



JOURNAL OF RESEARCH  
IN EDUCATION AND SOCIETY  
EĞİTİM VE TOPLUM  
ARAŞTIRMALARI DERGİSİ  
ISSN: 2458 - 9624 (Online)



*Eğitim ve Toplum Araştırmaları Dergisi/JRES, 7(1),323-348, 2020*

## REHBERLİK VE ARAŞTIRMA MERKEZLERİNDE SIKLIKLA KULLANILAN PSİKOLOJİK ÖLÇEKLERİN GEÇERLİĞİNİN ÖLÇME STANDARTLARINA GÖRE İNCELENMESİ

### EXAMINING THE VALIDITY OF THE PSYCHOLOGICAL SCALES FREQUENTLY USED IN GUIDANCE AND RESEARCH CENTERS WITH RESPECT TO MEASUREMENT STANDARDS OF VALIDITY

Buket KEMER<sup>1</sup> ve Mehtap ÇAKAN<sup>2</sup>

<sup>1</sup> Gazi Üniversitesi, Eğitimde Ölçme ve Değerlendirme Bölümü, Ankara, Türkiye. buketkemer94@gmail.com

<sup>2</sup> Gazi Üniversitesi, Eğitimde Ölçme ve Değerlendirme Bölümü, Ankara, Türkiye. mehtapcakan@gazi.edu.tr

*Gönderim Tarihi: 04.05.2020*

*Düzeltilme Tarihi: 13.06.2020*

*Kabul Tarihi: 18.06.2020*

#### Öz

Bu araştırmada rehberlik ve araştırma merkezlerinde sıklıkla kullanılan 4 psikolojik ölçeğin APA, AERA ve NCME tarafından 2014 yılında yayınlanan standartlara göre geçerlik açısından incelenerek eksikliklerin saptanması ve bu doğrultuda Milli Eğitim Bakanlığı ile ölçekleri uygulayan psikolojik danışmanlara yol gösterilmesi amaçlanmıştır. Bu ölçekler RAM yetkilileri tarafından sıklıkla kullanıldığı raporlanan Wechsler Zekâ Ölçeği Geliştirilmiş Formu, Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'dir. Araştırmada APA, AERA ve NCME tarafından yayınlanan psikolojik ölçek standartlarının geçerlik bölümüne dayalı olarak kontrol listesi oluşturulmuş ve ölçekler 3 uzman tarafından bu listeye göre incelenmiştir. İnceleme sonucunda; WISC-R'nin incelendiği toplam 9 standartın 4'ü ile tamamen uyumlu olduğu, 3'ü ile kısmen ve 2'si ile hiç uyumlu olmadığı belirlenmiştir. Stanford- Binet'in ise incelendiği toplam 7 standartın 2'si ile tamamen uyumlu olduğu, 1'i ile kısmen ve 4'ü ile hiç uyumlu olmadığı görülmüştür. Gesell'in 2 standartla tamamen uyumluyken, 2'si ile kısmen ve 2 standart ile hiç uyumlu olmadığı görülmüştür. Leiter'in ise incelendiği toplam 9 standartın 2'siyle tamamen uyumlu olduğu, 4'üyle kısmen ve 3'ü ile hiç uyumlu olmadığı görülmüştür. Ölçeklerle kısmen ve hiç uyumlu olmayan kriterlerin geliştirilmesi gerektiği önerilmiştir.

*Anahtar Kelimeler: Geçerlik standartları, WISC-R, Stanford-Binet, Leiter, Gesell, APA.*

### Abstract

In this study, it was aimed to determine the validity deficiencies with respect to the standards published in 2014 by APA, AERA, and NCME, which are frequently used in guidance and research centers, and hence to guide the psychological counselors who apply the scales with the Ministry of National Education. These scales are Wechsler Intelligence Scale Revised Form, Stanford-Binet Intelligence Scale, Leiter International Performance Scale, and Gesell Development Scale. In the research, a checklist was created based on the validity part of the psychological scale standards published by APA, AERA, and NCME, and the scales were examined according to this list by 3 experts. In terms of validity, it has been determined that WISC-R is fully compatible with 4 of the standards examined, partially compatible with 3, and not compatible with 2. It has been observed that Stanford-Binet is fully compatible with 2 of the standards, partially compatible with 1, and not compatible with 4. In Gesell, it was observed that it was fully compatible with 2 standards, partially compatible with 2, but not compatible with 2. It was observed that Leiter is fully compatible with 2 of the standards, partially compatible with 4 and not compatible with 3. It was recommended that the items which are partially compatible and fully incompatible with the scales need to be developed.

*Keywords: Validity Standards, WISC-R, Stanford-Binet, Leiter, Gesell, APA.*

---

### Giriş

Bireylerin eğitim ya da meslek yaşamları süresince kendi yetenek ve ilgilerine dayalı olarak yönlendirilmeleri önemlidir. Bu anlamda, Milli Eğitim Bakanlığı bireylerin rehberlik hizmetleri tarafından farklı ilgi ve yeteneklerine göre nesnel ölçme ve değerlendirme yöntemleri yardımıyla farklı okul ya da programlara yönlendirildiklerini belirtmiştir (Millî Eğitim Temel Kanunu, 1973). Bunu gerçekleştirmek amacıyla Rehberlik ve Araştırma Merkezlerinde öğrencileri tanıma, tarama ve sorunları saptama amacıyla çeşitli zekâ, yetenek ve başarı ölçekleri; kişilik ölçekleri; akademik ve meslekî envanterler; gelişim tarama envanterleri; kontrol listeleri ve anketler uygulanmaktadır (Millî Eğitim Bakanlığı Rehberlik ve Psikolojik Danışma Hizmetleri Yönetmeliği, 2017). Uygulanan ölçekler ve öğrencilerle yapılan görüşmeler sonucunda öğrencilere ilişkin karar verilmektedir.

İlk yıllarda, Rehberlik ve Araştırma Merkezlerinde, özel eğitime gereksinim duyan ve ilköğretim çağındaki öğrencilere zekâ ve özel yetenek ölçekleri uygulanmıştır. Ayrıca, yurt dışında sıklıkla kullanılan ölçeklerin ülkemize uyarlama çalışmaları yapılmıştır (Öner, 2008). Günümüzde Rehberlik ve Araştırma Merkezleri (RAM); rehberlik hizmetleri bölümü ile özel eğitim hizmetleri bölümünden oluşmaktadır. Rehberlik hizmetleri bölümü hem okul tarafından yönlendirilen öğrencilere psikolojik danışmanlık sağlanması hem de gerekli durumlarda ilgili ölçme araçlarının öğrencilere uygulanmasından sorumlu bulunmaktadır. Okullarda uygulanan psikolojik ölçme araçlarının tespiti, temini ve dağıtımı da bu birimin görevleri arasında bulunmaktadır (Millî Eğitim Bakanlığı Rehberlik ve Psikolojik Danışma Hizmetleri Yönetmeliği, 2017).

Özgüven (2014) psikolojik ölçekleri; pekçok özelliğe göre kişilerarası farklılıkları ayırt etmek amacıyla kullanılan ölçme araçları olarak tanımlamaktadır. Kullanım amaçları ise aynı koşullarda kişiler arası farkları ölçme ya da farklı koşullarda aynı kişiyi ölçerek seçme, sınıflama, uygulanan yöntemin etkisini ölçme, hipotezleri kontrol etme, teşhis veya yordama işlemlerinden birini uygulamaktır. Seçmede kriterlere en uygun kişiyi bulmak, sınıflamada ise kişiyi en uygun kategoriye yerleştirmek amaçlanmaktadır (Özgüven, 2014).

Psikolojik ölçeklerde aracın işlevlerini anlamlı ve doğru gerçekleştirebilmesi için bulundurma gereken birtakım psikometrik özellikler vardır ve bunların başında ölçeğin geçerliği gelmektedir. Kelly (1927) geçerliği ölçülmek istenen özelliğin ölçülebilme derecesi olarak açıklarken; Cronbach ve Meehl (1955) ise ölçeğin kullanım amacına uygunluğunu gösteren kanıtları toplama olarak tanımlamaktadır. Temel üç geçerlik türünün ilki olan kapsam geçerliği, ölçülmesi beklenen davranışları kapsama seviyesi olarak tanımlanmaktadır. Bu geçerlik türünün saptanmasında uzman kanısına başvurma, aynı kapsamı ölçtüğü bilinen ölçekle korelasyon gibi yaklaşımlardan yararlanılmaktadır. Diğer bir geçerlik türü olan ölçüte dayalı geçerlikte ise yordayıcı değişkenin ölçütle olan korelasyonu esas almaktadır. Bu geçerlik türü ölçüt puanlarının yordayıcı puanlarla aynı anda ya da öncesinde elde edilme durumuna göre uygunluk geçerliği ve yordayıcı puanlardan sonra elde edilme durumuna göre ise yordama geçerliği olarak ikiye ayrılmaktadır. Diğer bir geçerlik türü olan yapı geçerliğinde ise bireylerin o yapıyla ilgili özelliklerini ölçek sonuçlarına yansıtması beklenmektedir (Baykul, 2010). APA yayınladığı standartlarda geçerliği, ölçek puanlarının amaçlanan yorumu ve önerilen kullanımı lehinde ve aleyhinde argümanlar oluşturma ve değerlendirme süreci olarak tanımlamaktadır.

