
GAZİ ÜNİVERSİTESİ
Gazi
1926

REHBERLİK VE ARAŞTIRMA MERKEZLERİNDE MAKSİMUM PERFORMANSI ÖLÇMEDE KULLANILAN PSİKOLOJİK TESTLERİN ÖLÇME STANDARTLARI AÇISINDAN İNCELENMESİ

Buket Kemer

YÜKSEK LİSANS TEZİ

EĞİTİM BİLİMLERİ ANABİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

AĞUSTOS,2020

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU

YAZARIN

Adı : Buket

Soyadı : Kemer

Bölümü : Eğitimde Ölçme ve Değerlendirme

İmza :

Teslim tarihi :

TEZİN

Türkçe Adı : Rehberlik Ve Araştırma Merkezlerinde Maksimum Performansı Ölçmede Kullanılan Psikolojik Testlerin Ölçme Standartları Açısından İncelenmesi

İngilizce Adı : Examining The Psychological Scales Used To Measure Maximum Performance In Guidance And Research Centers According To Measurement Standards

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Buket Kemer

İmza:……..

JÜRİ ONAY SAYFASI

Buket Kemer tarafından hazırlanan "Rehberlik ve Araştırma Merkezlerinde Maksimum Performansı Ölçmede Kullanılan Psikolojik Testlerin Ölçme Standartları Açısından İncelenmesi" adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap Çakan

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi **……**

Başkan: Doç. Dr. Celal Deha Doğan

Eğitimde Ölçme ve Değerlendirme, Ankara Üniversitesi ……

Üye: Prof. Dr. İsmail Karakaya

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi ……

Tez Savunma Tarihi: 04/08/2020

Bu tezin Eğitim Bilimleri Anabilim Dalı'nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Selma YEL

Eğitim Bilimleri Enstitüsü Müdürü ……

TEŞEKKÜR

Tez sürecimde bana bilgi ve tecrübeleriyle rehberlik eden değerli hocam ve tez danışmanım Prof. Dr. Mehtap ÇAKAN ’a,

Araştırma ve tez savunma sürecimdeki önemli katkıları ve desteklerinden dolayı Arş. Gör. Tuba GÜNDÜZ’e,

Bugünlere gelmemde çok büyük emeği olan fedakâr anneme, babama ve abime sonsuz teşekkür ederim.

REHBERLİK VE ARAŞTIRMA MERKEZLERİNDE MAKSİMUM PERFORMANSI ÖLÇMEDE KULLANILAN PSİKOLOJİK TESTLERİN ÖLÇME STANDARTLARI AÇISINDAN İNCELENMESİ

(Yüksek Lisans Tezi)

Buket Kemer

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Ağustos 2020

ÖZ

Bu araştırmada Rehberlik ve Araştırma Merkezlerinde maksimum performansı ölçmede kullanılan psikolojik testlerin geçerlik, güvenirlik ve test uyarlama standartları açısından incelenmesi, eksikliklerin belirlenmesi ve böylece hem Milli Eğitim Bakanlığı'na hem de ilgili ölçekleri uygulayan uzmanlara yol gösterilmesi amaçlanmıştır. Rehberlik Araştırma Merkezlerinde sıklıkla uygulandığı belirlenen ve araştırma kapsamında incelemeye alınan ölçekler; Wechsler Zekâ Ölçeği Geliştirilmiş Formu (WISC-R), Stanford Binet Zekâ Ölçeği, Gesell Gelişim Ölçeği ve Leiter Uluslararası Performans Ölçeği'dir. Bu ölçekler APA, AERA ve NCME tarafından yayınlanan geçerlik- güvenirlik standartlarına ve Uluslararası Test Komisyonu (ITC) tarafından yayınlanan ölçek uyarlama standartlarına göre 3 puanlayıcı tarafından değerlendirilmiştir. Yapılan incelemeler sonucunda ölçeklerin ilgili standartları ne düzeyde karşıladığı ve geliştirilmesi önerilen özellikleri belirlenmiştir. Geçerlik açısından ele alındığında WISC-R'ın incelendiği 9 standartın 4'ü ile tamamen, 3'ü ile kısmen uyumlu olduğu ve 2'si ile hiç uyumlu olmadığı belirlenmiştir. Stanford- Binet'in ise incelendiği 7 standartın 2'si ile tamamen, 1'i ile kısmen uyumlu olduğu ve 4'ü ile hiç uyumlu olmadığı görülmüştür. Gesell'in 2 standartla tamamen, 2'ü ile kısmen uyumluyken,

2 standart ile hiç uyumlu olmadığı görülmüştür. Leiter'in ise incelendiği toplam 9 standartın 2'siyle tamamen, 4'üyle kısmen uyumlu olduğu ve 3'ü ile hiç uyumlu olmadığı görülmüştür. Güvenirlik açısından ele alındığında ise WISC-R'ın 8 standartla tamamen uyumlu olduğu ve 1 standartla hiç uyumlu olmadığı görülmüştür. Stanford- Binet'nin ise güvenirlik çalışması yapılmadığı için 27 standartın hiçbirini karşılamadığı görülmüştür. Gesell'in ise 3 standartla tamamen, 2'si ile kısmen uyumluyken, 7 standartla hiç uyumlu olmadığı görülmüştür. Leiter'in ise 8 standartın 2'si ile tamamen uyumlu, 2'si ile kısmen uyumlu olduğu; 4'ü ile ise hiç uyumlu olmadığı belirlenmiştir. Ölçek uyarlama standartlarına göre ise WISC-R'ın 5 standartla tamamen, 2'si ile kısmen uyumlu olduğu ve 4 standartla hiç uyumlu olmadığı görülmüştür. Stanford- Binet'in çevirisinin yapıldığı bilinmesine rağmen yazılı bir kaynakta bulunamadığı için uyarlama standartlarına göre incelenmesi yapılamamıştır. Gesell ve Leiter'de herhangi bir uyarlama ya da çeviri yapılmadığı için 2 ölçeğin de 12 standartla hiç uyumlu olmadığı görülmüştür. Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nun geçerlikte 5, güvenirlikte 1, ölçek uyarlamada ise 6 standarta göre geliştirilmesi önerilmiştir. Stanford-Binet Zekâ Ölçeği'nin ise geçerlikte 6, güvenirlikte 27 standarta göre geliştirilmesi önerilmiştir. Gesell Gelişim Ölçeği'nde ise geçerlikte 4, güvenirlikte 8, ölçek uyarlamada ise 12 standarta göre geliştirilmesi önerilmiştir. Diğer yandan Leiter Uluslararası Performans Ölçeği'nin geçerlikte 7, güvenirlikte 6, ölçek uyarlamada ise 12 standarta göre geliştirilmesi önerilmiştir.

Anahtar Kelimeler :Psikolojik Testler, Geçerlik Standartları, Güvenirlik Standartları, Test Uyarlama, APA, ITC

Sayfa Adedi : 87

Danışman :Prof. Dr. Mehtap ÇAKAN

EXAMINING THE PSYCHOLOGICAL SCALES USED TO MEASURE MAXIMUM PERFORMANCE IN GUIDANCE AND RESEARCH CENTERS ACCORDING TO MEASUREMENT STANDARDS

M.S Thesis

Buket Kemer

GAZI UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

August 2020

ABSTRACT

In this research, the 4 psychological scales which were determined to be used frequently in Guidance and Research Centers are examined in terms of validity, reliability and test adaptation standards and deficiencies are determined, as well. It is aimed to guide both the authorities and the the scale users. The selected scales were Wechsler Intelligence Scale Revised Form (WISC-R), Stanford Binet Intelligence Scale, Gesell Developmental Scale and Leiter International Performance Scale. These scales were examined seperately by 3 raters according to the validity, reliability and test adaptation standards. As a result of the examinations, the extent to which the scales meet the relevant standards and the features to be developed were determined. 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 of them. It has been observed that Stanford-Binet is fully compatible with 2 of the standards, partially compatible with 1 and not compatible with 4 of them. In Gesell, it was observed that it was fully compatible with 2 standards and partially with 2, but it was

not compatible with 2 standards at all. It was observed that Leiter is fully compatible with 2 of the standards, partially with 4 and not at all with 3. In terms of reliability, WISC-R was found to be fully compliant with 8 standards and not at all with 1 standard. It was observed that Stanford-Binet was not compatible with the 27 standards since there was no reliability study. In Gesell, it was found that it ise fully compliant with 3 standards; was partially compatible with 2, but not at all with 7 standards. Leiter is fully compatible with 2 standards and partially compatible with 2 and not compatible with 4 of them at all. According to test adaptation standards, WISC-R is fully compatible with 5 standards, partially compatible with 2 and not compatible with 4 standards. Although it is known that the translation of Stanford-Binet was made, it has not been examined according to the adaptation standards since it is not in a written source. Because of the fact that there isn't any study of adaptations or translations about Gesell and Leiter, it was observed that both scales were not compatible with 12 standards, at all. Wechsler Intelligence Scale Improved Form has been suggested to be improved according to 5 standards in validity, 1 standards in reliability, and 6 in scale adaptation. It has been suggested to improve the Stanford-Binet Intelligence Scale according to 6 standards in validity and 27 standards in reliability. Gesell Development Scale was suggested to be improved according to 4 standards in validity, 8 in reliability and 12 in scale adaptation. It has been suggested to improve the Leiter International Performance Scale according to 7 standards in validity, 6 in reliability, and 12 standards in scale adaptation.

Keywords **:** Psychological scales, Validity Standards, Reliability Standards, Test Adaptation, Standards, APA, ITC

Page Number : 87

Supervisor : Prof. Dr. Mehtap ÇAKAN

İÇİNDEKİLER

TABLOLAR LİSTESİ

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, problem ifadesi, araştırma soruları, araştırmanın amacı, önemi, sınırlılıkları ve varsayımları yer almaktadır.

Problem Durumu

Milli Eğitim Temel Kanunu (1973) eğitimin; vatandaşların istekleri, kabiliyetleri ve toplumun ihtiyaçlarına göre düzenlendiğini belirtmektedir. Bireyler farklı ilgi ve yeteneklerine göre rehberlik hizmetleri ile objektif ölçme ve değerlendirme metotlarından yararlanılarak çeşitli programlara veya okullara yönlendirilmektedirler (Millî Eğitim Temel Kanunu, 1973). Bu yasalarla paralel olarak Rehberlik Araştırma Merkezleri'nde öğrencileri tanıma, tarama, sorunlarını saptama gibi amaçlarla çeşitli zekâ, yetenek ve başarı ölçekleri; kişilik ölçekleri; ilgi envanterleri; gelişim tarama envanterleri; akademik ve meslekî benlik envanterleri; 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 öğrenciye en uygun eğitim şekli belirlenmektedir.

Ölçekler amacına ve kurallara uygun kullanıldığında öğrencilerin ve ailelerinin hayatına olumlu yönde etki edebilirken yanlış uygulamalar ise tam tersi bir etkiye sebep olabilir (Sattler & Hoge, 2006).

Günümüzde eğitimde ve psikolojik danışmanlıkta sıklıkla kullanılan psikolojik ölçekler hakkında pek çok tanım yapılmıştır. İlk olarak Cansever (1988) psikolojik ölçekleri

bireylerin davranışlarını gözlemlemek ve tanımlamak amacıyla uygulanan bir yöntem olarak tanımlamıştır. Murphy ve Davidshofer (2005)'e göre psikolojik ölçekler; davranış örneklemi, standart ölçme koşulları ve puanlama standartları olan ölçme araçlarıdır. Diğer yandan psikolojik ölçekleri, Öner (2008) standart koşullarda yetenek, beceri, tutum gibi alanlarda sorular yardımıyla bireyler hakkında bilgi edinilmesi olarak açıklarken; Özgüven (2014) ise belli bir özelliği ölçmek için bu özelliği temsil eden standart uyarıcılar grubu şeklinde tanımlamıştır.

Psikolojik ölçeklerin kullanım amacı seçme, sınıflama, hipotezleri kontrol etme, uygulanan yöntemin etkisini ölçme, yordama ve teşhistir. Yapılan seçmenin amacı ölçütlere en uygun kişiyi bulmakken, sınıflamanın amacı ise kişiyi en uygun kategoriye yerleştirerek teşhis ve yordamanın geçerliği arttırmaktır. Psikolojik ölçekler yardımıyla çalışmalarda uygulanan farklı teknik ya da tedavi yöntemin etkisini ölçmek amaçlanmaktadır. Bilimsel araştırmalarda ise hipotezlerin geçerliğini kontrol etmek ve değişkenler hakkında bilgi sahibi olmak için kullanılmaktadır (Özgüven, 2014).

Türkiye'de psikolojik ölçme çalışmaları ilk olarak İbrahim Alaattin Gövsa'nın 1915 yılında Binet-Simon Zekâ Ölçeği'nin Türkçe'ye çevrilmesiyle başlamıştır. Öğrenci başarılarının nesnel değerlendirmesine olan ilginin artmasıyla Türkiye'deki psikolojik ölçek kullanımı çoğunlukla başarı ölçekleri alanında olmuştur. Psikolojik ölçek alanında kurumsal olarak sayılabilecek ilk adım da bu gelişmelerle paralel olarak 1946 yılında Ankara Erkek öğretmen okulunda psikoteknik laboratuvarı kurulmasıdır. Bu laboratuvarda öğrencilerin yeteneklerini ölçebilmek amacıyla uygulamalı ölçekler kullanılmıştır (Özgüven, 2014). Millî Eğitim Bakanlığının kararnamesi doğrultusunda 1953 yılında kurulan Test Araştırma Bürosu ise; ölçek geliştirme, ölçme uzmanı yetiştirme, eğitim kurumlarına ölçme değerlendirme alanında danışmanlık hizmetleri götürme gibi alanlarda hizmet vermiştir. Günümüzdeki Rehberlik ve Araştırma Merkezleri'nin temeli ise Demirlibahçe İlkokulu'nda 1955 yılında açılan "Psikolojik Servis Merkezi" ile atılmıştır. Bu merkezde özel eğitime ihtiyaç duyan öğrencilere zekâ ve özel yetenek ölçekleri uygulanmıştır. Farklı ülkelerde yaygın olarak kullanılan ölçeklerin de uyarlamaları yapılmıştır. Bu kurum zaman içerisinde günümüzdeki Rehberlik ve Araştırma Merkezlerine dönüşmüştür (Öner, 2008).

