

Test Eşitleme Çalışmaları Üzerine Bir Bibliyometrik Analiz

Ceren MUTLUER 

Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Bolu, Türkiye cmutluer@yandex.com (Sorumlu Yazar/Corresponding Author)

Makale Bilgileri

ÖZ

Makale Geçmişi

Geliş: 23.08.2023

Kabul: 08.10.2023

Yayın: 29.10.2023

Anahtar Kelimeler:

Test Eşitleme,
Bibliyometrik Analiz,
Eşitleme Yöntemleri,
Web of Science Veri
Tabanı,

Bu çalışmada Web of Science veri tabanında yer alan İngilizce veya Türkçe dergilerde yayımlanan test eşitleme anahtar kelimesinin geçtiği araştırmalarının genel yönelimibibliyometrik analiz yoluyla incelenmek istenmiştir. Web of Science veri tabanında anahtar kelime olarak "test eşitleme" kavramı taranmıştır. Uygunluk kriterlerini karşılayan 247 çalışma, araştırmaya dahil edilmiştir. R-Studio, bibliyometrik paketi kullanılmıştır. Veriler WoS'tan BibTeX dosyası olarak elde edilmiştir. 1972-2023 yılları arasında yapılan çalışmalar bu araştırma için belirtilen anahtar kavramı içerme durumlarına göre incelenmiştir. Bibliyometrik analiz yapılarak incelenen bu kavram en fazla 2011 yılında 30 çalışmada yer aldığı ve Web of Science veri tabanında yayımlandığı görülmüştür. Ülkelere göre kıyaslama yapıldığında USA'da tamamlanan çalışma sayısı diğer ülkelere göre daha fazladır. Atıflar incelendiğinde en fazla atıf von Daiver'e (2004) aittir. En fazla çalışma Applied Psychological Measurement dergisinde yayımlanmıştır. Üretilen çalışmalardaki konu ve temalar incelendiğinde ise "linking (İlişkilendirme/bağlama)" kavramı, "Madde Tepki Kuramı", "Puan", "Modeller" gibi kavramlar anahtar kelime olarak test eşitleme kavramıyla birlikte ele alınan kavramlardır.

A Bibliometric Analysis on Test Equating Studies

Article Info

ABSTRACT

Article History

Received: 23.08.2023

Accepted: 08.10.2023

Published: 29.10.2023

Keywords:

Test Equating,
Bibliometric
Analysis,
Test Equating
methods,
Web of Science
database

In this study, it was aimed to examine the researches published in English or Turkish journals in the Web of Science database with the keyword test equalization by means of general orientation bibliometric analysis. The concept of "test synchronization" was scanned as a keyword in the Web of Science database. 247 studies that met the eligibility criteria were included in the study. R-Studio used bibliometric package. The data was obtained from WoS as a BibTeX file. Studies conducted between 1972 and 2023 were examined according to their inclusion of the key concept specified for this research. This concept, which was examined by bibliometric analysis, was included in 30 studies in 2011 and published in the Web of Science database. When compared by country, the number of studies completed in the USA is higher than in other countries. When the citations are examined, the most citations belong to von Daiver (2004). The largest number of studies has been published in the journal Applied Psychological Measurement. When the topics and themes in the produced studies are examined, concepts such as the concept of "linking / linking", "Item Reaction Theory", "Score", "Models" are the concepts that are discussed together with the concept of test equalization as a keyword.

Atıf/Citation: Mutluer, C. (2023). Test eşitleme çalışmaları üzerine bir bibliyometrik analiz. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF) Dergisi*, 5(3), 1451-1463.



"This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0)"

GİRİŞ

Eğitim kurumlarında bireylerin mevcut durumları, yeterlilikleri ve özellikleri en doğru ve hatasız şekilde ölçülmesi beklenir. Bu süreçte kullanılan ölçme araçlarının sonuçlarına bakılarak mevcut durum değerlendirilir. Bu sonuçlar, bireylerin meslek seçiminde, bir üst kuruma yerleştirilmesinde, başarı düzeyinin tespitinde kullanılmaktadır. Bazı sınavların sonuçları birkaç yıl geçerli olup farklı zamanlarda da kullanılır. Aynı sınava girmeyen kişilerin sınav sonuçlarının sadece niceliksel olarak karşılaştırmak doğru bir yaklaşım değildir. Sınav sonuçlarına göre doğru bir karşılaştırma yapabilmek için güvenilir ve geçerli testlerin kullanılması gerekmektedir. Sınav sonuçlarının karşılaştırılabilir, birbiri yerine kullanılabilir olması için istatistiksel bir yaklaşım önerilir. Bu istatistiksel yöntem, test eşitleme yöntemidir.

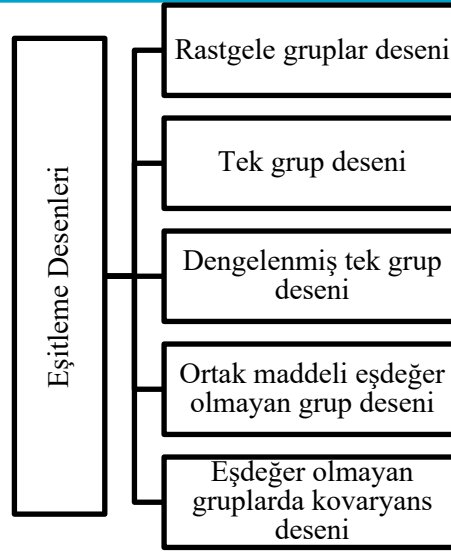
Test eşitleme, temelde iki ya da daha fazla test formundaki ham veya ölçeklenmiş puanlar arasındaki ilişkiyi belirleme sürecidir (Skaggs & Lissitz, 1986). Başka bir tanımla test eşitleme, iki veya daha fazla testten elde edilen puanlar arasındaki ilişkileri kurmak veya iki veya daha fazla testi ortak bir ölçüğe yerleştirmek için istatistiksel bir prosedür (Hambleton & Swaminthan, 1985).

Eşitleme sürecinde Eğitimsel ve Psikolojik Test Standartları (Standards for Educational and Psychological Testing (Amerikan Psikoloji Derneği, 1985) kapsamında karşılaştırılabilir puanlar “ölçeklenmiş puanlar” “karşılaştırılabilir puanlar” olarak kullanılsa da eşitleme sürecinde bu ifadeye “eşitlenmiş puanlar” ifadesi kullanılmaktadır. Eşitlenmiş puanların hesaplanması için geçen süreçte aynı kavramsal yapı için kullanılan X formu ve Y formunun birbiri yerine kullanılması için birtakım aşamalar takip edilmelidir. Bu aşamalar şunlardır:

1. Eşitlemek için amaç tespiti
2. Testin aynı içeriğine ve benzer istatistiksel özelliklerine uygun alternatif formların oluşturulması
3. Eşitleme için uygun desen seçilmesi
4. Seçilen uygun desenin uygulamasını yapılması
5. Eşitleme için bir ya da daha fazla operasyonel tanım seçilmesi
6. Bir ya da daha fazla istatistiksel tahmin yöntemi seçilmesi
7. Eşitleme sonuçlarını değerlendirilmesidir (Kolen ve Brennan, 2014).

Aşamalar olarak takip eden bu süreçte temelde eşitleme iki kategoriye ayrılmaktadır. Bunlardan ilki yatak diğeri de dikey eşitlemedir (Hambleton ve Swaminathan, 2013; Kolen, 1988). Yatay eşitlemede aynı kavramsal yapı için hazırlanmış farklı formlardan elde edilen puanların birbirleri ile kıyaslanması ve karşılaştırılmasına odaklanılırken; dikey eşitlemede içerik olarak aynı olmaksızın çeşitli eğitim kademelerinde değişen içeriklerin ve farklılaşan madde güçlük düzeylerinin olduğu test formlarının karşılaştırılması amaçlanır.

Eşitleme sürecinde formlardan elde edilen puanların karşılaştırılmasında desen seçimi çok önemlidir. Test eşitlemede veri toplama süreci “eşitleme deseni” olarak adlandırılmakta ve beş alt başlıkta açıklanmıştır (Kolen ve Brennan, 2014).



Şekil 1. Eşitleme desenleri

Rastgele gruplar deseninde bireylerin aldıkları formlar random bir şekilde verilir. Formların uygulanma süreci random olarak bireylere verildiği için ikinci formun gücüne göre bireylere sınav zor ya da kolay gelebilir. Bu nedenle rastgele gruplar deseninde sıra etkisi çalışma sonucunu etkileyen bir unsurdur.

Tek grup deseninde aynı kişilere uygulanan formlar için yapılan bir eşitleme sürecidir. Aynı kişilerin farklı formlardan aldıkları sonuçlar test eşitleme ile karşılaştırılabilir.

Dengelenmiş tek grup deseni, rastgele gruplar desenindeki sıra etkisini yok etmek için uygulanmaktadır. Birinci gruba verilen X formundan sonra Y formu, ikinci gruba verilen Y formundan sonra X formu uygulanır. Tan (2015) denek sayısının az olduğu ve test formları çok uzun olmadığında bu desenin kullanımını önermiştir.

Test formlarında ortak maddeler kullanılarak da test puanları eşitlenebilir. Bu eşitleme deseni, ortak maddeli eşdeğer olmayan grup deseni olarak açıklanır. Test formlarında aynı sırada ve ortak maddeler (ankor) kullanılarak farklı form puanları eşitlenmektedir. Bu ankor maddeler test formunun en az %20'sini kapsamaktadır (Petersen vd., 1989). Ortak maddeli eşdeğer olmayan gruplar deseninde farklı yetenek düzeyindeki bireylerin puanlarının eşitlenmesinde kullanılmaktadır. Ortak maddeli eşdeğer olmayan gruplar deseni için ortak maddeler üzerinden test puanlarının eşitlenmesi istenir. Uygulanan sınavlarda ortak maddeli formların bulunma durumu yaygın değildir. Farklı yetenekteki bireylerin puanları bu desen ile eşitlenmeye çalışılır. Bir yapının zaman içerisindeki değişimini incelemek için ortak test kullanılıyorsa ve bu ortak testler sınırlı sayıda test konu alanını kapsıyorsa yine problem teşkil etmektedir. Ortak testin yokluğunda, eşdeğer olmayan gruplarda kovaryans desen önerilmektedir (Wiberg ve Branberg, 2015). Bu desende, ortak test yerine test puanları ile yüksek korelasyon gösteren değişken kullanılması önerilmektedir.

Açıklanan bu desenlere göre test kuramlarına göre eşitleme yöntemleri çeşitlenmektedir. Ölçme kuramlarından ilki olarak Klasik Test Kuramı (KTK) bünyesinde temelde üç eşitleme yöntemi bulunmaktadır. Bunlar ortalama, lineer ve eşityüzelikli eşitleme yöntemleridir. Diğer ölçme kuramı ise Madde Tepki Kuramı (MTK) bünyesinde gerçek ve gözlenen puan eşitleme yöntemlerini barındırır. MTK'da eşitleme prosedüründen önce puanları aynı ölçek düzeyine dönüştürmek için kalibrasyon yapılmaktadır. MTK'ya dayalı momentler (ortalama-ortalama, ortalama-standart sapma) ve karakteristik eğri (Haebara, Stocking-Lord) yöntemleri kullanılarak kalibrasyon yapılmaktadır.

Testleri eşitleme yöntemlerine yönelik araştırmalar, eğitimsel ve psikolojik ölçüm alanının kendi haline geldiği 1920'lerin başından beri (Flanagan, 1982) devam eden bir süreçtir. Literatür tarandığında KTK bünyesinde eşitleme yöntemlerinden "en az hata ile eşitlenmiş puan elde edilen yöntem hangisidir?" sorusuna yanıt aramaya yönelik tamamlanmış eşitleme çalışmaları (Cope, 1987; Klein ve

Jarjova, 1985; Mutluer ve Nartgün, 2017; Özdemir, 2017; Pektaş ve Kılınç, 2016; Sezer Başaran, 2023; Sezer Başaran vd., 2023; Tan, 2015; von Davier ve Kong, 2005) yapılmıştır. MTK'ya dayalı eşitleme çalışmaları da (Brossman ve Lee, 2013; Demirus, 2015; Gök ve Kelecioğlu, 2014; Gündüz, 2015; Salmaner Doğan ve Tan, 2022; Yıldırım Seheryeli ve diğerleri, 2021; Yıldırım, Gündüz ve İnce Araacı, 2021; Yurtçu & Güzeller, 2018) kalibrasyon süreçlerine dayalı olarak eşitlenmiş puanların elde edilme sürecini ortaya koymaktadır. Her iki kuramında kullanılıp karşılaştırma yapan çalışmalara da (Han ve diğerleri, 1997; Kolen, 1981; Mutluer, 2021; Tanberkan Suna ve Tan, 2017) rastlanmıştır.

Farklı değişkenler (testlerin normalligi, değişen madde fonksiyonu gibi) açısından test eşitleme çalışmaları yapılmaktadır. Test eşitleme konusunda çalışma yelpazesinin geniş olması da bu konuda yapılan çalışmalardaki yönelimi izlemek ve geleceğe yönelik çalışma konuları önermek için daha kapsamlı incelemelere ihtiyaç vardır. Mevcut çalışma, bu geniş çalışma alanına kapsayıcı bir genel bakış sunmak amacıyla incelemenin kapsamını genişletmek için bibliyometrik analizden yararlanılmıştır.

Bibliyometrik Analiz

Bibliyometrik analiz, ilgili alanda yapılmış çalışmaların (kitap, kitap bölümü, makale, henüz yayımlanmamış erken görünümlü yayınları) yer aldığı literatürü değerlendirmek amacıyla yapılan bir analiz tekniğidir. Donthu ve diğerlerine (2020) göre araştırmacılar, bibliyometrik analizi makale ve dergi performansında ortaya çıkan eğilimleri, işbirliği modellerini ve araştırma bileşenlerini ortaya çıkarmak ve mevcut literatürdeki belirli bir alanın entelektüel yapısını keşfetmek gibi çeşitli nedenlerle kullanırlar. Bibliyometrik analizde merkezde yer alan veriler, çok büyük ve doğası gereği objektif (örneğin, alıntı ve yayın sayısı, anahtar kelimelerin ve konuların varlığı) olma eğilimindedir (Donthu vd., 2021). Ancak yorumları genellikle her iki nesnel veriye de dayanır. Büyük hacimli yapılandırılmamış verileri titizlikle anlamlandırarak, birikimli bilimsel bilginin ve köklü alanların evrimsel nüanslarının şifresini çözmek ve haritalamak için kullanışlıdır (Pritchard, 1969). Bu nedenle, iyi yapılan bibliyometrik çalışmalar, bir alanı yeni ve anlamlı yollarla ilerletmek için sağlam temeller oluşturabilir.

Araştırmanın Amacı ve Alt Problem Cümleleri

Bu çalışmada test eşitleme çalışmalarına ilişkin genel bir yönelim tespiti amaçlanmıştır. Bu çalışmada bibliyometrik analiz kullanılarak test eşitleme konularındaki trend konular ve çalışma önerilerine odaklanılmıştır.

Araştırma kapsamında aşağıdaki alt problem cümlelerine yanıt aranmıştır:

- 1- Test eşitleme konusunda yapılan çalışmaların yıllara göre dağılımı nasıldır?
- 2- Test eşitleme konusunda yapılan çalışmaların yazarların buldukları ülkelere göre dağılımı nasıldır?
- 3- Test eşitleme çalışmalarına yönelik atıflar incelendiğinde en fazla atıf hangi yazar(lar)a yapılmıştır?
- 4- Test eşitleme konulu çalışmalar en çok hangi dergide yayımlanmıştır?
- 5- Test eşitleme konusunda en güncel konular nelerdir?

YÖNTEM

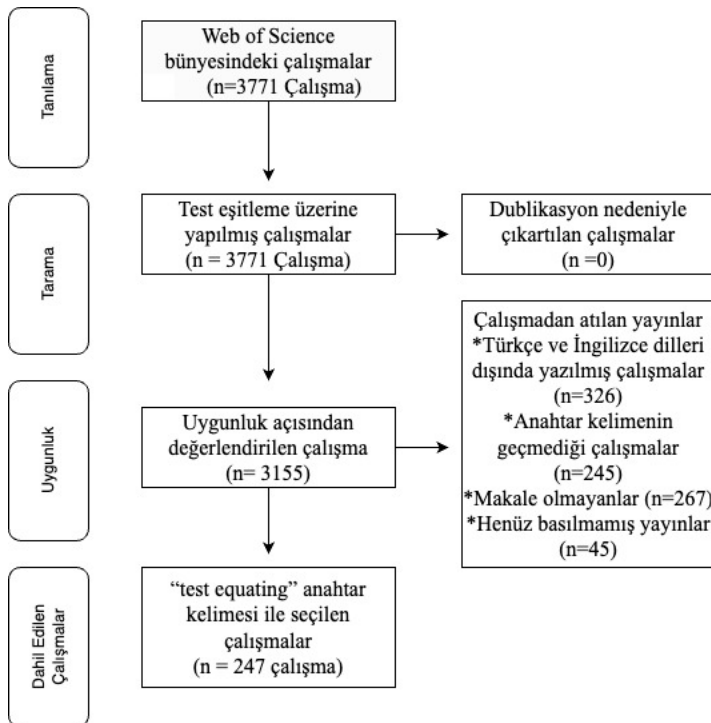
Araştırma Modeli

Araştırmada test eşitleme çalışmalarının Web of Science (WoS) veri tabanı bibliyometrik analizle incelenmiştir. Bibliyometrik analiz, özellikle bilimsel ve uygulamalı alanlarda araştırma değerlendirme metodolojisinin önemli bir parçasıdır (Ellegaard & Wallin, 2015). Bibliyometrik analiz, makale ve dergi performansında ortaya çıkan eğilimleri, işbirliği modellerini ve araştırma bileşenlerini ortaya çıkarmak ve mevcut literatürdeki belirli bir alanın entelektüel yapısını keşfetmek gibi çeşitli nedenlerle kullanılmaktadır

(Donthu vd., 2021). Web of Science veri tabanında 1972-2023 yılları arasında yayımlanan akademik çalışmaların betimsel analizini içermesi nedeniyle kesitsel türde araştırma olarak da tanımlanmıştır (Büyüköztürk vd., 2008).

Yayınların Seçimi

Literatür taramalarında nitelik olarak bilim alanında en iyi çalışmalar WoS veri tabanındaki dergilerde yayımlanmaktadır. Bu veri tabanında “test eşitleme” temalı çalışmalar taranmıştır. İlk aşamada 3771 çalışmaya rastlanırken, makale olarak yayımlanmak, dil seçeneği (Türkçe veya İngilizce) olarak kısıtlandığında 3253 çalışmaya rastlanmıştır. Bu çalışmaların 10 tanesi Türkçe geri kalan 3243 çalışma ise İngilizce olarak yazılmıştır. Bu çalışmada yer alan çalışmaların seçimi Şekil 1’de sunulmuştur. Araştırmada “test equating” kelimesi ile filtreleme yapıldığında ise bu çalışma için 247 bilimsel çalışma ayıklanmıştır. Çalışmadaki verilerin seçiminde Page ve diğerleri (2010) PRISMA akış diagramı Şekil 2’de verilmiştir.



Şekil 2. Araştırmaya seçilen çalışmaların akış şeması

Verilerin Analizi

Araştırma kapsamında test eşitleme çalışmalarındaki genel yapıyı derinlemesine incelemek için WoS veri tabanındaki veriler öncelikle BibTeX formatına dönüştürülmüştür. Analiz sürecinde R-Studio’da (R Core Team, 2021) mevcut “bibliometrix” paketi (Aria & Cuccurullo, 2017) kullanılmıştır. Elde edilen veriler yüzde ve frekans değerleri açısından değerlendirilmiştir.

BULGULAR

Yıllara Göre Yayın Dağılımı

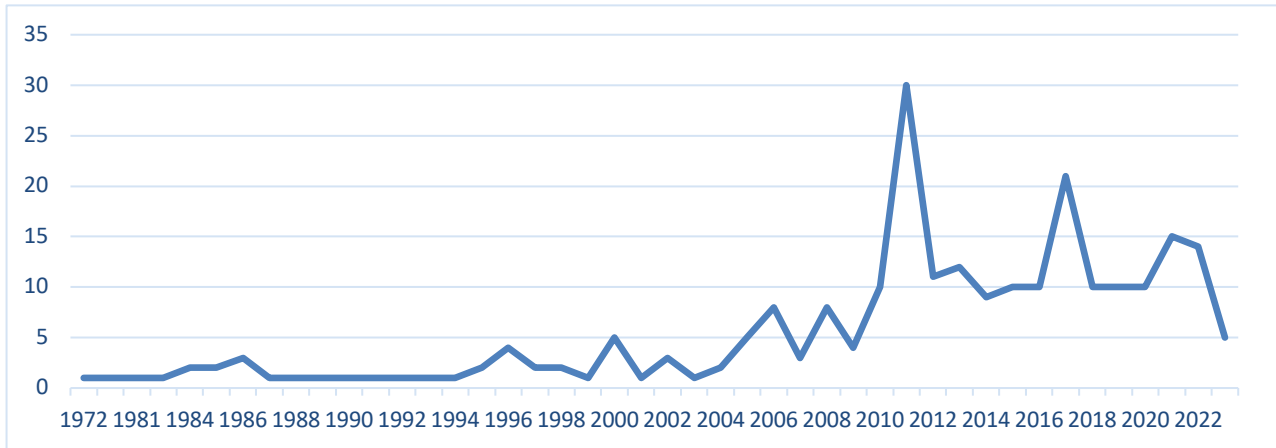
Araştırmada yıllara göre WoS veri tabanında test eşitleme konu başlıklı çalışmaların dağılımı aşağıdaki tabloda verilmiştir.

Tablo 1. Test Eşitleme Konularına İlişkin Yayınların Yıllara Göre Dağılımı

Yıllar	Yayınlar	Yüzde	Yıllar	Yayınlar	Yüzde
1972	1	0.405	2002	3	1.215
1978	1	0.405	2003	1	0.405

1981	1	0.405	2004	2	0.81
1983	1	0.405	2005	5	2.024
1984	2	0.81	2006	8	3.239
1985	2	0.81	2007	3	1.215
1986	3	1.215	2008	8	3.239
1987	1	0.405	2009	4	1.619
1988	1	0.405	2010	10	4.049
1989	1	0.405	2011	30	12.146
1990	1	0.405	2012	11	4.453
1991	1	0.405	2013	12	4.858
1992	1	0.405	2014	9	3.644
1993	1	0.405	2015	10	4.049
1994	1	0.405	2016	10	4.049
1995	2	0.81	2017	21	8.502
1996	4	1.619	2018	10	4.049
1997	2	0.81	2019	10	4.049
1998	2	0.81	2020	10	4.049
1999	1	0.405	2021	15	6.073
2000	5	2.024	2022	14	5.668
2001	1	0.405	2023	5	2.024

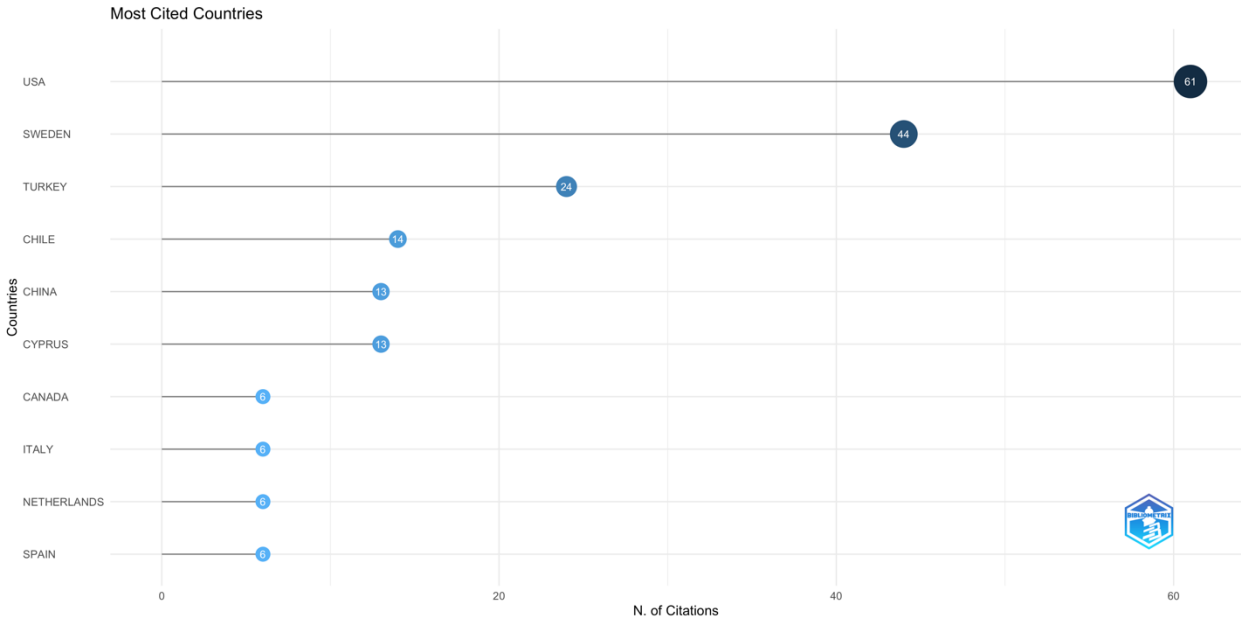
Test eşitleme kapsamında WoS veri tabanındaki yıllara göre dağılımı yukarıdaki tabloda sunulmuştur. Tablo incelendiğinde en fazla yayın 30 çalışma ile (% 12.146) 2011 yılında yapılmıştır. En fazla ikinci yıl 21 çalışma ile (% 8.502) 2017 yılında gerçekleşmiştir. 1972 ile 2023 yılları arasında 1972, 1978, 1981, 1983, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1999 ve 2001 yıllarında 1 çalışma ile (% 0.405) en az yayın yapılan yıllardır. Grafikselsel olarak gösterim Şekil 3'te verilmiştir.



Şekil 3. Yıllara göre yayın dağılımı

Ülkelere Göre Yayın Dağılımı

Test eşitlemeye ilişkin kaynak kitapların ve yönelimin çıkış noktası ABD olmasıyla birlikte yapılan yayınların ülkelere göre dağılımı araştırılmıştır. WoS veri tabanlı yapılan bibliyometrik analizde en fazla yayın yapmış ilk on ülkenin yayın sayıları Şekil 4'te verilmiştir.

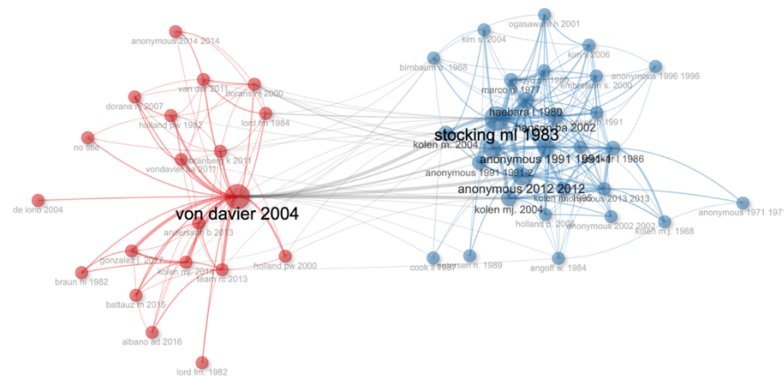


Şekil 4. Ülkelere göre yayın dağılımı

Şekil 4'te sıralamadaki ilk on ülke verileri görselleştirilmiştir. Yapılan bibliyometrik analiz sonucunda ülkelere göre yayın dağılımında en fazla yayın 61 çalışma ile (%24.70) Amerika'da yapılmıştır. Test eşitleme üzerine yapılan çalışmalar ülkelere göre sıralamasında ikinci sırada 44 çalışma ile (%17.81) İsveç, üçüncü sırada 24 çalışma ile (% 9.72) Türkiye, dördüncü sırada 14 çalışma ile (%5.67) Şili, beşinci sırada 13 çalışma ile (%5.26) Çin, altıncı sırada 13 çalışma ile (%5.26) Kıbrıs, yedinci sırada 6 çalışma ile (%2.43) Kanada, sekizinci sırada 6 çalışma ile (%2.43) İtalya, dokuzuncu sırada 6 çalışma ile (%2.43) Hollanda ve onuncu sırada 6 çalışma ile (%2.43) İspanya yer almaktadır.

Atıf Yapılan Yazarlara Göre Yayın Dağılımı

Kuram temelli eşitleme yöntemlerinin karşılaştırılması, incelenen değişken çeşitliliği açısından pek çok araştırmada atıf yapma yönelimleri Şekil 5'te gösterilmiştir.

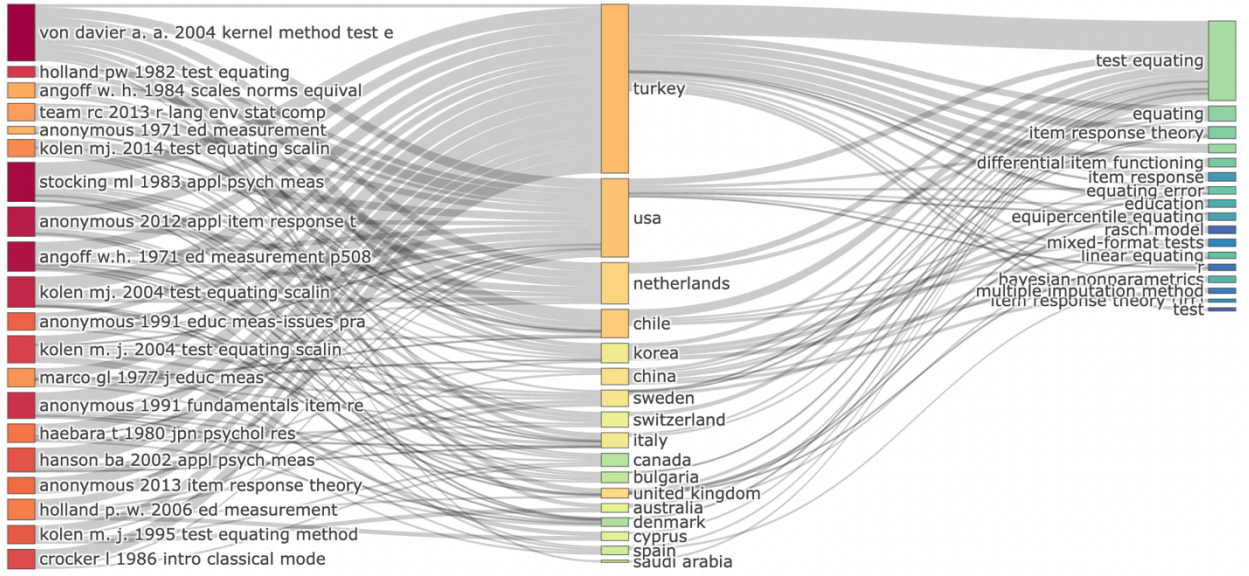


Şekil 5. Atıf Yapılan Yazarlara göre yayın dağılımı

Şekil 5'te bibliyometrik analiz sürecine dahil edilen çalışmalardaki atıflar incelendiğinde von Davier (2004) ve Stocking (1983) çalışmalarını sıklıkla referans olarak gösterilmiştir. Şekil 5'te bu iki çalışma dışında alana katkı getiren Kolen (2004) gibi çalışmaların da belitilen bu iki çalışma kadar atıf almasa da atıf aldığı

görülmektedir.

Detaylı olarak gösterim Şekil 6'da verilmiştir.



Şekil 6. Atıf yapılan yayın, ülke ve anahtar kelimeler

Test eşitleme çalışmaları arasında en fazla atıf von Davier (2004) çalışmasına yapıldığı Şekil 6'da gösterilmiştir. Şekil 6'daki sol kısım atıf yapılan yayınları, orta kısımda atıf yapan ülkeleri, sağ kısımda ise atıf yapılırken seçilen anahtar kelimeleri belirtmektedir. Şekil 6 incelendiğinde von Davier (2004) çalışmasına en fazla Amerika'dan atıf yapıldığı görülmektedir. Von Davier (2004) çalışmasına yapılan atıflar incelendiğinde anahtar kelime olarak "test eşitleme", "eşitleme", "Madde Tepki Kuramı", "Değişen madde fonksiyonu", "eşityüzelikli eşitleme" gibi anahtar kelimelerin seçildiği görülmektedir. Türkiye'de en fazla Stocking (1983) çalışmasına atıf yapıldığı ve anahtar kelimeler olarak "test eşitleme", "Madde Tepki Kuramı", "eşitleme hatası" gibi kavramlar seçilmiştir.

Dergilere Göre Yayın Dağılımı

Çalışmaların yayımlandığı dergiler incelenmiş ve Tablo 2'de sunulmuştur.

Tablo 2. Dergilere Göre Yayın Dağılımları

Sıra	Dergi Adı	f	%
1	Applied Psychological Measurement	29	11.741
2	Statistics For Social and Behavioral Sciences	22	8.907
3	Statistical Models for Test Equating Scaling and Linking	21	8.502
4	Psychometrika	17	6.883
5	Applying Test Equating Methods Using R	12	4.858
6	Methodology of Educational Measurement and Assessment	12	4.858
7	Educational and Psychological Measurement	9	3.644
8	Journal of Measurement and Evaluation in Education and Psychology	9	3.644
9	Applied Measurement in Education	8	3.239
10	Journal of Educational and Behavioral Statistics	8	3.239

WoS veri tabanında yer alan çalışmalar incelendiğinde test eşitleme çalışmalarının en fazla ($f=29$, %11.741) “Applied Psychological Measurement” dergisinde yayımlanmıştır. Yayımlanan çalışmalarda ikinci sırada ($f=22$, %8.907) “Statistics For Social and Behavioral Sciences” dergisi yer almaktadır. Spesifik olarak test eşitleme çalışmalarına vurguda bulunan “Statistical Models for Test Equating Scaling and Linking” adlı dergide yayımlanan çalışmalar ise üçüncü sırada ($f=21$, %8.502) bulunmaktadır. Diğer dergilerdeki yayın sayıları sırasıyla şu şekildedir; dördüncü sırada ($f=17$, %6.883) “Psychometrika”, beşinci sırada ($f=12$, %4.858) “Applying Test Equating Methods Using R”, altıncı sırada ($f=12$, %4.858) “Methodology of Educational Measurement and Assessment”, yedinci sırada ($f=9$, %3.644) “Educational and Psychological Measurement”, sekizinci sırada ($f=9$, %3.644) “Journal of Measurement and Evaluation in Education and Psychology”, dokuzuncu sırada ($f=8$, %3.239) “Applied Measurement in Education” ve onuncu sırada ($f=8$, %3.239) “Journal of Educational and Behavioral Statistics” dergisi yer almaktadır.

Güncel Konu Başlıklarına Göre Yönelimler

“Test eşitleme” anahtar kelimesiyle birlikte açıklanan ve güncel konu başlıkları incelendiğinde Şekil 7’de yer alan güncel konu başlıklarının sıklıkla çalışma konusuna dönüştüğü görülmüştür.



Şekil 7. Sıklıkla kullanılan anahtar kelimeler

Çalışmada çeşitli veri tabanlarında araştırma konusu hakkında tarama yaparken sıklıkla anahtar kelimeler araştırılmaktadır. WoS veri tabanından elde edilen veriler ayıklandığında test eşitleme kavramıyla birlikte çalışılan konular Şekil 7’de renklendirilerek sunulmuştur. Eşitleme kavramı ile “linking (İlişkilendirme/bağlama)” kavramı, “Madde Tepki Kuramı”, “Puan”, “Modeller” gibi kavramların sıklıkla kullanıldığı görülmektedir.

SONUÇ

Bu çalışma, WoS veritabanında yer alan 247 makaleye dayanarak test eşitleme araştırmalarına genel bir bakış açısı sunmak için yapılmıştır. İncelenen veri tabanındaki çalışmaların yıllar içinde bilimsel üretkenliği, test eşitleme literatürüne en çok katkıda bulunan ülkelere, en çok atıf yapılan yazarlara, dergilere ve alandaki etkili konu ve temalara ilişkin bulgular sunulmuştur.

Analiz verileri “test equating” anahtar kelimesinin WoS veri tabanı kullanılarak elde edilmiştir. 1972 ile 2023 yılları arasındaki çalışmalar BibTeX olarak R-Studio’da bibliyometrix paketi ile analiz edilmiştir. 2010 yılından itibaren bu konuda yapılan çalışma sayısında artış göstermiştir. 1994 yılında dahil olmak üzere

her sene bir yayın WoS veri tabanında yayımlandığı, 2000’li yıllardan sonra bu konuya yönelimin arttığı görülmektedir. Kuramsal olarak MTK kapsamında eşitleme yöntemlerinin daha az hata ile sonuçlanması, geliştirilen R-equate (Albano, 2016) gibi ek analiz paketlerinin kullanıcı kolaylığı sağlaması da bu sayıyı arttırmada etken olabilir. Bu durumla birlikte ALES gibi farklı zamanlarda yapılan sınavların geçerlilik sürelerinin uzatılması da elde edilen sonuçların karşılaştırılabilirliği açısından yeni test eşitleme çalışmalara ışık tutmuştur.

Yayımlanan çalışmalar ülkelere göre incelendiğinde test eşitleme alanında en fazla çalışma Amerika’da gerçekleşmiştir. Bu sırayı sırasıyla İsveç, Türkiye, Şili, Çin, Kıbrıs, Kanada, İtalya, Hollanda ve İspanya izlemektedir.

WoS veri tabanında yapılan çalışmalardaki atıflar incelendiğinde en fazla atıf von Davier ‘ın (2004) yapmış olduğu ‘Kernel eşitleme’ konusundaki çalışmaya yapıldığı görülmektedir. MTK’daki kalibrasyon yöntemlerinde de sıklıkla vurgulanan Stocking’in (1983) çalışması, von Davier’in (2004) gerçekleştirdiği çalışmadan sonra en fazla atıf gösterilen çalışma olmuştur. Atıflarla birlikte anahtar kelimeler incelendiğinde “test eşitleme”, “MTK”, “eşitleme”, “değişen madde fonksiyonu”, “eşitleme hatası”, “eğitim”, “eşityüzelikli eşitleme”, “karma test formu”, “Rasch modeli” gibi anahtar kelimeler sıklıkla kullanılmıştır.

Dergilerdeki test eşitleme başlıklı çalışmalar incelendiğinde en fazla yayın “Applied Psychological Measurement” dergisinde yayımlanmıştır. Diğer yayımlanan dergiler sırasıyla “Statistics For Social and Behavioral Sciences”, “Statistical Models for Test Equating Scaling and Linking”, “Psychometrika”, “Applying Test Equating Methods Using R”, “Methodology of Educational Measurement and Assessment”, “Educational and Psychological Measurement”, “Journal of Measurement and Evaluation in Education and Psychology”, “Applied Measurement in Education” ve “Journal of Educational and Behavioral Statistics” dergileridir. Dergilerin impact faktör değerleri incelendiğinde ise atıf sayısının oldukça fazla olduğu görülmektedir.

İncelenen veri tabanında test eşitleme kavramıyla birlikte çalışılan diğer konu ve tema başlıklarına bakıldığında “linking (İlişkilendirme/bağlama)” kavramı, “Madde Tepki Kuramı”, “Puan”, “Modeller” gibi kavramların sıklıkla kullanıldığı görülmektedir. İncelenen temalar arasında Klasik Test Kuramının dezavantajlı durumları için pratik çözümler üreten MTK modelleri kullanılarak eşitleme yöntemleri çeşitli eşitleme desenleri bünyesinde denenmektedir. Son yıllarda rasch modeline dayalı yetenek parametrelerinin tahmini için maksimum olabilirlik algoritması kullanılmaya yönelik çalışmalar mevcuttur (Wolfe, 2000).

MTK bünyesindeki üç parametrelilik için marjinal maksimum olabilirlik (Mislevy ve Bock, 1982) ve Bayesian (Swaminathan ve Gifford, 1983) algoritmaları geliştirilmiştir. Bu yaklaşımların bir tartışması ve karşılaştırması Lord’da (1984) bulunabilir. Günümüzde parametre tahminine yönelik eşitleme araştırmaların çoğu üç parametrelilik modelle ilgili olduğu görülmektedir.

Bu çalışma kapsamında yer alan çalışmalarda MTK’ya dayalı eşitleme yöntemlerine doğru yönelimin olduğu görülmektedir. Ayrıca eşitleme hatalarının hesaplanmasına dayalı olarak standart hataların hesaplanması konusunda farklı fikirler ortaya atılmıştır. Ogasawara (2000) MTK’ya dayalı eşitlemede asimptotik standart hatalara odaklanmıştır. Andersson (2016) zincirleme eşitleme çerçevesinde çok kategorili MTK modelleri kullanarak gözlenen puan eşitlemenin asimptotik standart hata değerlerini hesaplamaya çalışmıştır. Battauz (2017) yaptığı çalışmada eşitlemedeki standart hatayı daha kesin bir şekilde hesaplamak için delta yönteminin daha doğru sonuç vereceğini belirtmiştir.

Şu ana kadar yapılan eşitleme araştırmasının sonuçları incelendiğinde eşitleme yöntemlerinin kullanım durumları karıştırmacı görünmektedir. Burada incelenen eşitleme çalışmalarından, tüm uygulamalarda bir yöntemin diğerlerinden üstün olmasını beklemenin gerçekçi olmadığı açıktır. Burada incelenen çalışmaların çoğu bu açılarından bir veya daha fazlası açısından farklılık gösterdiğinden, çalışmalar arasında genelleme yapmak zordur.

KAYNAKÇA

- Albano, A. D. (2016). Equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74, 1-36. <http://doi.org/10.18637/jss.v074.i08>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Andersson, B. (2016). Asymptotic standard errors of observed-score equating with polytomous IRT models. *Journal of Educational Measurement*, 53(4), 459–477. <https://www.jstor.org/stable/45148403>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Battauz, M. (2017). Multiple equating of separate IRT calibrations. *Psychometrika*, 82(3), 610–636. <http://doi.org/10.1007/s11336-016-9517-x>
- Brossman, B. G., & Lee, W.-C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement*, 37(6), 460-481. <https://doi.org/10.1177/014662161348408>
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Pegem Akademi Yayıncılık.
- Cope, R. T. (1987). How well do the Angoff Design V linear equating methods compare with the Tucker and Levine methods? *Applied Psychological Measurement*, 11(2), 143-149. <https://doi.org/10.1177/014662168701100202>
- Demirus, K. B. (2015). *Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi*. [Yayımlanmamış Doktora Tezi]. Hacettepe Üniversitesi.
- Donthu, N., Gremler, D. D., Kumar, S., & Pattnaik, D. (2020). Mapping of Journal of Service Research themes: A 22-year review. *Journal of Service Research*. Available at <http://doi.org/10.1177/1094670520977670032>
- Donthu, N., Kumar, S., Pattnaik, D., & Lim, W. M. (2021). A bibliometric retrospection of marketing from the lens of psychology: Insights from Psychology & Marketing. *Psychology & Marketing*, 38(5), 834–865. <https://doi.org/10.1002/mar.21472>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Fischer, GH, & Formann, AK (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement*, 6, 397-416. <https://doi.org/10.1177/014662168200600403>
- Flanagan, J. C. (1982). Discussion of "Some issues in test equating.". Test equating. New York: Academic Press.
- Fraenkel, R.J. & Wallen, E. N. (2006). How to Design and Evaluate Research in Education. McGraw-Hill, New York.
- Gök, B. & Kelecioğlu, H. (2012). Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 10(1), 120-136.
- Gündüz, T. (2015). *Test eşitlemede Madde Tepki Kuramına dayalı yetenek parametresine yönelik ölçek dönüştürme yöntemlerinin karşılaştırılması*. [Yayımlanmamış Yüksek Lisans Tezi]. Gazi Üniversitesi.
- Hambleton, R.K., & Swaminathan, H. (1985). A look at psychometrics in the Netherlands. (Tech. Rep. No TM860514) Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL. (ERIC Document Reproduction Service No. ED273665).
- Hambleton, R. K., & Swaminathan, H. (2013). Item response theory: Principles and applications. New York: Springer Science & Business Media.
- Han, T., Kolen, M., & Pohlmann, J. (1997). A comparison among IRT true-and observed- score equatings and traditional equipercentile equating. *Applied Measurement in Education*, 10(2), 105-121. https://doi.org/10.1207/s15324818ame1002_1
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18(1), 1-11. <https://www.jstor.org/stable/1434813>
- Kolen, M. J. (1988). Traditional equating methodology. *Educational measurement: Issues and practice*, 7(4),

- 29-37. <https://doi.org/10.1111/j.1745-3992.1988.tb00843.x>
- Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices. New York: Springer Science and Business Media.
- Klein, L. W., & Jarjoura, D. (1985). The importance of content representation for common- item equating with nonrandom groups. *Journal of Educational Measurement*, 22(3), 197-206. <https://doi.org/10.1111/j.1745-3984.1985.tb01058.x>
- Lord, F. M. (1984). Maximum likelihood and Bayesian parameter estimation in item response theory (ETS Tech. Rep. No. RR-84-30-DNR). Princeton, NJ: Educational Testing Service.
- Mislevy, R. J., & Bock, R. D. (1982). BILOG—Maximum likelihood item analysis and test scoring: LOGISTIC model. Chicago: International Educational Services.
- Mutluer, C., & Nartgün, Z. (2017). Test equating study concerning to ALES (Academic Personnel And Postgraduate Education Entrance Exam) scores obtained at different times in a year. *European Journal of Education Studies*, 3(12), 96-120. <http://dx.doi.org/10.46827/ejes.v0i0.1270>
- Mutluer, C. (2021). Klasik Test Kuramına ve Madde Tepki Kuramına dayalı test eşitleme yöntemlerinin karşılaştırması: Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 matematik testi örneği. [Yayımlanmamış Doktora Tezi]. Gazi Üniversitesi.
- Özdemir, B. (2017). Equating TIMSS mathematics subtests with nonlinear equating methods using neat design: circle-arc equating approaches. *International Journal of Progressive Education*, 13(2), 116-132.
- Pektaş, S. & Kılınç, M. (2016). PISA 2012 Matematik Testlerinden İki Kitapçığın Gözlenen Puan Eşitleme Yöntemleri İle Eşitlenmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 40, 432-444. <http://doi.org/10.21764/efd.49376>
- 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. *International Journal of Surgery*, 10(89), 1-11. <https://doi.org/10.1186/s13643-021-01626-4>
- R Core Team. (2021). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing.
- Petersen, N. S., Kolen, M. J. ve Hoover, H. D. (1989). Scaling, norming, and equating. *Educational Measurement*, 3, 221-262.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 25(4), 348–349.
- Salmaner-Doğan, R. & Tan, Ş. (2022). Madde tepki kuramında eşitleme hatalarının belirlenmesinde kullanılan delta ve bootstrap yöntemlerinin çeşitli değişkenlere göre incelenmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 42(2), 1053-1081. <https://doi.org/10.17152/gefad.913241>
- Sezer Başaran, E. (2023) Farklı ortak değişkenlerle test eşitlemenin ortak maddeli test eşitlemeyle karşılaştırılması. [Yayımlanmamış Doktora Tezi]. Gazi Üniversitesi.
- Sezer Başaran, E., Mutluer, C. & Çakan, M. (2023). A Comparison of Covariates, Equating Designs, and Methods in Equating TIMSS 2019 Science Tests. *Participatory Educational Research*, 10(5), 41-63. <https://doi.org/10.17275/per.23.74.10.5>
- Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research*, 56(4), 495-529.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied psychological measurement*, 7(2), 201-210.
- Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175-191.
- Yıldırım Seheryeli, M., Yahsi SARI, H. & Kelecioğlu, H. (2021). Comparison of Kernel Equating and Kernel Local Equating in Item Response Theory Observed Score Equating. *Journal of Measurement and Evaluation in Education and Psychology*, 12(4), 348-357. <https://doi.org/10.21031/epod.900843>
- Yıldırım, Y., Gündüz, T., & İnce Aracı, F. G. (2021, Eylül). Madde tepki kuramına dayalı test eşitlemede ortak madde oranının ve madde ayırt ediciliğinin eşitleme hatasına etkisi. 7. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sunulan bildiri, Gazi Üniversitesi, Ankara.
- Tanberkan Suna, H., Tan, Ş. (2017). The Investigation of the Group Invariance Property on Diverse Equating Methods. *Karaelmas Journal of Educational Sciences*, 5, 316-327.
- Tan, Ş. (2015). Küçük örneklerde beta4 ve polynomial loglineer öndüzdünleştirme ve kübik eğri öndüzdünleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151.
- Von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. Springer.
- Von Davier, A. A., & Kong, N. (2005). A unified approach to linear equating for the nonequivalent groups design. *Journal of Educational and Behavioral Statistics*, 30(3), 313-342.

<https://doi.org/10.3102/10769986030003313>

- Wiberg, M., & Bränberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement, 39*(5), 349-361. <https://doi.org/10.1177/014662161456793>
- Wolfe, E. W. (2000). Equating and item banking with the Rasch model. *Journal of Applied Measurement, 1*(4), 409-434.
- Yurtçu, M., & Güzeller, C. O. (2018). Investigation of equating error in tests with differential item functioning. *International Journal of Assessment Tools in Education, 5*(1), 50-57. <https://doi.org/10.21449/ijate.316420>

EXTENDED ABSTRACT

Introduction: Considering that the results of the exams are used in situations such as making decisions about the performance of the individuals and the general education level of a country, ensuring the transition to business life, ensuring placement in a higher institution, it is expected that the exams will measure the abilities of the individuals participating in a less erroneous, more accurate and objective way. For the stated objective measurement, the measurement process should first be started by converting the application conditions of the tests to standard conditions. Since standardized conditions will not be possible in every exam application, an analysis method was needed to make the scores obtained comparable. This analysis has been brought to the literature as a test equating method. Test equating is basically the process of determining the relationship between raw or scaled scores on two or more test forms (Skaggs & Lissitz, 1986). In the literature, many studies have been carried out to compare various equating methods according to the design and measurement theories (Classical Test Theory, Item Response Theory) selected in the test equating process. In this study, it was aimed to determine a general orientation regarding test equating studies. In this study, using bibliometric analysis, trending topics and study recommendations on test equating are focused on.

Materials and Methods: In the research, the Web of Science (WoS) database of test synchronization studies was examined by bibliometric analysis. Since the research aims to describe all the features of an existing situation, the scanning model was used. In the literature review, the best studies in the field of science are published in the journals in the WoS database. Studies with the theme of "test equating" were searched in this database. The data in the study was reduced to 247 studies with the PRISMA flow diagram.

Findings: This study was conducted to provide an overview of test-equating research based on 247 articles in the WoS database. The data reviewed provides findings on the most productive, most-cited authors, journals, and influential topics and themes published over the years, contributing to the test-equating literature. Analysis data were obtained using "test equating", WoS database in the world. Studies between 1972 and 2023 were analyzed as BibTeX with the bibliometric package in R-Studio. Since 2010, the performance of studies on this subject has increased. It is seen that a publication WoS data is published every year, including 1994, and this issue has been focused on after the 2000s. According to the progress of published studies, the most work in the field of regular test synchronization management is in USA. This is followed by Sweden, Turkey, Chile, China, Cyprus, Canada, Italy, Netherlands, and Spain, respectively. When the citations in the studies in the WoS database are examined, it is seen that the most citations are made to the study on 'Kernel synchronization' made by von Davier (2004). Stocking's (1983) work, which is also frequently emphasized in the calibration methods in ITC, has been the most cited work after von Davier's (2004). When the keywords are examined together with the citations, keywords such as "test equating", "MTK", "equalization", "changing item function", "synchronization error", "training", "equal-percentage equating", "mixed test form", "Rasch model" words are used frequently. When the studies titled test equalization in the journals are examined, most publications were published in the "Applied Psychological Measurement" journal. Other published journals are respectively "Statistics For Social and Behavioral Sciences", "Statistical Models for Test Equating Scaling and Linking", "Psychometria", "Applying Test Equating Methods Using R", "Methodology of Educational Measurement and Assessment", "Educational and Psychological Measurement", "Journal of Measurement and Evaluation in Education and Psychology", "Applied Measurement in Education" and "Journal of Educational and Behavioral Statistics". When the impact factor values of the journals are examined, it is seen that the number of citations is quite high. When we look at the other topics and themes that are studied together with the concept of test synchronization in the analyzed database, it is seen that the concepts such as the concept of "linking", "Item Response Theory", "Score", and "Models" are frequently used.

Discussion: Comparing test results and using them interchangeably is possible with statistical conversion and test synchronization processes in the most accurate and most accurate way. The general orientation was examined through the studies in which the keyword "test equating" was used in the WoS database scanned in the research. In the research, it was observed that most studies on this subject were produced in 2011. When the study subjects in the countries are examined, most publications on the key concept were made by scientists in the USA. When the citations are examined, the study on Kernel equating by von Davier (2004) is the most cited work. When we look at the other topics and themes that are studied together with the concept of test equating in the analyzed database, it is seen that the concepts such as the concept of "linking", "Item Response Theory", "Score", "Models" are frequently used.