Güvenirlilik ise Karasar (2012) tarafından tanımlanırken ölçmenin tesadüfi hatalardan arınık olması şeklinde tanımlanırken; Mehrens ve Lehman (1987) tarafından ise aynı şeylerin bağımsız ölçümleri arası tutarlılık olarak açıklanmaktadır.

Ölçeklerde geçerliğin ön koşulu güvenirlilik olmasına rağmen tek başına güvenirlilik iyi bir ölçme aracının işareti değildir. Temel olarak 3 çeşit güvenirlilik ölçütü ölçmedeki zamana göre değişmezlik, bağımsız gözlemciler arası uyum ve ölçeğin iç tutarlılığıdır. Zamana göre değişmezlikte benzer koşullarda ve farklı zamanda yapılan ölçümler arası değişmezliğe bakılmaktadır. Bağımsız gözlemciler arası uyumda, aynı şeylerin ölçümünde birbirinden bağımsız gözlemciler arasındaki tutarlılık incelenmektedir. İç tutarlılık ise ölçme aracının bir bütün oluşturmak için eşit ağırlıklı ve bağımsız bölümlerden oluşması varsayımıdır.

Bu araştırma kapsamında 10 RAM yetkilisiyle görüşmeler yapılmış ve en sık kullandıkları 3'er psikolojik ölçek öğrenilmiştir. Bu görüşmeler sonucunda RAM'larda kullanılan psikolojik ölçme araçları arasında kullanım sıklığı açısından 4'ü öne çıktığı görülmüştür. Bunlardan biri, Wechsler Zekâ Ölçeği Geliştirilmiş Formu (WISC-R)'dur. Bu ölçek David Wechsler tarafından 1949 yılında hazırlanan Wechsler Zekâ Ölçeği'nin 1974 yılında bazı revizyonlarla geliştirilmiş şeklidir. Ölçek, sözel becerileri ve performans dayalı becerileri ölçmeye yönelik 6 ila 16 yaş arası çocuklara bireysel olarak uygulanan bir zekâ ölçeğidir. WISC-R, Şavaşır ve Şahin tarafından 1995 yılında Türkiye'ye uyarlanmıştır. Yapı geçerliği sözel beceriler için .34, performans dayalı beceriler için .33 ile .68 ve tüm ölçek için .38 ile .74 değerleri arasında raporlanmıştır. George ve Mallery (2003) Cronbach Alfa korelasyon katsayılarının .50'den düşükse kabul edilemez çok zayıf, .50 ile .60 arası zayıf, .60 ile .70 arası kabul edilebilir, .70 ile .90 arası iyi ve .90 üstü çok mükemmel şeklinde yorumlanabileceğini belirtmiştir. Test- tekrar test güvenilirliği ise sözel bölümde .87 ile .95, performans bölümünde .88 ile .93 ve toplamda .89 ile .95 arasında raporlanmıştır. Sözel bölüme ilişkin güvenilirlik katsayısı .94 iken, performans bölümü için .96 olarak raporlanmıştır (Şavaşır ve Şahin, 1974). Ölçeğin 2003 yılında revize edilen versiyonu WISC-IV için 243 çocuk üzerinde yaklaşık 1 ay arayla test-tekrar test güvenilirliği raporlanmış ve bu değerler alttestlerde .70 ile .90 arasında değişirken tüm ölçek için .97 olarak raporlanmıştır. Çalışmada değerlerin zaman içerisinde tutarlı olduğu raporlanmıştır (Williams, Weiss ve Rolfhus, 2003). Ölçeğin korelasyonu WPPSI III ile .79 ile .89 arası; WISC III ile .72 ile .89 arası; WAIS III ile .73 ile .89 arası ve WASI ile ise .78 ile .86 olarak belirtilmiştir. (Wechsler, 2004).

Sık kullanılan diğer bir ölçek ise Stanford-Binet Ölçeğidir. Bu ölçek ilk olarak 1905 yılında Paris'teki çocukların başarısızlık nedenlerini araştırmayı da içeren amaçlarla geliştirilmiştir. Stanford – Binet 2 ile 18 yaş arasındaki kişilere uygulanan bireysel bir zekâ ölçeğidir. Ölçek; cümleler, tersine sayma, okuma ve anlatma, nedenleri bulma, muhakeme, şifreler, eksik resimleri ve küpleri sayma gibi hem sözel hem de performans yönelik becerileri gerektirmektedir (Özgüven, 2014). Yapılan pek çok revizyonun sonuncusu olan Stanford-Binet Zekâ Testi-V 2003 yılında yayınlanmıştır. RAM'larda Uğurel-Şemin (1980) tarafından dilimize çevrilen ölçeğin 1972 yılı revizyonu kullanılmaktadır. Ülkemizde 5- 8 yaş arası öğrencilere uygulanan ölçeğin test- tekrar test güvenilirlik katsayısı .90 olarak raporlanırken alttestlerin güvenilirlikleri ise .70 ile .80 arasında raporlanmıştır. Ölçeğin kapsam geçerliğine ilişkin olarak, kelime alttestleri puanları ile yaş arası korelasyonlar .70'den büyük bulunmuştur, Ölçüt geçerliğini saptamak amaçlı ise sözel dersler ve ölçek puanları arası

korelasyon hesaplanarak .60 ila .70 arasında raporlanmıştır. Sayısal derslerle olan korelasyon ise .40 olarak belirlenmiştir (akt. Öner, 2008). Stanford-Binet V (SB-V), ABD’de 4800 kişi üzerinde standartlaştırılmıştır. Ölçeğin iç tutarlılık güvenirlik katsayısı .95 ile .98 arasında raporlanırken, iki yarı güvenirlik katsayıları alt testlerde .84 ile .89 arasında raporlanmıştır. Ölçeğin kapsam geçerliği için uzmanlardan yardım alınarak ölçeğin zekâ testine uygunluğu ile gerekli konuları kapsama durumları incelenmiştir. Ölçeğin eşzamanlı geçerliği Stanford Binet-IV ile .90, Wechsler Okulöncesi Zekâ Ölçeği ile .83 ve Wechsler Zekâ Ölçeği-III ile ise .84 raporlanmıştır (Roid, 2003).

Diğer bir ölçek ise, Rusell Leiter tarafından 1939 yılında geliştirilen Leiter Uluslararası Performans Ölçeği’dir. Ölçek 5-16 yaş arasındaki bireylere uygulanmakta ve her yaş grubu için farklı alttest içermektedir. Sözel olmayan bir zekâ ölçeği olduğu için ölçek uygulayıcıları talimatları katılımcılara hareketleriyle anlatmaktadır. Ölçekte, katılımcının verilen blokları alttestlerdeki örneğe benzer şekilde yerleştirmesi beklenmektedir. Puanlayıcı öğrencinin uygulamasının doğru olup olmadığına bakarak puanlama yapmaktadır (Leiter, 1936). RAM’larda orijinal hali kullanılan ölçeğin Türkiye’deki alanyazında herhangi bir geçerlik ya da güvenirlik çalışması bulunmamaktadır. Yurt dışında 3 ile 6 yaş arasındaki çocuklar üzerinde yapılan araştırmalara bakıldığında, ölçeğin test-tekrar test güvenirliği için .70 ile .80 arasında değerler raporlanırken Cronbach alfa katsayısı olarak ise .75 ile .83 arasında değerler raporlanmıştır. Yapı geçerliği için ise Vineland Ölçeği’yle yakınsak geçerliğine bakılmış ve anlamlı bulunmuştur (Kashani & Faramarzi, 2017). Ölçeğin 1994 yılında yayınlanan Gözden Geçirilmiş Leiter Uluslararası Ölçeği versiyonu, yaşları 2 ila 20 arasında değişen toplam 1719 kişiye uygulanarak geliştirilmiştir. Bu çalışmada iç tutarlılık için Cronbach alfa katsayısı 2 ile 11 yaş arasındaki bireyler için VR bölümü alttestlerinde .75 ile .90 arasında; 12 ila 20 yaş için ise MA bölüm alttestlerinde .67 ile .87 arasında raporlanmıştır. Toplam puan güvenirlikleri ise .91 ile .93 arasında raporlanmıştır. Ölçeğin kritere bağlı geçerliğinde ise bilişsel bozuklukları belirlemede % 96 oranında doğruluk payı olduğu belirtilmiştir. Eşzamanlı geçerlikte ise WISC–III ile korelasyon katsayısı .96 olarak raporlanmıştır (Bradley-Johnson, 1998).

Araştırmada ele alınan diğer bir ölçek ise Gesell Gelişim Ölçeği’dir. Bu ölçek “School readiness: behavior tests used at the Gesell Institute” adıyla Arnold Gesell’in iş arkadaşları olan Frances Jig ve Louise Bate Ames tarafından 1939 yılında yayımlanmıştır. Öğrencilerin gelişim düzeylerini ölçmeyi amaçlayan ölçek, 4-10 yaş arası çocuklara uygulanabilmektedir. RAM’larda Gesell Gelişim Ölçeği olarak sadece 1939 yılında yayınlanan versiyonunun

“kopya formları” adlı alttest uygulanmaktadır. Sonuçların yorumlanmasında yaş aralıkları 6 aylık sürelerle bölünerek her alttest için ayrı kriterler belirtilmiştir. Çocuğun bu sınırlar içinde olması “normal gelişim gösterdiği”, sınırların altında olması “gelişiminde gerilik olduğu”, sınırların üstünde olması ise “gelişiminin yaşlıtlarına göre ileri olduğu” şeklinde yorumlanmaktadır. Gesell’in 1939 versiyonunda uluslararası alanda herhangi bir geçerlik ve güvenilirlik çalışmasına yer verilmemiştir (Ilg & Ames, 1972). Türkiye’de ise 5-6 yaş çocukları ile yapılan güvenilirlik çalışmasında iç tutarlılık katsayısı .44 ve test- tekrar test güvenilirlik katsayısı ise .46 ile .80 arası raporlanmıştır (Evirgen, Kayhan ve Erden, 2015). Ölçekteki yetersiz geçerlik ve güvenilirlik çalışmaları sebebiyle 1978 yılında okula hazırlık ölçeği olarak uyarlanmıştır. Ölçek anaokuluna giden ve yaşları 5.5 ile 7 arasında değişen 46 çocuğa 5 hafta aralıkla uygulanmıştır. Bu çalışmada ölçeğin korelasyon katsayısı Kaufman Çocuklar İçin Değerlendirme Batarya’sı ile .62, test-tekrar test uygulamasıyla ise .73 ve son olarak 10 puanlayıcı arasındaki tutarlılık ise .73 olarak raporlanmıştır (Lichtenstein, 1990).

Araştırma kapsamında ele alınan ölçeklerin ilgili versiyonlarıyla ilişkili yapılan çalışmalarda ölçeklerin hem eski versiyonlarına kıyasla geliştirilmiş özellikleri hem de ölçme standartları açısından eksik yanları belirtilmiştir.

Uluç ve diğerleri (2011) ise WISC-IV’ün yeniliklerini ve geliştirilen özelliklerini açıklamışlardır. Bunların ilki norm çalışmasının coğrafi bölge, cinsiyet ve sosyoekonomik seviyelere göre yeterli düzeyde temsil edilen 2225 çocuk üzerinde ve yakın tarihte yapılmasıdır. Test yapısındaki değişiklik ise toplam test puanıyla korelasyonu düşük olan 3 alttestin çıkarılıp 6 yeni alttestin eklenmesidir. Ayrıca tüm alttestlere, madde güclüğü düşük ve yüksek maddeler eklenip WISC-R’daki tavan-taban etkisi sorunları çözümlenmiştir. Test yönergeleri daha açık hale getirilmiştir. Test yorumlamada ise 5 yeni birleşik zekâ puanı oluşturulmuş ve oluşturulma şekilleri detaylıca açıklanmıştır. Yiğit, Çelik ve Erden (2017) çalışmalarında 50 üstün yetenekli öğrenci üzerinde WISC-R ve WISC-IV ölçeklerini uygulamışlardır. Korelasyon katsayıları; performans puanı ile algısal akıl yürütme dönüştürülmüş puanı arasındaki .42, sözel puan ile sözel kavrama dönüştürülmüş puanı arasında .23 ve toplam ölçek puanları arasında .47 olarak raporlanmıştır. Araştırmacılar WISC-R’ın norm grubunu yenilemede geç kaldığını bu sebeple geçerlik ve güvenilirlik açısından daha güncel olan WISC-IV’ün kullanılmasını önermişlerdir.

Neale (1962) Stanford-Binet Zekâ Ölçeği’ni değerlendirdiği çalışmasında ölçeğin bazı özelliklerinden bahsetmiştir. Bunların ilki el kitapçığının hem ölçek uygulama talimatları hem de puanlama açısından net ve detaylı olmasıdır. IQ tablolarının yeniden düzenlenmesinin de

sonuçların verimli ve hızlı olmasına katkıda bulunduğu belirtilmiştir. Becker (2003) ise çalışmasında Stanford-Binet Zekâ Ölçeği'nin tarihsel gelişiminden bahsetmiş ve her bir revizyonu değerlendirmiştir. RAM'larda kullanılan revizyonunun avantajları el kitapçığının net puanlama talimatlarının olması, 1937 versiyonuna göre en iyi maddeleri içermesi, her yaş seviyesi için alternatif maddelerden oluşması, uygun olmayan maddelerin ölçekten çıkarılmış olması ve sabit bir standart sapmasının olması şeklindedir. Ölçeğin dezavantajları arasında ise genel zekâ olarak tek faktörün ölçülmesi, sözel maddelerin ağırlıklı olması ve yetişkin ya da zekâ düzeyi çok yüksek bireyler için tavan puanının yetersiz kalması şeklinde gösterilmiştir. Stanford-Binet Zekâ Ölçeği'nin 3 versiyonunun karşılaştırıldığı araştırmalarında Tyler-Wood ve arkadaşları (2014) ilgili versiyona yönelik bazı eleştirilerde bulunmuşlardır. Alttestlerin tek tek puanlanmaması, kelime alttesindeki madde türlerinin yaş seviyeleri arasındaki tutarsızlığı, puanlamada maddelerin yaş seviyelerine göre gruplanması ve ölçek bırakma prosedürleri bu eleştirilerden bazılarıdır.

Kaplan ve Saccuzzo (2013) çalışmalarında Gesell Gelişim Ölçeği'nin standardizasyon örnekleminin hedeflenen evreni temsil etmemesini, yönergelerin belirsiz olmasını ve puanlama sisteminin kesin olmamasını eleştirmişlerdir. Ölçeğin 2010'dan önceki versiyonlarındaki psikometrik özelliklerinin kabul edilebilir düzeyde olmadığı da belirtilmiştir. Huelsman (1965) Gesell Gelişim Ölçeği'ni incelediği çalışmasında ölçeğe yönelik bazı eleştirilerde bulunmuştur. Bunlardan bazıları, yaş ya da sınıf seviyesine göre normlarının bulunmaması, standart sapmanın belirtilmemiş olması, toplam zekâ puanı belirtilmemesi, ölçeğin kullanım yararlarının yeterince açıklanmaması ve geçerlik- güvenirlilik çalışmalarının yapılmamış olmasıdır.

Benzer şekilde, Arthur (1949), Leiter Uluslararası Performans Ölçeği'nin normlarının ortalama bir çocuk için çok yüksek seviyede olduğunu ve katılımcıların zekâ seviyelerini yeterince gösteremediklerini belirtmiştir. Her yaş seviyesi için verilen 4 testten birindeki katılımcı hatasının toplam test puanı üzerinde çok etkili olabileceğine bunun da ölçeğin geçerliğini düşüreceğine değinmiştir. Bessent (1949) çalışmasında yaşları 4 ile 58 arasında değişen 20 katılımcıya Stanford- Binet Zekâ Ölçeği ve Leiter Uluslararası Performans Ölçeği'ni uygulamış ve uygulamalar arasındaki Pearson korelasyon katsayı değerini .92 olarak raporlamıştır.

Dirlik (2013) yaptığı çalışmada, bazı psikolojik ölçeklerin özelliklerini incelemek amacıyla APA tarafından 1999 yılında yayınlanan standartları dikkate alarak geçerlik ve güvenirlilik kontrol listesi oluşturmuştur. Çalışmada ülkemizde sıklıkla kullanılan bazı psikolojik ölçekler

bu listeye göre incelemiştir. Çalışma sonuçlarına göre; Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı 10 geçerlik standardından birini, 9 güvenilirlik standardından dördünü karşılamaktadır. Akademik Benlik Kavramı Ölçeği'nin ise geçerlik standartlarının 3'ünü karşılarken güvenilirlik standartlarının hiçbirini karşılayamadığı bulunmuştur.

Bu sonuçları destekleyen bir diğer çalışmada, Yurtsever (2013) Rehberlik ve Araştırma Merkezlerinde çalışan psikolojik danışmanların karşılaştıkları sorunları incelemiş ve danışmanların karşılaştıkları ana sorunlardan biri olarak, kullandıkları ölçme araçlarının güncel olmadığını ifade etmiştir. Benzer şekilde Erol (2010) RAM'larda çalışan psikolojik danışmanların ve velilerin kullanılan ölçme araçlarının eğitsel değerlendirme ve tanılamada yetersiz olduğu görüşünde olduğunu raporlamıştır.

Alanyazın tarandığında, Rehberlik ve Araştırma Merkezleri'ndeki ölçeklerin uluslararası alandaki revizyonlara göre güncellenmediği ve hala eski versiyonlarının kullanıldığı gözlemlenmektedir. Uluslararası alanyazında güncel versiyonları kullanıldığı için ilgili versiyonların çalışmalarının eski tarihli olduğu ve geçerlik açısından önemli eksiklikleri olduğu görülmektedir. Bu çalışma alanyazından hareketle ölçeklerin önemli bir özelliği olan geçerliklerinin incelenmesi gerektiği ihtiyacından esinlenilmiştir.

Dolayısıyla, bu araştırmanın amacı ülkemizde sıklıkla uygulanan Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nun, Stanford-Binet Zekâ Ölçeği'nin, Leiter Uluslararası Performans Ölçeği'nin ve Gesell Gelişim Ölçeği'nin ilgili versiyonlarını APA, AERA ve NCME tarafından yayınlanan ölçek standartlarını dikkate alarak geçerlik açısından incelemek, eksiklikleri saptamak ve bu doğrultuda Milli Eğitim Bakanlığı ve ölçekleri uygulayan psikolojik danışmanlara yol göstermektir. Bu amaç doğrultusunda aşağıdaki sorulara cevap aranmıştır:

1. Belirtilen ölçeklerin APA, AERA ve NCME tarafından 2014 yılında yayınlanan "Standards for Educational and Psychological Testing"te belirtilen geçerlik standartlarına dayalı olarak oluşturulan kontrol listesindeki kriterleri karşılama düzeyleri nedir?

2. Ölçeklerin ilgili geçerlik kriterlerine göre geliştirilmesi gereken özellikleri nelerdir?



## Yöntem

### Araştırmanın Deseni

Bu araştırma nitel bir durum çalışmasıdır. Nitel durum çalışmasında bir ya da birden fazla durum detaylıca incelenir (Yıldırım & Şimşek, 2011). Bu çalışmada RAM'larda sıklıkla uygulanan 4 ölçeğin belirlenen geçerlik kriterlerine dayalı incelemesi yapılmıştır.

### Evren ve Örneklem

İncelenecek olan psikolojik ölçeklerin seçiminde amaçsal örnekleme yöntemlerinden ölçüt örnekleme yöntemi kullanılmıştır ve RAM larda kullanılan ölçekler arasında en sık kullanılan 4 psikolojik ölçek belirlenmiştir. Bu amaçla, ölçeklerin belirlenmesi amacıyla İstanbul ve Ankara'da bulunan 10 RAM yetkilisiyle görüşülmüş ve sık kullandıkları ölçeklerin neler olduğu sorulmuştur. Bu süreçte kullanım sıklığı ölçüt kabul edilmiş ve Wechsler Çocuklar İçin Zekâ Ölçeği'nin Gözden Geçirilmiş Hali (WISC-R), Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim ölçekleri en sık kullanılan 4 ölçek olarak örnekleme alınmıştır.

### Veri Toplama Aracı

Veri toplama aracı olarak 21 standarda yönelik olarak hazırlanan 36 kriterden oluşan bir kontrol listesi geliştirilmiştir. Kontrol listesi hazırlanırken APA, AERA ve NCME tarafından ortak yayınlanan psikolojik ölçeklerdeki geçerlik standartları araştırmacı tarafından dilimize çevrilmiş ve dilsel geçerliği Eğitimde Ölçme ve Değerlendirme alanında doktorasını yapan uzman tarafından kontrol edilmiştir. Listedeki standartların bazıları birden fazla kriter içerdiği için bu standartlar kriterlere ayrılıp ayrı ayrı değerlendirilmiştir. Uyarılama sonucunda kontrol listesi toplamda 36 kriter maddesini içermiştir.

Ölçeklerin kontrol listesindeki kriterlere uyma derecesi 3 kategoride ifade edilmiştir. Buna göre ölçek "Kriterle hiç uyumlu değil" ise 0, "Kriterle kısmen uyumlu değil" ise 1 ve "Kriterle tamamen uyumlu" ise 2 verilmiştir.

Araştırma kapsamında 36 kriterden oluşan kontrol listesi kullanılarak sözkonusu 4 ölçek araştırmacı ve Eğitimde Ölçme ve Değerlendirme alanında doktorasını yapan 2 uzman tarafından puanlanmıştır. Ölçekler kontrol listesine göre incelendikten sonra puanlayıcılar arası tutarlılık Fleiss'in Kappa istatistiği yöntemi ile hesaplanmıştır. Kappa ( $\kappa$ ) ve ağırlıklı Kappa ( $\kappa_w$ ), Cohen tarafından geliştirilen ve hakemlerin tesadüfen benzer kararlar vermiş olabileceklerini dikkate alınan bir düzeltmedir. Bu iki yöntemde de 2 hakem ve 2 veri

kategorisi bulunması gerekmektedir (Şencan, 2005). Fleiss (1971), ikiden fazla hakem ve ikiden fazla kategorinin incelenebilmesi için kappa istatistiğini düzenlemiştir. Araştırmada puanlayıcılar arası uyum düzeyleri Landis ve Koch (1977b) tarafından önerilen aralıklara göre değerlendirilmiştir. Buna göre .00 ile .20 arası “Önemsiz”, .21 ile .40 arası “Düşük düzeyde uyum”, .41 ile .60 arası “Orta düzeyde uyum”, .61 ile .80 arası “Yüksek düzeyde uyum” ve .81 ile 1 arası “Çok yüksek düzeyde uyum” şeklinde yorumlanmıştır. Araştırmalarda çoklu puanlayıcı ya da hakem kullanılması durumlarında kullanılan yaklaşımlardan biri çoğunluğun kararının kabul edilmesidir (Landis ve Koch, 1977a). Bu araştırmada puanlayıcılar arası ortak görüşün olmadığı maddelerde çoğunluğun kararı kabul edilmiştir. Detaylar bulgular bölümünde her bir ölçek sunulurken verilmiştir.

## Bulgular

Araştırma sorularına ilişkin bulgular aşağıda sunulmuştur.

### Birinci Araştırma Sorusuna İlişkin Bulgular:

Ölçekler kontrol listesindeki 36 geçerlik kriterine göre incelenmiştir. Bazı kriterler bazı ölçeklerde uygulanabilir olmadığı için kriter sayısı her ölçekte farklılık göstermektedir. Ölçekle ilgili olmayan kriterler dikkate alınmamıştır. Her ölçekte 1A, 1B ve 2 şeklinde numaralandırılan kriterler incelenmiştir. Ölçeklerin geçerlik kriterlerini karşılama düzeyleri Tablo 1’de özetlenmiş ve devamında ayrıntılı olarak verilmiştir.

Tablo 1

#### *Ölçeklerin Geçerlik Kriterlerini Karşılama Düzeyi*

	Stanford-Binet	WISC-R	Gesell	Leiter
Hiç Uyumlu Değil	4	2	2	3
Kısmen Uyumlu Değil	3	3	2	4
Tamamen Uyumlu	1	4	2	2

Tabloda da belirtildiği üzere Stanford-Binet Zekâ Ölçeği’nin incelendiği 8 kriterin 1’i ile tamamen uyumlu olduğu, 3’ü ile de kısmen ve 4’ü ile de hiç uyumlu olmadığı görülmüştür. WISC-R’in ise incelendiği 9 kriterin 4’ü ile tamamen uyum gösterdiği, 3’ü ile de kısmen, 4’ü ile de hiç uyum göstermediği görülmüştür. Gesell Gelişim Ölçeği’nin ise 6 kriterin 2’si ile tamamen uyumlu olduğu, 2’si ile kısmen uyumlu olmadığı ve 2 si ile de hiç uyumlu olmadığı görülmüştür. Leiter Uluslararası Zekâ Ölçeği’nin ise incelendiği 9 kriterin 2’si ile tamamen

uyum gösterdiği, 4'ü ile de kısmen uyum göstermediği, 3'ü ile de hiç uyum göstermediği görülmüştür.

### *Stanford-Binet Zekâ Ölçeği*

Tablo 2’de Stanford-Binet Zekâ Ölçeği’nin puanlayıcılara dayalı olarak kriterleri karşılama düzeylerine yer verilmiştir. Bu ölçek için puanlayıcılar arası Fleiss Kappa değeri .62 bulunmuştur. Puanlayıcılar arası uyumun yüksek düzeyde olduğu görülmektedir.

Tablo 2

#### *Stanford-Binet Zekâ Ölçeğinin Geçerlik Kriterlerini Karşılama Düzeyi*

	Kriterler	Araştırmacı	Uzman 1	Uzman 2	Karar
1A	Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını açıkça belirtmiştir.	1	1	1	1
1B	Ölçeğin uygulanacağı evren ve ölçeğin değerlendirmeyi amaçladığı yapı veya yapılar açıkça tanımlanmıştır.	2	1	1	1
2	Belirli bir kullanım için verilen ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur	0	0	1	0
3	Belirli bir kullanım için genel veya olası yorumların geçerliği değerlendirilmediyse ya da böyle bir yorum mevcut kanıtlarla tutarsız ise, bu durum açıklanmış ve potansiyel kullanıcılar desteklenmeyen yorumlar konusunda uyarılmıştır.	0	0	0	0
7	Geçerlik kanıtı elde edilen her bir katılımcı örneklemleri ilgili sosyo demografik ve gelişimsel temel özellikleri de içeren, pratik ve izin verilen ölçüde ayrıntılı olarak tanımlanmıştır.	2	1	2	2
9A	Geçerlik kanıtı tek başına ya da diğer değişkenlerin verileriyle ölçek sonuçlarının istatistiksel analizlerini içerdiğinde, verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmıştır.	0	0	0	0
9B	Tipik ölçek koşullarından farklı olması muhtemel olan ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmiştir.	0	0	0	0

“0” Kriterle hiç uyumlu değil

“1” Kriterle kısmen uyumlu değil

“2” Kriterle tamamen uyumlu

Geçerlik kanıtı elde edilen katılımcı örneklemlerinin ayrıntılı olarak tanımlanması gerektiğini açıklayan 7. kriter ile tamamen uyumlu kabul edilmiştir.

Ölçek, puan yorumlaması ve kullanım şeklinin net olarak açıklanması gerektiğini belirten 1A kriteri ile kısmen uyumlu olmadığı kabul edilmiştir. Ölçek, evreninin ve yapılarının açıkça tanımlanması gerektiğini belirten 1B kriteri ile de kısmen uyumlu olmadığı kabul edilmiştir.

Ölçek, ölçek puanlarının yorumlanmasında dayandığı kanıtların ve teorilerin ayrıntılı özetleriyle sunulması gerektiğini vurgulayan 2. kriterle hiç uyum göstermemektedir. Yorumların geçerliği değerlendirilmediyse ya da yorum kanıtlarla tutarsız ise, bu durumun açıklanması ve kullanıcıların uyarılması gerektiğini belirten 3. kriter ile de hiç uyumlu bulunmamıştır. Geçerlik verilerin toplandığı koşulların, istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı açıklanması gerektiğini belirten 9A kriteri ile de benzer şekilde hiç uyumlu bulunmamıştır. Benzer şekilde, ölçek tipik ölçek koşullarından farklı ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmesi gerektiğini vurgulayan 9B kriteri ile de hiç uyum göstermemektedir.

Stanford-Binet Zekâ Ölçeği'nin incelenen kriterlerin 1'i ile tamamen uyumluysen, 2'si ile kısmen ve 4'ü ile hiç uyumlu olmadığı değerlendirilmiştir.

### ***Wechsler Zekâ Ölçeği Geliştirilmiş Formu (WISC-R)***

Tablo 3'de Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nun puanlayıcılara göre dayalı olarak kriterleri karşılama düzeylerine yer verilmiştir. Bu ölçek için puanlayıcılar arası Fleiss Kappa değeri .39 bulunmuştur. Puanlayıcılar arası uyumun istenilen düzeyde olmadığı görülmektedir. Feinstein ve Cicchetti (1990) çalışmalarında Kappa İstatistiğinde oluşabilen belli problemleri açıklamışlardır. Tabloda bulunan bir sütun toplamının diğerlerine oranla fazla olması sonucunda Kappa değerinin çok düşük çıkması da bahsedilen problemlerden birdir. Araştırmada WISC-R'in incelendiği kriterin çoğunda puanlayıcılar ölçeği "kriterle tamamen uyumludur" şeklinde puanlamışlardır. Bu sebeple tablodaki ilgili sütun toplamı diğer sütunlara göre sayısal olarak çok daha büyük olduğu için WISC-R'in Fleiss Kappa istatistik değerinin düşük çıktığı görülmektedir.

Tablo 3

*Wechsler Zekâ Ölçeği Geliştirilmiş Formunun Geçerlik Kriterlerini Karşılama Düzeyi*

Kriterler	Araştırmacı	Uzman 1	Uzman 2	Karar
1A Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını açıkça belirtmiştir.	1	1	2	1
1B Ölçeğin uygulanacağı evren ve ölçeğin değerlendirilmesini amaçladığı yapı veya yapılar açıkça tanımlanmıştır.	2	2	2	2
2 Belirli bir kullanım için verilen ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur.	2	1	2	2
7 Geçerlik kanıtı elde edilen her bir katılımcı örneklemeleri ilgili sosyo demografik ve gelişimsel temel özellikleri de içeren, pratik ve izin verilen ölçüde ayrıntılı olarak tanımlanmıştır.	2	2	2	2
9A Geçerlik kanıtı tek başına ya da diğer değişkenlerin verileriyle ölçek sonuçlarının istatistiksel analizlerini içerdiğinde, verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmıştır.	0	0	1	0
9B Tipik ölçek koşullarından farklı olması muhtemel olan ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmiştir.	0	0	0	0
13A Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmıştır.	0	0	2	0
13B Birleşik puanların geliştirildiği durumlarda, bu puanı oluşturmak için temel ve gerekçeler verilmiştir.	2	2	2	2
17 Belirli bir ölçek performansı seviyesinin yeterli veya yetersiz kriter performansını öngördüğü belirtilmişse, kriter performansının seviyeleri hakkında bilgi verilmiştir.	1	1	2	1

Ölçek, evreninin ve yapılarının açıkça tanımlanması gerektiğini belirten 1B kriteri ile tamamen uyumlu bulunmuştur. Ölçek, ölçek puanlarının yorumlanmasında dayandığı kanıtların ve teorilerin ayrıntılı özetleriyle sunulması gerektiğini vurgulayan 2. kriterle tamamen uyumlu kabul edilmiştir. Ölçeğin, geçerlik kanıtının elde edildiği katılımcı örneklemelerinin ayrıntılı olarak tanımlanması gerektiğini açıklayan 7. kriterle tamamen uyumlu olduğu görülmüştür. Ölçeğin birleşik puanların temel ve gerekçelerinin verilmesi gerektiğini açıklayan 13B kriteri ile de tamamen uyumlu olduğu görülmüştür.

Ölçek, puanlarının yorumlanma ve kullanım şeklinin açıkça belirtilmesi gerektiğini açıklayan 1A kriteri ile kısmen uyumlu olmadığı bulunmuştur. Yine, ölçek, kriter performansının seviyeleri hakkında bilgi verilmesi gerektiğini belirten 17. kriterle kısmen uyumlu olmadığı kabul edilmiştir.

Ölçek, geçerlik verilerin toplandığı koşulların, istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı açıklaması gerektiğini belirten 9A kriteri ile hiç uyumlu değildir şeklinde kabul edilmiştir. Ölçek ayrıca, tipik ölçek koşullarından farklı ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmesi gerektiğini vurgulayan 9B kriteri ile hiç uyum göstermemektedir. Ölçek, alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanması gerektiğini açıklayan 13A kriteri ile kısmen uyumlu olmadığı kabul edilmiştir.

Genel bakıldığında, wechsler zekâ ölçeği geliştirilmiş formu incelendiği kriterlerin 4'ü ile tamamen uyumluyken, 2'si ile kısmen ve 3'ü ile hiç uyumlu değildir.

### *Gesell Gelişim Ölçeği*

Tablo 4'de Gesell Gelişim Ölçeği'nin puanlayıcılara dayalı olarak kriterleri karşılama düzeylerine yer verilmiştir. Bu ölçek için puanlayıcılar arası Fleiss Kappa değeri .70 bulunmuştur. Puanlayıcılar arası uyumun yüksek düzeyde olduğu görülmektedir.

Tablo 4

#### *Gesell Gelişim Ölçeğinin Geçerlik Kriterlerini Karşılama Düzeyi*

Kriterler	Araştırmacı	Uzman 1	Uzman 2	Karar
1A Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını açıkça belirtmiştir.	1	1	1	1
1B Ölçeğin uygulanacağı evren ve ölçeğin değerlendirmeyi amaçladığı yapı veya yapılar açıkça tanımlanmıştır.	1	1	1	1
2 Belirli bir kullanım için verilen ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur	2	2	1	2
3 Belirli bir kullanım için genel veya olası yorumların geçerliği değerlendirilmediyse ya da böyle bir yorum mevcut kanıtlarla tutarsız ise, bu durum açıklanmış ve potansiyel kullanıcılar desteklenmeyen yorumlar konusunda uyarılmıştır.	0	0	0	0
13A Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmıştır.	0	0	0	0
17 Belirli bir ölçek performansı seviyesinin yeterli veya yetersiz kriter performansını öngördüğü belirtilmişse, kriter performansının seviyeleri hakkında bilgi verilmiştir.	2	1	2	2

Ölçek, ölçek puanlarının yorumlanmasında dayandığı kanıtların ve teorilerin ayrıntılı özetleriyle sunulması gerektiğini vurgulayan 2. kriterle tamamen uyumlu kabul edilmiştir. Kriter performansının seviyeleri hakkında bilgi verilmesi gerektiğini belirten 17. kriterle de tamamen uyumlu kabul edilmiştir.

Ölçek, puanlarının yorumlanma ve kullanım şeklinin açıkça belirtilmesi gerektiğini açıklayan 1A kriteri ile kısmen uyumlu değildir. Ölçeğin evreninin ve yapılarının açıkça tanımlanması gerektiğini belirten 1B kriteri ile de kısmen uyum göstermemektedir.

Ölçek, yorumların geçerliği değerlendirilmedi ise ya da yorum kanıtlarla tutarsız ise, bu durumun açıklanması ve kullanıcıların uyarılması gerektiğini belirten 3. kriter ile hiç uyum göstermemektedir. Ölçek, alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanması gerektiğini açıklayan 13A kriteri ile de hiç uyumlu bulunmamıştır.

Gesell Gelişim Ölçeği incelendiği kriterlerin 2'si ile tamamen uyumluysen, 2'si ile kısmen ve 2'si ile hiç uyumlu olmadığı görülmüştür.