Türkiye'de ölçme ve değerlendirme ile ilgilenen 3 temel kurum bulunmaktadır ve bunlar aşağıda detaylı olarak sunulmuştur.

a. Test ve Araştırma Bürosu

Yabancı eğitim uzmanlarının öğrenciler arası farklılıkların objektif olarak ölçülmesi için test ve araştırma konularına yönelik bir kurum açılmasını önermeleri üzerine 1953 yılında Test ve Araştırma Bürosu kurulmuştur. Bu kurum eğitimde ölçek konusu ve objektif ölçmenin tanıtımı ve yaygınlaştırılmasında önemli rol oynamıştır. Bu kurumda çeşitli başarı ölçekleri geliştirilmiş ve farklı yetenek ölçeklerinin tercümesi yapılmıştır. Okula giriş sınavlarında, devlet sınavlarında ve okulların rehberlik servislerinde bu ölçeklerden yararlanılmıştır. İlk defa açık uçlu soruların bulunduğu sınavlar yerine çoktan seçmeli sorularla sınav yapılmıştır. Bu kurum ölçek, sınav, ölçme araçları gibi konularda yayınladıkları belgelerle ve düzenledikleri seminerlerle kişilerin bu konularda bilgilenmelerini sağlamıştır (Tan, 1992).

b. Türk Psikologlar Derneği

İstanbul'da 1956 yılında "Psikoloji Derneği" kurulmasıyla temeli atılmıştır. Sonraki yıllarda artan psikolojik ihtiyaçlar sonucunda bu alanda hizmet veren kişilere mesleki olarak yardımcı olmak için 1976 yılında Ankara'da "Psikologlar Derneği" kurulmuştur. Daha fonksiyonel ve bütünsel bir dernek oluşturmak amacıyla 1996 yılında bu dernekler birleşerek Türk Psikologlar Derneği'ni oluşturmuşlardır. Derneğin temel amaçları arasında psikolojiyi Türkiye'de çağdaş düzeye ulaştırmak, psikologların haklarını korumak, psikolojinin Türkiye'de tanıtılmasını ve geliştirilmesini sağlamak, mesleğin standartlarını yükseltmek ve etik ilkelere aykırı davrananlar hakkında ilgili birimlerle iletişime geçmek bulunmaktadır (Türk Psikologlar Derneği, 2020). Bu kurumun psikolojik ölçeklere yönelik çalışmaları; konferanslarında Türkiye'deki psikolojik ölçek kullanımına ve bu süreçte karşılaşılan problemlere değinmesi, psikolojik ölçeklere yönelik yayınların ve bazı ölçeklerin dağıtımını yapıp bu ölçeklere yönelik eğitim vermesi olarak belirtilmektedir (Koç, 2008).

c. Rehberlik ve Araştırma Merkezleri (RAM)

Bu alanla ilişkili kurumların en işlevseli olarak değerlendirilebilecek olan kurumdur. Ankara Demirlibahçe İlkokulunda "Psikolojik Servis Merkezi" olarak açılan bu merkezler ilerleyen yıllarda farklı şehirlerde de açılmaya ve hizmet vermeye başlamıştır. Rehberlik ve Araştırma Merkezi'nin yönetmeliklerinin 1968 yılında yayınlanmasıyla bu merkezlerin çalışmaları netlik kazanmıştır (Özgüven, 2014).

Rehberlik ve Özel Eğitim olmak üzere iki bölümden oluşan bu merkezler günümüzde her ilde ve merkez ilçelerde hizmet vermektedir. Millî Eğitim Bakanlığı 2017 yılında yayınlanan

Rehberlik Hizmetleri Yönetmeliği'nde bu bölümlerin görevleri genel olarak aşağıdaki şekilde açıklamaktadır.

Rehberlik bölümünde;

i. Rehberlik hizmetlerinden yararlanmak üzere rehberlik ve araştırma merkezine başvuran ya da yönlendirilen bireylere randevu verilir. Danışan dosyası açılır ve gerekli psikolojik yardım hizmeti verilir.
ii. Çalışmalarda kullanılan psikolojik ölçme araçları, danışan dosyaları ve diğer kayıtların güvenliği ve gizliliği sağlanır.
iii. Hizmetin niteliğine ve bireyin ihtiyaçlarına göre psikolojik ölçme araçları, bilimsel standartlara ve etik ilkelere göre uygulanır.
iv. Bireye ihtiyacı doğrultusunda rehberlik hizmetleri verilir.
v. Hizmetlerde kullanılacak psikolojik ölçme araçları ile diğer araç ve tekniklerin tespiti, temini, geliştirilmesi ve rehberlik servislerine dağıtımı için yapılabilecek çalışmalar planlanır ve yürütülür.
vi. Bireylere, ailelere, öğretmenlere ve eğitim kurumlarına yönelik rehberlik hizmetlerine ilişkin yayınlar oluşturulur ve eğitim kurumlarına ulaştırılır.
vii. Eğitim kurumlarında görevli rehberlik öğretmenlerine bilgi ve beceri artırıcı eğitim etkinlikleri düzenlenir. Gerektiğinde ilgili kuruluşlardan uzman desteği alınır

Özel eğitim hizmetleri bölümünde ise;

i. Özel eğitim ihtiyacı olan bireylerin tanılanması amacıyla tarama faaliyetleri planlanır.
ii. Eğitsel değerlendirme ve tanılama hizmetlerinden yararlanmak üzere rehberlik ve araştırma merkezine başvuran ya da yönlendirilen bireylere; randevu verilir, dosya açılır ve gerekli hizmetler sunulur.
iii. Özel eğitim ihtiyacı olan öğrencilerin gelişimi öğrencinin devam ettiği eğitim kurumunun rehberlik servisi ile iş birliği yapılarak takip edilir.
iv. Hizmetlerde kullanılacak psikolojik ölçme araçları ile diğer araç ve tekniklerin tespiti, temini, geliştirilmesi ve rehberlik servislerine dağıtımı için yapılabilecek çalışmalar planlanır ve yürütülür.
v. Özel eğitim ihtiyacı olan bireylere, ailelere, öğretmenlere ve eğitim kurumlarına yönelik özel eğitim hizmetlerine ilişkin yayınlar oluşturulur ve eğitim kurumlarına ulaştırılır (Millî Eğitim Bakanlığı Rehberlik Hizmetleri Yönetmeliği, 2017).

Psikolojik ölçeklere dayalı anlamlı ve doğru kararlar verilebilmesi için bu araçların belli psikometrik özellikler sahip olması gerekmektedir. Bu psikometrik özelliklerin başında ölçeğin güvenilir puanlar üretmesi gelmektedir. Turgut (1977) güvenirliği ölçek sonuçlarının seçkisiz hatadan arınmış olma derecesi olarak tanımlarken; Tekin (1979) ise güvenilir ölçme aracının ilgilenilen özelliğin tekrarlı ölçümlerinde benzer sonuçları vermesi şeklinde ifade etmiştir. Crocker ve Algina (1986) ise güvenilirliği bireylerin sonuçlarının aynı ya da alternatif formların tekrarlı kullanımı sonucu tutarlı kalması olarak betimlemiştir.

Ölçme araçlarında güvenirliğin incelenmesi için üç özelliğe bakılmalıdır. Bunların ilki, ölçmek istenen özelliğin ne kadar hassas ölçüldüğüyle ilişkili olan duyarlılıktır. Madde sayısının artması duyarlılığı arttırmaktadır. Farklı zamanlardaki benzer şartlarda yapılan ölçümler arası tutarlılık ise, ölçeğin kararlılık özelliğiyle ilişkilidir. Ölçmedeki karasızlık ölçme aracından, ölçmeyi yapan kişiden ya da ölçülmek istenen özelliğin zaman içinde farklılaşmasından kaynaklanabilmektedir. Tutarlılık; madde puanları ile toplam puanın

anlamlı pozitif korelasyona sahip olması şeklinde açıklanmaktadır. Maddelerin ölçülmesi hedeflenen nitelik açısından benzer olması ölçeğin tutarlılığını arttırmaktadır (Seçer, 2015).

Güvenirliği etkileyen 3 temel faktörün ilki ölçme aracına ilişkin etmenlerdir. Madde sayısının fazla olması, anlaşılır uygulama yönergesi ile madde ifadeleri, maddelerin ölçülmek istenen özellik açısından benzeşik olması ve madde puanlamasının nesnel olması ölçeğin güvenirliğini arttırmaktadır. Ölçeğe katılan kişi ve gruplara bağlı değişkenler de güvenirlikte önemli rol oynamaktadır. Kişilerin ölçek uygulanırken olumsuz duygular hissetmesi ya da bu süreçteki sağlık problemleri, kişinin performansını etkileyeceği için hesaplanan puan gerçek puandan uzaklaşmaktadır. Ölçülecek nitelik bakımından normal dağılım gösteren gruplarda ise güvenirliğin artması beklenmektedir. Uygulama zamanının az olması kişilerin gerçek performanslarını yansıtmalarını engelleyebilmekte; fazla olması ise kopya ihtimali artmaktadır (Büyüköztürk, Kılıç, Akgün, Karadeniz, ve Demirel, 2014).

Ölçeklerin içermesi gereken bir başka psikometrik özellik ise geçerliktir. Cronbach (1990), geçerliği ölçek puanlarına dayanarak yapılan çıkarımların doğruluğuna yönelik araştırma olarak tanımlamaktadır. Turgut (1977) geçerliği ölçeğin kullanım amacına hizmet etme seviyesi olarak, Karasar (2012) ise ölçmek istenen özelliğin ölçülebilme derecesi şeklinde betimlemektedir.

Pek çok geçerlik türü olsa da üç temel geçerlik türü tanımlanmaktadır ve bunların ilki kapsam geçerliğidir. Kapsam geçerliğinde ölçeği uygulayan kişi katılımcıların sonuçlarına dayanarak daha geniş maddelere çıkarım yapmaktadır. Amaç maddelerin performans alanını veya belirli ilgi yapısını yeterli ölçüde temsil edip etmediğini belirlemektir.

Diğer bir geçerlik türü olan kritere dayalı geçerlikte katılımcıların sonuçları belli bir dış ölçütle karşılaştırılmaktadır. Bu geçerlik türünde kriterler ölçeği hazırlayan ya da bu ölçüm sonuçlarını kullanacak kişiler tarafından belirlenmektedir. Kritere dayalı geçerlik, ölçülmek istenen özelliğin o anki durumu gösteren eşzaman (mevcut durum) geçerliği ve gelecekteki durumunu yordaması amaçlanan yordama geçerliği olarak ikiye ayrılmaktadır.

Son geçerlik türü olan yapı geçerliğinde ise ölçeğin ölçmeyi amaçladığı yapıyı ne düzeyde ölçtüğüyle ilgilenilmektedir. Lord ve Novick (1968) bir yapının hem işevuruk hem de sözdizimsel olmak üzere iki düzeyde tanımlanması gerektiğini belirtmiştir. Bir yapı için operasyonel yani işevuruk tanımı yapılırken ölçülmek istenen konudaki teorik yapı ile bu yapıyla ilişkili olan gözlemlenebilir davranışlar arasında ilişkisel bir kural belirlenmesi

gerekmektedir. Böylece yapının önemi diğer değişkenlerle ilişkilendiren tanım ile açıkça belirtilmelidir (Crocker ve Algina, 1986). Psikolojik ölçeklerde yapı genelde soyut bir kavramla ilişkili olduğu için bunların tanımı zaman içinde ve farklı veri kaynaklarından yararlanarak yapılmalıdır. Bu üç temel geçerlik türünün yanında yakınsak geçerlik, ayırma geçerliği, çapraz geçerlik ve görünüm geçerliği gibi farklı geçerlik türleri de bulunmaktadır (Öner, 2008).

Ölçeğin geçerliğini etkileyen 4 temel faktörün başında ölçme sonuçlarının güvenirliği gelmektedir. Ölçekte geçerlikten bahsedebilmek için öncelikle güvenirliğin sağlanmış olması gerekmektedir. Geçerlikte etkili bir diğer faktör ölçekteki maddelerin ölçülmek istenen özelliği yansıtan madde örneklemini içermesidir. İlgili özelliği yansıtan madde örneklemi için ise ölçekte örneklemi yansıtan madde sayıları arttırılması gerekmektedir. Ölçeğin standart şekilde herkese uygulanmaması ise uygulama hatası olarak belirtilen ve geçerliği etkileyen diğer bir faktördür (Büyüköztürk vd., 2014).

Ölçeklerde bulunması gereken başka bir psikometrik özellik ise ölçeğin kullanışlılığıdır. Kullanışlılıkla belirtilen durumlar; ölçek hazırlama, uygulama ve puanlama aşamalarındaki kolaylık ve ekonomikliktir. Ölçeği uygulayacak kişi, ölçme araçlarının uygulamadaki sınırlıklarına göre ihtiyacına en uygun ölçeği seçmelidir. Bu sınırlıklardan bazıları: ölçeğin uygulama zamanı; puanlama şekli; uygulamanın mali giderleri; ölçme aracının etik sorunlardan arınık olma seviyesi; ölçeği uygulamada yasal izin gerekliliği; ölçeği hazırlama, uygulama ve puanlama aşamalarındaki uzman kullanımı şeklinde belirtilmektedir (Tekindal, 2008). Uygulanacak ölçek seçiminde yüksek geçerlik ve güvenirliğin öncelikli olması gerekmektedir (Özçelik, 2013).

Dünya genelinde son 25 yılda ölçek çevirisi ve uyarlamaları çok gelişmiştir. Bunun sebepleri arasında dünya çapında yapılan TIMSS ya da PISA gibi öğrencileri değerlendirme sınavlarının artması, kültürlerarası psikolojiye yönelik çalışmaların artması, ülkeler arası sertifika programlarının yaygınlaşması ve çok dilli ülkelerde öğrencilere sınav için farklı dil seçeneklerinin olması gösterilebilir (International Test Commission, 2017).

Ölçek uyarlama sürecinde gözlemlenen farklı türlerde hata ve geçersizlik kaynakları; kültürel ve dil farklılıkları, çevirmenle ilgili sorunlar, teknik sorunlar ve sonuçların yorumlanması olarak sıralanmaktadır. Bu hata ve geçersizlik kaynaklarından herhangi biri

varsa gruplar eşdeğer ölçeği almayacakları için sonuçlarının yorumlarının doğruluğu sorgulanmaktadır (Hambleton, Merenda ve Spielberger, 2005).

Hambleton ve Patsula (1999) Uluslararası Test Komisyonu'nun yayınladığı standartlardan yola çıkarak ölçek uyarlama adımları yayınlamışlardır. Bu adımlar sırasıyla açıklanmıştır:

- Madde uyarlanma yapılmak istenen popülasyonlarda yapısal denkliğin olup olmadığını sorgulamak ilk adımdır. Yapıların denkliğini sorgulamayı iki kültüre de hakim olan araştırmacılar yapmalıdır.
- Uyarlama yapılacak kültürdeki insanlar gözlemlenerek ve onlara soru sorularak yapıların denkliğine bakılmalıdır. Yapılar denk değilse yapının iki popülasyonda da aynı olacak şeklide değiştirilmesi gerekmektedir.
- Sonraki adımda mevcut ölçeğin uyarlamasının yapılacağına ya da yeni ölçek geliştirileceğine karar verilmektedir. Bu süreçte uyarlama ölçeğin olumlu ve olumsuz yanları gözden geçirilmelidir. Ölçeğin orijinal dilinin ve kültürünün uyarlanacağı dil ve kültüre benzer olmasına da dikkat edilmelidir.
- Sonraki aşama ise uyarlama için nitelikli tercümanlar seçmektir. Tercümanlara karar verilirken iki dilde akıcı, iki kültürü de tanıyan ve ölçülen yapı hakkında bilgili olan kişiler tercih edilmelidir.
- Sonraki basamak ise ölçeğin tercümesi ve uyarlamasıdır. Ölçeği uyarlamanın yapılacağı dile çevrilirken sistematik bir yöntem izlenmelidir. Tercümede oluşan hataları fark edebilmek için ölçeğin orijinal diline geri çeviri yapılmalıdır.
- Ölçeğin uyarlanmış halinin incelenmesi ve gerekli düzeltmelerin yapılması gerekmektedir. Ölçek çevrildikten sonra farklı bir çevirmen grubu uyarlamada iki dil arasında anlam farklılıkları olup olmadığı incelemelidir.
- Ölçeğin uyarlanmış halinin küçük ölçekli bir denemesi yapılmalıdır. Bu aşama uyarlamanın sadece uzmanlar tarafından gözden geçirmesine göre daha çok bilgi vereceği için daha gereklidir. Bu denemeden sonra daha çeşitli bilgi elde etmek amacıyla ölçek, daha büyük ölçekli ve evreni yeterince temsil eden bir örneklem üzerinde de uygulanmalıdır.
- Ölçeğin orijinal ve uyarlanmış versiyonları arası ilişkiyi bulmak için istatistiksel desen seçmek bu süreçteki bir sonraki adımdır. Ölçek puanlarını ortak bir ölçekte birleştirmek için desen kullanılmalıdır.

- Kültürlerarası karşılaştırmalar söz konusu olduğunda ölçeğin uyarlanmış hali ile orijinal hali arasında dilsel eşitlik sağlanmalıdır. Ölçeğin orijinal hali ölçeğin ilk hedeflendiği popülasyonun büyük bir örnekleminde uygulanmalı ve madde yanlılığı çalışması ile istatiksel analiz yapılmalıdır.
- Ölçeğin uygulanması ve geçerlik çalışmasının yapılması da bir sonraki basamaktır.
- Süreci raporlaştırmak ve ölçeğin uyarlanmış halini kullanacak kişiler için bir el kitabı hazırlamak gerekmektedir.
- Her ne kadar el kitabındaki bilgiler yol gösterici olsa da mümkün olduğunca bu ölçeği uygulayacak kişilere eğitim de verilmelidir.
- Son adım ise ölçeklerin uyarlanmış halleri sürekli izlenmesi ve güncelliğinin sağlanmasıdır. Yetenek, kişilik ve zekâ ölçekleri gibi ölçekler birden çok kullanım için tasarlanmıştır. Bu sebeple araştırmacılar sürekli olası hatalara karşı dikkatli olmalı ve ölçek puanlarının güvenirlilik ve geçerliğini sürekli kontrol etmelidir

Özellikle son yıllarda ülke çapında pek çok psikolojik ölçek geliştirilmiş olmasına rağmen yurt dışında geliştirilen ve uyarlaması yapılan ölçekler de yadsınamaz çoğunluktadır. Uyarlama ölçeklerin kullanımıyla birlikte süreçte çok zaman alan madde yazma adımına gerek kalmaması bu olumlu özelliklerinin başında gelmektedir; fakat uyarlama ölçek kullanımının bazı olumsuz tarafları da vardır. Bunların ilki, farklı kültürde geliştirilen ölçeğin uyarlanacak kültüre uygun olmayan maddelerinin bulunabilmesidir. Örneğin; Uzak Doğu kültüründe geliştirilmiş olan bir ölçekte "Çubukla yemek yerken zorlanır mısınız?" maddesinin Türk kültürüne uyarlanmadan kullanılması uygun olmayacaktır. Ölçek uyarlamalarındaki diğer bir sorun ise orijinal ölçekteki faktör yapısı ve kesme noktalarının uyarlanacak olan dil veya kültürde farklı olmasıdır (Seçer, 2015). Diğer bir olumsuz özellik ise çevirilerde karşılaşılan sorunlardır. Orijinal metin hem psikolojik hem de dilsel özelliklerini koruyarak uyarlanmak istenen dile çevrilebiliyorsa anlamlı kabul edilmektedir. Dilsel yönler metnin anlam, anlaşılabilirlik gibi konularına odaklanırken; psikolojik özellikler ise anlatılmak istenen kelimenin kültürel anlamına odaklanmaktadır. Bu iki özelliği korumak ise her ölçek için mümkün olmadığı için bazı ölçeklerin çevirisi yapılamamaktadır (Hambleton vd., 2005).

Bu genel sorunların yanı sıra, Türkiye'de kullanılan uyarlanmış ölçeklerin çoğunda çeviriler uyarlama olarak kabul edilmiştir. Ölçekler Türkiye'de uygulanmamış olup psikometrik özellikleri yorumlamada orijinal el kitabındaki normlar kullanılmıştır (Gülgöz, 1994).

Uyarlama ölçeklerdeki bu olumsuzlukları önlemek için maddelerin ayrıntılı şekilde incelenmesi, uyarlanacak dille kültüre uygun düzenlemelerin yapılması ve ölçeğin uyarlanacağı bireylerin normlarına göre standardize edilmesi gerekmektedir (Öner, 2008).

Çüm ve Koç (2013) Türkiye'deki dergilerde yayımlanan ölçek geliştirme ve uyarlama çalışmalarında belli başlı eksiklikler olduğunu ve bunlardan ilkinin, ölçülecek değişkenin soyut kavramsal tanımlar şeklinde ele alınması sonucunda gözlenebilir ve ölçülebilir bir işevuruk tanımının yapılmamasının geldiğini belirtmiştir. Yapılacak bütün işlem ve süreçler işevuruk tanıma göre şekillenecek olduğu için ölçek geliştirme için önem arz etmektedir. Ölçek geliştirme çalışmalarında gözlenen bir diğer eksik ise ön deneme uygulamasının yapılmamasıdır. Ön deneme uygulaması anlaşılmayan maddelerin belirlenmesi, ölçeğin ortalama yanıtlama süresi gibi açılardan önem taşımaktadır. Araştırmaya dahil edilen çalışmaların yaklaşık üçte birinde madde ve ölçek analizlerinin rapor edilmemesi de karşılaşılan bir diğer eksiklik olan göze çarpmaktadır. Çalışmalar geçerlik açısından incelendiğinde çoğunda sadece uzman görüşüne başvurulduğu görülmektedir.

Dirlik (2013), Türkiye'deki eğitim kurumlarında sıklıkla kullanıldığı belirlenen Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı, Akademik Benlik Kavramı Ölçeği, Temel Kabiliyetler Testi ve Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nu APA tarafından 1999 yılında yayınlanan psikolojik standartlara göre oluşturulan kontrol listesine göre incelemiştir. Araştırma kapsamında oluşturulan kontrol listesi toplam 10 geçerlik standartı, 9 güvenirlik standartı, 12 ölçek geliştirme- revizyon standartı ve 15 uyarlama standartı içermektedir. Bu incelemeye göre; Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı 1 geçerlik, 4 güvenirlik ve 6 ölçek geliştirme-revizyon standardını tamamen karşıladığı raporlanmıştır. Akademik Benlik Kavramı Ölçeği ise 3 geçerlik ve 6 ölçek geliştirme-revizyon standardını tamamen karşıladığı belirtilmiştir. Temel Kabiliyetler Testi ve Wechsler Zekâ Ölçeği Geliştirilmiş Formu yurtdışında geliştirildiği için sadece test uyarlama standartlarına göre incelenmiştir. Temel Kabiliyetler Testi'nin 8 uyarlama standardını, WISC-R'ın ise 14 standartı tamamen karşıladığı raporlanmıştır.

Türkiye'de kullanılan ölçeklerin güncelliğine yönelik olarak alanyazın tarandığında genelde uyarlanan ölçeklerin güncel olmadığı görülmüştür. Yurtsever (2013) Rehberlik ve Araştırma Merkezlerinde çalışan psikolojik danışmanların sorunlarını incelediği çalışmasında katılımcıların yaklaşık %76'sı en az günde bir kere ölçme araçlarındaki içeriklerin güncel olmamasıyla ilgili bir sorunla karşılaştıklarını bildirmiştir. Pişkin (2006) ise Türkiye'deki

ölçme araçlarının çoğunun başka ülkelerde geliştirilen ölçeklerin eksik uyarlamaları olduğunu, ölçeklerin faktör analizi ve ülke normları içermediğini raporlanmıştır.

Bu araştırmada Türkiye'de sıklıkla kullanılan 4 ölçek geçerlik, güvenirlik ve uyarlama standartlarına uygunlukları bakımından ele alınmıştır. Aşağıda bu ölçeklere ilişkin detaylı tanıtım yapılmıştır.

Wechsler Çocuklar İçin Zekâ Ölçeği Geliştirilmiş Formu (WISC-R)

Wechsler Zekâ Ölçeği Geliştirilmiş Formu, David Wechsler tarafından 1949 yılında oluşturulan Wechsler Zekâ Ölçeği'nin 1974 yılında bazı maddeleri ekleyip çıkararak ya da değiştirilerek geliştirilmiş haldir. Ölçek 6 ve 16 yaş arası çocuklara bireysel olarak uygulanmaktadır. Ölçek ilk yayınlandığı zamanki koşullarda zekâ ölçeklerine pek çok yenilik getirmiştir. Bu yenilikler; alt ölçek kullanımı, soruların tek tek puanlanması, alt ölçekler sayesinde çocuğa ait pek çok farklı alanda puan elde edilmesi ve zekâ ölçeklerinde performans alanını kullanmasıdır. Performansa dayalı becerilerin ölçümlerinin yapılması zekâ ölçümünde bireylerin sözel beceri ile eğitsel farklılıkların etkilerini azaltmasını sağlamıştır. Uluslararası alanda Wechsler Zekâ Ölçeği'nin son revizyonu (WISC-V) 2014 yılında yayınlanmıştır. WISC-V; Sözel Anlama Dizini, Görsel Mekansal İndeks, Akışkan Akıl Yürütme İndeksi, Çalışma Belleği Dizini ve İşleme Hızı Dizini olarak temel 5 indeks puanı ölçümü yapmaktadır.

WISC-R, sözel beceriler ve performansa dayalı beceriler olarak 2 kısma ayrılır.

Sözel becerileri ölçmekte kullanılan alttestler;

- Genel Bilgi,
- Benzerlikler,
- Aritmetik,
- Sözcük Dağarcığı,
- Yargılama ve
- Sayı Dizisidir.

Performansa dayalı becerileri ölçmekte kullanılan alttestler ise;

- Resim Tamamlama,
- Resim Düzenleme,

- Küplerle Desen,
- Parça Birleştirme,
- Şifre
- Labirentlerdir.

Bir altölçekten alınan ham puanı için o alt ölçekte verilen bütün puanları toplanır. Sonra çocukların yaşlarına göre hazırlanmış tablolardan yararlanılarak ham puan standart puana çevrilir. Sözel ve sayısal alt ölçek puanları toplanarak çocuğun ölçeğe ait toplam ölçek puanı hesaplanır (Kaufman, 1979).

Wechsler Zekâ Ölçeği Geliştirilmiş Formu'nun Türkiye'ye uyarlaması 1995 yılında Şavaşır ve Şahin tarafından yapılmıştır. Ölçeğin yapı geçerliği katsayıları sözel beceriler bölümününde 0,34; performansa dayalı beceriler bölümünde 0,33 ile 0,68 arasında ve tüm ölçekte 0,38 ile 0,74 değerleri arasında raporlanmıştır. Spearman Brown güvenirlik katsayısı sözel beceriler için 0,94; performansa dayalı beceriler için ise 0,96 olarak raporlanmıştır. Standart hata ise 1,15 ile 1,70 arasında raporlanmıştır (Savaşır ve Şahin, 1995).

Testin standardizasyonu ABD'nin farklı bölgelerinden gelen yaşları 6 ila 16 arasında değişen 2200 katılımcı üzerinde yapılmıştır. Her yaş seviyesi için hesaplanan iki-yarı güvenirlik katsayıları sözel beceriler için 0,94; performansa dayalı beceriler için 0,90 ve tüm ölçek için 0,96 olarak raporlanmıştır. Alt testler arası korelasyonlar ise 0,38 ile 0,74 arasında raporlanmıştır. Stanford-Binet Zekâ Ölçeği L-M formları ile WISC-R 118 katılımcı üzerinde uygulandığında alttestler arası korelasyon katsayıları 0,51 ile 0,77 arasında raporlanmıştır (Wechsler, 1974).

Stanford – Binet Zekâ Ölçeği

Stanford ve Binet bu ölçeği Paris'teki çocukların başarısızlık nedenlerini araştırmak için 1905 yılında geliştirmiştir. Ölçeğe 1908 yılında yaş çizelgesi eklenmiş, 1911 yılında ise yaş aralığı genişletilmiştir. Termann tarafından 1917 yılında İngilizceye çevirip uyarlanmıştır. Sonrasında 1937, 1960, 1972, 1986 ve 2003 yıllarında ölçeğe yeni formlar ekleyip düzeltmeler yaparak yeni versiyonları geliştirmişlerdir. Bireysel bir zekâ ölçeği olan Stanford – Binet 2 ile 18 yaş arası bireylere uygulanmaktadır. Uluslararası düzeyde Stanford-Binet Zekâ Ölçeği'nin son hali Stanford-Binet- V 2003 yılında yayınlanmış olup son hali

bilgi, nicel akıl yürütme, görsel -uzamsal işleme, çalışma belleği ve akıl yürütme olarak 5 faktörden oluşmaktadır.

Ölçeğin 1972 yılındaki versiyonu 1937'de yayınlanandan farkı L ve M adlı formlardaki en başarılı maddeleri alıp birleştirmesidir. Aynı zamanda önceden çoğu yaş seviyesine özgü maddeler varken bu versiyonda her yaş seviyesine özgü maddeler oluşturulmuştur. Hala önceki haliyle benzer maddeler içermesine rağmen uygun olmayan maddeler çıkarılmıştır (Becker, 2003). Türkçe'ye Şemin tarafından 1972 yılında çevrilmiş olan ölçeğin RAM'larda 1978'deki ikinci versiyonu kullanılmaktadır.

Sayı kavramı, problem durumları, kağıt kesmek, hikaye hatırlamak, cümle hatırlamak, hayvanın ismini bilmek, saçma cümleler I, saçma cümleler II, soyut kelimeler, zıtlardaki benzeyişler, atasözleri, doğrultu, esas farklar, 4 nokta hatırlamak, ezbere boncuk dizmek, resimlere cevap vermek, para saymak, para bozdurmak haftanın günlerini saymak, şekil hatırlamak, karışık cümleler, tersine saymak, okumak ve anlatmak, nedenleri bulmak, muhakeme, şifreler, bir parçadaki fikrin tekrarı, eksik resimler, intikal, iç içe kutular ve küpleri saymak Stanford Binet Zekâ Ölçeği'nde bulunan alt ölçeklerdir. Ölçekteki alttestler hem sözel becerilere hem de performansa dayalı becerilere yöneliktir. Yaş arttıkça sözel maddelerin sayısı artarken performans maddeleri azalır. Bu alt alanlarda elde edilen puanlar toplanarak genel zekâ faktörü g'yi oluşturur. Uygulama yapılacak kişinin takvim yaşına göre uygun alttest seçilerek kişiye uygulanır. Ölçek puanlanırken tamamen başarılı olunan alttest taban zekâ, tamamen başarısız olunan alttest ise tavan zekâ yaşını gösterir. Diğer alttestlerdeki zihin yaşları da toplanarak öğrencinin toplam zihin yaşı elde edilir (Şemin, 1972).

Ölçüt geçerliğinde ölçek puanları ile sözel dersler arası korelasyon 0,60 ile 0,70 arasında sayısal derslerle arasındaki korelasyon ise 0,40 olarak raporlanmıştır. Kapsam geçerliğinde yaş ile kelime alttestleri arasındaki korelasyonlar 0,70'den büyük raporlanmıştır. Ölçeğin test- tekrar test güvenirliği 0,90; alttest güvenirlikleri ise 0,70 ile 0,80 arasında raporlanmıştır. Paralel form güvenirlik çalışmasında 1937 versiyonundaki L ve M formları arası korelasyon; yaşı küçük zekâsı yüksek kişilerde düşük, yaşı büyük zekâsı düşük kişiler arasında yüksek olarak belirlenmiştir (akt. Öner, 2008).

Leiter Uluslararası Performans Ölçeği

Leiter Uluslararası Performans Ölçeği, Rusell Leiter tarafından 1936 yılında geliştirilmiş olan sözel olmayan bir zekâ ölçeğidir. O zamanlarda kullanılan ve sözel odaklı olan zekâ ölçeklerine alternatif olarak geliştirilmiştir. Sözel uygulama içermediği için ölçeği uygulayan kişiler talimatları katılımcılara en basit halde, hareketleriyle anlatırlar. Ölçeğin uygulandığı yaş aralığı 5 ile 16 olarak belirlenmiş olup her yaş grubuna yönelik farklı alttestler bulunmaktadır. Katılımcının verilen blokları örneğe uygun olarak yerleştirmesi beklenmektedir. Puanlayıcı uygulamanın doğruluğuna göre puanlama yapar. Bu puanlardan faydalanarak çocuğun zekâ katsayısı, taban ve tavan yaşları elde edilir. Ölçeğin kültürler arası uygulanabilirliğini sağlamak amacıyla ölçekteki kültürlere özgü resimler değiştirilmiştir. Zamanındaki diğer ölçeklere göre uygulamasının basit olması, objektif puanlama yapılabilmesi, sözel etkileşimi kaldırarak çevresel etkileri azaltması, yenilikçi materyallerin kullanılması, zaman sınırının olmaması ve yapılması istenen görevlerin basit olması gibi avantajları bulunmaktadır (Leiter, 1936). Uluslararası alanda Leiter Uluslararası Performans Ölçeği'nin üçüncü revizyonu olan Leiter-3 kullanılmaktadır. Yaş aralığı 3 ile 75 arasında genişletilen ölçekte küpün yanı sıra çerçeve ve köpükle yapılan oyun benzeri uygulamalar yapılmaktadır (Roid & Koch, 2017).

Türkiye'de Leiter Uluslararası Ölçeği'inin 1936 yılındaki orijinal versiyonu kullanılmaktadır. Türkiye'deki alanyazında ölçeğe ilişkin herhangi bir geçerlik ya da güvenirlik çalışması bulunmamaktadır. Yurt dışındaki araştırmalarda ölçeğin yapı geçerliği Vineland Ölçeği'yle yakınsak geçerliğine göre anlamlı bulunmuştur. Test-tekrar test güvenirliği 0,70 ile 0,80 arasında ve Cronbach alfa katsayısı ise 0,75 ile 0,83 arasında raporlanmıştır (Kashani & Faramarzi, 2017).

Gesell Gelişim Ölçeği

Gesell Gelişim Ölçeği; Arnold Gesell'in iş arkadaşları olan Frances Jig ve Louise Bate Ames tarafından 1939 yılında "School Readiness: Behavior Tests Used At The Gesell Institute" adlı kaynakta yayınlanmıştır. Amacı çocukların gelişim düzeylerini belirlemek olan ölçek 4 ila 10 yaş arası çocuklara uygulanabilmektedir. Uluslararası alanda ölçeğin 2011 yılında revize edilmiş hali olan GDO-R kullanılmaktadır. GDO-R'da; çocuğun bilişsel, dil, motor ve sosyal-duygusal tepkilerini değerlendirmek için gelişimsel, harfler/ sayılar, dil / anlama,

görsel / mekansal ve sosyal / duygusal olmak üzere 5 alttestt uygulanmaktadır (Gesell Institute, 2020)

Gesell Gelişim Ölçeği çocuğun yaşı, doğum tarihi, varsa kardeşlerinin isimleri, yaşları ve ebeveyn mesleklerinin sorulduğu giriş görüşmesi ile başlar. Sonraki aşamada ise sırayla harfleri yazması, 1'den 20'ye kadar olan sayıları yazması, verilen geometrik şekillerin kopyalarını yapması ve verilen eksik adamı tamamlaması istenen kalem - kağıt testleri uygulanır. Çocuğun vücudunun sağında ve solunda bulunan uzuvlarını söyleyip sözel olarak söylenen uzuvlarını da göstermesi istenen Sağ-sol alttestinde ise bir sonraki aşamadır. Sonraki adımda ise belli formlar verilerek bu formları kağıdın üzerine yerleştirmesi söylenen form alttesti uygulanır. Hayvan sayma adlı alttestte ise çocuktan 60 saniyede sayabildiği kadar çok hayvan ismi söylemesi istenir. Çocuğa okulda ve evde neler yapmayı sevdiği sorularak uygulama sonlandırılır.

Ölçek yorumlanmasında her 6 aylık fark için alttestlerde ayrı kriterler belirtilmiştir. Çocuğun kriterler içindeki performansı "normal gelişim" gösterdiği, kriterlerin altında performans göstermesi "gelişiminde gerilik" olduğu, kriterlerin üstünde performans göstermesi ise "gelişiminin yaşıtlarına göre ileri" olduğu şeklinde yorumlanmaktadır (Ilg & Ames, 1972).

Türkiye'de ise Gesell Gelişimsel Ölçeği adı altında sadece kopya formları alttesti uygulanmaktadır. Kopya formları alttestinde öğrenciye üstlerinde farklı şekiller olan kartlar gösterilerek bunları A4 kâğıdına çizmesi istenilir. Düzeltme yapılması istenmediği için uygulama esnasında çocuğa silgi verilmez. Kartlardaki şekiller kolaydan zora doğru sıralanmıştır ve ölçek puanlamasında çocuğun yaşına dikkat edilmektedir. Şeklin bozukluğu, döndürülmesi, birleştirmesi gibi durumlarda hata puanı verilir ve ölçekteki tüm hata puanları toplanarak toplam hata puanı oluşturulur.

Ölçeğin ilgili versiyonu için uluslararası alanyazında geçerlik ve güvenirlik çalışması bulunmamaktadır. Türkiye'de ise yaşları 5 ila 6 olan çocuklar üzerinde uygulanan güvenirlik çalışmasında iç tutarlılık katsayısı 0,44 ve test- tekrar test güvenirlik katsayısı ise 0,46 ile 0,80 arasında raporlanmıştır (Evirgen, Kayhan ve Erden, 2015).

Alanyazın incelendiğinde RAM'larda kullanılan ölçeklerin çoğunluğunu uyarlama ölçeklerin oluşturduğu ve ilgili ölçeklerin uluslararası alandaki revizyonlara göre güncellenmediği görülmektedir. Reise, Waller, ve Comrey (2000) ölçeklerin revize edilmesiyle ölçeklerdeki psikometrik özelliklerin iyileşeceğini özellikle iç tutarlılık

katsayısının artacağını, yapıların daha iyi temsil edileceğini belirtmiştir. Bu bilgilere dayanarak öğrencilerin eğitim hayatını önemli ölçüde etkileyebilecek olan psikolojik ölçeklerin sonuçlarının istenilen düzeyde ve doğru bilgiler vermesi için ölçme araçlarının belli evrensel özellikleri taşıması gerekliliği görülmektedir. Bu çalışmada alanyazından hareketle RAM'larda sıklıkla kullanılan 4 ölçeğin geçerlik, güvenirlik ve ölçek uyarlamaya ilişkin standartları ne düzeyde taşıdıkları açısından incelenmesi gerektiği ihtiyacından hareketle planlanmıştır.

Problem İfadesi

Rehberlik ve Araştırma Merkezleri'nde kullanılan maksimum performansı ölçmede kullanılan testlerin (Wechsler Çocuklar için Zekâ Ölçeği Gözden Geçirilmiş Hali , Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği) American Psychological Association (APA), American Educational Research Association (AERA) ve National Council on Measurement in Education (NCME) tarafından 2014 yılında ortak olarak yayınlanan psikolojik ölçek standartlarına ve 2017 yılında Uluslararası Test Komisyonu (ITC) tarafından yayınlanan "Guidelines for Translating and Adapting Tests" yönergesindeki ölçek uyarlama adımlarına göre incelenmesi.

Araştırma Soruları

Araştırmada aşağıdaki sorulara cevap aranmıştır:

1) Wechsler Çocuklar için Zekâ Ölçeği'nin Gözden Geçirilmiş Hali, Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nin APA, AERA ve NCME tarafından 2014 yılında yayınlanan "Standards for Educational and Psychological Testing" kaynağındaki geçerlik ve güvenirlik standartları ile Uluslararası Test Komisyonu (ITC) tarafından 2017 yılında yayınlanan "Guidelines for Translating and Adapting Tests" adlı kaynaktaki psikolojik ölçek uyarlama standartlarına dayalı olarak oluşturulan kontrol listesindeki;
    a) Geçerlik,
    b) Güvenirlik,
    c) Psikolojik ölçek uyarlama kriterlerini karşılama düzeyi nedir?

2) Wechsler Çocuklar için Zekâ Ölçeği'nin gözden geçirilmiş hali (WISC-R), Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nin bu kriterlere göre geliştirilmesi gereken özellikleri nelerdir?

Araştırmanın Amacı

Bu çalışmanın amacı Rehberlik ve Araştırma Merkezleri'nde maksimum performansı ölçmede kullanılan psikolojik testleri belirleyip bu ölçekleri APA, AERA ve NCME tarafından 2014 yılında ortak yayınlanan ölçek standartlarına göre geçerlik ve güvenirlik açısından incelerken 2017 yılında Uluslararası Test Komisyonu tarafından yayınlanan "Guidelines for Translating and Adapting Tests" belgesindeki psikolojik ölçek uyarlama adımlarına göre incelemek, bu standartlardan hareketle eksik yönlerini belirlemek ve hem Milli Eğitim Bakanlığı'na hem de ilgili ölçekleri uygulayan psikolojik danışmanlara yol göstermektir.

Araştırmanın Önemi

Türkiye'de psikolojik ölçekleri farklı amaçlar doğrultusunda uygulayan pek çok farklı kurum olmasına karşın ölçeklerin korunup geliştirilmesi, dağıtımı ve yeni ölçek gelişiminin desteklenmesi konusunda etkin rol alan belli bir kurum bulunmamaktadır. Bu sebeple kullanılan ölçekler için izleme ve geliştirme çalışmaları bireysel çabalarla sınırlı kalmakta olup sistematik şekilde yapılamamaktadır (Dirlik, 2013).

Bu araştırma, APA, AERA ve NCME tarafından 2014 yılında ortak yayınlanan psikolojik ölçek standartlarına ve 2017 yılında Uluslararası Test Komisyonu tarafından yayınlanan "Guidelines for Translating and Adapting Tests" yönergesindeki psikolojik ölçek uyarlama adımlarına göre oluşturulan kontrol listesi aracılığıyla Rehberlik ve Araştırma Merkezleri'nde maksimum performansı ölçmede kullanılan psikolojik testlerin incelenmesi hem Dirlik'in (2013) APA tarafından 1999 yılında yayınlanan standartlara göre oluşturduğu kontrol listesinin güncellenmesi hem de kullanılan psikolojik ölçekler için geçerlik, güvenirlik, ölçek uyarlama çalışmalarına yönelik standartları ortaya koyan toplu b kaynak olarak kullanılabilmesi açısından önemlidir.

Psikolojik ölçekleri geliştiren, uygulayan ya da uyarlayan kişiler bu kontrol listesi yardımıyla ölçekleri inceleyip listedeki kriterlere göre eksik kısımlarını belirleyebileceklerdir. Bu bulgular ölçeklerin daha doğru yorumlanmasına katkı sağlayacaktır.

Sınırlılıklar

Bu araştırma,

1. Rehberlik Araştırma Merkez'lerinde maksimum performansı ölçme amacıyla kullanılan ölçme araçlarından olan Wechsler Çocuklar için Zekâ Ölçeği'nin Gözden Geçirilmiş Hali, Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği ile sınırlıdır.
2. Ölçekler incelenirken APA, AERA ve NCME tarafından 2014 yılında ortak yayınlanan "Standards for Educational and Psychological Testing" kaynağındaki geçerlik ve güvenirlik bölümleri ile 2017 yılında Uluslararası Test Komisyonu (ITC) tarafından yayınlanan "Guidelines for Translating and Adapting Tests" yönergesinin ölçek uyarlama bölümü ile sınırlıdır.
3. Stanford-Binet Zekâ Ölçeği'nin çevirisinin Şemin (1972) tarafından yapıldığını; fakat uyarlamasının yapılmadığı belirtilmektedir. Çeviri süreci yazılı bir kaynakta bulunmadığı için Stanford-Binet ölçek uyarlama adımlarında göre incelenememiştir.

Varsayımlar

Bu araştırmada,

Psikolojik ölçekleri incelemek amacıyla geliştirilen kontrol listesinin dayandırıldığı APA, AERA ve NCME tarafından 2014 yılında ortak yayınlanan "Standards for Educational and Psychological Testing" kaynağı ile 2017 yılında Uluslararası Test Komisyonu (ITC) tarafından yayınlanan "Guidelines for Translating and Adapting Tests" yönergesindeki standartlar ile uyarlama adımlarının evrensel kriterleri temsil ettiği varsayılmıştır.

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırma konusuna ilişkin alanyazın taranmış ve aşağıda sunulmuştur

Bu alanda Türkiye'de yapılan ilk çalışma 1971 yılında Millî Eğitim Bakanlığı Planlama-Araştırma ve Koordinasyon Dairesi tarafından yapılmıştır. Bu kurum "Psikolojik Testler: Genel Bilgi Rehberi" isimli bir broşür yayınlayarak rehberlik ve psikolojik danışmanlık alanında kullanılan 20 ölçeği genel olarak tanıtmıştır. Bu ölçeklerin hepsinin Amerika Birleşik Devletleri kökenli olup direkt tercüme ya da uyarlama yapılarak Türkiye'de kullanıldığını belirtilmiştir. Ayrıca, bu ölçeklerde norm, geçerlik ve güvenirlik çalışması da yapılmadığı da rapor edilmiştir (MEB'den aktaran: Koç, 1993).

Yiğit, Çelik ve Erden (2017) çalışmalarında WISC-R ve WISC-IV ölçeklerini Ankara ilindeki ikamet eden 23 erkek 50 üstün yetenekli öğrenci üzerinde uygulamışlardır. Çalışmada sırasıyla WISC-IV'ün tüm ölçek puanı, sözel kavrama ve algısal akıl yürütme puanları ile WISC-R'daki toplam puan, sözel ve performans puanları arasındaki Pearson Momentler korelasyon katsayıları hesaplanmıştır. Katsayılar; toplam ölçek puanları arasında 0,47; sözel puan ile sözel kavrama dönüştürülmüş puanı arasında 0,23 ve performans puanı ile algısal akıl yürütme dönüştürülmüş puanı arasındaki 0,42 olarak raporlanmıştır. Çalışmada Benzerlikler, Küplerle Desen ve Şifre alt test puanlarının WISC-R'da daha yüksek olduğu; sözcük dağarcığı alt test puanlarının ise WISC-IV'de daha yüksek olduğu raporlanmıştır. Araştırmacılar, norm grubu daha güncel olduğu için WISC-IV'ün kullanılmasını önermişlerdir. Ölçeklerin belli aralıklarla yenilenmesinin de geçerlik ve güvenirlik açısından önemli olduğu vurgulamışlardır.

Uluç, Öktem, Erden, Gençöz ve Sezgin (2011) ise WISC-IV'ün yeniliklerini; norm grubu, ölçeğin yapısı ve sonuçlarının yorumlaması olarak 3 kategoride incelemiştir. Norm grubundaki yeniliklerin bir parçası olarak ölçek, Amerika'da 2200 ve İngiltere'de 780 katılımcı üzerinde özel gruplara da dikkat edilerek standartlandırılmıştır. Türkiye'de ise ölçek Türkiye'deki coğrafi bölge, yaş dilimi, cinsiyet ve sosyoekonomik düzeyi temsil eden 2225 katılımcı üzerinde standartlaştırılmıştır. Ölçeğin yapısıyla ilgili olarak ise toplam puan ile korelasyonu düşük olan ve uygulama süreleri istenenden uzun olan resim düzenleme, parça birleştirme ve labirent alt testleri ölçekten çıkarılmıştır. Bunların yerine resim kavramları, harf-rakam dizileri, mantık yürütme kareleri, simge arama, çiz çıkar ve sözcük bulma isimli 6 yeni alttest eklenmiştir. Alttestlere madde güçlük düzeyi düşük ve yüksek yeni maddeler eklenerek madde sayısı arttırılmıştır. Bu durum ölçekteki performans aralığını genişletilerek tavan-taban etkisi sorunlarının ortadan kalkmasını sağlamıştır. Ölçekteki yönergeler de daha açık hale getirilmiş ve ölçeğin uygulamasının bırakılmasında kullanılan hata sayıları arttırılmıştır. Sonuçlarının yorumlamasındaki değişikliklerde ise sözel kavrama birleşik puanı, algısal akıl yürütme birleşik puanı, çalışma belleği birleşik puanı, işlem hızı birleşik puanı ve tüm test zekâ puanı olarak 5 birleşik puan oluşturulmuştur. Oluşturulan bu birleşik puanlar hem kuramsal temele dayandırılmış hem de doğrulayıcı faktör analizleri yapılmıştır.

Flanagan ve Kaufman (2004) ise WISC-R (1974), WISC-III (1991) ve WISC-IV (2003)'ü karşılaştırmışlardır. Bu çalışmada WISC-R'ın önceki versiyona göre; yeni bir norm çalışması olduğunu, grafiklerin düzenlendiğini, madde içeriklerinin yenilendiğini, yanlı maddelerin ölçekten çıkarıldığını, puanlama ve uygulama talimatlarını geliştirildiğini raporlamışlardır. WISC-III'ün ise WISC-R'a göre norm çalışmasının daha temsili olduğunu, farklı indekslerin eklendiğini, madde güçlüğü seviyelerinin çeşitlendirildiğini, yeni alt test eklendiğini, puanlama ve uygulama talimatlarının geliştirildiğini belirtmişlerdir. WISC-IV'ün ise WISC-III 'e göre daha yeni ve daha temsili bir norm çalışması içerdiğini, yeni indeks eklendiğini, 5 yeni alt test eklendiğini ve indekslerin düzenlemelerinin yapıldığını raporlamışlardır.

Neale (1962) çalışmasında Stanford-Binet Zekâ Ölçeği'ni değerlendirmiştir. Çalışmada ölçeğin el kitapçığının puanlama ve uygulama talimatları açısından açık ve ayrıntılı olduğu raporlanmıştır. İlgili revizyonunda IQ tablolarının düzenlenmesiyle birlikte sonuçların daha hızlı ve verimli bir şekilde edildiği belirtilmiştir. Alttestlerdeki değişikliklerin ölçeğin

uygulanmasına yararlı olduğu ve resimli alttestlerin eklenmesi ile küçük yaş gruplarının iş birliğini kolaylaştırdığını da raporlamıştır.

Waddell (1980) çalışmasında Stanford-Binet Zekâ Ölçeği'nin 1972 revizyonundaki teknik verileri incelemiştir. Ölçekte standartlaştırılmanın yapıldığı örneklemin sosyoekonomik düzey, etnik yapı, cinsiyet ve zekâ seviyeleri açısından yeterince detaylandırılmadığı raporlanmıştır. Anadili İngilizce olmayan katılımcıların örnekleme alınmaması eleştirilmiştir. İlgili versiyon için IQ puan dönüşümünü içeren bir tablo olmadığı ve güvenirlik çalışmasının yapılmadığı da raporlanmıştır. Ölçeğin geçerlik verilerinin çoğunun diğer zekâ ölçekleriyle olan karşılaştırmalarına dayanmasının yapı ve konu geçerliği için yeterli olduğu; fakat yordayıcı geçerlik çalışmalarının da yapılması gerektiği belirtilmiştir.

Becker (2003) ise Stanford-Binet Zekâ Ölçeği'nin tüm revizyonlarını değerlendirmiştir. RAM'larda kullanılan revizyonunun olumlu yanları; açık puanlama talimatları içermesi, geçerliği düşük olan maddelerin çıkarılmasıyla önceki versiyondaki sağlam maddeleri içermesi, her yaş aralığına uygun alternatif maddeler içermesi, sabit bir standart sapmasının olması ve ölçeğin temellendiği ayrıntılı alanyazın içermesi olarak raporlanmıştır. Olumsuz yanları ise genel zekâ olarak tek faktörün ölçülmesi, ölçeğin genelinde sözel maddelerin fazla olması ve tavan puanının bazı kullanıcılar için yetersiz kalması şeklinde açıklanmıştır.

Tyler-Wood, Knezek, Christensen, Moreales, ve Dunn-Rankin (2014) ise araştırmalarında Stanford-Binet Zekâ Ölçeği'nin 1973, 1986 ve 2003 versiyonlarını karşılaştırmışlardır. Bu karşılaştırma sonucunda 1973 versiyonundaki kültürel yanlılık, puanlamada zorluk, sonuçların yorumlanmasında yanlılık, puanlama prosedüründe alttestlerin ayrı puanlamasının olmaması, kelime alttesindeki madde türlerinin her yaş seviyesinde yeterli oranda temsil edilmemesi, puanlamada maddelerin yaş seviyelerine göre gruplanması ve ölçek bırakma prosedürleri konularını eleştirmişlerdir.

Huelsman (1965) Gesell Gelişim Ölçeği'ni ilgili versiyonunu incelediği çalışmasında ölçeğin yaş ya da sınıf seviyesine göre normlarının bulunmamasını, toplam zekâ puanı belirtilmemesi, ölçeğin kullanım yararlarının açıklanmaması, standart sapmanın belirtilmemesini, geçerlik- güvenirlik çalışmalarının yapılmamış olmasını, ölçeğin kullanım yararları hakkındaki açıklamaların yetersizliğini ve alanyazın taramasında yazarların çoğunlukla kendi çalışmalarına atıf yapmalarını eleştirmiştir.

Lichtenstein (1990) Gesell Gelişim Ölçeği sonuçlarını incelediğinde puanlayıcılar arası farklılıklar olduğunu ve bu durumun ölçekte net bir puanlama sistemi olmamasından kaynaklandığını belirtmiştir.

Shepard (1997) ise çalışmasında Gesell Gelişim Ölçeği el kitabında test- tekrar test ve puanlayıcılar arası güvenirlik katsayısı gibi önemli psikometrik özelliklerin bulunmamasını eleştirmiştir.

Kaplan ve Saccuzzo (2013); Gesell Gelişim Ölçeği'nde standardizasyon örnekleminin hedeflenen evreni temsil etmemesini, puanlama sistemi, yönergelerin de net olmamasını olmasını ve ölçeğin psikometrik açıdan istenilen seviyede olmamasını eleştirmişlerdir. Ayrıca, ölçeği sadece çocuk gelişimi alanında ileri düzeyde eğitimi olan kişilerin doğru şekilde uygulayabileceğini belirtmişlerdir.

Arthur (1949), Leiter Uluslararsı Performans Ölçeği'nde bulunan norm seviyelerinin ortalama bir çocuk için çok yüksek olduğunu ve bu sebeple katılımcıların zekâ seviyelerini doğru ölçülemediğini açıklamıştır. Her yaş seviyesi için sadece 4 test verilmesi sonucunda bir hatanın bile toplam puan üzerinde çok etkili olabileceğine bunun da ölçeğin geçerliğini düşüreceğine değinmiştir.

Savaşır (1994) çalışmasında Türk Psikiyatri Derneği ve Psikoloji Dergisi'nde yayınlanan 33 psikolojik ölçek uyarlama makalesi incelemiştir. Bu çalışmalardan yola çıkarak belli ölçek uyarlama sorunları; çevirmenlerin seçimi, çeviri süreci, ölçek uygulama yönergesi ve çevirinin denetlenmesi şeklinde 4 kategoride toplanmıştır. İncelenen çalışmaların 12'sinde çeviriyi kimin yaptığı belirtilmemiştir. Çeviri çalışmaların 4'ünde tek kişi, 5'inde ise 3-5 kişilik gruplar tarafından yapıldığı raporlanmıştır. Çeviriyi yapan ve denetleyen iki farklı gruptan yararlanıldığını belirten ise 6 çalışma bulunmaktadır. Çeviri sürecinde ise dilin ölçeğin uygulanacağı kişilere uygun şekilde çevrilmediği görülmüştür. Yazar; çevirideki kelime ve ifadelerin ölçeğin hedeflediği gruptaki herkes tarafından kolayca anlaşılabilir olmayabileceğini farklı çalışmalardan örneklerle açıklamıştır. Ölçek uygulama yönergesi kısmında dereceli ölçeklerde sıklık kavramlarının çevirisindeki farklılıkların araştırmanın sonucunu önemli ölçüde değiştirdiği görülmüştür. Ayrıca, dilimizin yapısından dolayı cümlelerdeki çift olumsuzlukların da madde analizleri sırasında sorun yarattığı belirtilmiştir. Çeviri denetlenme kısmında ise çalışmaların 7'sinde geri çevirme yöntemi 12'sinde ise tek yönde çeviri kullanıldığı raporlanmıştır.

Erkuş (2007) ölçek geliştirme ve uyarlama aşamasındaki sorunları incelediği araştırmada, akademisyenlerin hızlı yayın çıkarabilmek için ölçek geliştirme veya uyarlama çalışmalarını tercih ettiklerini ve bunun da pek çok hata veya noksana yol açtığını belirtmiştir. Bu hatalar; çalışmanın başlığı, örtük değişkenin işevuruk tanımı, madde üretilmesi, deneme yapılması, madde analizi, güvenirlik, geçerlik, puanların yorumlanması ve çalışmanın raporlaştırılması şeklinde 8 kategoride değerlendirilmiştir. Çalışmanın başlığı; içeriği özetlemeli, anlatım bozukluğu içermemeli ve uygun uzunlukta olmaması gerekmektedir. Örtük değişkenin işevuruk tanımının detaylı ve kuramsal şekilde tanımlanması ölçek geliştirmenin temelini oluşturmaktadır. Örtük değişken doğrudan gözlemlenemediği için, kurama dayalı gözlemler ve benzer kavramların ortak niteliklerini içermesinin gerekliliğini vurgulanmıştır. Madde üretimi aşaması ise kavramsal yapıya uygun davranışsal göstergeler belirlenmesi ile başlaması gerektiğini açıklanmaktadır. Bu aşamada aynı zamanda maddelerin kimler tarafından yazıldığı, hakemler tarafından incelenip incelenmediği, kimler tarafından çevrildiği, uyarlama geçerliği için hangi istatistiksel çalışmalar yapıldığı gibi sorulara da cevap vermenin önemi vurgulanmıştır.

Çüm (2013) TÜBİTAK ULAKBİLİM veri tabanında bulunan 2005 ile 2013 yılları arası psikoloji ve eğitim alanında yayınlanan 50 adet ölçek uyarlama ve geliştirme makalesi incelemiştir. Ölçek uyarlama açısından bütün araştırmalarda ölçeğin güvenirlik analizlerinin yapıldığı, deneysel- korelasyonel geçerlik analizlerinin yapıldığı ve araştırmaların çoğunluğunda çevirmenlerin iki dile de hakim olduğu raporlanmıştır. Diğer taraftan ise, çoğu araştırmada ölçek uyarlama gerekçeleri belirtilmemesi, çoğunda madde analizine ilişkin bilgi vermemesi, araştırmaların yarısına yakınında uygulama için alınması gereken izinlere yer verilmemesi, sadece 3 çalışmada çeviri sürecinde ölçme değerlendirme uzmanı bulunması, yalnız 2 çalışmada tercümanların iki kültüre de hakim olduğunun belirtilmesi ve hiçbir çalışmada dilsel ya da kültürel açıdan yapısal eşitlikten emin olunmaması çalışmaların ölçek uyarlama açısından yetersiz kısımlarıdır. Araştırmada genel olarak çalışmaların ölçek uyarlama ilkelerine %26 oranında uygunluk gösterdiği raporlanmıştır.

Boztunç-Öztürk, Eroğlu ve Kelecioğlu (2015) SSCI ve ULAKBİM'de 2005 – 2014 yılları arasında yayınlanan 108 ölçek uyarlama makalesi incelemiştir. Bu inceleme sonucunda ilgili çalışmalarda genel olarak ölçek uyarlama adımlarına uyulduğu gözlenmiştir. Ölçüt geçerliğinin çalışmaların yarısından fazlasında uygulanmaması, madde analizinde yalnızca madde – toplam puan korelasyonu yapılması, güvenirliğe sadece iç tutarlılık açısından

bakılması, sadece istatistiksel sonuçlara dayanarak ölçekten madde çıkarılıp yeni madde eklenmesi bu araştırmalarda gözlemlenen bazı eksikliklerdendir. Ayrıca, dilsel eşitliği kontrol etmek için çalışmaların yarısına yakınında hedef gruba yalnızca Türkçe'ye çevrilmiş olan ölçek uygulanırken; yaklaşık üçte birine de hem Türkçe'ye çevrilmiş ölçek hem de ölçeğin orijinal hali beraber uygulanmıştır.

Acar-Güvendir ve Özer-Özkan (2015) ise 2006- 2014 yılları arasında Eğitim ve Bilim Dergisi, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi ve Türk Dünyası Sosyal Bilimler Dergisi'nde yayınlanan 59 ölçek geliştirme ve uyarlama makalelerini incelemiştir. Bu inceleme sonucunda, ölçek geliştirilirken madde havuzu oluşturma sürecinde araştırmacıların tamamına yakını literatür taramasından faydalanırken yaklaşık üçte birinin buna ek olarak hedef gruba yazdırdığı kompozisyonlardan ya da yaptığı görüşmelerden faydalandığı görülmüştür. Yine madde havuzu aşamasında araştırmacıların tamamına yakını alan uzmanlarından görüş alırken, yaklaşık üçte biri ayrıca ölçme uzmanlarından görüş aldığı belirtilmiştir. Çalışmaların büyük bir çoğunluğunda ise maddeler bu geribildirimler sonucunda değiştirildiği açıklanmıştır. Ölçek için uygulama yönergesi ise bu çalışmaların çok azında yazılmıştır. Çalışmaların yaklaşık yarısında ölçeğin küçük grupla ön denemesi yapılmıştır. Ölçek uyarlama çalışmaları incelendiğinde ise bunların yarısından fazlası uyarlama izinlerinin alındığı görülür. Çalışmaların yaklaşık yarısında ölçeğin Türkiye'ye göre uyarlanıp uyarlanamayacağı ile ilgili alan uzmanlarına danışmışken; çok az bir kısmı ölçme uzmanlarına danışmıştır. Çeviriler yapılırken ise araştırmaların tamamına yakını her iki dile hakim en az iki tercümandan faydalanmıştır. Çalışmaların yarısından fazlasında dil eşdeğerliğinin sağlandığı rapor edilmiştir. Diğer taraftan ise hiçbir çalışma ölçeklerin uygulama yönergesini uyarlandığı belirtilmemiştir. Ön deneme uygulaması çoğu çalışmalarda raporlanmamıştır. Genel olarak bakıldığında Türkiye'de ölçek geliştirme ve uyarlama adımlarında bir birliğin olmadığı görülmüştür.

Dirlik (2013) çalışmasında, eğitim kurumlarında kullanılan psikolojik ölçekleri belirleyip bunları geçerlik, güvenirlik ve ölçek uyarlama standartları açısından incelemiştir. Bu inceleme yapılırken APA'nın 1999 yılında yayınladığı standartlardan faydalanmıştır. Çalışma sonuçlarına göre; Gazi Erken Çocukluk Gelişimi Değerlendirme Aracı 10 geçerlik standardından birini, 9 güvenirlik standardından dördünü karşılamadığı belirtilmiştir. Akademik Benlik Kavramı Ölçeği'nin ise geçerlik standartlarının 3'ünü karşılarken

güvenirlik standartlarının hiçbirini karşılayamadığı bulunmuştur. WISC-R'da ise 15 uyarlama standardının 14'ünü tamamen karşıladığı raporlanmıştır.

Alanyazın incelendiğinde Türkiye'de kullanılan psikolojik ölçeklerin daha güncel versiyonları bulunduğu ve uluslararası alanda bu versiyonların kullanıldığı görülmüştür. Bu sebeple Türkiye'de kullanılan versiyonlarına yönelik çalışmaların eski ve kısıtlı olduğu görülmektedir. Bu çalışmalardan yola çıkarak bu versiyonların psikometrik özelliklerinde ve uyarlamalarında eksikliklerin olduğu ve bunların tartışıldığı görülmektedir.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın deseni, evren ve örneklem, veri toplama araçları ve verilerin analizine yer verilmiştir.

Araştırma Deseni

Araştırma bir durum çalışmasıdır. Durum çalışması, güncel sınırlanmış bir durumun ya da sınırlı zaman içerisindeki birden çok sınırlandırılmış durumun farklı bilgi kaynakları yardımıyla betimlenmesidir. Bu araştırma deseninde konular kronolojik şekilde tasarlanıp aralarındaki farklılıklar ya da benzerliklere göre analizi yapılabilir. Böylece araştırmacının konuyu derinlemesine araştırmasına olanak sağlamaktadır. Araştırmada doküman incelemesi de yapılmıştır. Döküman incelemesi, nitel araştırmalarda kullanılması gereken veri çeşitlendirmesi ve araştırmanın geçerliğini arttırması önemli bir bilgi kaynağıdır (Yıldırım ve Şimşek, 2011). Bu araştırmada RAM'larda maksimum performansı ölçmede kullanılan 4 test belirlenip bu ölçekler derinlemesine incelenmiştir.

Çalışma Grubu

Bu araştırmada incelenecek olan psikolojik ölçekler seçkisiz olmayan örnekleme türlerinden biri olan amaçsal örnekleme yöntemiyle seçilmiştir. Amaçsal örnekleme, çalışmanın bilgisel açıdan zenginleşmesi için ona yönelik durum, olay, kişi ya da kurumların seçilmesidir. Bu çalışmada amaçsal örnekleme türlerinden olan ölçüt örnekleme yöntemi kullanılmıştır. Ölçüt örnekleme, çalışmanın problemine göre belli ölçüt seçilmesi ve durum, olay, kişi ya da

kurumların bu ölçüte göre seçilmesidir (Büyüköztürk vd., 2014). Bu araştırma kapsamında İstanbul'da bulunan 8 RAM ile Ankara'da bulunan 2 RAM yetkilileriyle görüşmeler yapılmış ve en sık kullandıkları 3'er psikolojik ölçek öğrenilmiştir (bkz. Tablo 1). Bu görüşmeler sonucunda rapor edilen kullanım sıklığı ölçüt olarak kabul edilmiştir. Çalışma grubuna WISC-R Zekâ Ölçeği, Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği alınmıştır.

Tablo 1

RAM'larda Kullanılan Psikolojik Testler ve Kullanım Sıklıkları

Ölçeğin Adı	Uygulanma Sıklığı
WISC-R Zekâ Ölçeği	8
Stanford Binet Zekâ Ölçeği	5
Anadolu Sak Zekâ Ölçeği (ASİS)	4
Leiter Uluslararası Performans Ölçeği	4
Gesell Gelişim Ölçeği	3
Peabody Resim Kelime Ölçeği	2
Porteus Labirentleri Zekâ Ölçeği	1
Gazi Gelişim Envanteri	1
Bender Gestalt Görsel Motor Algılama Ölçeği	1
Denver 2 Gelişimsel Tarama Envanteri	1

Veri Toplama Aracı

Kontrol listesi hazırlanırken hem APA, AERA ve NCME tarafından yayınlanan psikolojik ölçek standartlarının geçerlik ve güvenirlik bölümleri hem de Uluslararası Test Komisyonu tarafından hazırlanan ölçek uyarlama yönergesi araştırmacı tarafından Türkçe'ye çevrilmiştir. Sonrasında çevirilerin dilsel geçerliği Eğitimde Ölçme ve Değerlendirme alanında doktorasını yapan bir uzman tarafından kontrol edilip düzenlenmiştir. Hazırlanan bu listede birden fazla öğe içeren standartların daha anlaşılır olması için ilgili standartlar ayrı kriterler halinde eklenerek kontrol listesinin son hali oluşturulmuştur.

APA, AERA ve NCME tarafından yayınlanan 21 adet geçerlik ve 21 adet güvenirlik standardı içinde birden fazla konu içeren standartlar ayrılmıştır. Bu işlem sonucunda oluşan geçerlik kontrol listesi 36, güvenirlik listesi ise 28 kriter içermektedir. Uluslararası Test

Komisyonu'nun yayınladığı 11 ölçek uyarlama standartı ise yapılan düzenleme ile 12 kriter içeren bir liste haline getirilmiştir.

Seçilen psikolojik ölçeklerin orijinal formlarındaki yönergeleri ve uyarlama süreçleri hazırlanan kontrol listesine göre incelenip kodlanmıştır. Ölçekler kriterlere göre incelenirken incelenirken; "0" kriterle hiç uyumlu değil, "1" kriterle kısmen uyumlu, "2" kriterle tamamen uyumlu ve "X" kritere göre incelenmedi şeklinde değerlendirilmiştir.

Veri Toplama Süreci

WISC-R, Stanford Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nin incelenmesi araştırmacı ve Eğitimde Ölçme ve Değerlendirme alanında doktorasını yapan 2 uzman tarafından yapılmıştır. Geçerlik, güvenirlik ve ölçek uyarlama kontrol listeleri her ölçek için ayrı ayrı hazırlanmıştır. Bu incelemeler sırasında yararlanılan kaynaklar aşağıda ayrıntılı olarak sunulmuştur.

Wechsler Çocuklar için Zekâ Ölçeği'nin gözden geçirilmiş hali (WISC-R) incelenirken;

- Kaufman (1979) "Intelligent Testing With the WISC-R"
- Öner (2008) "Türkiye'de Yazılan Psikolojik Testlerden Örnekler"
- Savaşır ve Şahin (1995) "Wechsler Çocuklar İçin Zekâ Ölçeği (WISC-R) El Kitabı"
- Savaşır ve Şahin (1978) "WISC uyarlama çalışmaları: Ön Rapor – I"
- Savaşır ve Öktem (1979) "WISC uyarlama çalışmaları: Ön Rapor – II" adlı kaynaklar kullanılmıştır.

Stanford Binet Zekâ Ölçeği incelenirken;

- Roid ve Barram (2004) "Essentials of Stanford-Binet Intelligence Scale"
- Sattler (1965) "Analysis of Functions of the 1960 Stanford-Binet Intelligence Scale Form L-M"
- Becker (2003) "History of the Stanford-Binet Intelligence Scales: Content and Psychometrics"
- Öner (2008) "Türkiye'de Yazılan Psikolojik Testlerden Örnekler" adlı kaynaklardan yararlanılmıştır.

Öner (2008) tarafından da belirtildiği gibi Stanford-Binet'in uyarlama çalışması Şemin tarafından yapılmıştır; fakat bu uyarlama yazılı bir kaynakta bulunmamaktadır.

Leiter Uluslararası Performans Ölçeği için ise geçerlik ve güvenirlik açısından incelemesi Leiter (1936) tarafından yazılan "The Leiter International Performance Scale" adlı orijinal kaynağa göre yapılmıştır. Leiter Uluslararası Performans Testi için alanyazında herhangi bir uyarlama çalışmasına rastlanmamıştır.

Gesell Gelişim Ölçeği geçerlik standartlarına uygunluğu Ilg ve Ames (1972) tarafından yayınlanan "School Readines: Behavior Tests Used at the Gesell Institude" adlı orijinal el kitabına göre incelenirken güvenirlik standartlarına uygunluğu Evirgen, Kayhan ve Erden (2015) tarafından yazılan "Gesell Gelişim Figürleri'nin Anasınıfı Çocuklarında Güvenirliğine Yönelik Bir Ön Çalışma" adlı çalışmaya göre yapılmıştır. Gesell Gelişim Ölçeği için alanyazında herhangi bir uyarlama çalışmasına rastlanmamıştır.

Verilerin Analizi

Araştırma kapsamında puanlayıcılar arası güvenirlik düzeyleri ölçümü için Fleiss Kappa ve Genellenebilirlik kuramına dayalı varyans bileşenlerinin ANOVA kestirimi kullanılmıştır. Fleiss (1971) ikiden fazla hakem veya ikiden fazla kategorinin incelenebilmesi için Kappa istatistiğinde düzenleme yapmıştır. Genellenebilirlik kuramında tüm hata kaynaklarının eş zamanlı ve birlikte değerlendirilebilmektedir (Brennan, 2001).

Araştırmada puanlayıcılar arası güvenirlik düzeyleri Landis ve Koch (1977b) tarafından önerilen aralıklara göre değerlendirilmiştir. Buna göre 0 ile 0,20 arası "Önemsiz"; 0,21 ile 0,40 arası "Düşük düzeyde uyum"; 0,41 ile 0,60 arası "Orta düzeyde uyum"; 0,61 ile 0,80 arası "Yüksek düzeyde uyum" ve 0,81 ile 1 arası "Çok yüksek düzeyde uyum" şeklinde yorumlanmıştır. Genel olarak 0,40'ün üstü kabul edilebilir düzeyde uyum olarak yorumlanmaktadır.

Araştırmalarda çoklu puanlayıcı kullanılması durumlarında kullanılan yaklaşımlardan biri çoğunluğun kararının kabul edilmesidir (Landis ve Koch, 1977a). Araştırmada puanlayıcılar arası ortak görüşün olmadığı kriterlerde çoğunluğun kararı kabul edilmiştir.

Geçerlik kontrol listesi için puanlayıcılar arası Fleiss Kappa değeri Stanford-Binet Zekâ Ölçeği'nde 0,62; WISC-R'da 0,39; Leiter Uluslararası Performans Ölçeği'nde 0,64 ve Gesell Gelişim Ölçeği'nde 0,70 bulunmuştur. Puanlayıcılar arası uyumun Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nde yüksek düzeyde olduğu; WISC-R'da ise düşük düzeyde olduğu görülmektedir.

Güvenirlik kontrol listesi için ise Fleiss Kappa değeri Gesell Gelişim Ölçeği'nde 0,53; Leiter Uluslararası Performans Ölçeği'nde 0,59 ve WISC-R'da 0,03 bulunmuştur. Stanford-Binet Zekâ Ölçeği'nde ilgili versiyonunda güvenirlik çalışması yapılmadığı için puanlayıcılar tarafından ortak alınan kararla ölçek hiçbir kriterle uyumlu değildir şeklinde kabul edildiği için Fleiss Kappa analizi yapılmamıştır. Puanlayıcılar arası uyumun Leiter Uluslararası Performans Ölçeği'nde ve Gesell Gelişim Ölçeği'nde orta düzeyde; WISC-R'da ise önemsiz düzeyde olduğu görülmektedir.

Feinstein ve Cicchetti (1990) Kappa İstatistiğinde görülen bazı paradokslardan bahsetmişlerdir. Bunlardan biri tablodaki bir sütun toplamının diğerlerine oranla fazla olması sonucunda Kappa değerinin çok düşük çıkması olarak açıklanmıştır. WISC-R'ın incelendiği çoğu kriterde puanlayıcılar ölçeği kriterle tamamen uyumlu şeklinde değerlendirdikleri için tablodaki bu sütununun toplamı diğer sütunlara göre çok daha büyüktür. Bu sebeple WISC-R'ın Kappa istatistik değerinin çok düşük çıktığı görülmektedir.

Ölçek uyarlama kontrol listesi için ise puanlayıcılar arası Fleiss Kappa değeri WISC-R'da 0,53 bulunmuştur. Buna dayanarak puanlayıcılar arası uyumun yüksek düzeyde olduğu görülmektedir. Leiter Uluslararası Zekâ Ölçeği ve Gesell Gelişim Ölçeği'nin alanyazında herhangi bir çeviri ya da uyarlaması bulunmamaktadır. Puanlayıcıların ortak kararıyla 2 ölçekte de hiçbir kriterle uyumlu değildir şeklinde kabul edildiği için bu ölçek puanlamalarında Fleiss Kappa analizleri yapılmamıştır. Öner (2008) Stanford-Binet Zekâ Ölçeği için ölçeğin Şemin tarafından çevirisinin yapıldığını; fakat uyarlamasının yapılmadığı belirtilmektedir. Bu çeviri süreci yazılı bir kaynakta bulunmadığı için Stanford-Binet Zeka Ölçeği ölçek uyarlama adımlarında göre incelenememiştir.

Geçerlik kontrol listesi için genellenebilirlik kuramına dayalı olarak hesaplanan puanlayıcılardan kaynaklanan varyans değeri Stanford-Binet Zekâ Ölçeği'nde %1,3; WISC-R'da %8,3; Leiter Uluslararası Performans Ölçeği'nde %1,9 ve Gesell Gelişim Ölçeği'nde %0,0 bulunmuştur. Güvenirlik kontrol listesi için ise Gesell Gelişim Ölçeği'nde ve WISC-R'da %0,0 iken Leiter Uluslararası Performans Ölçeği'nde %3,0 bulunmuştur. Ölçek uyarlama kontrol listesi için ise WISC-R'da %4,3 bulunmuştur. İlgili varyans yüzde değerlerine bakıldığında Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği için geçerlik ve güvenirlik kontrol listelerinde puanlayıcıların birbirleriyle benzer puanlama yaptıklarını görülmektedir. WISC-R 'da ise puanlayıcıların

geçerlik kontrol listesinde benzer puanlama yaptıklarını, güvenirlik ve ölçek uyarlama kontrol listelerinde ise benzer puanlama yapmadıkları görülmektedir.

Genellenebilirlik kuramı G katsayısı ise geçerlik kontrol listesinde Stanford-Binet Zekâ Ölçeği'nde 0,91; WISC-R'da 0,89; Leiter Uluslararası Performans Ölçeği'nde 0,95 ve Gesell Gelişim Ölçeği'nde 0,93 olarak bulunmuştur. Güvenirlik kontrol listesi için ise Gesell Gelişim Ölçeği ve Leiter Uluslararası Performans Ölçeği'nde 0,95 iken WISC-R'da ise 0,65 bulunmuştur. Ölçek uyarlama kontrol listesi için ise WISC-R'da 0,96 bulunmuştur. Geçerlik kontrol listesi için en yüksek G katsayısının 0,95 ile Leiter Uluslararası Performans Ölçeği'nde iken en düşük katsayının ise 0,89 ile WISC-R'da olduğu görülmektedir. Güvenirlik kontrol listesi için en yüksek G katsayısının 0,95 ile hem Gesell Gelişim Ölçeği hem de Leiter Uluslararası Performans Ölçeği'nde olduğu, en düşük G katsayısının ise 0,65 ile WISC-R'da olduğu görülmektedir.

Genellenebilirlik kuramı Phi katsayısı ise geçerlik kontrol listesinde Stanford-Binet Zekâ Ölçeği'nde 0,90; WISC-R'da 0,86; Leiter Uluslararası Performans Ölçeği'nde 0,95 ve Gesell Gelişim Ölçeği'nde 0,93 olarak bulunmuştur. Güvenirlik kontrol listesi için ise Gesell Gelişim Ölçeği'nde 0,95; Leiter Uluslararası Performans Ölçeği'nde 0,94 ve WISC-R'da ise 0,65 bulunmuştur. Ölçek uyarlama kontrol listesi için ise WISC-R'da 0,95 bulunmuştur. Phi katsayısının göreceli olarak istatistiksel olarak anlamlı kabul edilebilmesi için değerlerin 0,80 ve üstünde olması gerekmektedir (Brennan, 2001). Araştırmada Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nin Phi katsayılarının geçerlik ve güvenirlik kontrol listesi için anlamlı olduğu görülmektedir. WISC-R'da ise Phi katsayılarının güvenirlik ve ölçek uyarlama kontrol listelerinde anlamlı olduğu güvenirlikte ise anlamlı olmadığı görülmektedir.

BÖLÜM IV

BULGULAR

Bu bölümde araştırma sonucunda elde edilen bulgular, her bir araştırma sorusu için ayrı ayrı sunulmuştur.

Ölçeklerin Kontrol Listesindeki Standartları Karşılama Düzeyine İlişkin Bulgular

Bu bölümde birinci araştırma sorusuna ilişkin bulgulara yer verilmiştir. Kontrol listesinde yer alan geçerlik, güvenirlik ve ölçek uyarlama standartları ayrı başlık altında ele alınmıştır.

Ölçeklerin Kontrol Listesindeki Geçerlik Standartlarını Karşılama Düzeyine İlişkin Bulgular

APA, AERA ve NCME'nın standartlarına dayalı olarak oluşturulan geçerlik kontrol listesi toplamda 36 kriter içermektedir. Yapılan incelemeler sonucunda puanlayıcılar arası Fleiss Kappa değeri Stanford-Binet Zekâ Ölçeği'nde 0,62; WISC-R'da 0,39; Leiter Uluslararası Performans Ölçeği'nde 0,64 ve Gesell Gelişim Ölçeği'nde 0,70 bulunmuştur. Puanlayıcılar arası uyumun Stanford-Binet Zekâ Ölçeği, Leiter Uluslararası Performans Ölçeği ve Gesell Gelişim Ölçeği'nde yüksek düzeyde, WISC-R 'da ise düşük düzeyde olduğu görülmektedir. İncelemeye alınan ölçeklerin geçerlik kriterlerini karşılama düzeyine ilişkin bilgiler Tablo 2'de verilmiştir.

Tablo 2

Ölçeklerin Geçerlik Kriterlerini Karşılama Düzeyine İlişkin Sayılar

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

Tablo 2'de de belirtildiği üzere WISC-R'ın ise incelendiği 9 kriterin 4'ü ile tamamen, 2'si ile de kısmen uyum gösterdiği, 3'ü ile de hiç uyum göstermediği görülmüştür. Stanford-Binet Zekâ Ölçeği'nin incelendiği 7 kriterin 1'i ile tamamen, 2'si ile de kısmen uyumlu olduğu, 4'ü ile de hiç uyumlu olmadığı görülmüştür. Gesell Gelişim Ölçeği'nin ise 6 kriterin 2'si ile tamamen uyumlu olduğu, 2'si ile kısmen uyumlu olduğu ve 2 si ile de hiç uyumlu olmadığı görülmüştür. Leiter Uluslararası Zekâ Ölçeği'nin ise incelendiği 9 kriterin 2'si ile tamamen, 4'ü ile de kısmen uyum gösterdiği, 3'ü ile de hiç uyum göstermediği görülmüştür.

İncelemeye alınan ölçeklerin 3 puanlayıcı tarafından geçerlik kriterlerini karşılama düzeyi Tablo 3'de verilmiştir. Tablo 3'ün devamında ise ölçekler bu her bir kriteri karşılamaları açısından incelenmiştir.

Tablo 3

Ölçeklerin Geçerlik Kriterlerini Göre Uyum Durumları

	Kriterler	WISC-R	S.Binet	Gesell	Leiter
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	2	0	2	2
3	Belirli kullanım için genel veya olası yorumların geçerliği değerlendirilmediyse ya da yorum mevcut kanıtlarla tutarsız ise, bu durum açıklanmış ve potansiyel kullanıcılar	X	0	0	X

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

"0" Kriterle hiç uyumlu değil
"1" Kriterle kısmen uyumlu
"2" Kriterle tamamen uyumlu
"X" Kriter ilgili ölçek için incelenmedi

Kontrol listesindeki 1A kriterinde ölçeklerde, puanlarının yorumlanma ve kullanım şeklinin açıkça belirtilmesi gerektiğini açıklamaktadır. İncelenen tüm ölçeklerin bu kriterle kısmen uyum sağladığı kabul edilmiştir.

Kontrol listesindeki 1B kriteri ölçek evreninin ve yapılarının açıkça tanımlanması gerektiğini belirtmektedir. Bu kriterin WISC-R ile tamamen uyumlu; Stanford-Binet, Gesell ve Leiter ile kısmen uyumlu olduğu kabul edilmiştir.

Listedeki 2. kriter ise ölçek puanlarının yorumlanmasında dayandığı kanıtların detaylıca açıklanması gerektiğini belirtmektedir. Ölçeğin önceki çalışmalardan yararlanarak, yapılan araştırma sonuçlarına dayanarak ya da istatistiksel analizler yardımıyla ampirik kanıtları sağlanmalıdır. Bu kriter WISC-R, Leiter ve Gesell ile tamamen uyumlu; Stanford-Binet ile hiç uyumlu değil şeklinde kabul edilmiştir.

Listedeki 3. kriter ölçekte yorumların geçerliği değerlendirilmedi ise ya da yorum kanıtlarla tutarsız ise, bu durumun açıklanması ve kullanıcıların uyarılması gerektiğini belirtmektedir. Eğer ölçeğin herhangi bir karar için ya da belli katılımcılarda kullanımı için uygun olmadığı görüldüyse bu konular hakkında uyarıda bulunulmalıdır. Hem Gesell hem de Stanford-Binet ilgili kriterle hiç uyumlu değildir şeklinde kabul edilmiştir. WISC-R, Leiter ise bu kritere göre incelenmemiştir.

Kontrol listesinde yer alan 7. kriter ise geçerlik kanıtının elde edildiği tüm örneklemlerin pratik ve verilen izinler çerçevesinde detaylıca tanımlanması gerektiğini açıklamaktadır. Katılımcıların sosyodemografik bilgileri, eğitim durumu, ikamet edilen yer bilgisi, dilsel özellikleri ve örneklemin seçme prosedürleri gibi özellikleri tanımlanmalıdır. İlgili kriterle WISC-R ve Stanford-Binet tamamen uyumlu Leiter'le de kısmen uyumlu kabul edilmiştir. Gesell ise bu kritere göre incelenmemiştir.

Kontrol listesindeki 9A kriteri veri toplama koşullarının, istatistiksel bulguların yerel koşullarla ilişkisinin karşılaştırılabileceği şekilde detaylıca açıklanması gerektiğini belirtmektedir. Bu koşullar katılımcıların motivasyon düzeyini, ön hazırlık durumlarını, ölçek puan aralıklarını, ölçek için verilen süreyi, ölçek uygulama şeklini (kâğıt- kalem ya da bilgisayarda), ölçek uygulayıcısının özelliklerini, veri toplama süreleri arası zaman gibi veriler toplandıktan sonra değişmesi muhtemel olan durumları içermelidir. İlgili kriter açısından Stanford-Binet, WISC-R ve Leiter'in hiç uyumlu olmadığı kabul edilmiştir. Gesell ise bu kritere göre incelenmemiştir.

Listedeki 9B kriteri ise 9A kriteri ile ilişkili şekilde veri toplama sürecinde tipik test koşullarına göre farklılık gösteren ve performansı önemli şekilde etkileyebilecek durumlara dikkat edilmesi gerektiğini vurgulamaktadır. Stanford-Binet, WISC-R ve Leiter'in ilgili kriter açısından hiç uyumlu olmadığı kabul edilmiştir. Gesell ise bu kritere göre incelenmemiştir.

Listedeki 13A kriteri alt bölüm puanları, puan farklılıkları ya da profil yorumlamasını destekleyen gerekçe ile ilgili kanıt sağlanması gerektiğini belirtmiştir. Ölçeğin puanlama sürecinde birden fazla puan varsa her puanın ayırt edilebilirliği ve güvenirliği belirtilmeli ve puanlar arası ilişkinin de yapı ile tutarlı olduğu gösterilmelidir. Gesell, Leiter ve WISC-R'ın ilgili kriter açısından hiç uyumlu olmadığı kabul edilmiştir. Stanford-Binet ise bu kritere göre incelenmemiştir.

Bir önceki kriter ile ilişkili olarak 13B kriteri de ise birleşik puanların temel ve gerekçelerinin verilmesi gerektiğini açıklamaktadır. Leiter ve WISC-R'ın ilgili kriter açısından hiç uyumlu olmadığı kabul edilmiştir. Stanford-Binet ve Gesell ise bu kritere göre incelenmemiştir.

Listedeki 17. kriterde ise kriter performansının seviyeleri hakkında bilgi verilmesinin gerekliliği açıklanmaktadır. Kriter performans seviyelerinde istatistiksel veriler ve katılımcıların belirtilen aralıklardaki performans dağılımı sunulmalıdır. Bu kriterin Gesell ile tamamen uyumlu olduğu; Leiter ve WISC-R ile ise hiç uyumlu olmadığı kabul edilmiştir. Stanford-Binet ise bu kritere göre incelenmemiştir.

İncelenen Ölçeklerin Kontrol Listesindeki Güvenirlik Standartlarını Karşılama Düzeyine İlişkin Bulgular

APA, AERA ve NCME'nın standartlarına dayalı olarak oluşturulan güvenirlik kontrol listesi toplamda 28 kriter içermektedir. Yapılan incelemeler sonucunda puanlayıcılar arası Fleiss Kappa değeri Gesell Gelişim Ölçeği'nde 0,53; Leiter Uluslararası Performans Ölçeği'nde 0,59 ve WISC-R'da 0,03 bulunmuştur. Stanford-Binet Zekâ Ölçeği'nde ilgili versiyonunda güvenirlik çalışması yapılmadığı için puanlayıcılar tarafından ortak alınan kararla ölçek hiçbir kriterle uyumlu değildir şeklinde kabul edilmiştir. Bu sebeple Fleiss Kappa analizi yapılmamıştır. Puanlayıcılar arası uyumun Leiter Uluslararası Performans Ölçeği'nde ve Gesell Gelişim Ölçeği'nde yüksek düzeyde; WISC-R'da ise önemsiz düzeyde olduğu görülmektedir.

İncelemeye alınan ölçeklerin güvenirlik kriterlerini karşılama düzeyine ilişkin bilgiler Tablo 4'de verilmiştir.

Tablo 4

Ölçeklerin Güvenirlik Kriterlerini Karşılama Düzeyine İlişkin Sayılar

	WISC-R	Stanford-Binet	Gesell	Leiter
Tamamen Uyumlu	8	0	3	2
Kısmen Uyumlu	0	0	1	2
Hiç Uyumlu Değil	1	28	7	4

Tablo 4'de de belirtildiği üzere WISC-R'ın incelendiği 9 kriterin 8'i ile tamamen uyumlu, 1'i ile de hiç uyumlu değil şeklinde kabul edilmiştir. Gesell Gelişim Ölçeği'nin ise incelendiği 11 kriterin 3'ü ile tamamen uyumlu olduğu, 1'i ile kısmen uyumlu olduğu ve 7'si ile de hiç uyumlu olmadığı görülmüştür. Leiter Uluslararası Zekâ Ölçeği'nin ise incelendiği 8 kriterin 2'si ile tamamen, 2'si ile de kısmen uyum gösterdiği, 4'ü ile de hiç uyum göstermediği görülmüştür. Stanford- Binet Zekâ Ölçeği'nin kullanılan versiyonunun güvenirlik çalışması yapılmamış olup 1937'de oluşturulan versiyonunun güvenirliğinin kullanıldığı belirtilmiştir. Böyle bir uygulama güvenirlik ilkelerine aykırı olduğu için puanlayıcılar tarafından ortak alınan kararla Stanford-Binet güvenirlik kriterlerinin hepsinde kriterle hiç uyumlu değildir şeklinde kabul edilmiştir.

İncelemeye alınan ölçeklerin 3 puanlayıcı tarafından güvenirlik kriterlerini karşılama düzeyi Tablo 5'de verilmiştir. Tablo 5'in devamında ise ölçekler bu her bir kriteri karşılamaları açısından incelenmiştir.

Tablo 5

Ölçeklerin Güvenirlik Kriterlerine Göre Uyum Durumları

	Kriterler	WISC-R	S-Binet	Gesell	Leiter
1	Her amaçlanan puan kullanımının yorumlanmasına uygun güvenilirlik kanıtı sağlanmıştır.	2	0	0	1
2	Güvenilirliğin değerlendirildiği tekrarlama aralığı, bu aralığın seçilme gerekçesiyle birlikte açıkça belirtilmiştir.	X	0	0	X
3	Puanların güvenirliği için sağlanan kanıtlar, ölçek prosedürleriyle ilgili tekrarlar ve ölçek skorlarının kullanımı için planlanmış yorumlarla tutarlıdır	X	0	0	X

4	Her bir toplam puan, alt puan veya yorumlanacak puanların kombinasyonu için, uygun güvenilirlik indekslerinin kestirimleri rapor edilmiştir.	2	0	2	X
6	Güvenilirlik tahmin prosedürleri ölçeğin yapısı ile tutarlıdır.	2	0	1	2
7	Bir çeşit değişkenliği ele alan bir güvenilirlik veya genellenebilirlik katsayısı (veya standart hata), ölçüm hatası tanımları eşdeğer olarak kabul edilemezse, diğer değişkenlik türlerini ele alan indekslerle değiştirilebilir olarak yorumlanmamıştır.	2	0	2	2
12	Ölçeğin önerildiği her ilgili alt grup için mümkün olan en kısa sürede (2 yıl içinde) güvenilirlik kestirimleri sağlamıştır.	2	0	0	0
13	Bir ölçeğin birkaç sınıf seviyesinde veya belli yaş aralığında kullanılması öneriliyorsa ve her sınıf seviyesi veya her yaş aralığı için ayrı normlar sağlanmışsa, her yaş veya sınıf seviyesi alt grubu için güvenilirlik verileri sağlanmıştır.	2	0	0	0
14	Hem genel hem de koşullu (eğer varsa) standart ölçüm hatası, rapor edilen her puanın birimlerinde sağlanmıştır.	2	0	0	0
15A	Uygun olduğunda ve standart hatanın puan seviyeleri arasında sabit olduğuna dair kanıt olmadığı sürece, koşullu standart ölçüm hataları birkaç puan seviyesinde rapor edilmelidir.	0	0	0	0
20A	Puanların güvenilirliğini ölçmekte kullanılan her yöntem açıkça tanımlanmış ve yönteme uygun istatistiklerle ifade edilmiştir.	2	0	2	1

Kontrol listesindeki 1. kriter tüm amaçlanan puan kullanımının yorumlaması için kanıtı sağlanması gerektiğini belirtmektedir. İlgili kriterle WISC-R'ın tamamen, Leiter'in kısmen uyumlu olduğu, Stanford-Binet ve Gesell'in de hiç uyumlu olmadığı görülmüştür.

Listedeki 2. kriter ise güvenilirlik çalışmasının tekrarlama aralığı ve aralığın seçilme gerekçesinin belirtilmesi gerektiğini belirtmektedir. Bu tekrarlamalarda farklı koşullar kabul edilebilir düzeyde tutulduğu sürece ölçek uygulanabilmektedir. Bu sebeple farklı olması

muhtemel koşulların sınırları belirlenmelidir. İlgili kriterle Gesell ve Stanford-Binet'in hiç uyumlu olmadığı görülmüştür. WISC-R ve Leiter ise bu kritere göre incelenmemiştir.

Listenin 3. kriteri ise güvenirlik kanıtlarını, prosedür tekrarları ve ölçek puanlarının kullanımın yorumlarıyla tutarlı olması gerektiğini belirtmektedir. Özellikle birden çok uygulamaya dayalı çalışmalarda farklı hata kaynakları tek katsayı, ayrı ayrı ya da standart hata olarak raporlanıp kaynakları belirtilmelidir. İlgili kriterle Gesell ve Stanford-Binet'in hiç uyumlu olmadığı görülmüştür. WISC-R ve Leiter ise bu kritere göre incelenmemiştir.

Kontrol listesindeki 4. kriter ise her toplam puan, alt puan veya puan kombinasyonları için güvenilirlik indeks kestirimlerinin rapor edilmesi gerektiğini vurgulamaktadır. Ölçekte alttest ya da madde birleşimlerinin puanları yorumlanıyorsa güvenirlik kestirimleri bu puanlar için de raporlanmalıdır. İlgili kriterle WISC-R ve Gesell'in tamamen uyumlu olduğu; Stanford-Binet'in ise hiç uyumlu olmadığı kabul edilmiştir. Leiter ise bu kritere göre incelenmemiştir.

Kontrol listesindeki 6. kriter kullanılan güvenilirlik analizlerinin ölçek yapısı ile tutarlı olması gerektiğini belirtmektedir. İç tutarlılık katsayısı iki yarı yöntemi kullanılarak hesaplanıyorsa iki yarının istatistiksel olarak ve içerik yönünden karşılaştırılabilir olduğu kanıtlanmalıdır. İlgili kriterle WISC-R ve Leiter'in tamamen uyumlu olduğu, Gesell'in kısmen uyumlu olduğu; Stanford-Binet'in ise hiç uyumlu olmadığı görülmüştür.

Listenin sonraki kriteri, bir çeşit değişkenliğe yönelik olan güvenilirlik veya genellenebilirlik katsayısının diğer değişkenlik türlerinin indeksleriyle değiştirilebilir olarak yorumlanmaması gerektiğini belirten 7. kriterdir. İç tutarlılık, test- tekrar test katsayısı ve alternatif formların üçü de farklı ölçme hatası kaynaklarına sahip olduğu için aynı kabul edilmemelidirler. Bu süreçte belirtilen ölçme hatası kaynaklarını ve güvenirlik ya da genellenebilirlik katsayılarını raporlanmalıdır. İlgili kriterle WISC-R, Leiter ve Gesell'in tamamen uyumlu olduğu; Stanford-Binet'in ise hiç uyumlu olmadığı görülmüştür.

Listedeki 12. kriter ölçeğin her alt grup için mümkün olan en kısa sürede güvenilirlik kestirimlerinin açıklanması gerektiğini belirtmektedir. Kriterde belirtilen en kısa süre, araştırma kapsamında 2 yıl olarak belirlenmiştir. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell, Stanford-Binet ve Leiter'in ise hiç uyumlu olmadığı görülmüştür.

Listedeki 13. kriter ilgili tüm sınıf seviyesi ya da yaş aralığı için farklı normlar sağlanmışsa, tüm alt gruplar için güvenilirlik verilerinin sağlanması gerektiğini belirtmektedir. İlgili

kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell, Stanford-Binet ve Leiter'in ise hiç uyumlu olmadığı görülmüştür.

Kontrol listesindeki 14. kriter, genel ve koşullu standart ölçüm hatasının tüm puan birimlerinde hesaplanması gerektiğini açıklamaktadır. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell, Stanford-Binet ve Leiter'in ise hiç uyumlu olmadığı görülmüştür.

Kontrol listesindeki 15A kriteri koşullu standart ölçüm hatalarının birkaç puan seviyesinde rapor edilmesinin gerekliliğini vurgulamaktadır. İnceleme sonucunda ilgili kriterle Leiter, WISC-R, Stanford-Binet ve Gesell'in hiç uyumlu olmadığı görülmüştür.

Listedeki 20A kriteri puanların güvenilirliğinin ölçümünde kullanılan tüm yöntemlerin tanımlamasının ve uygun istatistiklerle açıklanmasının gerekliliğini açıklamaktadır. Bu süreçte; ortalama, standart sapma, ölçeğin uygulandığı grupların demografik özellikleri, veri toplama yöntemi ve örneklem büyüklükleri gibi prosedürlerin raporlanması gerekmektedir. İlgili kriterle WISC-R ile Gesell'in tamamen uyumlu olduğu, Leiter'in kısmen uyumlu olduğu, Stanford-Binet'in ise hiç uyumlu olmadığı kabul edilmiştir.

İncelenen Ölçeklerin Kontrol Listesindeki Ölçek Uyarlama Standartlarını Karşılama Düzeyine İlişkili Bulgular

Uluslararası Test Komisyonu'nun standartlarına dayalı olarak oluşturulan ölçek uyarlama kontrol listesi toplamda 12 kriter içermektedir.

Yapılan incelemeler sonucunda puanlayıcılar arası Fleiss Kappa değeri WISC-R'da 0,53 bulunmuştur. Puanlayıcılar arası uyumun yüksek düzeyde olduğu görülmektedir. Leiter Uluslararası Zekâ Ölçeği ve Gesell Gelişim Ölçeği'nin alanyazında herhangi bir çeviri ya da uyarlaması bulunmamaktadır. Puanlayıcıların ortak kararıyla 2 ölçek de hiçbir kriterle uyumlu değildir şeklinde kabul edildiği için bu ölçek değerlendirmesi için Fleiss Kappa analizleri yapılmamıştır. Öner (2008) Stanford-Binet Zekâ Ölçeği için ölçeğin Şemin tarafından çevirisinin yapıldığını; fakat uyarlamasının yapılmadığı belirtilmektedir. Bu çeviri süreci yazılı bir kaynakta bulunmadığı için test ölçek uyarlama adımlarında göre incelenmemiştir.

İncelemeye alınan ölçeklerin ölçek uyarlama kriterlerini karşılama düzeyine ilişkin bilgiler Tablo 6'da verilmiştir.

Tablo 6

Ölçeklerin Ölçek Uyarlama Kriterlerini Karşılama Düzeyine İlişkin Sayılar

	WISC-R	Stanford-Binet	Gesell	Leiter
Tamamen Uyumlu	5	*	0	0
Kısmen Uyumlu	2	*	0	0
Hiç Uyumlu Değil	4	*	12	12

* İncelemesi yapılmamıştır

Yukarıda verilen Tablo 6'da da görüldüğü üzere WISC-R'ın incelendiği 11 ölçek uyarlama kriterinin 5'i ile tamamen, 2'si ile kısmen uyum gösterdiği; 4'ü ile ise hiç uyum göstermediği kabul edilmiştir. Leiter Uluslararası Zekâ Ölçeği ve Gesell Gelişim Ölçeği'nde alanyazında herhangi bir çeviri ya da uyarlama kaynağı bulunmadığı için 2 ölçekte tüm kriterler kriterle hiç uyumlu değildir şeklinde değerlendirilmiştir.

Stanford-Binet Zekâ Ölçeği'nin Şemin tarafından çevirisinin yapıldığı belirtilmektedir; fakat bu çeviri süreci yazılı bir kaynakta bulunmadığı için test uyarlama kriterlerine göre incelemesi yapılamamıştır.

İncelemeye alınan ölçeklerin 3 puanlayıcı tarafından ölçek uyarlama kriterlerini karşılama düzeyi Tablo 7'de verilmiştir. Tablo 7'in devamında ise ölçekler bu her bir kriteri karşılama durumları açısından incelenmiştir.

Tablo 7

Ölçeklerin Ölçek Uyarlama Kriterlerine Göre Uyum Durumları

	Kriterler	WISC-R	S-Binet	Gesell	Leiter
1	Çalışmanın ana amaçları için önemli olmayan kültürel farklılıkların etkileri, mümkün olduğu kadar en aza indirilmiştir.	2	*	0	0
2	İlgili evrenlerde ölçek ile ölçülen yapıdaki örtüşme miktarı değerlendirilmiştir.	1	*	0	0
3	Uyarlama sürecinin, amaçlanan evrende dilsel ve kültürel farklılıkları tam olarak dikkate alması sağlanmıştır.	1	*	0	0
4	Ölçek talimatlarında, rubriklerde ve maddelerde kullanılan dilin ölçeğin kültürel ve	2	*	0	0

	dil açısından hedef kişiler için uygun olduğuna dair kanıt sağlanmıştır.				
5	Ölçek tekniklerinin seçimi, madde formatları, ölçek kuralları ve diğer prosedürlerin amaçlanan tüm evrenlere uygun olduğuna dair kanıt sağlanmıştır.	2	*	0	0
6	Madde içeriğinin ve uyarıcı malzemelerinin amaçlanan tüm evrenlere uygun olduğuna dair kanıt sağlanmıştır.	2	*	0	0
7A	Uyarlama sürecinin doğruluğunu artırmak için dilsel ve psikolojik eleştirel kanıtlar derlenmiştir.	0	*	0	0
7B	Farklı dillere olan tüm uyarlamaların eşdeğerliği ile ilgili kanıtlar derlemiştir.	0	*	0	0
8	Ölçeğin dillere uyarlanmış hallerinin denkliğini belirlemek ve ölçeğin uygulanması amaçlanan bir ya da daha fazla evrende yetersiz olabilecek problemli bileşenleri veya yönlerini tanımlamak için uygun istatistiksel teknikler uygulanmıştır.	0	*	0	0
9	Ölçeğin uyarlanmış versiyonlarının hedeflenen evrendeki geçerliğine ilişkin bilgi sağlanmıştır.	2	*	0	0
10	Hedef evrendeki madde eşdeğerliği hakkında istatistiksel kanıt sağlamıştır.	0	*	0	0
11	Hedef evrendeki eşdeğer olmayan maddeler, ölçeğin uyarlanmış versiyonlarını ortak bir puan raporlama ölçeğine “bağlamak” için kullanılmamıştır.	X	*	0	0

* İncelemesi yapılamamıştır

Kontrol listesindeki 1. kriter, araştırmanın esas amaçları ile ilgisi olmayan kültürel farklılık etkilerinin olabildiğince azaltılması gerektiğini belirtmektedir. İlgili kriterle WISC-R’ın tamamen uyumlu olduğu; Gesell ve Leiter’in ise hiç uyumlu olmadığı kabul edilmiştir.

Listedeki 2. kriter ise ölçek ile ölçülen yapıdaki örtüşme miktarının değerlendirilmesi gerektiği vurgulamaktadır. Kişiler arası kültürel ve dilsel farklılıklar farklı yorumlamalara sebep olabileceği için ölçekteki yapının uyarlaması yapılmak istenen diğer grupta da aynı biçim ve sıklıkta bulunması gerekmektedir. İlgili kriterle WISC-R’ın kısmen uyumlu olduğu; Gesell ve Leiter’in ise hiç uyumlu olmadığı kabul edilmiştir.

Kontrol listesindeki 3. kriterde uyarlama sürecinde, evrendeki dilsel ve kültürel farklılıkların dikkate alınması gerektiği belirtilmektedir. Çevirmenlerin iki dilde ve kültürde uzman olması, ölçeğin konusunda ve ölçek geliştirme ilkeleri hakkında ise bilgi ile deneyim sahibi olmaları gerekmektedir. İlgili kriterle WISC-R'ın kısmen uyumlu olduğu; Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

Listenin 4. kriterinde ölçek talimatlarındaki ve maddelerdeki dilin hedeflenen evren için uygun olmasına dikkat edilmesi gerektiği vurgulanmaktadır. Ölçeğin dilin zorluk derecesi, okunabilirliği, dil bilgisi ve noktalama işaretleri açısından karşılaştırılabilir olması gerekmektedir. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

Sonraki kriterde ise ölçek tekniklerinin seçimi, madde formatları, ölçek kuralları ve diğer prosedürlerin hedeflenen evren için uygun olduğuna dair kanıt sağlanması gerektiği belirtilmektedir. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

Kontrol listesinde 6. kriterde madde içeriğinin ve malzemelerin hedeflenen evrene uygun olduğuna dair kanıt sağlaması gerektiği açıklanmaktadır. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

Sonraki kriter olan 7A'da ise uyarlama sürecinin doğruluğunu artırmak için dilsel ve psikolojik eleştirel kanıtları derlenmesi gerektiği vurgulanmaktadır. Çeviri denkliğindeki eleştirel kanıtlar, her maddenin çeviri denkliği derecesine dayanmalıdır. İlgili kriterle Gesell, WISC-R ve Leiter'in hiç uyumlu olmadığı kabul edilmiştir.

Listedeki 7B kriteri ise farklı dillerdeki tüm uyarlamaların eşdeğerliği ile ilgili kanıtları derlenmesi gerektiğini açıklamaktadır. İlgili kriterle WISC-R, