

## Testlerde Adillik ve Uyarlamalar: Bibliyometrik ve Sistematik Bir Analiz (2000–2024)

Sevda ÇETİN  
Hacettepe Üniversitesi  
tsevda@hacettepe.edu.tr  
ORCID ID: [0000-0001-5483-595X](https://orcid.org/0000-0001-5483-595X)

Zeynep Neveser KIZILÇİM  
Hacettepe Üniversitesi  
[Kizilcim.zeynep@hacettepe.edu.tr](mailto:Kizilcim.zeynep@hacettepe.edu.tr)  
ORCID ID: [0000-0002-0164-5682](https://orcid.org/0000-0002-0164-5682)

|                          |                               |
|--------------------------|-------------------------------|
| Araştırma Makalesi       | DOI: 10.31592/aeusbed.1773492 |
| Geliş Tarihi: 28.08.2025 | Revize Tarihi: 01.11.2025     |
|                          | Kabul Tarihi: 27.11.2025      |

### Atıf Bilgisi

Çetin, S. ve Kızılçim, Z.N. (2025). Testlerde adillik ve uyarlamalar: Bibliyometrik ve sistematik bir analiz (2000–2024). *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(3), 890-909.

### ÖZ

Bu çalışma, test adillliği ve sınav uyarlamaları üzerine 2000-2024 yılları arasında Scopus veritabanında yayımlanan 63 makaleyi bibliyometrik ve sistematik inceleme yöntemleriyle analiz etmiştir. Amaç, test adillliği çalışmalarının yıllara, ülkelere, dergilere, anahtar kelimelere ve temalara göre dağılımını incelemek ve alana katkıları değerlendirmektir. PRISMA yöntemiyle seçilen makaleler, “test fairness”, “test accommodations”, “access arrangements” ve “disabilities” anahtar kelimeleriyle taranmıştır. Bulgular, test adillliği çalışmalarının 2013 sonrası arttığını, 2019 ve 2023’te sekizer makaleyle zirve yaptığını göstermektedir. ABD (n=101), en fazla atıf yapan ülke olarak öne çıkmaktadır; onu İngiltere (n = 17) ve Avustralya (n = 10) izlemektedir. Türkiye 4 atıfla katkı sağlamıştır. Educational Assessment (n=5) en çok yayın yapılan dergidir. “Assessment”, “fairness” ve “accommodations” en sık kullanılan anahtar kelimelerdir. Sistemantik inceleme, çalışmaları beş temada sınıflandırmıştır: dilsel faktörler, sınav uyarlamaları, özel gereksinimli öğrenciler ve erişilebilirlik, adillik-geçerlik çerçeveleri ve politika-tarihsel analizler. Sınav uyarlamaları (n=18) en çok çalışılan tema olup, ek süre ve teknolojik araçların erişimi artırdığı, ancak bazen geçerliği tehdit ettiği belirtilmiştir. Dilsel karmaşıklık, özellikle İngilizce öğrenen öğrencilerde performansı olumsuz etkilerken, evrensel tasarım ve understandardization-standartlaştırma eksikliği- gibi kavramlar teorik katkılar sunmuştur. Öneriler, evrensel tasarıma dayalı esnek uyarlamalar, daha fazla kültürel çeşitlilik odaklı araştırma ve pilot testlerdir. Bu çalışma, eğitimcilere ve politika yapıcılara daha kapsayıcı sınav uygulamaları için rehber sunmayı amaçlamaktadır.

**Anahtar Kelimeler:** Test adillliği, sınav uyarlamaları, özel gereksinimli öğrenciler, erişilebilirlik.

## Fairness and Accommodations in Tests: A Bibliometric and Systematic Analysis (2000–2024)

### ABSTRACT

This study analyzes 63 articles published between 2000 and 2024 in the Scopus database on test fairness and accommodations using bibliometric and systematic review methods. The aim is to examine the distribution of studies by year, country, journal, keywords, and themes, and to evaluate their contributions to the field. Articles were selected using the PRISMA method with keywords "test fairness," "test accommodations," "access arrangements," and "disabilities". Findings show an increase in test fairness studies post-2013, peaking in 2019 and 2023 with eight articles each. The United States (n=101) leads in contributions, followed by the United Kingdom (n=17) and Australia (n=10). Turkey contributed four publications. Educational Assessment (n=5) is the top journal. "Assessment," "fairness," and "accommodations" are the most frequent keywords. The systematic review categorized studies into five themes: linguistic factors, test accommodations, students with disabilities and accessibility, fairness-validity frameworks, and policy-historical analyses. Test accommodations (n=18) are the most studied, with findings indicating that extended time and technological tools enhance access but may sometimes compromise validity. Linguistic complexity negatively impacts English learners' performance, while concepts like universal design and understandardization offer theoretical advancements. Recommendations include flexible accommodations based on universal design, increased research on cultural diversity, and pilot testing. This study aims to guide educators and policymakers toward more inclusive assessment practices.

**Keywords:** Test fairness, test accommodations, students with disabilities, accessibility.

## Giriş

Bir testin değerlendirilmesi öncesinde dikkate alınması gereken temel unsurlardan biri, test maddelerinin tüm katılımcılar için aynı anlamı taşımasıdır. Özellikle geniş ölçekli sınavlarda test adilliğinin sağlanamaması, sonuçların yorumlanmasında yanlılığa yol açabilir ve farklı alt gruplar arasında haksız sonuçlara neden olabilir (AERA, APA ve NCME, 2014). Test adilliği, sınav sürecinde tüm katılımcılara eşit koşullar sunmayı ifade eder; bu, uygulama ve puanlama prosedürlerinin standartlaştırılmasını gerektirir. Adil bir test, ölçülmek istenen yapılar dışındaki unsurları içermemeli ve bireysel özelliklere karşı duyarlı olmalıdır (Camilli, 2006). Bu kavram, geçerlik kapsamında ele alınır ve cinsiyet, kültür, etnik köken, sosyoekonomik statü gibi faktörleri dikkate alır (AERA vd., 2014).

Test adilliğinin sağlanamaması, ölçüm yanlılıklarına yol açabilir. Örneğin, yüksek riskli sınavlarda düşük performans gösteren öğrencilerin bilgi eksikliğinden ziyade engel durumları nedeniyle test materyallerine eşit erişim sağlayamaması, adaletsizliklere neden olur (Dembitzer ve Kettler, 2023). Eğitsel ve Psikolojik Test Standartları'na göre (AERA vd., 2014), test adilliği iki temel boyut içerir: erişilebilirlik ve evrensel tasarım. Erişilebilirlik, öğrencilerin engelsiz bir şekilde teste ulaşmasını yasal bir zorunluluk haline getirirken, evrensel tasarım ise testin tüm alt gruplar için üst düzeyde erişilebilir olmasını sağlar. Standartlarda ele alınan bu iki boyut, test uygulamalarında öğrencilerin özelliklerine duyarlı olarak test uyarlamalarının yapılmasını gerekli kılar.

Test uyarlamaları, adilliği artırmak için engel durumu, dil yetersizliği veya öğrenme güçlüğü gibi bireysel ihtiyaçlara yönelik değişiklikleri kapsar (AERA vd., 2014). Bu uyarlamalar basit uyumlar (örneğin, ekstra süre, okuma yardımı) olabileceği gibi daha kapsamlı düzenlemeleri de içerebilir (Ofqual, 2020). Ancak, uyarlamalar ters etki yaratmamalı ve haksız avantaj sağlamamalıdır; temel test yapısını değiştirmeden uygulanmalıdır (Rodeiro ve Macinska, 2022). Pek çok ülkede, özel gereksinimli öğrencilerin ihtiyaçlarını karşılamak için ekstra süre, kelime işlemci, yazma yardımı gibi önlemler alınmıştır. Bu bağlamda, standart ve uyarlanmış uygulamalar arasındaki farklar, geçerliliği artırır mı yoksa tehdit eder mi gibi sorular, eğitim paydaşlarını, uzmanları ve politika yapıcılarını yakından ilgilendirmektedir (Sireci vd., 2005).

Bu araştırma, test adilliği alanında son yirmi beş yılda ortaya çıkan eğilimleri bütüncül biçimde inceleyerek, alandaki kavramsal ve yöntemsel gelişmeleri sistematik olarak ortaya koyması açısından önemlidir. Bibliyometrik ve sistematik inceleme yöntemlerinin birlikte kullanılması hem nicel eğilimlerin hem de nitel temaların eşzamanlı olarak değerlendirilmesine olanak tanımaktadır. Böylece çalışma, adil ve erişilebilir ölçme uygulamalarının geliştirilmesi sürecinde politika yapıcılar, araştırmacılar ve uygulayıcılar için kapsamlı bir çerçeve sunmaktadır.

Araştırmada, 2000 yılı ve sonrasında İngilizce olarak yayımlanan çalışmalar incelenmiştir. 2000 yılı ve sonrasında İngilizce yayımlanan çalışmalar incelenmiştir. Araştırma, testlerin daha erişilebilir ve adil hale getirilmesi için çıkarımlar sunmayı amaçlamakta olup şu sorulara yanıt aramaktadır:

1. Test adilliğine ilişkin araştırmaların yıllara göre dağılımı nasıldır?
2. Bu konudaki çalışmalarda öne çıkan dergiler hangileridir?
3. En sık atıf alan makaleler nelerdir?
4. En sık kullanılan anahtar kelimeler nelerdir ve eğilimleri nasıl bir seyir izlemektedir?
5. En fazla bilimsel üretim yapan ülkeler hangileridir ve üretim dağılımı nasıldır?
6. Ele alınan başlıca araştırma konuları nelerdir?
7. Kullanılan yöntemler ve bu yöntemlerin alana katkıları nelerdir?

