x>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8. <https://doi.org/10.3102/0013189X005010003>
- Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins. <https://doi.org/10.1016/j.annemergmed.2008.06.461>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181-217. <https://doi.org/10.1002/jrsm.1378>
- Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: a practical guide. *Management Review Quarterly*, 1-19. <https://doi.org/10.1007/s11301-021-00247-4>
- Ho, R. (2014). *Univariate and multivariate data analysis with IBM SPSS*. CRC Press.
- Hoon, C. (2013). Meta-synthesis of qualitative case studies: An approach to theory building. *Organizational Research Methods*, 16(4), 522-556. <https://doi.org/10.1177/1094428113484969>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
- Hür, G. (2021). PRISMA kontrol listesi 2020 güncellemesi. *Online Turkish Journal of Health Sciences*, 6(4), 603-605. <https://doi.org/10.26453/otjhs.1001606>
- Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *The Lancet*, 361(9357), 567-571. [https://doi.org/10.1016/S0140-6736\(03\)12516-0](https://doi.org/10.1016/S0140-6736(03)12516-0)
- Lau, J., Ioannidis, J. P., & Schmid, C. H. (1997). Quantitative synthesis in systematic reviews. *Annals of Internal Medicine*, 127(9), 820-826. <https://doi.org/10.7326/0003-4819-127-9-199711010-00008>
- McMillan, J. H. & Schumacher, S. (2001). *Research in education: A conceptual introduction*. Longman.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Prisma-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1-9. <https://doi.org/10.1186/2046-4053-4-1>
- Mor, E. (2024). Ölçme süreç ve sonuçlarının taşınması gereken temel nitelikler. In Ö. Yıldırım & K. S. Kartal (Eds.), *Eğitimde ölçme ve değerlendirme* (pp. 22-62). Lisans Yayıncılık.
- Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, 17(1), 1-23. <https://doi.org/10.1146/annurev.pu.17.050196.000245>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research, and Evaluation*, 17(15), 1-8.

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know?. *International Business Review*, 29(4), 101717. <https://doi.org/10.1016/j.ibusrev.2020.101717>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.) (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. <https://doi.org/10.1002/0470870168>
- Rudolph, C. W., Chang, C. K., Rauvola, R. S., & Zacher, H. (2020). Meta-analysis in vocational behavior: A systematic review and recommendations for best practices. *Journal of Vocational Behavior*, 118, 103397. <https://doi.org/10.1016/j.jvb.2020.103397>
- Sánchez-Meca, J., Marín-Martínez, F., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-López, J. A., & López-García, J. J. (2019, Temmuz). Reporting practices in reliability generalization meta-analyses: assessment with the REGEMA checklist. *XVI Congress of Methodology of the Social and Health Sciences*. Madrid, Spain.
- Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., ... & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, 12(4), 516-536. <https://doi.org/10.1002/jrsm.1487>
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7(1), 10.
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., ... & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013-1020. <https://doi.org/10.1016/j.jclinepi.2008.10.009>
- 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. *Jama*, 283(15), 2008-2012. <https://doi.org/10.1001/jama.283.15.2008>
- 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. *JAMA*, 283(15), 2008-2012.
- Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. Sage.
- Traub, R. E. (1994). *Reliability for the social sciences: Theory and applications*. Sage Publications.
- Tuttle, B. D., Isenburg, M. V., Schardt, C., & Powers, A. (2009). PubMed instruction for medical students: Searching for a better way. *Medical Reference Services Quarterly*, 28(3), 199-210. <https://doi.org/10.1080/02763860903069839>
- Vacha Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20. <https://doi.org/10.1177/0013164498058001002>
- Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. John Wiley & Sons. <https://doi.org/10.1002/0470854200>
- Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, 3(1), 94-123. <https://doi.org/10.1177/2515245919885611>
- Yörük, S., & Sen, S. (2023). A reliability generalization meta-analysis of the creative achievement questionnaire. *Creativity Research Journal*, 35(4), 714-729. <https://doi.org/10.1080/10400419.2022.2148073>

EK MATERYAL*REGEMA kontrol listesi*

BAŞLIK	Evet	Hayır	Belirsiz	Uygulanabilir Değil
1. Başlık				
Başlık şunları içerir: a) 'güvenirlilik genellemesi' veya 'meta-analiz' terimi, güvenirliliğe (iç tutarlılık, test-tekrar test, puanlayıcılar arası veya puanlayıcılar içi) açık bir atıfla birlikte ve (b) ölçeğin adı veya birden fazla ölçek olma durumunda ölçeklerin ölçtüğü yapı.				
ÖZET				
2. Özet				
Özette açıkça şunları belirtin: (a) amacın, bir veya birkaç ölçeğin güvenirlilik genellemesi (RG) meta-analizini gerçekleştirmek olduğu; (b) çalışmaların uygunluk kriterleri; (c) veri kaynakları ve ulaşılan zaman aralığı; (d) analiz edilen güvenirlilik katsayılarının türleri; (e) uygulanan istatistiksel model; (f) ana bulgular (örneğin, birleştirilmiş güvenirlilik katsayısı ve %95 güven aralığı, güvenirlilikle ilişkili moderatör değişkenler); ve (g) ana sonuçlar. Alan (kelime) sınırlaması durumunda, (b) ve (c) kriterleri çıkarılabilir.				
GİRİŞ				
3. Çalışmanın temeli				
Çalışmanın temeli anlatılırken şunlardan bahsedin: (a) ölçek/ler tarafından değerlendirilen özelliğin/yapının kavramsal tanımı; (b) ölçeğin/ölçeklerin uygulandığı hedef kitlenin/kitlelerin ve bu ölçeklerin ne amaçla kullanıldığının (örneğin, tarama, klinik tanı) açıklanması; (c) farklı versiyonlarını ve diğer dillere/kültürlere uyarlamalarını da içeren ölçekle/rle ilgili açıklamalar (uzunluk, kategori sayısı); ve (d) ölçeğin/ölçeklerin önceki psikometrik çalışmalarında elde edilen güvenirlilik tahminlerinin kısa bir açıklaması. İsteğe bağlı olarak, ölçeklerin geçerlik çalışmalarına (örneğin, açıklayıcı/doğrulayıcı faktör analizleri, eşzaman/yakınsak/ayırıcı geçerlilik, duyarlılık (boylamsal geçerlik)) dair kısa bir inceleme de eklenebilir.				
4. Amaçlar				
Meta-analizin amacının, daha kesin genel bir güvenirlilik katsayısı tahmini elde etmek mi ve/veya ölçeklerin farklı uygulamaları arasındaki güvenirlilik katsayılarının nasıl değiştiğini araştırmak mı olduğunu belirtin. İsteğe bağlı olarak, meta-analizin amaçlarından birinin ölçeğin/ölçeklerin güvenirlilik türetme (güvenirlilik indüksiyon) oranlarını tahmin etmek olup olmadığını belirtin.				
YÖNTEM				
5. Seçim kriterleri				
Şu noktalardan bahsederek dahil edilme kriterlerini belirtin: (a) dahil edilen versiyonlar ve/veya uyarlamalarla birlikte RG meta-analizinde analiz edilen ölçeğin/ölçeklerin adı/adları; (b) coğrafi ve/veya kültürel kısıtlamalar; (c) dikkate alınan yıllar; (d) makalenin dili; (e) yayın durumu; (f) çalışmaya özgü örnekleme/örneklemlere dayalı herhangi bir güvenirlilik tahmininin rapor edilmesi; (g) dikkate alınan güvenirlilik türü/türleri (örn, iç tutarlılık, kararlılık, puanlayıcılar arası/içi güvenirlilik...); (h) hedef örneklemler (örneğin, toplum, klinik, subklinik/analog, üniversite...); ve (i) gerekli minimum örneklem büyüklüğü.				

6. Arama stratejileri	Çalışmaların nasıl bulunduğunu belirtin: (a) başvuru elektronik veri tabanları; (b) diğer resmi arama yöntemleri (örneğin, belirli dergilerde elle arama, seçilen çalışmalarda listelenen kaynaklardan geriye doğru arama); ve (c) resmi olmayan arama yöntemleri (örneğin, internet aramaları, ek çalışmaları belirlemek için çalışma yazarlarıyla iletişime geçme). Elektronik aramalar için, kullanılan anahtar kelimeler ve bunların nasıl birleştirildiğini, ve arama sınırlılıklarını (örneğin, anahtar kelimelerin arandığı alanlar - başlık, özet, tam metin -, zaman aralığı, dil) arama stratejisinde açıklayın.
7. Veri ayıklama	Şu konuları da belirterek çalışmalardan ayıklanan özellikleri tanımlayın: (a) örneklem büyüklüğü/büyüklikleri, toplam test puanlarının ve alt ölçeklerin (eğer varsa) ortalaması/ları ve standart sapması/ları; (b) örneklem özellikleri (örneğin, hedef evren, ülke, ortalama yaş, yaşın standart sapması, cinsiyet dağılımı, etnik dağılım, önceki hastalıklar -yıl cinsinden ortalama ve SD); (c) test versiyonu (örneğin, uyarılma/sürüm, madde sayısı, raporlama biçimi -kendi kendine raporlama, klinisyenin raporlaması); (d) yöntemler (örneğin, çalışmanın türü, çalışmanın amacı -psikometrik veya uygulamalı-, kalite kontrol listesi); (e) dışsal özellikler (örneğin, yayın durumu, araştırmacıların bağlı olduğu kurumlar, fon kaynağı).
8. Raporlanan güvenilirlik	RG meta-analizine dahil edilen güvenilirlik katsayılarının türlerini belirtin: İç tutarlılık (örneğin, Cronbach Alfa, KR-21, paralel formlar, omega), kararlılık (test-tekrar test), puanlayıcılar arası ve puanlayıcılar içi güvenilirlik (örneğin, sınıf içi korelasyon, kappa katsayısı). Her bir güvenilirlik katsayısı türü için ayrı meta-analizler yapıldığını açıkça belirtin. Çok değişkenli/MASEM yaklaşımının uygulanması durumunda, çalışmalardan toplanan istatistiksel bilgilerin türünü belirtin (örneğin, madde-madde korelasyon/kovaryans matrisleri, faktör yükleri, vb.)
9. Güvenirlik türetme ve diğer yanlılık kaynaklarının tahmin edilmesi	Meta-analizde güvenilirlik türetmesini tahmin etmek amaçlanıyorsa, güvenilirlik türetmesi türlerini belirtin: ihmal yoluyla türetme (test güvenilirliğinden hiçbir şekilde bahsedilmemesi) veya raporlamaya dayalı türetme (belirsiz ifade veya katsayı belirterek raporlama). Diğer yanlılık kaynaklarının nasıl değerlendirildiğini açıklayın (örneğin, güvenilirlik katsayısının varsayımları, ölçüm modelinin yeterliliği, vb.)
10. Türetme yapılan çalışmalardan verilerin ayıklanması	Türetme yapan çalışmaların özelliklerinin mi ya da tam tersi yalnızca güvenilirliği rapor eden çalışmaların özelliklerinin mi dahil edildiğini belirtin.
11. Veri ayıklama işleminin güvenilirliği	Veri ayıklama sürecinin güvenilirliğinin nasıl değerlendirildiğini açıklayın: kaç kodlayıcıya hangi uyum katsayıları uygulandı (örneğin, Kappa katsayısı, sınıf içi korelasyon), hangi değerler elde edildi ve anlaşmazlıklarla nasıl başa çıkıldı.
12. Veri dönüştürme yöntemi	Meta-analitik birleştirme için güvenilirlik katsayılarının dönüştürülüp dönüştürülmediğini belirtin. Eğer yapıldıysa, dönüştürme yöntemlerini belirtin: Korelasyon katsayıları için Fisher'in Z'si (örneğin, test-tekrar test katsayıları), iç tutarlılık katsayıları (örneğin, Cronbach'ın alfa) için Bonett ve Hakstian ile Whallen'in dönüşüm yöntemleri, güvenilirlik indeksi, ölçüm hatası (örneğin, ölçmenin standart hatası) veya diğer (belirtin).

13. İstatistiksel model	Analiz çerçevesinin (sıklıkçı ya da Bayesyen) yanı sıra ortalama güvenilirlik katsayısını tahmin etmek ve moderatör değişkenlerin etkisini analiz etmek için meta-analitik birleştirmede varsayılan istatistiksel model(ler)i tanımlayın (örneğin, sabit etki(ler), rastgele etkiler, karışık etkiler, değişen katsayı modelleri, genelleştirilmiş doğrusal modeller). Çok değişkenli/MASEM yaklaşımı uygulanıyorsa, madde korelasyon/kovaryans matrislerinin veya faktör yüklerinin nasıl sentezlendiğini açıklayın.				
14. Ağırlıklandırma yöntemi	Meta-analitik birleştirmede uygulanan ağırlıklandırma yöntemini belirtin: ağırlıklandırma yapılmadan, örneklem büyüklüğüne göre ağırlıklandırma, ters varyansa göre ağırlıklandırma veya diğer ağırlıklandırma yöntemleri.				
15. Heterojenlik değerlendirilmesi	Güvenirlik katsayıları arasındaki heterojenliğin nasıl değerlendirildiğini açıklayın (örn. standart sapma, Q istatistiği, I2 indeksi, çalışmalar arası varyans, Hunter-Schmidt'in %75 kuralı). Eğer kullanıldıysa, güven aralığının, güvenilirlik aralığının veya yordama aralıklarının nasıl hesaplandığını yanı sıra çalışmalar arası varyans tahmincisini belirtin (DerSimonian ve Laird, Maksimum Likelihood, Kısıtlanmış Maksimum Likelihood, Ampirik Bayes, Paule ve Mandel).				
16. Moderatör analizler	Eğer kullanıldıysa, moderatör değişkenlerinin etkisinin nasıl değerlendirildiğini açıklayın (örneğin, alt grup analizleri, meta-regresyon analizleri, korelasyon analizleri).				
17. Ek analizler	Duyarlılık analizleri gibi gerçekleştirilen diğer ek analizleri açıklayın (örneğin, dönüştürülmüş ve dönüştürülmemiş güvenilirlik katsayılarıyla yapılan istatistiksel analizler, güvenilirlik katsayılarının tek tek çıkarılması, yayın yanlılığının, raporlama yanlılıklarının ve diğer yanlılık kaynaklarının değerlendirilmesi).				
18. Yazılım	İstatistiksel analizleri gerçekleştirmek için kullanılan yazılım ve versiyonunu belirtin (örneğin, R'de Metafor, SAS'ta Proc MIXED, Comprehensive Meta-Analysis).				
SONUÇLAR		Evet	Hayır	Belirsiz	Uygulanabilir Değil
19. Çalışma seçim sürecinin sonuçları	Her arama kaynağında belirlenen çalışma sayısını, seçilen çalışmaları ve seçilme nedenlerini, test puanlarının güvenilirliğini raporlayan ve türetme yapan çalışma sayısını bir akış diyagramı kullanarak çalışmaların seçim sürecini açıklayın. Güvenirlik türetmesi ile ilgili olarak, "ihmal" ve "raporlama" yollarıyla yapılan türetmeleri birbirinden ayrı vererek türetme oranlarını rapor edin (bkz. Örneğin REGEMA akış şeması). Ayrıca, güvenilirlik türetme oranlarını yayın yılı, ülke/kıta ve çalışma amacı (psikometrik vs. uygulamalı) gibi değişkenler açısından karşılaştırmanız önerilir.				
20. Ortalama güvenilirlik ve heterojenlik	Ölçeğin (ve varsa alt ölçeklerin) ve her tür güvenilirlik için (örneğin, iç tutarlılık, kararlılık, puanlayıcı/lar arası ve iç uyum) birleştirilmiş güvenilirlik katsayılarını ve güven aralıklarını belirtin. Güvenirlik katsayılarına herhangi bir dönüştürme uygulanmışsa, sonuçlar yorumlamayı kolaylaştırmak için orijinal ölçü birimine geri dönüştürülmelidir. Güvenirlik katsayılarının dağılımını grafik teknikleriyle (örneğin, orman grafikleri, kutu grafikleri, gövde ve yaprak grafikleri, histogramlar) gösterin ve heterojenliğin derecesini bir veya daha fazla heterojenlik ölçüsü ile tanımlayın (bkz. Madde 15).				
21. Moderatör analizler	Kategorik moderatörlerde, moderatörün her kategorisi için birleştirilmiş güvenilirlik katsayısını, güven aralığını ve diğer heterojenlik ölçülerini belirtin. Sürekli moderatörlerde, regresyon katsayılarını, standart hataları ve güven sınırlarını dahil edin. Her iki tür moderatör için, istatistiksel anlamlılık testlerinin sonuçlarını, model uygunluğu testlerini ve				

	açıklanan varyans oranını rapor edin. Ek bir adım olarak, en ilişkili moderatör değişkenlerini içeren bir yordayıcı/açıklayıcı model oluşturulması tavsiye edilir.				
22. Duyarlılık analizleri	Gerçekleştirilen herhangi bir duyarlılık analizinin sonuçlarını rapor edin veya açıklayın (bkz. Madde 17).				
23. Türetme ve raporlama yapan çalışmaların karşılaştırılması	Eğer yapılmışsa, türetme ve raporlama çalışmalarının özelliklerini (örneğin, örneklemelerin sosyodemografik ve klinik özellikleri) karşılaştırmanın sonuçlarını sunun.				
24. Veri seti	Güvenirliği rapor eden bireysel çalışmaların özelliklerini tablolaştırın (Bkz. Madde 7). Tablolar ekler veya ek dosyalar olarak sunulabilir. Ayrıca, RG meta-analizine dahil edilen tüm çalışmaların bir listesini ya kaynaklar bölümünde ya da ek dosya olarak verin.				
TARTIŞMA		Evet	Hayır	Belirsiz	Uygulanabilir Değil
25. Sonuçların özeti	Ölçek/testten elde edilen ortalama güvenilirlik ve güvenilirlik katsayılarının moderatörleri gibi ana sonuçları sunun. Varsa, sonuçları önceki kanıtlar ışığında tartışın.				
26. Sınırlılıklar	Meta-analizin sınırlılıklarını tartışın. Güvenirlik türetme oranlarını ve güvenilirliği türeten ve raporlayan çalışmaların örneklem özellikleri açısından ne kadar karşılaştırılabilir olduğunu açık ve net bir şekilde belirtin.				
27. Uygulama için öneriler	Ölçeğin/testin farklı ortamlarda ve hedef kitlelerde kullanılabilirliğine ilişkin profesyonel uygulamalar için yönlendirici önerilerde bulunun.				
28. Gelecekteki araştırmalar için öneriler	Ölçeğin/testin hangi koşullar altında uygulanması gerektiğine ilişkin olarak araştırmacılara yönelik tavsiyeler verin.				
FİNANSMAN		Evet	Hayır	Belirsiz	Uygulanabilir Değil
29. Finansman	Meta-analizin finansal kaynaklarını ve yazarların potansiyel çıkar çatışmalarını belirtin.				
PROTOKOL		Evet	Hayır	Belirsiz	Uygulanabilir Değil
30. Protokol	Meta-analiz protokolünün daha önce yayınlanıp yayınlanmadığını veya bir web sitesinde (örneğin Prospero'da) erişilebilir hale getirilip getirilmediğini belirtin.				



A Guide to Reliability Generalization Meta-Analyses: An Adaptation Study of the REGEMA Checklist and Flowchart

Özen Yıldırım¹Endam Düzyol-Türk²

ARTICLE INFO

DOI: 10.29299/kefad.1658099

Received: 14.03.2025

Revised: 11.11.2025

Accepted: 27.11.2025

Keywords:

Reliability,
Reliability Generalization,
Meta-Analysis,
REGEMA

ABSTRACT

Reliability Generalization (RG) is a specific form of meta-analysis that examines how reliability estimates derived from test scores differ when the same measurement instrument is applied to different samples. Although various reporting guidelines such as PRISMA, AMSTAR, MOOSE, and MARS were developed for meta-analyses examining intervention effects or relationships between variables to aid in meta-analysis reporting, they are also frequently used for reporting reliability generalization meta-analyses. The Reliability Generalization Meta-Analysis (REGEMA) checklist and flowchart is a tool designed for the proper development and reporting of reliability generalization meta-analyses. REGEMA seems to be the only checklist specifically designed to guide the reporting of reliability generalization meta-analyses. In this adaptation study, the aim is to introduce the REGEMA Checklist and Flowchart to researchers in Türkiye and discuss its content. The study includes the REGEMA Checklist and corresponding explanations, and the Flowchart is presented and illustrated with examples. It is hoped that through this study, researchers in Türkiye can easily access and make use of REGEMA in their work. Researchers conducting meta-analyses can use REGEMA to follow the process and prevent any shortcomings in their reporting.

1. Introduction

Although access to information today is practical and easier than in the past, the inability to reach accurate information leads to various consequences for individuals and society. Therefore, accurate information is a crucial requirement for society. In quantitative research, obtaining controlled and objective information from studied samples and evaluating data in a valid and reliable manner are fundamental principles. Glass (1976) states that although information is easily accessible, the fact that independent studies on the same topic may yield different findings raises doubts about the reliability of knowledge and creates a need for re-examination. To address these needs, he has proposed meta-analysis, which involves synthesizing research results under a single research framework.

1.1. Meta-Analysis

Meta-analysis is an approach that systematically collects the findings of previously conducted studies on a research topic and reanalyzes them (Paul & Criado, 2020). Similarly, Sander Greenland (2008) describes meta-analysis as a method that synthesizes the results of multiple independent studies on a given topic and subjects them to statistical analysis. By statistically integrating numerical

data obtained from quantitative studies on a particular topic, meta-analysis provides a holistic perspective, allowing the development of a new model or theoretical framework (McMillan & Schumacher, 2001). Meta-analysis, defined as the quantitative integration of findings drawn from multiple studies (Lau et al., 1997), plays a critical role in assessing the reliability of research evidence by combining results across independent investigations (Ioannidis, 2003). By statistically examining research findings of the same type, meta-analysis increases the overall sample size, strengthens statistical significance, estimates effect sizes, addresses questions that may not have been anticipated at the outset of a study, and helps make decisions regarding uncertainty when results are inconsistent (Whitehead, 2002). Meta-analytic methods are widely used to inform researchers and practitioners, particularly in the social and health sciences (Hunter & Schmidt, 2004; Wiernik & Dahlke, 2020). Because individual studies are often conducted with limited sample groups, it is not always easy to obtain reliable findings from a single study. Through meta-analysis, findings from different studies can be combined, increasing statistical power and yielding stronger evidence. By synthesizing the results of multiple studies, it provides educators, academics, policymakers, and society with generalizable

and robust data (Borenstein et al., 2009). Because this study aims to introduce and examine the Reliability Generalization Meta-Analysis (REGEMA) Checklist and Flowchart proposed by Sánchez-Meca et al. (2021) for researchers in Türkiye, a concise overview of the main stages of meta-analysis is first presented to provide context and facilitate a clearer understanding of the REGEMA procedures.

1.2. Steps of Meta-Analysis

To conduct a meta-analysis, a sequence of steps must be followed systematically. The literature includes various sources outlining these steps (Borenstein et al., 2009, 2010; Hansen et al., 2022; Hoon, 2013; Page, 2021; Tuttle et al., 2009). Drawing on these references, the steps of meta-analysis are summarized below:

Step 1: Defining the Research Question

The initial stage of a meta-analysis involves clearly defining the research question. This step specifies which interventions' effects will be evaluated or determines the conceptual domain of the constructs under investigation (Hansen et al., 2022). For instance, a research question could be formulated as: "What is the overall effect size of outdoor activities on the scientific process skills of children aged 60–72 months?"

Step 2: Literature Search

A literature search can be carried out by using relevant keywords and applying Boolean operators across major databases and search platforms such as Web of Science, ScienceDirect, Scopus, ERIC, EBSCO, PsycINFO, ProQuest, PubMed/Medline, Google Scholar, Ulakbim Discovery, and the Council of Higher Education National Thesis Center (Tuttle et al., 2009). Within Boolean logic, connectors such as "AND," "OR," and "NOT" are used to refine search results by either expanding or restricting the scope of retrieved studies (Basu, 2017).

In a meta-analysis, the literature search must be conducted in a systematic, transparent, and reproducible manner, aiming to identify and include all studies relevant to the research topic (Hansen et al., 2022). For instance, a study aiming to examine the reliability generalization of scales measuring scientific process skills in young children could use a search code such as: ("preschool" OR "early childhood" OR "kindergarten") AND ("scientific process skills") AND ("cronbach alpha").

Step 3: Coding, Selection of Effect Size Measure, and Analytic Method

The coding process begins with the development of a coding form. Because its structure is shaped by the study's design, analytical approach, and software used, a single standardized template does not exist (Hansen et al., 2022). Prior to coding the dataset, it is typically essential to take into account the proposed research model and any variables that are not central to the main analysis. Decisions

about including relevant moderator or control variables should be grounded in theoretical rationale explaining how these variables might influence the effect under investigation (Bernerth & Aguinis, 2016). A long-standing debate in conducting meta-analyses concerns whether to use only one effect size per primary study for the same construct or to include all available effect sizes. Hansen et al. (2022) emphasize taking advantage of all available effect size observations but caution that their statistical dependence must be carefully addressed using appropriate methods such as multilevel models, panel regression models, or robust variance estimation.

The effect size—a standardized, common statistical value calculated from the obtained results—is used in meta-analytic coding. Commonly used effect size measures include standardized mean differences (Cohen's *d*, Hedges' *g*, Glass's delta), correlation coefficients (Pearson's *r*), and binary outcome metrics (odds ratio, risk ratio, and risk difference) (Page, 2021). After defining the primary effect size measure, it may be necessary to convert findings reported in other formats into the selected effect size during the coding process (Hansen et al., 2022). Meta-analyses typically employ two primary statistical models: the fixed-effect model and the random-effects model (Borenstein et al., 2009).

Directly linked to the meta-analysis research question, the selection of the meta-analytic method and software is made at this stage. Research questions may focus on the relationship between constructs, the effect of an intervention, or the moderation or mediation of effects. Various analytical approaches can be applied, including traditional univariate meta-analysis, meta-regression, meta-analytic structural equation modeling, and qualitative meta-analysis (Hoon, 2013).

Step 4: Analysis

Prior to conducting the primary analysis, it is generally important to perform preliminary sensitivity checks to evaluate the robustness of the findings (Rudolph et al., 2020). Influential outliers may distort the results, particularly when the number of effect sizes is limited. A range of statistical techniques can be employed to identify such outliers in meta-analytic datasets. Among these, forest plots and funnel plots are commonly used diagnostic tools (Aguinis et al., 2013). As noted in the literature search phase, potential bias is another key concern and should be systematically examined. Several approaches are available for this purpose, including funnel plots, the widely used Trim-and-Fill procedure, as well as the Begg and Egger tests (Rothstein et al., 2005).

Once influential outliers or potential biases have been detected and addressed, the next step is to select between the two primary meta-analytic models—fixed-effect and random-effects—each based on distinct underlying assumptions (Borenstein et al., 2010).

Step 5: Reporting the Results

The last step in a meta-analysis is the reporting of its findings. All procedures and methodological choices should be communicated clearly and transparently so that readers can follow the process. Several protocols have been developed to standardize the reporting of meta-analyses. Among these, the most widely used today is known as the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) (Moher et al., 2015).

Based on the stages described above—defining the research question, conducting the literature search, coding, selecting the effect size measure and analytic method, performing the analysis, and reporting the results—a meta-analysis can be carried out systematically.

1.3. Reliability

A distinctive feature of the meta-analytic method is that it relies on quantitative techniques rather than subjective judgments when evaluating research findings. Through this analysis, researchers aim to estimate parameters that are reliable, valid, and minimally variable for the topic under investigation (Mosteller & Colditz, 1996). For a scale to be standardized, several criteria must be met, with reliability and validity being indispensable conditions. Reliability refers to the stability of the measurement values obtained by an instrument when the measurement is repeated under the same conditions, while validity refers to the degree to which the instrument measures the intended construct without being confounded by other attributes (Carmines & Zeller, 1979; Crocker & Algina, 1986). Consistency across repeated measurements represents reliability (Ho, 2014). Reliability also affects validity and plays a critical role in accurately interpreting scores obtained from measurement instruments. Reliability estimates may vary across different studies using the same instrument. Therefore, reliability must be re-estimated for each individual study. Various reliability estimation methods exist, based on either multiple administrations or a single administration. For example, the parallel-forms method and the test-retest method are based on multiple administrations, whereas the split-half method and Cronbach's alpha are based on a single administration. In these methods, the reliability coefficient is estimated as a correlation value. In the test-retest method, reliability is estimated as a stability coefficient, whereas in Cronbach's alpha, the coefficient reflects internal consistency (Mor, 2021). For instance, the Cronbach's alpha reliability coefficient is frequently used in determining the reliability of scales designed to measure psychological constructs. Cronbach's alpha is calculated as a function of item variances, obtained by relating the sum of the variances of the k items in a scale to the total variance. A separate α value can be calculated for each item, or an average α value can be obtained for all items. The α value reflects the overall reliability of the scale, and a value of .70 or higher indicates

high internal consistency, meaning that the scores obtained from the scale are reliable (Cronbach, 1951).

1.4. Reliability Generalization

Reliability Generalization (RG) is a specific form of meta-analysis. Vacha-Haase (1998) introduced this approach to synthesize and generalize the reliability coefficients of a given scale across multiple studies. RG centers on analyzing how reliability estimates based on test scores may differ when the same instrument is administered to various samples. In this regard, it constitutes a specialized meta-analytic approach that enables the systematic examination of reliability coefficients across studies (Thompson, 2003). The main aim of a RG meta-analysis is to provide researchers and practitioners with information about the expected reliability of scores derived from a test or scale, and to identify which test formats and administration conditions may affect these reliability estimates (Vacha-Haase, 1998). While traditional meta-analyses typically focus on the validity or effectiveness of measurement instruments—whether they accurately measure the constructs they aim to measure—RG focuses specifically on the consistency of the instrument across studies. This helps researchers make more informed choices when selecting scales for future studies (Osborne & Fitzpatrick, 2012).

A common assumption among researchers and practitioners is that reliability is a fixed property intrinsic to a test and remains stable across administrations. However, psychometric theory shows that reliability is not a permanent characteristic of the test itself; it is a property of the test scores and therefore varies with each administration. Reliability generalization based on meta-analysis is the only type of meta-analysis in which a common effect size index consists of reliability coefficients (e.g., internal consistency, test-retest reliability, inter-rater correlations, etc.) reported in primary studies applying the same test (Badenes-Ribera, 2023; Traub, 1994). RG can also be described as the statistical synthesis of reliability estimates from previous studies in which reliability has been examined across various contexts and sample groups (Badenes-Ribera et al., 2023). Vacha-Haase (1998) defined the aims of a reliability generalization meta-analysis as follows:

- a) estimating the average reliability of test scores through repeated administration of the test across different samples and contexts,
- b) evaluating the generalizability of reliability from one test administration to another, and
- c) identifying potential study and sample characteristics that may be statistically associated with variability in reliability coefficients.

To ensure consistency and transparency in reporting meta-analyses, a number of established guidelines have been developed, including PRISMA (Moher et al., 2015), AMSTAR (Shea et al., 2009), MOOSE (Stroup et al., 2000),

and MARS (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). Among these, PRISMA offers a comprehensive framework for the planning, execution, and reporting of systematic reviews and meta-analyses (Page et al., 2021). The 2020 update of PRISMA has also been translated into Turkish (Hür, 2021), along with adaptations into 15 other languages. AMSTAR is a checklist used to assess the methodological quality, neutrality, and validity of systematic reviews (Shea et al., 2007). MOOSE offers recommendations and checklists specifically for analyzing patient records, case-control studies, and cohort studies (Stroup et al., 2000). MARS, developed by the American Psychological Association (APA), outlines standards regarding which statistical methods should be used and how they should be reported (Appelbaum et al., 2018). However, these guidelines were not developed for the reporting of RG meta-analyses; rather, they were designed for meta-analyses examining intervention effects or relationships between variables. The unique characteristics of reliability generalization meta-analyses make it difficult to adapt existing checklists efficiently to RG (Sánchez-Meca et al., 2021). A reporting checklist specifically tailored to RG meta-analyses is currently available only in the form of REGEMA (Sánchez-Meca et al., 2021). The Reliability Generalization Meta-Analysis (REGEMA) Checklist and Flowchart, developed by Sánchez-Meca et al. (2021), serves as a structured tool to support the proper conduct and transparent reporting of reliability generalization meta-analyses. Sánchez-Meca et al. (2021) created a flowchart similar to PRISMA but adapted to the unique aspects of RG meta-analyses, aiming to support researchers in searching for and selecting primary studies that meet the criteria for reliability generalization. In this study, the acronym REGEMA is used in order to increase the international recognition and consistent use of the term for reliability generalization meta-analysis.

Although the use of RG in Türkiye is limited, the REGEMA Checklist and Flowchart developed by Sánchez-Meca et al. (2021) is considered an important guide for the systematic execution of RG studies. Using REGEMA—an instrument standardized for reliability generalization—may contribute to establishing a common language and a more systematic evaluation of RG studies among researchers in Türkiye. Considering that each scientific field has its own technical language and concepts, adapting this flowchart and checklist into Turkish by experts and introducing it into the field is seen as valuable. In this context, the present study introduces and discusses REGEMA in order to guide the systematic implementation and reporting of reliability generalization research. This will enable researchers in Türkiye to easily access REGEMA and use it in their work. Using REGEMA in scientific research may help researchers follow the necessary procedures for RG, prevent reporting deficiencies, and assist journal editors and reviewers in evaluating RG meta-analyses.

2. Method

2.1. Design

In this study, the REGEMA Checklist and Flowchart were adapted from English into Turkish. Because this is an adaptation study, the research follows a basic research design. The study discusses the conformity of newly translated information to established standards.

2.2. Procedure

The REGEMA Checklist and Flowchart were examined in detail. Four academics who specialize in measurement and evaluation or English language education—and who have experience with meta-analytic studies—were consulted during the translation process. After the flowchart and checklists were translated into Turkish, a measurement and evaluation expert conducted a comparative review of the English and Turkish versions. In addition, the authors reviewed the concepts in the flowchart and checklist in relation to the literature and selected the terms that most accurately reflected their semantic meaning. For this purpose, international studies referencing REGEMA were examined and their definitions evaluated. The draft Turkish version was then sent back to three experts in measurement and evaluation, who provided feedback on conceptual and linguistic appropriateness, clarity, and terminological accuracy. Based on expert feedback, the text was revised to produce the final version.

2.3. Data Collection Tools

2.3.1. REGEMA checklist and flowchart

The REGEMA Checklist and Flowchart were developed by Sánchez-Meca et al. (2021). The checklist was designed to standardize the conduct and reporting of reliability generalization meta-analyses. Consisting of 30 items, it specifies, step by step, what information should be reported, at which stages, and how it should be presented. In addition, the REGEMA Flowchart provides a visual overview of the key steps to be followed when conducting an RG meta-analysis.

2.4. Data Analysis

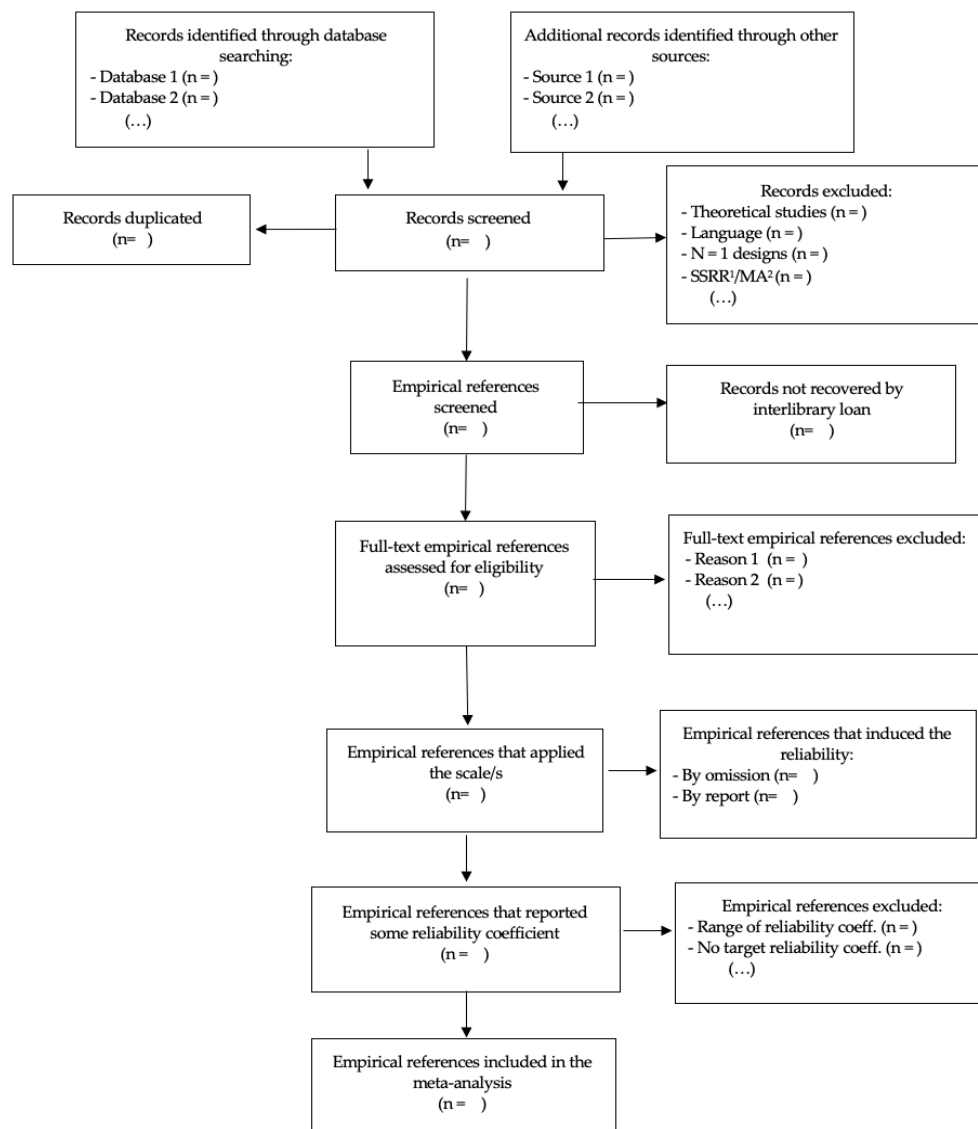
Feedback obtained during the translation, expert review, and back-translation stages was analyzed systematically, and the final version was produced.

3. Results

Below, the REGEMA Flowchart—which guides the data collection process in RG studies—and the REGEMA Checklist—which outlines the criteria to be considered when reporting RG analyses—are described.

3.1. REGEMA Flowchart

The REGEMA Flowchart is intended to visually depict the processes of literature searching, screening, and selecting studies for inclusion. Figure 1 displays the REGEMA Flowchart developed by Sánchez-Meca et al. (2021).



Note. ¹Systematic Reviews. ²Meta-analyses

Figure 1. REGEMA flow chart 2019

Explanations of the required steps in the REGEMA Flowchart and examples related to each step are presented below:

Records identified through database searching: The literature is reviewed based on the selected topic. Searches are conducted in selected databases using identified keywords. Databases such as WoS, Scopus, ScienceDirect, and ERIC are typically used. Boolean operators such as “AND” or “OR” may be used. For example, in WoS advanced search, one might use: TI=((“preschool” OR “pre-school” OR “early child” OR “early childhood” OR “kindergarten”) AND (“scientific process skills”) AND (“cronbach alpha”)).

Additional records identified through other sources: Searches using the same keywords may be performed

through alternative resources such as ResearchGate, OpenGrey, Google Scholar, or direct contact with authors.

Records screened: The total number of records retrieved from database searches and additional sources is recorded. At this stage, it is essential to record the number of records and the dates on which they were retrieved.

Records duplicated: Records obtained from all sources are combined, and duplicates are removed. Only one version of a repeated record is included.

Records excluded: Certain exclusion criteria may be applied. For example: theoretical papers, studies not meeting language criteria (e.g., only English), studies below a sample-size threshold, systematic reviews, or meta-analyses. The criteria used and the number of excluded records must be documented.

Empirical references screened: The number remaining after removing duplicates and excluded records is recorded here.

Records not recovered by interlibrary loan: Records that appear in searches but cannot be obtained through libraries are reported.

Full-text empirical references assessed for eligibility: The number of records remaining after subtracting those unavailable through interlibrary loan is reported here.

Full-text empirical references excluded: The number of studies that did not use the scale(s) should be reported in this section, along with the reasons for their exclusion. For example, sources that did not provide sufficient data to conduct the analyses.

Empirical references that applied the scale/s: This section should report the remaining number after subtracting the full-text empirical sources that were excluded from the full-text empirical sources assessed for eligibility.

Empirical references that induced the reliability: Studies that use the scale(s) are examined in detail. In these studies, researchers who do not estimate reliability based on their own sample are identified. Researchers may assume that scores obtained from the scale are already reliable because reliability was reported in a previous study, and therefore they may choose not to recalculate the reliability coefficient for their own data. Alternatively, they may report a reliability coefficient calculated by another researcher for the same scale as if it were the reliability estimate for their own study. This practice is referred to as reliability derivation (induction). It represents a form of publication bias specific to RG meta-analyses. Reliability derivation is considered one of the questionable measurement practices—referring to decisions that raise concerns about the validity of measurements and, in turn, the overall validity of study findings (Sánchez-Meca et al., 2021). There are two types of reliability induction. Reliability induction by reporting refers to a study reporting the reliability of test scores obtained from another study, whereas reliability induction by omission refers to a study using one or more psychometric instruments without mentioning reliability at all. The number of records that derive reliability through omission or reporting should be documented. It may even be necessary to examine and record the types of studies in which these derivations occur in more detail.

Empirical references that reported some reliability coefficient: The number obtained after subtracting the sources that relied on a reliability induction approach from the empirical sources using the scale(s) should be reported in this section.

Empirical references excluded: In this section, reliability coefficients are examined based on the purpose of the study. For example, in a meta-analysis focusing only on Cronbach's alpha, test-retest reliability, inter-rater reliability, or split-half coefficients may be excluded.

Coefficients presented only as ranges or without a specific value may also be excluded.

Empirical references included in the meta-analysis: The remaining number obtained by subtracting the number of "empirical sources with reliability coefficients that were not included" from "empirical sources reporting some reliability coefficients" should be reported in this section. The remaining studies in this section are included in the meta-analysis. If reliability coefficients were calculated more than once in the available sources — for example, calculations based on different samples or different administrations such as pretest-posttest or longitudinal applications — each coefficient should be treated as a separate study. For instance, in a study based on an experimental model, two coefficients calculated from administering the test to the same sample at two different time points should be included in the analysis as separate studies. This issue should be considered during data extraction and stated in this section.

3.2. REGEMA Checklist

Sánchez-Meca et al. (2021) developed the REGEMA Checklist to define the key criteria for reporting RG studies. In developing the checklist, the authors conducted a comprehensive literature review and drew on widely used reporting frameworks in meta-analysis, including PRISMA, MOOSE, AMSTAR, AMSTAR 2, and MARS, along with 80 methodological publications. The REGEMA Checklist is structured into eight sections—Title (one item), Abstract (one item), Introduction (two items), Method (fourteen items), Results (six items), Discussion (four items), Funding (one item), and Protocol (one item)—comprising a total of thirty items. Each item is accompanied by response options such as "Yes," "No," "Unclear," and "Not Applicable," as well as explanations clarifying the intent of each criterion. The sections, descriptions, and response options of the checklist are presented in Appendix.

4. Discussion

This study aimed to present the REGEMA Checklist and Flowchart, designed by Sánchez-Meca et al. (2021), to researchers in Türkiye and to discuss their content. The study includes the REGEMA Checklist and its explanations, and the Flowchart is presented with examples. The usefulness of meta-analyses depends on the quality of their reporting, and a meta-analysis that is inadequately reported provides limited benefit for future research. RG meta-analyses are also subject to this limitation (Nosek et al., 2015). In a systematic review by Sánchez-Meca et al. (2019), the reporting practices of 150 RG meta-analyses on psychological tests were evaluated, revealing substantial shortcomings in key methodological components. The authors later emphasized that inadequately reported meta-analyses compromise replicability and run counter to the transparency principles advocated by the Open Science Framework (Sánchez-Meca

et al., 2021). In response to these issues, they introduced the REGEMA Checklist and Flowchart as a structured solution. Accordingly, the present study aims to present REGEMA to researchers and examine its content in detail.

Reporting the reliability coefficient of a scale is considered important because it informs other researchers about whether the scale can be reliably used in different studies. When researchers calculate the reliability coefficient for the scale they use, reliability generalization can be conducted by synthesizing the reliability coefficients reported across multiple studies that have employed the same scale. This allows researchers to make comparisons between scales and gain insights (Deng et al., 2019). Reaching all studies that use the scale is crucial when conducting reliability generalization. Therefore, researchers who use a scale are encouraged to calculate and report its reliability coefficient (Yörük & Sen, 2023). REGEMA includes a reporting process that is based on the identification of reliability coefficients. Since it provides a comprehensive checklist and flowchart, it is expected that following the procedures in REGEMA carefully will allow researchers to obtain reliable results.

Ethical Considerations: In this study, all rules specified within the "Directive on Scientific Research and Publication Ethics of the Higher Education Institutions" were followed, and none of the actions listed under Section Two, titled "Actions Contrary to Scientific Research and Publication Ethics," were committed. The developer of the REGEMA Checklist and Flowchart (Julio Sánchez-Meca) was contacted, informed about the study, and permission for use was obtained. Since this was an adaptation study that did not involve experimental procedures or data collection from human or animal participants, ethical committee approval was not required.

Author Contributions : Both authors contributed equally.

Funding: No financial support was received for this study.

Conflict of Interest : There is no conflict of interest.

Data Availability : No implementation-based data were collected.

References

- Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16(2), 270–301. <https://doi.org/10.1177/1094428112470848>
- American Psychological Association, 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. <https://doi.org/10.1037/0003-066X.63.9.839>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Badenes-Ribera, L., Duro-García, C., López-Ibáñez, C., Martí-Vilar, M., & Sánchez-Meca, J. (2023). The adult prosocialness behavior scale: A reliability generalization meta-analysis. *International Journal of Behavioral Development*, 47(1), 59–71. <https://doi.org/10.1177/01650254221128280>
- Basu, A. (2017). How to conduct meta-analysis: A basic tutorial. *PeerJ Preprints*, 5, e2978v1. <https://doi.org/10.7287/peerj.preprints.2978v1>
- Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, 69(1), 229–283. <https://doi.org/10.1111/peps.12103>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage. <https://doi.org/10.4135/9781412985642>
- Cheung, M. W. L., & Vijayakumar, R. (2016). A guide to conducting a meta-analysis. *Neuropsychology Review*, 26, 121–128. <https://doi.org/10.1007/s11065-016-9319-z>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Deng, J., Wang, M. C., Zhang, X., Shou, Y., Gao, Y., & Luo, J. (2019). The inventory of callous unemotional traits: A reliability generalization meta-analysis. *Psychological Assessment*, 31(6), 765–779. <https://doi.org/10.1037/pas0000698>
- Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly*, 68, 103–106. <https://doi.org/10.1007/s11301-018-0142-x>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189X005010003>
- Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins.
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181–217. <https://doi.org/10.1002/jrsm.1378>
- Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: A practical guide. *Management Review Quarterly*, 72, 1–19. <https://doi.org/10.1007/s11301-021-00247-4>
- Ho, R. (2014). *Univariate and multivariate data analysis with IBM SPSS*. CRC Press.
- Hoon, C. (2013). Meta-synthesis of qualitative case studies: An approach to theory building. *Organizational Research Methods*, 16(4), 522–556. <https://doi.org/10.1177/1094428113484969>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
- Hür, G. (2021). PRISMA kontrol listesi 2020 güncellemesi. *Online Turkish Journal of Health Sciences*, 6(4), 603–605. <https://doi.org/10.26453/otjhs.1001606>
- Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *The Lancet*, 361(9357), 567–571. [https://doi.org/10.1016/S0140-6736\(03\)12516-0](https://doi.org/10.1016/S0140-6736(03)12516-0)
- Lau, J., Ioannidis, J. P., & Schmid, C. H. (1997). Quantitative synthesis in systematic reviews. *Annals of Internal Medicine*, 127(9), 820–826. <https://doi.org/10.7326/0003-4819-127-9-199711010-00008>
- McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction*. Longman.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1–9. <https://doi.org/10.1186/2046-4053-4-1>
- Mor, E. (2024). Ölçme süreç ve sonuçlarının taşınması gereken temel nitelikler. Ö. Yıldırım ve K. S. Kartal (Eds.), *Eğitimde ölçme ve değerlendirme* içinde (ss. 22–62). Lisans Yayıncılık.
- Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, 17(1), 1–23. <https://doi.org/10.1146/annurev.pu.17.050196.000245>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>

- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research, and Evaluation*, 17(15), 1–8.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 1–11. <https://doi.org/10.1186/s13643-021-01626-4>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know? *International Business Review*, 29(4), 101717. <https://doi.org/10.1016/j.ibusrev.2020.101717>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. <https://doi.org/10.1002/0470870168>
- Rudolph, C. W., Chang, C. K., Rauvola, R. S., & Zacher, H. (2020). Meta-analysis in vocational behavior: A systematic review and recommendations for best practices. *Journal of Vocational Behavior*, 118, Article 103397. <https://doi.org/10.1016/j.jvb.2020.103397>
- Sánchez-Meca, J., Marín-Martínez, F., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-López, J. A., & López-García, J. J. (2019, July). *Reporting practices in reliability generalization meta-analyses: Assessment with the REGEMA checklist* [Paper presentation]. XVI Congress of Methodology of the Social and Health Sciences, Madrid, Spain.
- Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, 12(4), 516–536. <https://doi.org/10.1002/jrsm.1487>
- Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7(1), 10. <https://doi.org/10.1186/1471-2288-7-10>
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013–1020. <https://doi.org/10.1016/j.jclinepi.2008.10.009>
- 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. *JAMA*, 283(15), 2008–2012. <https://doi.org/10.1001/jama.283.15.2008>
- Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. Sage.
- Traub, R. E. (1994). *Reliability for the social sciences: Theory and applications*. Sage Publications.
- Tuttle, B. D., Isenburg, M. V., Schardt, C., & Powers, A. (2009). PubMed instruction for medical students: Searching for a better way. *Medical Reference Services Quarterly*, 28(3), 199–210. <https://doi.org/10.1080/02763860903069839>
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6–20. <https://doi.org/10.1177/0013164498058001002>
- Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*. John Wiley & Sons. <https://doi.org/10.1002/0470854200>
- Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, 3(1), 94–123. <https://doi.org/10.1177/2515245919885611>
- Yörük, S., & Sen, S. (2023). A reliability generalization meta-analysis of the creative achievement questionnaire. *Creativity Research Journal*, 35(4), 714–729. <https://doi.org/10.1080/10400419.2022.2148073>

APPENDIX

REGEMA Checklist

TITLE		Yes	No	Unclear	NA
1. Title	In the title include: (a) the term “reliability generalization” or “meta-analysis” together with some explicit indication to reliability (internal consistency, test-retest, inter- or intra-rater) and (b) the name of the scale or, if more than one scale, the attribute/outcome measure that the scales are assessing.				
ABSTRACT		Yes	No	Unclear	NA
2. Abstract	In the abstract explicitly state: (a) that the objective was to carry out a reliability generalization (RG) meta-analysis of one or several scales; (b) eligibility criteria of the studies; (c) data sources with the temporal range covered; (d) types of reliability coefficients analyzed; (e) statistical model applied; (f) main results (e.g., pooled reliability coefficient and 95% CI, moderator variables related to reliability); and (g) main conclusions. In case of space limitation, (b) and (c) criteria can be omitted.				
INTRODUCTION		Yes	No	Unclear	NA
3. Background	In the background include: (a) a conceptual definition of the attribute/outcome measure assessed by the scale/s; (b) description of the target population/s to which the scale/s is/are applied and its/their purposes (e.g., screening, clinical diagnosis); (c) a complete description of the scale/s (length, number of categories), including the versions and adaptations to other languages/cultures; and (d) a brief presentation of reliability estimates obtained in previous psychometric studies of the scale/s. Optionally, a brief review of validation studies of the scale/s (e.g., exploratory/confirmatory factor analyses, concurrent/convergent/discriminant validity, responsiveness) could be included.				
4. Objectives	State whether the purpose of the meta-analysis was to obtain a more precise overall reliability coefficient estimate and/or investigate how reliability coefficients vary among different applications of the scales. Optionally, specify whether one objective of the meta-analysis is to estimate the reliability induction rates of the scale/s.				
METHOD		Yes	No	Unclear	NA
5. Selection criteria	Specify inclusion criteria: (a) name/s of the scale/s analysed in the RG meta-analysis, as well as the versions and/or adaptations included; (b) geographical and/or cultural restrictions; (c) years considered; (d) language of the paper; (e) publication status; (f) to report any reliability estimate based on the study-specific sample/s; (g) type/s of reliability considered (e.g., internal consistency, temporal stability, inter-/intra rater reliability...); (h) target population/s (e.g., community, clinical, subclinical/analog, university...); and (i) minimum sample size required.				

6. Search strategies	Specify how the studies were located: (a) electronic databases consulted; (b) other formal search procedures (e.g., manual search in specific journals, backward search from references listed in selected studies); and (c) informal search procedures (e.g., internet searches, contacting study authors to identify additional studies). For electronic searches, describe the search strategy, including the keywords used and how they were combined, and the search limits (e.g., fields where the keywords were searched - title, abstract, full-text -, temporal range, language).
7. Data extraction	Describe the characteristics extracted from the studies, including: (a) sample size/s, mean/s and standard deviation/s of total test scores and subscales (if applicable); (b) sample characteristics (e.g., target population, country, mean age, standard deviation of the age, gender distribution, ethnic distribution, disorder history –mean and SD in years); (c) test version (e.g., adaptation/version, number of items, reporting format –self-report, clinician); (d) methods (e.g., study design, purpose of the study –psychometric versus applied–, quality checklist); (e) extrinsic characteristics (e.g., publication status, researchers' affiliations, funding source).
8. Reported reliability	Identify the types of reliability coefficients included in the RG meta-analysis: internal consistency (e.g., Cronbach's alpha, KR-21, parallel forms, omega), temporal stability (test-retest), inter- and intra-rater reliability (e.g., intraclass correlation, kappa coefficient). Clearly state that separate meta-analyses were conducted for each type of reliability coefficient. In case of applying a multivariate/MASEM approach, specify the type of statistical information extracted from the studies (i.e., item-item correlation/covariance matrices, factor loadings, etc.).
9. Estimating the reliability induction and other sources of bias	In case that the meta-analysis intends to estimate the reliability induction, identify the types of reliability induction: induction by omission (no mention of test reliability whatsoever) or reporting induction (vague or precise reporting). Describe how other sources of bias were assessed (e.g. assumptions of the reliability coefficient, adequacy of the measurement model, etc.).
10. Data extraction of inducing studies	Declare whether characteristics of inducing studies were also extracted or if, on the contrary, only characteristics of studies that reported reliability were extracted.
11. Reliability of data extraction	Describe how the reliability of data extraction process was appraised: how many coders which agreement coefficients were applied (e.g., kappa coefficient, intraclass correlation), which values were obtained, and how disagreements were dealt with.
12. Transformation method	State whether or not the reliability coefficients were transformed for the meta-analytic integration. If relevant, specify the transformation methods: Fisher's Z for correlation coefficients (e.g., test-retest coefficients), Bonett's and Hakstian and Whallen's transformation for internal consistency coefficients (e.g., Cronbach's alpha), reliability index, measurement error (e.g., standard error of measurement), or other (specify).

13. Statistical model	Describe the statistical model(s) assumed in the meta-analytic integration for estimating the average reliability coefficient and for analysing the influence of moderator variables (e.g. fixed-effect(s), random-effects, mixed-effects, varying-coefficient models, generalized linear models), as well as the analysis framework (frequentist or Bayesian). In case of applying a multivariate/MASEM approach, describe how the item correlation/covariance matrices or factor loadings were synthesized.
14. Weighting method	Specify the weighting method applied in the meta-analytic integration: unweighted, weighting by sample size, weighting by inverse variance, or other weighting methods.
15. Heterogeneity assessment	Describe how heterogeneity among reliability coefficients was assessed (e.g., standard deviation, <i>Q</i> statistic, <i>I</i> ² index, between-studies variance, 75% rule of Hunter-Schmidt). If relevant, specify the between-studies variance estimator (DerSimonian and Laird, Maximum Likelihood, Restricted Maximum Likelihood, Empirical Bayes, Paule and Mandel), as well as how confidence intervals, credibility intervals, or prediction intervals were calculated.
16. Moderator analyses	If relevant, describe how the influence of moderator variables was assessed (e.g., subgroup analyses, meta-regression analyses, correlational analyses).
17. Additional analyses	Describe other additional analyses accomplished, such as sensitivity analyses (e.g., statistical analyses with transformed and untransformed reliability coefficients, one-to-one deleting of reliability coefficients, assessment of publication bias, reporting biases, and other sources of bias).
18. Software	Mention the software and version used to carry out the statistical analyses (e.g., metafor in R, Proc MIXED in SAS, Comprehensive Meta-analysis).
RESULTS	Yes No Unclear NA
19. Results of the study selection process	Describe, ideally with a flow chart, the selection process of the studies, specifying the number of studies identified from each search source, excluded studies and reasons why, and the number of studies that reported and induced reliability of test scores. Regarding reliability induction, report induction rates, distinguishing between induction “by omission” and “by report” (see e.g., REGEMA flowchart). Furthermore, it is advisable to compare the reliability induction rates as a function of variables such as publication year, country/continent and study purpose (psychometric vs. applied).
20. Mean reliability and heterogeneity	Present pooled reliability coefficients and confidence/credibility intervals for the scale (and subscales, if applicable) and for each type of reliability (e.g., internal consistency, temporal stability, inter- and intra-rater agreement). In case of applying any transformation of the reliability coefficients, results should be back-transformed to the original metric to facilitate interpretation. Illustrate the distribution of reliability coefficients with graphical techniques (e.g., forest plots, box plots, stem and leaf displays, histograms) and describe the degree of heterogeneity by one or more heterogeneity measures (see Item 15).

21. Moderator analyses	For categorical moderators, provide the pooled reliability coefficient, confidence interval and other heterogeneity measures for each category of the moderator. For continuous moderators, include the regression coefficients, standard errors and confidence limits. For both types of moderators, report results of the statistical significance tests, misspecification tests, and proportion of variance accounted for. As a further step, it is advisable to fit a predictive/explanatory model including the most relevant moderator variables.				
22. Sensitivity analyses	Report or describe the results of any sensitivity analyses conducted (see Item 17).				
23. Comparison of inducing and reporting studies	If performed, present the results of comparing the characteristics of inducing and reporting studies (e.g., sociodemographic and clinical characteristics of the samples).				
24. Data set	Tabulate the characteristics of the individual studies that reported reliability (see Item 7). Tables can be presented as appendices or supplementary files. In addition, list of all studies included in the RG meta-analysis, either in the reference section or as a supplementary file.				
DISCUSSION		Yes	No	Unclear	NA
25. Summary of results	Present the main results, such as mean reliability exhibited by the scale/test and moderators of the reliability coefficients. If available, discuss the results in the light of previous evidence.				
26. Limitations	Discuss the limitations of the meta-analysis. Include an explicit statement of the reliability induction rates and the extent to which inducing and reporting studies are comparable in terms of samples characteristics.				
27. Implications for practice	Provide guidelines for professional practice regarding the usefulness of the scale/test in different settings and target populations.				
28. Implications for future research	Include recommendations for researchers regarding the conditions under which the scale/test should be applied.				
FUNDING		Yes	No	Unclear	NA
29. Funding	State the financial sources of the meta-analysis, as well as potential conflict of interests of the authors.				
PROTOCOL		Yes	No	Unclear	NA
30. Protocol	State whether a protocol of the meta-analysis was previously published or made accessible in some web-site (e.g., in Prospero).				

Note. NA: Not Applicable.