### *Leiter Uluslararası Performans Ölçeği*

Tablo 5' de Leiter Uluslararası Performans Ölçeği'nin puanlayıcılara dayalı olarak kriterleri karşılama düzeylerine yer verilmiştir. Bu ölçek için puanlayıcılar arası Fleiss Kappa değeri .64 bulunmuştur. Puanlayıcılar arası uyumun yüksek düzeyde olduğu görülmektedir.

Tablo 5

#### *Leiter Uluslararası Performans Ölçeği'nin Geçerlik Kriterlerini Karşılama Düzeyi*

Kriterler	Araştırmacı	Uzman 1	Uzman 2	Karar
1A Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını açıkça belirtmiştir.	1	1	1	1
1B Ölçeğin uygulanacağı evren ve ölçeğin değerlendirmeyi amaçladığı yapı veya yapılar açıkça tanımlanmıştır.	1	1	1	1
2 Belirli bir kullanım için verilen ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur.	2	2	1	2

7	Geçerlik kanıtı elde edilen her bir katılımcı örneklemi ilgili sosyo demografik ve gelişimsel temel özellikleri de içeren, pratik ve izin verilen ölçüde ayrıntılı olarak tanımlanmıştır.	1	1	1	1
9A	Geçerlik kanıtı tek başına ya da diğer değişkenlerin verileriyle ölçek sonuçlarının istatistiksel analizlerini içerdiğinde, verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmıştır.	0	0	0	0
9B	Tipik ölçek koşullarından farklı olması muhtemel olan ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmiştir.	0	0	0	0
13A	Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmıştır.	0	0	0	0
13B	Birleşik puanların geliştirildiği durumlarda, bu puanı oluşturmak için temel ve gerekçeler verilmiştir.	2	2	1	2
17	Belirli bir ölçek performansı seviyesinin yeterli veya yetersiz kriter performansını öngördüğü belirtilmişse, kriter performansının seviyeleri hakkında bilgi verilmiştir.	1	1	2	1

Leiter Uluslararası Performans Ölçeği'nin, puanların yorumlanmasında dayandığı kanıtların ve teorilerin ayrıntılı özetleriyle sunulması gerektiğini vurgulayan 2. kriterle tamamen uyumlu olduğu görülmüştür. Ölçek, birleşik puanların temel ve gerekçelerinin verilmesi gerektiğini açıklayan 13B kriteri ile de tamamen uyumlu kabul edilmiştir.

Ölçek, puanlarının yorumlanma ve kullanım şeklinin açıkça belirtilmesi gerektiğini açıklayan 1A kriteri ile kısmen uyum sağlamamaktadır. Ölçek evreninin ve yapılarının açıkça tanımlanması gerektiğini belirten 1B kriteri ile de kısmen uyum sağlamamaktadır. Geçerlik kanıtı elde edilen katılımcı örneklemelerinin ayrıntılı olarak tanımlanması gerektiği açıklayan 7. kriterle de kısmen uyumlu değildir. Ölçek, kriter performansının seviyeleri hakkında bilgi verilmesi gerektiğini belirten 17. kriterle de kısmen uyumlu olmadığı kabul edilmiştir.

Ölçek, geçerlik verilerinin toplandığı koşulların, istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı açıklanması gerektiğini belirten 9A kriteri ile hiç uyum göstermemektedir. Ölçeğin, tipik ölçek koşullarından farklı ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmesi gerektiğini vurgulayan 9B kriteri ile de yine hiç uyumlu olmadığı görülmüştür. Ölçek, benzer şekilde alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı



destekleyen gerekçe ile ilgili kanıtlar sağlanması gerektiğini açıklayan 13A kriteri ile de hiç uyum göstermemektedir.

Leiter Uluslararası Performans Ölçeği'nin incelendiği kriterleri 2'si ile tamamen uyumluyken, 4'ü ile kısmen ve 3'ü ile hiç uyumlu olmadığı görülmüştür.

### **Ölçeklerin İlgili Geçerlik Kriterlerine Göre Geliştirilmesi Gereken Özelliklerine İlişkin Bulgular**

Ölçekler, kısmen ya da hiç uyumlu olmadıkları saptanan kriterler açısından geliştirilmelidir şeklinde yorumlanmıştır. Buna göre her bir ölçek aşağıda ayrı ayrı ele alınmıştır.

Stanford-Binet Zekâ Ölçeği'nde geliştirilmesi önerilen standartlar aşağıda sunulmuştur:

- Belirlenen standart puanın yorumlama şekli açıklanmalıdır (Kriter 1A).
- Ölçekte değerlendirilmesi amaçlanan yapılar açıkça belirtilmelidir (Kriter 1B).
- Puanların yorumlanması belli kanıt ya da teoriye dayandırılmalıdır (Kriter 2).
- Puanların her kullanım amacı için geçerliklerin değerlendirilmediği açıklanmalı ve sonuçlarla tutarlı olmayan geçerlikler belirtilmelidir (Kriter 3).
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplama koşulları detaylandırılmalıdır (Kriter 9A)
- Tipik ölçek koşullarından farklı olması olası ve ölçek performansını önemli şekilde etkileyebilecek veri toplama özellikleri açıklanmalıdır (Kriter 9B).

WISC-R'da geliştirilmesi önerilen standartlar aşağıda sunulmuştur:

- Ölçekte yeterli puan aralıkları dışında da puan kriterleri belirtilip nasıl yorumlanacağı açıklanmalıdır (Kriter 1A).
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplama koşulları detaylandırılmalıdır (Kriter 9A).
- Tipik ölçek koşullarından farklı olan ve bireylerin ölçekteki performansını belirgin şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmelidir (Kriter 9B).
- Sözel beceri ve performans dayalı beceri puanları ayrı ayrı yorumlanmalı ya da gerekçesi sunulmalıdır (Kriter 13A).
- Katılımcıların belirlenen puan aralıkları dışında puan alması durumuna yönelik olarak farklı kriter aralıkları ve yorumlamalar tanımlanmalıdır (Kriter 17).

Gesell Gelişim Ölçeği'nde geliştirilmesi önerilen standartlar aşağıda sunulmuştur:

- Aynı kişiler farklı alttestlerde farklı kategorilerde performans gösterirse ne yapılacağı belirtilerek puanlarının yorumlanması düzenlenmelidir (Kriter 1A).
- Ölçeğin değerlendirmeyi amaçladığı yapılar açıkça belirtilmelidir (Kriter 1B).
- Puanların her kullanım amacı için geçerlikleri değerlendirilmediyse ya da sonuçlarla tutarlı değilse bu durum açıkça belirtilmelidir (Kriter 3).
- Alttestlerinin ayırt edilebilirliği, güvenilirliği, aralarındaki ilişki ya da yapıyla olan ilişkileri raporlanmalıdır (Kriter 13A).

Leiter Uluslararası Performans Ölçeği'nde geliştirilmesi önerilen standartlar aşağıda sunulmuştur:

- Ölçekte yeterli puan aralıkları dışında da puan kriterleri belirtilip nasıl yorumlanacağı açıklanmalıdır (Kriter 1A).
- Ölçekte değerlendirilmesi amaçlanan yapılar açıkça belirtilmelidir (Kriter 1B).
- Katılımcıların yaş ve ırklarının yanı sıra diğer sosyo demografik ve gelişimsel temel özellikleri de izin verilen ölçüde açıklanmalıdır (Kriter 7).
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplanma koşulları detaylandırılmalıdır (Kriter 9A).
- Tipik ölçek koşullarından farklı olan ve bireylerin ölçekteki performansını belirgin şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmelidir (Kriter 9B).
- Alt bölüm puanlarının ayırt edilebilirliği, güvenilirliği ve alttest puanları arası ilişkinin yapı ile tutarlı olduğu kanıtlanmalıdır (Kriter 13A).
- Katılımcıların belirlenen puan aralıkları dışında puan alması durumuna yönelik olarak farklı kriter aralıkları ve yorumlanması tanımlanmalıdır (Kriter 17)

## **Sonuç ve Tartışma**

Araştırmada Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nun, Stanford-Binet Zekâ Ölçeği'nin, Leiter Uluslararası Performans Ölçeği'nin ve Gesell Gelişim Ölçeği'nin ilgili versiyonlarını geçerlik açısından incelemek, eksiklikleri saptamak ve bu doğrultuda Milli Eğitim Bakanlığı ve ölçekleri uygulayan psikolojik danışmanlara yol göstermek amaçlanmıştır.

Yapılan incelemeler sonucunda Stanford-Binet Zekâ Ölçeği'ne ilişkin verilerin toplandığı koşulların detaylıca açıklandığı görülmüştür. Ölçek puanlarının yorumlanma ve kullanım şekli, ölçeğin uygulanacağı evren ve değerlendirmeyi amaçladığı yapılar ise kısmen açıklanmıştır. Ancak, ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler

ayrıntılı özetleriyle açıklanmadığı, ölçek performansını etkileyebilecek olan veri toplama özelliklerine dikkat edilmediği, belirli bir kullanım için genel veya olası yorumların geçerliği değerlendirilmediği görülmüştür. Buna bağlı olarak potansiyel kullanıcıların bu noktaları dikkate alması önemli görülmektedir. Benzer şekilde Becker (2003), yetişkin ya da zekâ düzeyi çok yüksek bireylerin tavan puanının yetersiz kaldığını belirterek ölçek puanlamasının dayandığı kanıtların eksik olduğunu belirtmiştir. Alanyazına bakıldığında, Tyler-Wood ve arkadaşlarının (2014) puanlama yaparken maddelerin yaş seviyelerine göre gruplanmasına ilişkin eleştirerek kanıtların eksikliğini vurgulamışlardır.

Araştırma sonucunda WISC-R'da ölçeğin uygulanacağı evrenin ve ölçek yapıların açıkça tanımlandığı, puanların yorumlanmasında dayandığı kanıtların özetleriyle sunulduğu, geçerlik kanıtı elde edilen her bir katılımcı örneklemlerinin ayrıntılı olarak tanımlandığı ve birleşik puanları oluşturmak için gerekçeler verildiği görülmüştür. Ancak, ölçek puanlarının nasıl yorumlanıp kullanılacağı ve kriter performans seviyeleri hakkında kısmen bilgi verildiği görülmüştür. Yine, geçerlik verilerinin toplandığı koşullara ve veri toplama özelliklerine dikkat edilmediği saptanmıştır. Alt bölüm puanlarının yorumlamasını destekleyen gerekçeler ile ilgili kanıtların da sağlanmadığı görülmüştür. Buna bağlı olarak potansiyel kullanıcıların bu noktaları dikkate alması önemli görülmektedir. Benzer şekilde Uluç ve diğerleri (2011) de araştırmalarında verilerin toplandığı koşulların ve veri toplama özelliklerinin yetersiz olduğunu, alttestlerdeki taban-tavan etkisi sebebiyle yeterli performans seviyesi ile ilgili sorunlar olduğunu raporlamışlardır.

Araştırma sonucunda Gesell Gelişim Ölçeği'nde ölçek puanlarının yorumlanmasında dayanılan kanıtların ayrıntılı özetleriyle sunulduğu ve kriter performans seviyeleri hakkında bilgi verildiği görülmüştür. Ölçekte, ölçek puanlarının nasıl yorumlanacağı ve kullanılacağı ise kısmen belirtilmiştir. Ancak, ölçek yapılarının belirtilmediği, genel veya olası yorumların geçerliğinin değerlendirilmediği, potansiyel kullanıcıların desteklenmeyen yorumlar konusunda uyarılmadığı ve alt bölüm puanlarının yorumlamasını destekleyen gerekçe ile ilgili kanıtlar sağlanmadığı görülmüştür. Benzer şekilde Kaplan ve Saccuzzo (2013) Gesell Gelişim Ölçeği'ndeki ölçek uygulama yönergelerinin ve puanlama sisteminin net olmadığını belirtmişlerdir. Huelsman (1965) ise ölçekte yaşa ya da sınıf seviyesine dayalı normlar bulunmamasını, standart sapmanın olmamasını, toplam bir zekâ yaşı belirtilmemesini, ölçeğin kullanım yararlarının yeterince açıklanmamasını, geçerlik ve güvenirlik çalışmasının yapılmamasını eleştirmiştir.

Araştırma sonucunda Leiter Uluslararası Performans Ölçeği'nin ölçek puanlarının yorumlanmasında dayandığı kanıtların ayrıntılı özetleriyle sunulduğu ve birleşik puan oluşturulmasına yönelik gerekçelerin açıklandığı görülmüştür. Ölçeğin yapıları, geçerlik kanıtı elde edilen katılımcı örneklemi, kriter performansının seviyeleri ve puanlarının nasıl yorumlanacağı hakkında ise kısmen bilgi verildiği görülmüştür. Ancak, verilerin toplandığı koşulların ayrıntılı olarak açıklanmadığı; tipik ölçek koşullarından farklı olan veri toplama özelliklerine dikkat edilmediği; alt bölüm puanlarının yorumlamasını destekleyen gerekçe ile ilgili kanıtlar sağlanmadığı görülmüştür.

Benzer bir çalışmada, Arthur (1949) da ölçek normlarının ortalama bir çocuk için yüksek olduğunu ve her yaş seviyesi için uygulanan 4 testin yetersiz olduğunu belirtmiştir. Bu sebeplerin de ölçeğin geçerliğini düşürdüğüne değinilmiştir.

Genel olarak bakıldığında, gerek RAM çalışanları gerekse bireysel kullanıcıların, bu ölçekleri kullanırken ve sonuçları yorumlarken, yukarıda belirtilen eksiklikleri dikkate alması ve özellikle RAM'larda bu ölçeklerin en son versiyonlarının kültürümüze uyarlanarak kullanılması önem arz etmektedir.

## Kaynaklar

American Educational Research Association, American Psychological Association & National Council on Measurement in Education, (2014). *Standards for educational and psychological testing*. Washington: AERA.

Arthur, G. (1949). The Arthur adaptation of the Leiter International Performance Scale. *Journal of Clinical Psychology*, 5(4), 345-349. doi: 10.1002/1097-4679(194910)5:4<345::AID-JCLP2270050402>3.0.CO;2-0

Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme*. Ankara: Pegem.

Becker, K. A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics*. Itasca, IL: Riverside.

Bessent, T. E. (1950). A note on the validity of the Leiter International Performance Scale. *Journal of Consulting Psychology*, 14(3), 234. doi:10.1037/h0063338

Bradley-Johnson, S. (1998). Test reviews. *Psychology in the Schools*, 36(6), 523-525. doi:10.1002/(SICI)1520-6807(199810)35:4<409::AID-PITS16>3.0.CO;2-0

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- Dirlik; E. (2013). *Eğitim kurumlarında kullanılan psikolojik testlerin ölçme standartlarına göre incelenmesi*. Ankara Üniversitesi Yüksek Lisans Tezi, Eğitim Bilimleri Enstitüsü, Ankara.
- Evirgen, N., Kayhan, E., ve Erden, G. (2015, Mayıs). *Gesell Gelişim Figürleri'nin anasınıfı çocuklarında güvenilirliğine yönelik bir ön çalışma*. Uluslararası Katılımlı 3. Çocuk Gelişimi ve Eğitimi Kongresi'nde sunulan bildiri, Ankara.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382. doi: 10.1037/h0031619
- George, D., & Mallery, P. (2003). SPSS for Windows step by step: answers to selected exercises. *A Simple Guide and Reference*, 63, 1461-1470.
- Huelsman, C. B. (1965). A review: school readiness: behavior tests used at the Gesell Institute. *Journal of School Psychology*, 3(3), 78-80. doi: 10.1016/0022-4405(65)90045-2
- Ilg, F. L., & Ames, L. B. (1972). *School readiness : behavior tests used at the Gesell Institute*. New York: Harper & Row.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychology Testing: principles, applications, issues*. Belmont: Thomson Wadsworth.
- Karasar, N. (2012). *Bilimsel araştırma yöntemi*. Ankara: Nobel.
- Kashani, E. T., & Faramarzi, S. (2017). Construction and validation of the Leiter International Performance Scale software for 3 to 6 year old children. *International Journal of Academic Research*, 4(1), 16-33.
- Landis, J. R., & Koch, G. G. (1977a). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363-374. doi:10.2307/2529786
- Landis, J., & Koch, G. (1977b). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310

- Leiter, R. (1936) The Leiter international performance scale. *University Of Hawaii Bulletin*, 15(7), 4-33.
- Lichtenstein, R. (1990). Psychometric characteristics and appropriate use of the Gesell School Readiness Screening Test. *Early Childhood Research Quarterly*, 5(3), 359-378. doi:10.1016/0885-2006(90)90027-X
- Millî Eğitim Bakanlığı Rehberlik Hizmetleri Yönetmeliği. (2017, 10 Kasım). Resmi Gazete. Sayı: 30236. <https://resmigazete.gov.tr/eskiler/2017/11/20171110-2.htm>
- Millî Eğitim Temel Kanunu. (1973, 24 Haziran). Resmi Gazete. Sayı: 14574. <https://resmigazete.gov.tr/arsiv/14574.pdf> adresinden erişilmiştir.
- Mehrens, W. A., & Lehmann, I. J. (1987). *Using standardized tests in education*. Longman/Addison Wesley Longman.
- Neale, M. (1962) Book review: Stanford-Binet Intelligence Scale. *Australian Journal of Education*, 6(3), 233-234. doi:10.1177/000494416200600320
- Öner, N. (2008). *Türkiye'de kullanılan psikolojik testler bir başvuru kaynağı*. İstanbul: Boğaziçi.
- Özgüven, İ. E. (2014). *Psikolojik testler*. Ankara: Nobel Akademik.
- Roid, G. H., & Barram, R. A. (2004). *Essentials of Stanford-Binet Intelligence Scales (SB5) assessment*. New Jersey: John Wiley & Sons.
- Savaşır, I., ve Şahin, N. (1995). *Wechsler çocuklar için zekâ ölçeği (WISC-R) el kitabı*. Ankara: Türk Psikologlar Derneği.
- Şencan, H. (2005). *Güvenilirlik ve geçerlilik*. Ankara: Seçkin.
- Tyler-Wood, T., Knezek, G., Christensen, R., Moreales, C., & Dunn-Rankin, P. (2014). Scaling three version of the Stanford-Binet Intelligence Test: examining ceiling effects and identifying giftedness. *Educational Research Journal*, 5(2), 42-51. doi:10.14303/er.2011.481
- Uluç, S., Öktem, F., Erden, G., Gençöz, T., & Sezgin, N. (2011). Wechsler Çocuklar için Zekâ Ölçeği-IV: Klinik bağlamda zekânın değerlendirilmesinde Türkiye için yeni bir dönem. *Türk Psikoloji Yazıları*, 14(28), 49.

Wechsler, D. (2004). *WISC-IV: Wechsler intelligence scale for children: Technical and interpretive manual*. Psychological Corporation.

Williams, P. E., Weiss, L. G., & Rolfhus, E. (2003). Psychometric properties (WISC-IV Technical Report No. 2). <http://harcourtassessment.com/hai/Images/pdf/wisciv/WISCIVTechReport2.pdf> adresinden erişilmiştir.

Yıldırım, A., & Şimşek, H. (2011). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin.

Yiğit, İ., Çelik, C., & Erden, G. (2017). Üstün Yetenekli Çocuklarda WÇZÖ-R ve WÇZÖ-IV Zekâ Puanlarının Karşılaştırılması. *Türk Psikoloji Dergisi*, 32(79), 80.

Yurtsever, Ş. (2013). *Eğitsel tanılama ve değerlendirme sürecinde görev alan rehberlik ve araştırma merkezi personelinin karşılaştığı sorunların belirlenmesi*. Yüksek Lisans Tezi, Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul.

### EK 1. Kontrol Listesindeki Geçerlik Kriterleri

1A	Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını açıkça belirtmiştir.
1B	Ölçeğin uygulanacağı evren ve ölçeğin değerlendirmeyi amaçladığı yapı veya yapılar açıkça tanımlanmıştır.
2	Belirli bir kullanım için verilen ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur
3	Belirli bir kullanım için genel veya olası yorumların geçerliği değerlendirilmediyse ya da böyle bir yorum mevcut kanıtlarla tutarsız ise, bu durum açıklığa kavuşturulmuş ve potansiyel kullanıcılar desteklenmeyen yorumlar konusunda uyarılmıştır.
4	Belirli bir kullanım için önerilen ölçek puanı yorumunun belli sonucu olacağı açıkça belirtildiğinde ya da ima edildiğinde, bu sonuç ilgili kanıtlarla birlikte sunulmuştur.
5A	Ölçeğin veya ölçek programının kendisinin bazı dolaylı yararlarla sonuçlanacağı gerekçesiyle bir ölçek kullanımı önerildiğinde, ölçek puanlarının yorumlanmasından elde edilen bilgilerin faydalanmasına ek olarak, öneri, dolaylı yararı öngören için gerekçesiyle beraber verilmiştir.

5B	Dolaylı yarar için mantıksal ya da teorik argümanlar ve ampirik kanıtlar sağlanmalıdır.
5C	Bilimsel literatürdeki çelişkili bulgulara, öngörülenin dışındaki önemli dolaylı sonuçları gösteren bulgulara da yer verilmiştir.
6	Ölçek performansının veya bunlardan alınan bir kararın, pratik yapmak ve rehberlikten etkilenmediği iddia edildiyse, ölçek performansının bunlara göre değişme eğilimi belgelenmiştir.
7	Geçerlik kanıtı elde edilen her bir katılımcı örneklemleri ilgili sosyo demografik ve gelişimsel temel özellikleri de içeren, pratik ve izin verilen ölçüde ayrıntılı olarak tanımlanmıştır.
8A	Bir geçerleme kısmen uzmanların, gözlemcilerin veya puanlayıcıların görüşlerine veya kararlarına dayanıyorsa, bu tür uzmanları seçme kararları veya derecelendirme prosedürleri tam olarak açıklanmıştır.
8B	Uzmanların nitelikleri ve deneyimleri açıklanmıştır.
8C	Prosedürler tanımı verilen eğitim ve talimatları içerip, katılımcıların kararlarına bağımsız olarak ulaşip ulaşmadıklarını göstermiş ve ulaşılan anlaşma seviyesini belirtmiştir.
8D	Katılımcılar birbirleriyle etkileşime girdiyse veya bilgi alışverişinde bulduysa, birbirlerini etkileyebilecekleri prosedürler ortaya konmuştur.
9A	Geçerlik kanıtı tek başına ya da diğer değişkenlerin verileriyle ölçek sonuçlarının istatistiksel analizlerini içerdiğinde, verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmıştır.
9B	Tipik ölçek koşullarından farklı olması muhtemel olan ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmiştir.
10A	Belirli bir kullanım için ölçek puanı yorumlama gerekçesi, kısmen ölçek içeriğinin uygunluğuna dayandırıldıysa, ölçek içeriğinin belirlenmesi ve üretilmesinde izlenen prosedürler, ölçülecek amaçlanan evrene ve ölçeğin yapısına atıfta bulunarak açıklanmış ve gerekçelendirilmiştir.
10B	İçeriğin tanımı, önem, sıklık veya kritiklik gibi kriterler içeriyorsa, bu kriterler de açıkça açıklanmış ve gerekçelendirilmiştir.



11A	Belirli bir kullanım için puan yorumlama gerekçesi, ölçek katılımcılarının psikolojik süreçleri veya bilişsel işlemleriyle ilgili öncüllere dayanıyorsa, bu öncülleri destekleyen teorik veya deneysel kanıtlar sağlanmıştır.
11B	Gözlemciler veya puanlayıcıların süreçle ilgili ifadeleri geçerlikte önemliyse benzer bilgiler burada da sunulmuştur.
12	Belirli bir kullanım için bir ölçek puanı yorumlama gerekçesi, ölçek maddeleri veya ölçeğin bölümleri arasındaki ilişkinin öncellerine dayanıyorsa, ölçeğin iç yapısı ile ilgili kanıtlar sağlanmıştır.
13A	Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmıştır.
13B	Birleşik puanların geliştirildiği durumlarda, bu puanı oluşturmak için temel ve gerekçeler verilmiştir.
14	Belirli maddelerdeki ya da madde gruplarındaki performansın yorumlanması önerildiyse bu yorumlamayı destekleyen mantıklı ve ilgili kanıtlar sağlanmıştır.
15	Geçerlik kanıtı; ölçek maddelerine verilen yanıtların diğer değişkenlere ilişkin verilerle birlikte ampirik analizlerini içeriyorsa, ek değişkenlerin seçilmesi için gerekçe sağlanmıştır.
16	Geçerlik, ölçek puanlarının bir veya daha fazla kriter değişkeni ile ilgili olduğuna dair kanıtlara dayanıyorsa, kriterlerin uygunluğu ve teknik kalitesi hakkında bilgi verilmiştir.
17	Belirli bir ölçek performansı seviyesinin yeterli veya yetersiz kriter performansını öngördüğü belirtilmişse, kriter performansının seviyeleri hakkında bilgi verilmiştir.
18	Ölçek sonuçları; bazı sonuçları veya kriterleri tahmin etmek için diğer değişkenlerle birlikte kullanıldıysa, yordayıcı kriter ilişkisinin istatistiksel modellerine dayanan analizler, ölçek puanlarıyla birlikte bu ilgili ek değişkenleri de içermiştir
19A	Ranjdada daralma veya azalma olduğu durumlardaki olduğu gibi istatistiksel düzeltmeler yapıldıysa, düzeltilmiş ve düzeltilmemiş katsayılar ile kullanılan özel işlem ve düzeltmelerde kullanılan tüm istatistikler rapor edilmiştir.
19B	Ölçme hatasının ölçeğe etkilerini ortadan kaldıran yapı-ölçüt ilişkisinin kestirimleri düzeltilmiş kestirimler olarak açıkça rapor edilmiştir.
20A	Bir meta-analiz, ölçek kriterleri ilişkisinin gücünün kanıtı olarak kullanıldığında, ölçek ve kriter değişkenleri, özetlenen çalışmalarla karşılaştırılabilir olmalıdır.

---

20B	İlgili arařtırmalar, ölçek uygulamasının diđer spesifik özelliklerinin ölçek-kriter ilişkisinin gücünü etkileyebileceğine dair güvenilir kanıtlar içeriyorsa, yerel durumdaki ve meta-analizdeki bu özellikler arasındaki ilişki rapor edilmiştir.
20C	Meta-analiz bulguların o duruma uygulanabilirliğini sınırlayabilecek herhangi bir önemli farklılık açıkça belirtilmiştir.
21A	Belirli bir kullanım için amaçlanan ölçek puanı yorumlamasını desteklemek için kullanılan bütün meta-analitik kanıtlar açıkça tanımlanmıştır.
21B	Kriter güvenilirliğinin olmadığı ve ranjda daralma gibi yapı düzeltilmesi için yapılan varsayımlar sunulup bu varsayımların karşılandığı gösterilmiştir.
22	Eđer ölçek bireyleri alternatif yöntemlere yönlendirmede kullanılıyorsa ve bu yöntemlerin sonuçları ortak bir kritere göre karşılaştırılabiliriyorsa, farklı sonuçların destekleyen kanıtlar sunulmalıdır.

---