Özet olarak bu çalışma, 2000–2024 yılları arasında yayımlanan test adilliği araştırmalarını bibliyometrik ve sistematik yöntemlerle inceleyerek alandaki eğilimleri, boşlukları ve gelecek yönelimleri belirlemeyi amaçlamaktadır. Bu doğrultuda yapılan incelemelerin, eğitim politikalarının geliştirilmesine ve daha kapsayıcı değerlendirme pratiklerinin oluşturulmasına katkı sağlaması beklenmektedir.

## Yöntem

### Araştırma Modeli

Bu çalışma, nitel araştırma yöntemlerinden doküman analizini temel almaktadır. Doküman analizi, belirlenen konuyla ilgili yazılı materyallerin sistematik bir şekilde incelenmesini içerir (Yıldırım ve Şimşek, 2021). Bu yöntemde, test adilliği ve sınav uyarlamalarıyla ilgili makaleler analiz edilerek anlamlı bütünler oluşturulmuş ve literatürün genel eğilimleri ortaya konmuştur (Karacabey, 2021).

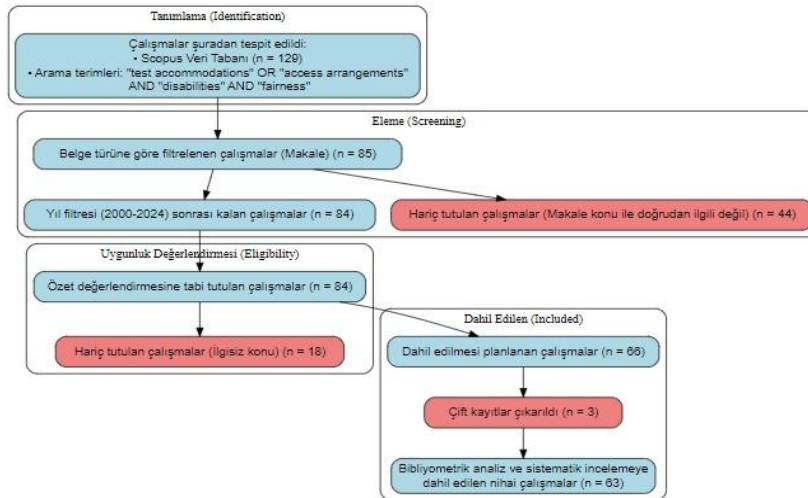
### İnceleme Nesnesi

Test adilliği, ölçme-değerlendirme literatüründe geçerlik konusunun bir alt başlığı olarak ele alınmakta olup bu alanda yürütülen araştırmalar nispeten yenidir ve gelişim sürecindedir. Bu çalışmada, 2000-2024 yılları arasında İngilizce yayımlanmış makaleler incelenerek 24 yıllık bir eğilim analizi hedeflenmiştir. Araştırma örneklemini, ölçüt örnekleme yöntemiyle belirlenmiştir (Yıldırım ve Şimşek, 2021). Ölçütler, makalelerin test adilliği ve sınav uyarlamalarıyla doğrudan ilişkili olması, açık erişimli olması ve Scopus veritabanında yer almasıdır.

### Veri Toplama Süreci

Araştırma, Scopus veritabanında "test fairness", "test accommodations", "access arrangements" ve "disabilities" anahtar kelimeleri kullanılarak yürütülmüştür. Scopus, farklı disiplinlerarası veritabanlarına erişim imkânı sunması, veritabanını yönetmek için araçlar sağlaması ve atıf sayısı ve erişilebilirlik gibi diğer kriterleri karşılaması nedeniyle bibliyografik bilgi kaynağı olarak tercih edilmiştir (Okagbue ve Silva, 2020). Elsevier Science, Scopus veritabanını 2004 yılında kullanıma sunmuştur Scopus, şu anda 5000'den fazla uluslararası yayıncı tarafından yayınlanan ve çoğu durumda 1996'dan bu yana yayınlanan yaklaşık 25.000 dergiyi aktif olarak kapsayan, hakemli literatürün özet ve atıflarının bulunduğu en büyük veritabanı olduğu için kısa sürede iyi bir alternatif haline gelmiştir. (Zhu ve Liu, 2020). Tarama sürecinde ise "AND" ve "OR" bağlaçları kullanılarak 2000-2024 yılları arasında yayımlanan açık erişimli makaleler filtrelenmiştir.

Toplam 129 makale ilk taramada tespit edilmiştir. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) yöntemi uygulanarak, katı seçim ve dışlama kriterleri doğrultusunda 44 makale elenmiştir. Kalan 85 makale, içerik, yöntem ve sonuç açısından değerlendirilmiş; yıl filtresi sonrası 84 makaleye düşmüştür. Konuyla doğrudan ilgisi olmayan 18 makale kapsam dışı bırakılmış, 3 çift kayıt tespit edilerek elenmiş ve nihai olarak 63 makale sistematik incelemeye dahil edilmiştir. PRISMA akış diyagramı Şekil 1'de sunulmuştur.



Şekil 1 PRISMA Akış Diyagramı

## Verilerin Analizi

Veri analizi, betimsel analiz yöntemiyle gerçekleştirilmiştir. Betimsel analiz, önceden belirlenmiş kodlamalara göre verilerin yüzeysel bir şekilde sınıflandırılmasını içerir ve içerik analizine kıyasla daha az ayrıntılıdır (Strauss ve Corbin, 1990; Yıldırım ve Şimşek, 2021). Analizde, makalelerin yayın yılı, araştırma soruları, kullanılan yöntemler ve alana katkıları açısından kodlanmıştır. Bulgular tablo ve şekillerle görselleştirilmiştir.

Araştırma iki aşamada yürütülmüştür:

**Bibliyometrik Analiz:** İlk aşamada, literatürdeki genel eğilimleri belirlemek için bibliyometrik analiz yapılmıştır. Bu analiz, yapılandırılmamış verileri anlamlı bilimsel bilgilere dönüştürmek için etkili bir yöntemdir (Donthu, Kumar, Mukherjee, Pandey ve Lim, 2021). Analiz, R programında Biblioshiny arayüzü ve bibliometrix paketi kullanılarak gerçekleştirilmiştir (Aria ve Cuccurullo, 2017). İncelenen unsurlar şunlardır:

- Sık kullanılan anahtar kelimeler
- En çok yayın yapılan dergiler
- En etkili yazarlar
- Alana en çok katkı sağlayan ülkeler
- En üretken araştırmacılar

**Sistematiik İnceleme:** İkinci aşamada, 63 makale; yayın yılı, araştırma amacı, yöntemi, örnekleme ve alana katkıları açısından sistematiik olarak incelenmiştir. Bu süreçte, yalnızca belirlenen ölçütlere uygun makaleler değerlendirilmiştir.

## Araştırmanın Sınırlılıkları

Bu çalışma, Scopus veritabanındaki açık erişimli makalelerle ve 2000-2024 yılları arasındaki yayınlarla sınırlıdır. Diğer veritabanları veya erişim kısıtlamalı makaleler incelemeye dahil edilmemiştir.

## Araştırma Etiği

Çalışmada, Scopus veritabanındaki açık erişimli makaleler kullanılmış olup bu makalelere herkes tarafından erişilebildiğinden, etik kurul izni gerekmemiştir.

## Bulgular

Bu bölümde, araştırma sorularına yanıt vermek amacıyla Scopus veritabanında taranan 63 makalenin bibliyometrik ve sistematiik analiz sonuçları sunulmuştur. Her bir araştırma sorusu ayrı bir alt başlık altında ele alınmış ve bulgular tablo ve şekillerle desteklenmiştir.

## Test Adillii Araştırmalarının Yıllara Göre Dağılımı

Test adillii konusundaki çalışmaların 2000-2024 yılları arasındaki dağılımı Tablo 1’de verilmiştir. Analiz, 2000-2007 yıllarında bu konuda yayın bulunmadığını, ilk çalışmanın 2008 yılında yayımlandığını göstermektedir.

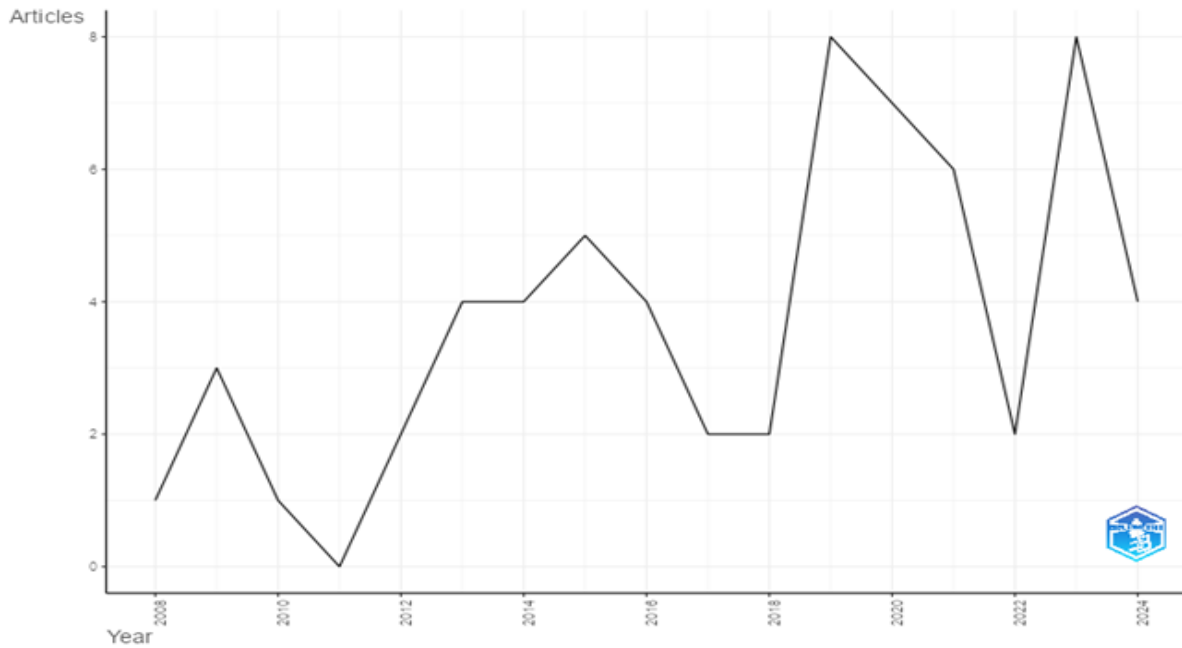
Tablo 1

*Scopus Veritabanında Taranan Makalelerin Yıllara Göre Dağılımı*

| Yıl  | Makale sayısı | Yıl  | Makale sayısı |
|------|---------------|------|---------------|
| 2008 | 1             | 2017 | 2             |
| 2009 | 3             | 2018 | 2             |
| 2010 | 1             | 2019 | 8             |
| 2011 | 0             | 2020 | 7             |

|      |   |      |   |
|------|---|------|---|
| 2012 | 2 | 2021 | 6 |
| 2013 | 4 | 2022 | 2 |
| 2014 | 4 | 2023 | 8 |
| 2015 | 5 | 2024 | 4 |
| 2016 | 4 |      |   |

Tablo 1'e göre, 2008-2012 yılları arasında yayın sayısı sınırlı olup (toplam 7 makale), 2013 yılından itibaren daha düzenli bir artış gözlenmiştir. 2019 ve 2023 yılları, sekizer makale ile en yüksek üretim dönemleridir. 2011'de hiç yayın bulunmazken, 2017 ve 2018 yıllarında yayın sayısında düşüş yaşanmıştır. Şekil 2, bu eğilimi görselleştirmektedir ve özellikle 2013 sonrası artışın istikrarlı olduğunu ortaya koymaktadır.



Şekil 2. Test adillliği konusundaki çalışmaların yıllık eğilimi

### Test Adillliği Konusundaki Öne Çıkan Dergiler

Test adillliği çalışmalarının yayınlandığı dergiler Tablo 2'de listelenmiştir.

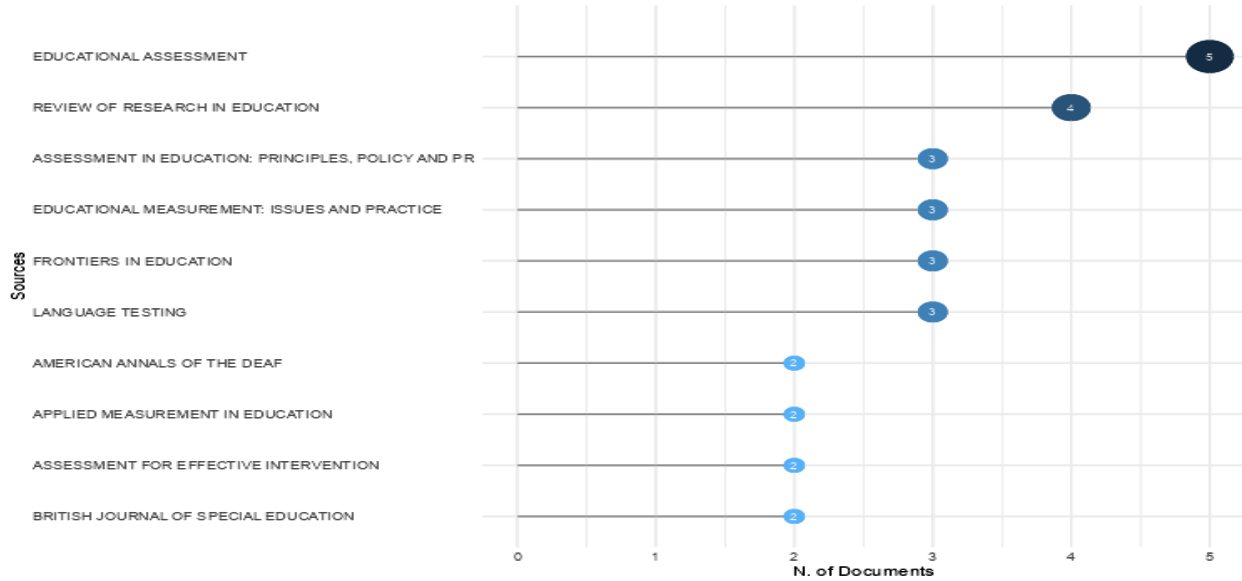
Tablo 2

#### Test Adillliği Konusunda Öne Çıkan Dergiler

| Dergi İsimleri                                           | Makale sayıları | Dergi İsimleri                              | Makale sayıları |
|----------------------------------------------------------|-----------------|---------------------------------------------|-----------------|
| Educational Assessment                                   | 5               | Ets Research Report Series                  | 1               |
| Review Of Research In Education                          | 4               | European Journal Of Special Needs Education | 1               |
| Assessment In Education: Principles, Policy And Practice | 3               | International Journal Of Special Education  | 1               |
| Educational Measurement: Issues And Practice             | 3               | Journal Of Applied School Psychology        | 1               |
| Frontiers In Education                                   | 3               | Journal Of Deaf Studies And Deaf Education  | 1               |
| Language Testing                                         | 3               | Journal Of Educational Measurement          | 1               |
| American Annals Of The Deaf                              | 2               | Journal Of Learning Disabilities            | 1               |

|                                                         |   |                                                       |   |
|---------------------------------------------------------|---|-------------------------------------------------------|---|
| Applied Measurement In Education                        | 2 | Journal Of Multilingual And Multicultural Development | 1 |
| Assessment For Effective Intervention                   | 2 | Journal Of School Psychology                          | 1 |
| British Journal Of Special Education                    | 2 | Journal Of Special Education                          | 1 |
| Education Sciences                                      | 2 | Language Assessment Quarterly                         | 1 |
| International Journal Of Testing                        | 2 | Learning And Instruction                              | 1 |
| International Multilingual Research Journal             | 2 | Oxford Review Of Education                            | 1 |
| Journal Of Psychoeducational Assessment                 | 2 | Practical Assessment, Research And Evaluation         | 1 |
| Advances In Health Sciences Education                   | 1 | Psychology In The Schools                             | 1 |
| Behaviormetrika                                         | 1 | Sage Open                                             | 1 |
| Bilingual Research Journal                              | 1 | Scandinavian Journal Of Educational Research          | 1 |
| Educational Assessment, Evaluation And Accountability   | 1 | Social Psychology Of Education                        | 1 |
| Educational Research And Evaluation                     | 1 | Teaching And Teacher Education                        | 1 |
| Empirical Research In Vocational Education And Training | 1 | Urban Education                                       | 1 |

Tablo 2, yayınların çeşitliliğini ve disiplinler arası doğasını ortaya koymaktadır. Educational Assessment (n=5) en fazla yayına sahip dergi olurken, Review of Research in Education (n=4) ikinci sıradadır. Assessment in Education: Principles, Policy and Practice, Educational Measurement: Issues and Practice, Frontiers in Education ve Language Testing (her biri n=3) diğer öne çıkan dergilerdir. Yayınların ölçme-değerlendirme, özel eğitim, çok dillilik ve psikoloji gibi alanlarda dağıldığı görülmektedir. Bu durum test adilliğinin giderek artan ölçüde disiplinler arası bir araştırma alanı olarak ele alındığı ortaya koymaktadır.



Şekil 3. Test Adilliği Konusundaki Yayınların Dergilere Göre Dağılımı

Şekil 3'te, test adilliği üzerine yayımlanan makalelerin dergilere göre dağılımları gösterilmektedir. En fazla yayının Educational Assessment (n = 5) dergisinde olduğu, bunu Review of Research in Education (n = 4) ve Assessment in Education: Principles, Policy and Practice (n = 3) ile takip ettiği görülmektedir.

### En Sık Atıf Alan Makaleler

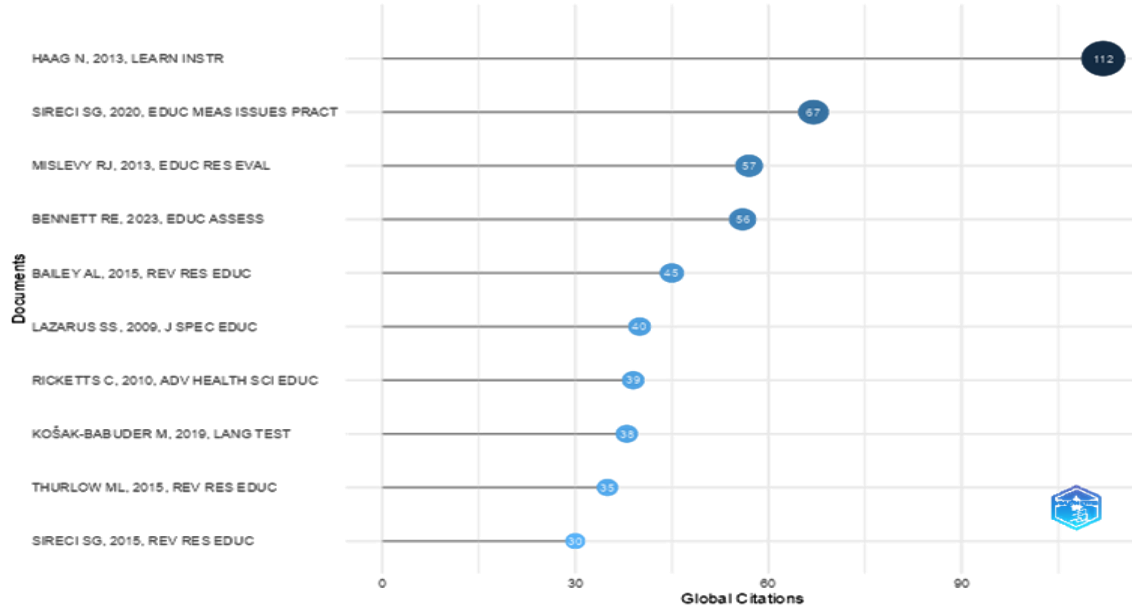
Test adilliği literatüründeki en çok atıf alan makaleler Tablo 3'te sunulmuştur.

Tablo 3

*Test Adilliğı Çalışmalarında En Çok Atıf Alan Makaleler*

| Makale                                                | Toplam Atıf | Makale                                       | Toplam Atıf |
|-------------------------------------------------------|-------------|----------------------------------------------|-------------|
| Haag N, 2013, Learn Instr                             | 112         | Yang X, 2020, Int Multiling Res J            | 9           |
| Sireci Sg, 2020, Educ Meas Issues Pract               | 67          | Guzman-Orth D, 2021, Ets Res Rep Series      | 9           |
| Mislevy Rj, 2013, Educ Res Eval                       | 57          | Woods K, 2019, Oxf Rev Educ                  | 8           |
| Bennett Re, 2023, Educ Assess                         | 56          | Lı H, 2012, Lang Assess Q                    | 6           |
| Bailey Al, 2015, Rev Res Educ                         | 45          | Woods K, 2018, Br J Spec Educ                | 6           |
| Lazarus Ss, 2009, J Spec Educ                         | 40          | Solano-Flores G, 2019, Int Multiling Res J   | 6           |
| Ricketts C, 2010, Adv Health Sci Educ                 | 39          | Camillerı S, 2019, Sage Open                 | 5           |
| Košak-Babuder M, 2019, Lang Test                      | 38          | Buzick Hm, 2019, Educ Assess                 | 5           |
| Thurlow Ml, 2015, Rev Res Educ                        | 35          | Şenel S, 2021, Educ Assess Eval Account      | 5           |
| Sireci Sg, 2015, Rev Res Educ                         | 30          | Witmer Se, 2023, Assess Eff Intervention     | 5           |
| Solano-Flores G, 2019, Front Educ                     | 28          | Şenel S, 2018, Europ J Spec Needs Educ       | 4           |
| Young Jw, 2008, Educ Assess                           | 27          | Oliverı Me, 2019, Front Educ                 | 4           |
| Cawthon S, 2013, Am Ann Deaf                          | 27          | Motteram J, 2023, Lang Test                  | 4           |
| Leppo Rht, 2014, J Deaf Stud Deaf Educ                | 25          | Wei X, 2024, J Learn Disabil                 | 4           |
| Abedi J, 2014, Appl Meas Educ                         | 24          | Stanczak A, 2024, Soc Psychol Educ           | 4           |
| Solano-Flores G, 2014, Appl Meas Educ                 | 24          | Lıu D, 2017, Empir Res Vocat Educ Tram       | 3           |
| Lane S, 2015, Rev Res Educ                            | 24          | Hjärne Ms, 2020, Scan J Educ Res             | 3           |
| Cawthon S, 2015, Am Ann Deaf                          | 22          | Hıpkiss A, 2021, Br J Spec Educ              | 3           |
| Finch H, 2009, Educ Assess                            | 21          | Fensham Pj, 2013, Educ Sci                   | 2           |
| De Backer F, 2019, J Multiling Multicult Dev          | 18          | Mathes Ne, 2020, J Appl Sch Psychol          | 2           |
| Rasooli A, 2021, Assess Educ Princ Policy Pract       | 16          | Hjärne Ms, 2020, Front Educ                  | 2           |
| Clark-Gareca B, 2016, Teach Teach Educ                | 15          | O'neill R, 2020, Psychol Sch                 | 2           |
| Abedi J, 2020, Educ Meas Issues Pract                 | 15          | Roschmann S, 2021, Int J Test                | 2           |
| Buono S, 2021, Educ Assess                            | 15          | Razmjoe M, 2023, Educ Sci                    | 2           |
| Cawthon Sw, 2009, Pract Assess Res Eval               | 12          | Alicia Kim A, 2023, Lang Test                | 2           |
| Cawthon Sw, 2012, J Sch Psychol                       | 12          | Huenergarde Mc, 2023, Biling Res J           | 2           |
| Cahan S, 2016, J Pshychoeduc Assess                   | 12          | Saka N, 2022, Assess Educ Princ Policy Pract | 1           |
| Svetina D, 2017, Behaviormetrika                      | 12          | Johnson A, 2024, Educ Meas Issues Pract      | 1           |
| Lee Hs, 2014, Int J Test                              | 11          | Seo Dg, 2016, J Pshychoeduc Assess           | 0           |
| Sınharay S, 2024, J Educ Meas                         | 11          | Dembitzer L, 2023, Assess Eff Intervention   | 0           |
| Turkan S, 2016, Urban Educ                            | 10          | Papuda-Dolińska B, 2023, Int J Spec Educ     | 0           |
| Vidal Rodeiro C, 2022, Assess Educ Princ Policy Pract | 10          |                                              |             |

En yüksek atıf alan çalışma, Haag ve diğerlerinin (2013) 112 atıfla Learning and Instruction dergisinde yayımlanan makalesidir. Bunu, Sireci (2020, n=67), Mislevy (2013, n=57) ve Bennett (2023, n=56) izlemektedir. Bu makaleler, test adilliğı literatüründe Sireci, Abedi, Solano-Flores ve Cawthon gibi araştırmacıların öncü rolünü ortaya koymaktadır. Şekil 4, test adilliğı alanında alanyazında en çok atıf alan çalışmaların dağılımını göstermektedir.



Şekil 4. Test Adillliği Üzerine Yayımlanmış Çalışmaların Atıf Dağılımı (İlk 10 Makale)

#### En Sık Kullanılan Anahtar Kelimeler ve Eğilimleri

Test adillliği literatüründeki anahtar kelimeler ve kullanım sıklıkları Tablo 4’te sunulmuştur.

Tablo 4

*Test Adillliği Alanyazınında Öne Çıkan Anahtar Kelimeler ve Kullanım Sıklıkları*

| Kelime                  | Sıklık | Kelime                        | Sıklık |
|-------------------------|--------|-------------------------------|--------|
| assessment              | 10     | learning disabilities         | 4      |
| fairness                | 10     | learning disorder             | 4      |
| accommodations          | 9      | students with disabilities    | 4      |
| accessibility           | 5      | differential item functioning | 3      |
| education               | 5      | dyslexia                      | 3      |
| educational measurement | 5      | female                        | 3      |
| human                   | 5      | learning disorders            | 3      |
| humans                  | 5      | male                          | 3      |
| test accommodations     | 5      | mathematics                   | 3      |
| english learners        | 4      | test accommodation            | 3      |

Tablo 4, “assessment” (n=10), “fairness” (n=10) ve “accommodations” (n=9) kelimelerinin en sık kullanıldığını göstermektedir. “Accessibility”, “education” ve “educational measurement” (her biri n=5) ile “english learners” ve “learning disabilities” (her biri n=4) gibi terimler, çalışmaların dil yetersizliği ve özel gereksinimli öğrenciler üzerine odaklandığını ortaya koymaktadır. “Differential item functioning”, “validity” ve “bias” gibi kavramlar da literatürde önemli yer tutmaktadır. Bu dağılım, test adillliği çalışmalarında hem kuramsal kavramların (ör. fairness, validity, bias) hem de uygulamaya dönük düzenlemelerin (ör. accommodations, extended time, universal design) birlikte ele alındığı söylenebilir. Şekil 5’te test adillliği literatüründe en sık kullanılan anahtar kelimeler görselleştirilmiştir.



Şekil 5. Test Adilliği Alanyazınında En Sık Kullanılan Anahtar Kelimeler

Şekil 5'te yer alan görselde öne çıkan *assessment* (değerlendirme), *fairness* (adillik) ve *accommodations* (uyarlamalar) terimleri literatürde en çok tekrar eden kavramlara yer verilmiştir. Bununla birlikte, *erişilebilirlik*, *eğitimde ölçme*, *İngilizce öğrenenler*, *öğrenme güçlükleri* ve *özel gereksinimli öğrenciler* gibi anahtar kelimelerin sıklıkla birlikte görülmesi, araştırmaların büyük ölçüde kapsayıcı değerlendirme ve test adilliği temalarına yöneldiğini göstermektedir.

#### En Fazla Bilimsel Üretim Yapan Ülkeler

Test adilliği literatüründeki bilimsel üretimlerin ülkelere göre dağılımı Tablo 5'te verilmiştir.

Tablo 5  
*Test Adilliği Alanyazınında En Fazla Bilimsel Üretim Yapan Ülkeler*

| Ülkeler    | Bilimsel Üretim Sıklığı |
|------------|-------------------------|
| ABD        | 101                     |
| İngiltere  | 17                      |
| Avustralya | 10                      |
| Fransa     | 6                       |
| İsrail     | 6                       |
| Almanya    | 5                       |
| Polonya    | 5                       |
| Belçika    | 4                       |
| Türkiye    | 4                       |
| Kanada     | 3                       |
| Malta      | 3                       |
| İsviçre    | 3                       |
| Çin        | 2                       |
| Slovenya   | 2                       |
| Güney Kore | 1                       |
| Toplam     | 172                     |

Tablo 5, ABD'nin (n=101) test adilliği literatürüne en fazla katkı sağlayan ülke olduğunu göstermektedir. İngiltere (n=17) ve Avustralya (n=10) diğer önde gelen ülkeler olarak öne çıkmaktadır. Türkiye ise dört yayınlı alana katkı sağlamaktadır. Toplam sıklığın 63 makaleden fazla olması, bir

makalede birden fazla ülkeden yazar bulunması nedeniyledir. Toplam sıklık, incelenen 63 makalenin toplam sayısından fazla görünmektedir. Bunun nedeni, bibliyometrik analizde kullanılan bibliometrix paketinin, bir makalede birden fazla ülkeden yazar bulunması durumunda her ülkeyi ayrı bir katkı olarak saymasıdır. Bu nedenle, toplam bilimsel üretim sıklığı, makale sayısından daha yüksek bir değer alabilir.

### Başlıca Araştırma Konuları, Yöntemler ve Katkılar

Test adilliği literatüründeki ana temalar, kullanılan yöntemler ve katkılar Tablo 6’da özetlenmiştir.

Tablo 6

*Test Adilliği İncelemeleri Özet Tablosu*

| Konu Alanı                                      | Çalışma Sayısı | Yöntemler                                                 | Öne Çıkan Bulgular / Katkılar                                                                                                                            | Örnek Çalışmalar                                   |
|-------------------------------------------------|----------------|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| Dilsel Faktörler ve Karmaşıklık                 | 12             | DIF analizi, regresyon, faktör analizi, kuramsal inceleme | Akademik dil, uzun cümleler ve dilsel karmaşıklık, ELL öğrencilerin matematik ve fen performansını olumsuz etkilemektedir.                               | Haag (2013), Abedi (2014), Bailey (2015)           |
| Sınav Uyarlamaları (Accommodations)             | 18             | Deneyisel tasarım, meta-analiz, DIF, faktör analizi       | Ek süre, sesli okuma, bilgisayar tabanlı araçlar kimi zaman adaleti artırıyor, kimi zaman dezavantaj yaratmaktadır.                                      | Sireci (2020), Leppo (2014), Abedi (2012)          |
| Özel gereksinimli Öğrenciler ve Erişilebilirlik | 14             | Nitel, nicel ve karma yöntemler                           | Görme ve işitme özel gereksinimi olan öğrenciler için standart testlerde DIF bulunuyor; teknolojik araçlar (CAT, okuma kalemi) erişimi artırmaktadır.    | Cawthon (2013), Leppo (2014), Şenel (2021)         |
| Adillik ve Geçerlik Çerçevesi                   | 10             | Teorik analiz, literatür inceleme, kavramsal modelleme    | Evrensel tasarım, koşullu adalet, UNDERSTANDARDIZATION gibi yeni kavramlar öneriliyor; test politikalarında daha kapsayıcı yaklaşımlar vurgulanmaktadır. | Mislevy (2013), Lane (2015), Bennett (2023)        |
| Politika ve Tarihsel Analizler                  | 9              | Zaman seri analizi, politika doküman incelemesi           | Eyalet politikaları, GCSE erişim düzenlemeleri, yüksek riskli sınavlardaki adillik tartışmaları sistematik olarak ele alınmış.                           | Thurlow (2009), Woods (2018), Vidal Rodeiro (2022) |

Tablo 6’da görüldüğü üzere, test adilliği alanyazınında ele alınan çalışmalar beş ana tema altında toplanmaktadır: *dilsel faktörler ve karmaşıklık*, *sınav uyarlamaları*, *özel gereksinimli öğrenciler ve erişilebilirlik*, *adillik ve geçerlik çerçeveleri* ile *politika ve tarihsel analizler*. Çalışmaların en çok yoğunlaştığı alan “sınav uyarlamaları” (n = 18) olup bu araştırmalar genellikle deneyisel tasarımlar, meta-analizler ve farklı değişen madde fonksiyonu (DMF) yöntemleri aracılığıyla yürütülmüş; elde edilen bulgular sınav düzenlemelerinin kimi zaman adilliğin sağlandığını, kimi zaman ise dezavantajlar yarattığını göstermiştir. “Dilsel faktörler ve karmaşıklık” (n = 12) teması altında yapılan çalışmalar; özellikle ikinci dil öğrenen öğrencilerde akademik dil, uzun cümle yapıları ve karmaşık terminolojinin test performansını olumsuz etkilediğini ortaya koymuştur. “Özel gereksinimli öğrenciler ve erişilebilirlik” konulu araştırmalar (n = 14) ise görme ve işitme özel gereksinimli öğrenciler için standart testlerin DMF içerdiğini; buna karşın bilgisayar destekli uyarlanabilir testler ve teknolojik araçların teste erişimi artırmada önemli bir rol oynadığı sonucuna ulaşıldığı görülmüştür. Buna ek olarak, “adillik ve geçerlik çerçeveleri” (n = 10) kapsamında yürütülen kuramsal çalışmalar, evrensel tasarım, koşullu adalet ve test koşullarına esneklik ekleyerek öğrencilerin gerçek becerilerini daha doğru bir şekilde değerlendirmeyi amaçlayan understandardization- standartlaştırma eksikliği (Sireci, 2020) gibi yeni

kavramsal yaklaşımlar geliştirerek alanyazına teorik katkılar sağlamış olduğu söylenebilir. Son olarak “politika ve tarihsel analizler” (n = 9), eyalet politikalarının, erişim düzenlemelerinin ve yüksek riskli sınavlardaki adalet tartışmalarının zaman içindeki değişimini sistematik olarak ortaya koymuştur.

### Sonuç, Tartışma ve Öneriler

Bu çalışmada, test adilliği ve sınav uyarlamaları üzerine 2000-2024 yılları arasında Scopus veritabanında yayımlanan 63 makale, bibliyometrik ve sistematik inceleme yöntemleriyle analiz edilmiştir. Araştırmamızın ilk beş sorusu bibliyometrik analizle, son sorusu ise sistematik incelemeyle yanıtlanmıştır. Bulgular, test adilliği çalışmalarının 2013 sonrası artış gösterdiğini, ABD’nin alana en fazla katkı sağlayan ülke olduğunu, ABD’nin en fazla katkı sağlayan ülke olarak öne çıkmasının, ölçme ve değerlendirme alanında uzun süredir etkin olan araştırma merkezleri ve geniş ölçekli sınav sistemleriyle (örneğin NAEP, PISA ve TIMSS gibi uluslararası sınavlarda aktif rol alması) ilişkili olduğu söylenebilir. Educational Assessment gibi dergilerin öne çıktığını ve “assessment”, “fairness” ile “accommodations” anahtar kelimelerinin sık kullanıldığını ortaya koymuştur. Sistematik inceleme, literatürü beş ana tema altında sınıflandırmıştır: dilsel faktörler ve karmaşıklık, sınav uyarlamaları, özel gereksinimli öğrenciler ve erişilebilirlik, adillik ve geçerlik çerçeveleri ile politika ve tarihsel analizler. Bu temalar, test adilliğinin hem teorik hem de pratik boyutlarını yansıtmaktadır.

Test adilliği, ölçme ve değerlendirme süreçlerinde eşitlik ve geçerlik açısından temel bir öneme sahiptir. Bibliyometrik analiz, 2019 ve 2023 yıllarında sekizer makale ile en yüksek üretimin gerçekleştiğini göstermektedir; bu, pandemi sonrası kapsayıcı eğitim tartışmalarının artmasıyla ilişkilendirilebilir (Sireci, 2020). Yayınların Educational Assessment (n=5) ve Review of Research in Education (n=4) gibi dergilerde yoğunlaşması, konunun ölçme-değerlendirme, özel eğitim ve çok dillilik gibi disiplinler arası alanlarda ele alındığını doğrular. Anahtar kelimeler arasında “fairness” (n=10) ve “accommodations” (n=9) ön planda olup, çalışmaların özellikle dil yetersizliği ve özel gereksinimli öğrenciler üzerine odaklandığını göstermektedir (AERA vd., 2014).

Sistematik inceleme, dilsel faktörlerin (n=12) test performansını olumsuz etkilediğini ortaya koymuştur. Örneğin, İngilizce öğrenen öğrencilerde akademik dil ve karmaşık cümle yapıları performans düşüşüne yol açmaktadır (Haag, Heppt, Stanat, Kuhl ve Pant, 2013; Solano-Flores, 2014, 2019). Bu durum, testlerin sadece hedeflenen becerileri değil, aynı zamanda dil yeterliliğini de dolaylı olarak ölçebildiğini göstermektedir. Sınav uyarlamaları (n=18), en çok incelenen tema olarak öne çıkmaktadır. Çift dilli testler ve sesli okuma gibi düzenlemelerin test performansını artırabileceği, ancak bazı durumlarda geçerliği zayıflatabileceği belirtilmiştir (Abedi, 2014; Cawthon ve Bond, 2014; Finch, Barton ve Meyer, 2009; Şenel, 2021). Bu bulgular arasındaki çelişkiler, uyarlamaların dikkatle ve geçerlik ilkeleri gözetilerek tasarlanması gerektiğini göstermektedir (AERA vd., 2014). Bu çelişkili etkiler, uyarlamaların dikkatle tasarlanması gerektiğini vurgulamaktadır (AERA vd., 2014).

Sınav uyarlamalarında (n=18) çalışma sayısının fazla olduğu görülmektedir. Araştırmalar incelendiğinde yapılan bazı dil bazlı uygulamaların (çift dilli testler, sesli okuma desteği gibi) test performansını olumlu etkilediği tespit edilmiştir (Abedi, 2014). Ancak sınav uyarlamalarında sağlanan kolaylıkların sınavların geçerliğini olumsuz yönde etkileyebileceğine ulaşılmıştır (Cawthon ve Bond, 2014). Örneğin yapılan sınav uyarlamalarının beklendiği gibi etkili olmadığı ve dezavantajlara neden olduğu görülmüştür (Finch vd., 2009; Şenel, 2021). Araştırmalar sonucunda daha adil sınav uygulamalarının yapılmasında ve değerlendirilmesinde bireysel ihtiyaçlara göre sınav uyarlamalarının yapılması, bu konuda tek bir eğitim politikasına bağlı kalınmaması önerilebilir. Bailey ve Carol (2015) da eğitim politikalarının dil bariyerlerini aşacak şekilde düzenlemesi gerekliliği üzerinde durmuştur. Özel gereksinimi olan öğrenciler ve erişilebilirlik teması (n=14), teknolojik araçların (örneğin, bilgisayar tabanlı testler) erişimi artırdığını, ancak standart testlerde farklı değişen madde fonksiyonu (DMF) sorunlarının devam ettiğini göstermektedir (Cawthon, 2013; Şenel, 2021). Adillik ve geçerlik çerçeveleri (n=10), understandardization- standartlaştırma eksikliği gibi yeni kavramlarla teorik katkılar sunarken, politika ve tarihsel analizler (n=9), eyalet politikalarının evrimini ve iş birliği ihtiyacını ele almıştır (Lazarus, Thurlow, Lail ve Christensen, 2009; Woods, James ve Hipkiss, 2018). İngiltere’de GCSE ve GCE sınavlarındaki erişim düzenlemeleri, 2010 Eşitlik Yasası’nın etkisini

yansıtmaktadır (Hipkiss, Woods ve Mccaldin, 2021). Bu bulgular, uluslararası standartların (AERA vd., 2014) daha adil değerlendirmeler için rehber olduğunu desteklemektedir.

Literatürdeki tartışmalar, test uyarlamalarının avantaj ve dezavantajlarını ortaya koyar (Sireci vd., 2005; Dembitzer ve Kettler, 2023). Bazı uyarlamalar erişimi artırırken, diğerleri haksız avantaj veya geçerlik sorunları yaratabilir. Örneğin, ekstra süre gibi düzenlemeler bazı gruplar için faydalı olsa da ölçümün amacını saptırabilir (Finch vd., 2009). Bu durum, evrensel tasarım ile bireysel ihtiyaçlara göre esnekliğin dengelenmesi gerektiğini gösterir (Bailey ve Carroll, 2015). Ayrıca, farklı kültürel bağlamlarda test adilliği çalışmalarının sınırlı olması, alandaki bir eksikliği işaret etmektedir.

Bu bulgular doğrultusunda şu öneriler sunulmaktadır:

**Eğitim Politikaları için Öneriler:** Ülkeler, evrensel tasarım ilkelerine dayalı ancak bireysel farklılıklara esneklik tanıyan sınav uyarlamaları geliştirmelidir. Örneğin, ekstra süre veya teknolojik araçlar, performans verilerine dayalı olarak standardize edilmelidir (Ofqual, 2020). Eğitim düzenleyicileri, okullar ve sınav kurumları arasındaki iş birliği güçlendirilmelidir (Woods vd., 2018).

**Araştırma için Öneriler:** Gelecek çalışmalar, test adilliğini sosyal adalet ve kültürel çeşitlilik bağlamında ele almalı, özellikle az temsil edilen ülkelerde (örneğin, Türkiye) araştırmalar artırılmalıdır. Boylamsal çalışmalar, politika değişikliklerinin uzun vadeli etkilerini incelemelidir.

**Uygulama için Öneriler:** Test geliştiriciler, dilsel karmaşıklığı azaltarak ve erişilebilirlik araçlarını entegre ederek test adilliğini artırabilir. Uyarlamaların geçerliliği tehdit etmemesi için pilot testler yapılmalıdır.

Sonuç olarak, test adilliği, geçerli ve güvenilir değerlendirmelerin temel taşıdır. Bu çalışma, eğitimcilere, politika yapıcılara ve araştırmacılara, daha kapsayıcı ve adil sınav uygulamaları için rehber sunmayı amaçlamaktadır. Alandaki çelişkili bulgular, evrensel ve bireyselleştirilmiş yaklaşımlar arasında denge kurulmasının önemini vurgulamaktadır.

## Araştırmacıların Katkı Oranı

Bu çalışmada yazarların katkı oranları eşittir.

## Çıkar Çatışması

Çalışmada çıkar çatışması oluşturabilecek herhangi bir durum bulunmamaktadır.

**Teşekkür ve/veya Açıklama:** Bu çalışma, 4–6 Ekim 2024 tarihlerinde Millî Eğitim Bakanlığı ve Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) Başkanlığı tarafından Ankara’da düzenlenen *Uluslararası Ölçme, Seçme ve Yerleştirme Sempozyumu*’nda sözlü bildiri olarak sunulmuştur.

## Kaynaklar

- Abedi, J. (2014). The use of computer technology in designing appropriate test accommodations for English language learners. *Applied Measurement in Education*, 27(4), 261–272. <https://doi.org/10.1080/08957347.2014.944310>
- Abedi, J., Zhang, Y., Rowe, S. E. and Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. *Educational Measurement: Issues and Practice*, 39(4), 41–52. <https://doi.org/10.1111/emip.12328>
- Alicia Kim, A., Yumsek, M., Kemp, J. A., Chapman, M. and Gary Cook, H. (2023). Universal tools activation in English language proficiency assessments: A comparison of grades 1–12 English learners with and without disabilities. *Language Testing*, 40(4), 877–903. <https://doi.org/10.1177/02655322221149009>

- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Aria, M. and 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>
- Bailey, A. L. and Carroll, P. E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education*, 39(1), 253–294. <https://doi.org/10.3102/0091732X14556074>
- Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment*, 28(2), 83–104. <https://doi.org/10.1080/10627197.2023.2202312>
- Buono, S., and Jang, E. E. (2021). The effect of linguistic factors on assessment of English language learners' mathematical ability: A differential item functioning analysis. *Educational Assessment*, 26(2), 125–144. <https://doi.org/10.1080/10627197.2020.1858783>
- Cahan, S., Nirel, R. and Alkoby, M. (2016). The extra-examination time granting policy: A reconceptualization. *Journal of Psychoeducational Assessment*, 34(5), 461–472. <https://doi.org/10.1177/0734282915616537>
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221–256). Praeger Publishers.
- Camilleri, S., Chetcuti, D. and Falzon, R. (2019). “They labeled me ignorant”: Narratives of Maltese youth with dyslexia on national examinations. *SAGE Open*, 9(2). <https://doi.org/10.1177/2158244019855674>
- Cawthon, S. W. (2015). From the margins to the spotlight: Diverse deaf and hard of hearing student populations and standardized assessment accessibility. *American Annals of the Deaf*, 160(4), 385–394. <https://doi.org/10.1353/aad.2015.0036>
- Cawthon, S. W. and Leppo, R. (2013). Assessment accommodations on tests of academic achievement for students who are deaf or hard of hearing: A qualitative meta-analysis of the research literature. *American Annals of the Deaf*, 158(3), 363–376. <https://doi.org/10.1353/aad.2013.0023>
- Cawthon, S. W., Ho, E., Patel, P. G., Potvin, D. C. and Trundt, K. M. (2009). Multiple constructs and effects of accommodations on accommodated test scores for students with disabilities. *Practical Assessment, Research and Evaluation*, 14(18). <https://doi.org/10.7275/15345084>
- Cawthon, S. W., Kaye, A. D., Lockhart, L. L. and Beretvas, S. N. (2012). Effects of linguistic complexity and accommodations on estimates of ability for students with learning disabilities. *Journal of School Psychology*, 50(3), 293–316. <https://doi.org/10.1016/j.jsp.2012.01.002>
- Clark-Gareca, B. (2016). Classroom assessment and English language learners: Teachers' accommodations implementation on routine math and science tests. *Teaching and Teacher Education*, 54, 139–148. <https://doi.org/10.1016/j.tate.2015.11.003>
- De Backer, F., Slembrouck, S. and Van Avermaet, P. (2019). Assessment accommodations for multilingual learners: Pupils' perceptions of fairness. *Journal of Multilingual and Multicultural Development*, 40(9), 833–846. <https://doi.org/10.1080/01434632.2019.1571596>
- Dembitzer, L. and Kettler, R. J. (2023). Universally designed accommodations on a reading comprehension test: What do they accomplish? *Assessment for Effective Intervention*, 49(1), 18–28. <https://doi.org/10.1177/15345084231170317>

- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and 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>
- Fensham, P. J. and Cumming, J. J. (2013). “Which child left behind”: Historical issues regarding equity in science assessment. *Education Sciences*, 3(3), 326–343. <https://doi.org/10.3390/educsci3030326>
- Finch, H., Barton, K. and Meyer, P. (2009). Differential item functioning analysis for accommodated versus nonaccommodated students. *Educational Assessment*, 14(1), 38–56. <https://doi.org/10.1080/10627190902816264>
- Guzman-Orth, D., Supalo, C. A., Smith, D. W., Lee, O. and King, T. (2021). *Equitable STEM instruction and assessment: Accessibility and fairness considerations for special populations*. ETS Research Report Series, 2021(1), 1–16. <https://doi.org/10.1002/ets2.12324>
- Haag, N., Heppt, B., Stanat, P., Kuhl, P. and Pant, H. A. (2013). Second language learners’ performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, 28, 24–34. <https://doi.org/10.1016/j.learninstruc.2013.04.001>
- Hipkiss, A., Woods, K. A. and Mccaldin, T. (2021). Students’ use of GCSE access arrangements. *British Journal of Special Education*, 48(1), 50–69. <https://doi.org/10.1111/1467-8578.12347>
- Hjärne, M. S. (2020). Just enough time to level the playing field: Time adaptation in a college admission test. *Scandinavian Journal of Educational Research*, 64(7), 1019–1033. <https://doi.org/10.1080/00313831.2020.1788143>
- Hjärne, M. S. and Lyrén, P. E. (2020). Group differences in the value of subscores: A fairness issue. *Frontiers in Education*, 5, 55. <https://doi.org/10.3389/feduc.2020.00055>
- Huenergarde, M. C., Floyd, R. G. and Harris, B. (2023). Current nationwide practices in U.S. test accommodations for active bilingual learners/users of English (ABLE) students. *Bilingual Research Journal*, 45(3–4), 380–400. <https://doi.org/10.1080/15235882.2022.2155885>
- Johnson, A., Barker, E. and Cespedes, M. V. (2024). Reframing research and assessment practices: Advancing an antiracist and anti-ableist research agenda. *Educational Measurement: Issues and Practice*, 43(3), 95–105. <https://doi.org/10.1111/emip.12601>
- Karacabey, M. F. (2021). *Nitel veri toplama teknikleri*. In S. Şen and İ. Yıldırım (Eds.), *Eğitimde araştırma yöntemleri* (pp. 407–424). Nobel Akademik Yayıncılık.
- Koşak-Babuder, M., Kormos, J., Ratajczak, M. and Piżorn, K. (2019). The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners. *Language Testing*, 36(1), 51–75. <https://doi.org/10.1177/0265532218756946>
- Lane, S. and Leventhal, B. (2015). Psychometric challenges in assessing English language learners and students with disabilities. *Review of Research in Education*, 39(1), 165–214. <https://doi.org/10.3102/0091732X14556073>
- Lazarus, S. S., Thurlow, M. L., Lail, K. E. and Christensen, L. (2009). A longitudinal analysis of state accommodations policies: Twelve years of change, 1993–2005. *Journal of Special Education*, 43(2), 67–80. <https://doi.org/10.1177/0022466907313524>

- Lee, H. S. and Geisinger, K. F. (2014). The effect of propensity scores on DIF analysis: Inference on the potential cause of DIF. *International Journal of Testing*, 14(4), 313–338. <https://doi.org/10.1080/15305058.2014.922567>
- Leppo, R. H. T., Cawthon, S. W. and Bond, M. P. (2014). Including deaf and hard-of-hearing students with co-occurring disabilities in the accommodations discussion. *Journal of Deaf Studies and Deaf Education*, 19(2), 189–202. <https://doi.org/10.1093/deafed/ent029>
- Li, H. and Suen, H. K. (2012). Are test accommodations for English language learners fair? *Language Assessment Quarterly*, 9(3), 293–309. <https://doi.org/10.1080/15434303.2011.653843>
- Liu, D. and Xu, B. (2017). Test anxiety: Perceptions of American community college nursing students. *Empirical Research in Vocational Education and Training*, 9(1), 4. <https://doi.org/10.1186/s40461-017-0048-1>
- Mathes, N. E., Witmer, S. E. and Volker, M. A. (2020). Middle school teachers' perceptions of academic and behavioral support testing accommodations. *Journal of Applied School Psychology*, 36(3), 293–323. <https://doi.org/10.1080/15377903.2020.1749202>
- Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., Debarger, A., Murray, E. and Vendlinski, T. (2013). A “conditional” sense of fairness in assessment. *Educational Research and Evaluation*, 19(2–3), 121–140. <https://doi.org/10.1080/13803611.2013.767614>
- Motteram, J., Spiby, R., Bellhouse, G. and Sroka, K. (2023). Implementation of an accommodations policy for candidates with diverse needs in a large-scale testing system. *Language Testing*, 40(4), 904–932. <https://doi.org/10.1177/02655322231166587>
- O'Neill, R., Cameron, A., Burns, E. and Quinn, G. (2020). Exploring alternative assessments for signing deaf candidates. *Psychology in the Schools*, 57(3), 344–361. <https://doi.org/10.1002/pits.22326>
- Ofqual. (2021). *Systematic divergence between teacher and test-based assessment: Literature review*. Office of Qualifications and Examinations Regulation Publishers.
- Okagbue, H. I. and Teixeira da Silva, J. A. (2020). Correlation between the CiteScore and Journal Impact Factor of top-ranked library and information science journals. *Scientometrics*, 124(1), 797–801. <https://doi.org/10.1007/s11192-020-03457-x>
- Papuda-Dolińska, B., Knopik, T., Krasowicz-Kupis, G. and Wiek, K. (2023). The psychological assessment of students with diverse developmental needs – Universal design approach. *International Journal of Special Education*, 38(3), 47–49. <https://doi.org/10.52291/ijse.2023.38.38>
- Rasooli, A., Razmjooe, M., Cumming, J., Dickson, E. and Webster, A. (2021). Conceptualising a fairness framework for assessment adjusted practices for students with disability: An empirical study. *Assessment in Education: Principles, Policy and Practice*, 28(3), 301–321. <https://doi.org/10.1080/0969594X.2021.1932736>
- Razmjooe, M., Cumming, J. and Wyatt-Smith, C. (2023). A case study of effective classroom assessment adjustments for a student with disability: The role of teacher pedagogical mobility in assessment adjustments. *Education Sciences*, 13(12), 1178. <https://doi.org/10.3390/educsci13121178>
- Ricketts, C., Brice, J. and Coombes, L. (2010). Are multiple choice tests fair to medical students with specific learning disabilities? *Advances in Health Sciences Education*, 15(2), 265–275. <https://doi.org/10.1007/s10459-009-9197-8>

- Roschmann, S., Witmer, S. E. and Volker, M. A. (2021). Examining provision and sufficiency of testing accommodations for English learners. *International Journal of Testing*, 21(1), 32–55. <https://doi.org/10.1080/15305058.2021.1884872>
- Saka, N., Kleper, D. and Kennet-Cohen, T. (2022). Assessing fairness in selection toward applicants who request accommodations in higher education admissions tests. *Assessment in Education: Principles, Policy and Practice*, 29(4), 441–461. <https://doi.org/10.1080/0969594X.2022.2122400>
- Şenel, S. (2021). Assessing measurement invariance of Turkish “Central Examination for Secondary Education Institutions” for visually impaired students. *Educational Assessment, Evaluation and Accountability*, 33(4), 621–648. <https://doi.org/10.1007/s11092-020-09345-5>
- Şenel, S. and Kutlu, Ö. (2018). Comparison of two test methods for VIS: Paper-pencil test and CAT. *European Journal of Special Needs Education*, 33(5), 631–645. <https://doi.org/10.1080/08856257.2017.1391014>
- Seo, D. G. and Hao, S. (2016). Scale comparability between nonaccommodated and accommodated forms of a statewide high school assessment: Using LZ person-fit. *Journal of Psychoeducational Assessment*, 34(3), 230–243. <https://doi.org/10.1177/0734282915596126>
- Sinharay, S., and Johnson, M. S. (2024). Computation and accuracy evaluation of comparable scores on culturally responsive assessments. *Journal of Educational Measurement*, 61(1), 5–46. <https://doi.org/10.1111/jedm.12381>
- Sireci, S. G. (2020). Standardization and understandardization in educational assessment. *Educational Measurement: Issues and Practice*, 39(3), 100–105. <https://doi.org/10.1111/emip.12377>
- Sireci, S. G. and Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education*, 39(1), 215–252. <https://doi.org/10.3102/0091732X14557003>
- Sireci, S. G., Scarpati, S. E. and Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457–490. <https://doi.org/10.3102/00346543075004457>
- Solano-Flores, G. (2014). Probabilistic approaches to examining linguistic features of test items and their effect on the performance of English language learners. *Applied Measurement in Education*, 27(4), 236–247. <https://doi.org/10.1080/08957347.2014.944308>
- Solano-Flores, G. (2019). Examining cultural responsiveness in large-scale assessment: The matrix of evidence for validity argumentation. *Frontiers in Education*, 4, 43. <https://doi.org/10.3389/feduc.2019.00043>
- Solano-Flores, G., Chia, M. and Kachchaf, R. (2019). Design and use of pop-up illustration glossaries as accessibility resources for second language learners in computer-administered tests in a large-scale assessment system. *International Multilingual Research Journal*, 13(4), 277–293. <https://doi.org/10.1080/19313152.2019.1611338>
- Stanczak, A., Aelenei, C., Pironom, J., Toczec-Capelle, M. C., Rohmer, O. and Jury, M. (2024). Can students with special educational needs overcome the “success” expectations? *Social Psychology of Education*, 27(3), 687–708. <https://doi.org/10.1007/s11218-023-09806-x>
- Strauss, A. and Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications.

- Svetina, D., Dai, S. and Wang, X. (2017). Use of cognitive diagnostic model to study differential item functioning in accommodations. *Behaviormetrika*, 44(2), 313–349. <https://doi.org/10.1007/s41237-017-0021-0>
- Thurlow, M. L. and Kopriva, R. J. (2015). Advancing accessibility and accommodations in content assessments for students with disabilities and English learners. *Review of Research in Education*, 39(1), 331–369. <https://doi.org/10.3102/0091732X14556076>
- Turkan, S. and Buzick, H. M. (2016). Complexities and issues to consider in the evaluation of content teachers of English language learners. *Urban Education*, 51(2), 221–248. <https://doi.org/10.1177/0042085914543111>
- Vidal Rodeiro, C. and Macinska, S. (2022). Equal opportunity or unfair advantage? The impact of test accommodations on performance in high-stakes assessments. *Assessment in Education: Principles, Policy and Practice*, 29(4), 462–481. <https://doi.org/10.1080/0969594X.2022.2121680>
- Wei, X. and Zhang, S. (2024). Extended time accommodation and the academic, behavioral, and psychological outcomes of students with learning disabilities. *Journal of Learning Disabilities*, 57(4), 242–254. <https://doi.org/10.1177/00222194231195624>
- Witmer, S. E. and Bouck, E. C. (2023). Predictors of accessibility tool use on a low-stakes computer-based math test. *Assessment for Effective Intervention*, 49(1), 7–17. <https://doi.org/10.1177/15345084231152477>
- Woods, K., James, A. and Hipkiss, A. (2018). Best practice in access arrangements made for England's General Certificates of Secondary Education (GCSEs): Where are we 10 years on? *British Journal of Special Education*, 45(3), 236–255. <https://doi.org/10.1111/1467-8578.12221>
- Woods, K., Mccaldin, T., Hipkiss, A., Tyrell, B. and Dawes, M. (2019). Linking rights, needs, and fairness in high-stakes assessments and examinations. *Oxford Review of Education*, 45(1), 86–101. <https://doi.org/10.1080/03054985.2018.1494555>
- Yang, X. (2020). Assessment accommodations for emergent bilinguals in mainstream classroom assessments: A targeted literature review. *International Multilingual Research Journal*, 14(3), 217–232. <https://doi.org/10.1080/19313152.2019.1681615>
- Yıldırım, A. ve Şimşek, H. (n.d.). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.
- Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J. and Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment*, 13(2–3), 170–192. <https://doi.org/10.1080/10627190802394388>
- Zhu, J. and Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>

## **Extended Abstract**

### **Introduction**

Ensuring fairness and equity in educational testing is a cornerstone of valid and inclusive assessment practices. Test fairness involves designing and administering assessments that provide equal opportunities for all examinees, regardless of gender, race, socioeconomic status, or disability, to demonstrate their knowledge and skills without bias (AERA, APA and NCME, 2014). However, barriers such as linguistic complexity, accessibility limitations, or inadequate accommodations can introduce measurement bias, disproportionately affecting subgroups like students with disabilities or English language learners (ELLs) (Dembitzer and Kettler, 2023). Test accommodations, such as extended time, read-aloud support, or technological aids, aim to mitigate these barriers while preserving the validity of test scores (Ofqual, 2020). Despite advancements, questions remain about the effectiveness, equitable distribution, and theoretical foundations of these accommodations, necessitating a comprehensive review of the research landscape.

This study conducts a bibliometric and systematic review of 63 English-language articles published between 2000 and 2024, sourced from the Scopus database, to analyze global trends in test fairness and accommodations research. The study addresses six key research questions to provide a holistic understanding of the field:

1. How are studies on test fairness distributed across years?
2. Which journals are most prominent in publishing test fairness research?
3. Which articles receive the highest citations?
4. What are the most frequently used keywords and their trends over time?
5. Which countries contribute most to the scientific output in this area?
6. What are the main research themes, methodologies, and contributions to the field?

By answering these questions, the study seeks to inform educators, policymakers, and researchers about trends, gaps, and best practices to foster equitable testing environments globally.

### **Method**

This study adopts a qualitative document analysis approach, integrating bibliometric and systematic review methodologies to examine the literature on test fairness and accommodations (Yıldırım and Şimşek, 2021). The research sample was selected using criterion sampling, focusing on open-access, English-language articles published between 2000 and 2024 in the Scopus database, chosen for its comprehensive coverage and robust citation metrics (Zhu and Liu, 2020). Search terms included “test fairness,” “test accommodations,” “access arrangements,” and “disabilities,” combined with Boolean operators (AND/OR) to identify relevant studies.

The data collection process adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure transparency and rigor. An initial search yielded 129 articles. After applying strict inclusion and exclusion criteria, 44 articles were removed for irrelevance, 85 were evaluated for content, methodology, and results, and an additional 22 were excluded (1 due to publication year, 18 for being off-topic, and 3 duplicates), resulting in a final sample of 63 articles.

Data analysis occurred in two phases. First, bibliometric analysis was conducted using the Biblioshiny interface and the bibliometrix package in R to map trends in publication years, journals, authors, countries, and keywords (Aria and Cuccurullo, 2017). This approach provided quantitative insights into the research landscape, such as publication frequency and citation patterns. Second, a systematic review categorized the 63 articles based on publication year, research objectives, methodologies, sample characteristics, and contributions to the field. Descriptive analysis was employed to classify data using predefined codes, with findings visualized through tables, figures, and thematic summaries (Strauss and Corbin, 1990). The study is limited to open-access articles in Scopus, excluding other databases or restricted-access publications..

## Findings

The bibliometric and systematic analysis of 63 articles provided comprehensive insights into test fairness and accommodations research, addressing the six research questions.

**Distribution of Studies by Year:** Publications were sparse from 2000 to 2012 (7 articles total), with no studies in 2011. A steady increase began in 2013, peaking in 2019 and 2023 with eight articles each. The upward trend post-2013 reflects growing interest in inclusive assessment practices.

**Prominent Journals:** Educational Assessment published the most articles (n=5), followed by Review of Research in Education (n=4), and Assessment in Education: Principles, Policy and Practice, Educational Measurement: Issues and Practice, Frontiers in Education, and Language Testing (n=3 each). The diversity of journals highlights the interdisciplinary nature of test fairness, spanning measurement, special education, and multilingualism.

**Highly Cited Articles:** The most cited article, Haag et al. (2013) in Learning and Instruction, received 112 citations, followed by Sireci (2020, n=67), Mislevy (2013, n=57), and Bennett (2023, n=56). These works underscore the influence of researchers like Sireci, Abedi, and Cawthon in shaping the field.

**Keyword Trends:** The most frequent keywords were “assessment” (n=10), “fairness” (n=10), and “accommodations” (n=9), followed by “accessibility,” “education,” and “educational measurement” (n=5 each). Terms like “English learners” and “learning disabilities” (n=4 each) indicate a focus on linguistic and disability-related challenges. Keywords like “differential item functioning” and “validity” reflect theoretical and practical concerns.

**Country Contributions:** The United States led with 101 contributions, followed by the United Kingdom (n=17) and Australia (n=10). Turkey contributed four publications. The higher total frequency (exceeding 63 articles) results from multi-country authorship, as bibliometrix counts each country’s contribution separately.

**Main Themes, Methods, and Contributions:** The systematic review identified five key themes: Linguistic Factors and Complexity (n=12): Studies using differential item functioning (DIF) analysis, regression, and theoretical reviews found that academic language and complex sentence structures negatively impact ELLs’ performance in math and science (Haag, 2013; Abedi, 2014).

Test Accommodations (n=18): Experimental designs, meta-analyses, and DIF analyses showed that accommodations like extended time and read-aloud tools sometimes enhance fairness but may introduce disadvantages (Sireci, 2020; Leppo, 2014).

Students with Disabilities and Accessibility (n=14): Qualitative, quantitative, and mixed-methods studies revealed DIF in standard tests for visually and hearing-impaired students, with technological tools like computer-adaptive testing improving access (Cawthon, 2013; Şenel, 2021).

Fairness and Validity Frameworks (n=10): Theoretical analyses proposed concepts like universal design and understandization, advocating for inclusive assessment frameworks (Mislevy, 2013; Bennett, 2023).

Policy and Historical Analyses (n=9): Time-series and policy document analyses explored state policies and high-stakes testing fairness, such as GCSE accommodations (Thurlow, 2009; Vidal Rodeiro, 2022).

## Discussion

The findings reveal a growing emphasis on test fairness, particularly after 2013, likely driven by global policy shifts such as the UK’s 2010 Equality Act, which prioritized inclusive education (Hipkiss et al., 2021). The prominence of journals like Educational Assessment and the interdisciplinary spread of publications reflect the topic’s relevance across educational measurement, special education, and multilingual contexts. The frequent use of keywords like “fairness” and “accommodations” underscores a dual focus on theoretical constructs and practical interventions, particularly for ELLs and students with disabilities (AERA et al., 2014).

Systematic findings highlight that linguistic complexity often measures language proficiency rather than content knowledge for ELLs, leading to biased outcomes in subjects like math and science (Haag et al., 2013; Solano-Flores, 2014). Test accommodations, the most studied theme, show mixed

results: while extended time and read-aloud tools improve access for some groups, they may create unfair advantages or compromise score validity, posing challenges for standardized testing (Abedi, 2014; Finch et al., 2009). For students with disabilities, accessibility remains a significant barrier, with DIF persisting in standard test formats. However, innovations like computer-adaptive testing and assistive technologies offer promising solutions (Cawthon, 2013). Theoretical advancements, such as “understandardization,” propose balancing universal design principles with individualized flexibility to enhance fairness (Sireci, 2020). Policy analyses reveal progress in regulatory frameworks, such as those governing GCSE accommodations, but gaps persist in addressing cultural and regional variations, particularly in underrepresented countries like Turkey.

The contradictory effects of accommodations—improving access while potentially undermining validity—highlight the need for careful design and empirical validation. The limited research from non-Western contexts suggests a critical gap in understanding how cultural and regional factors influence test fairness, necessitating further investigation.

### **Recommendations**

Based on the findings, countries should adopt universal design principles while allowing flexible accommodations, such as standardized extended time or technological aids, validated by performance data (Ofqual, 2020), and strengthen collaboration among schools, testing agencies, and regulators to ensure equitable implementation (Woods et al., 2018). Future research should explore test fairness through social justice and cultural diversity lenses, particularly in underrepresented regions like Turkey, and conduct longitudinal studies to assess the long-term impact of policy changes on fairness outcomes. Practically, test developers should reduce linguistic complexity, integrate accessibility tools like read-aloud features, and conduct pilot tests to maintain validity, while providing educator training for effective accommodation implementation. This study offers a roadmap for educators, policymakers, and researchers to advance fair, inclusive testing practices, balancing universal standards with individualized needs.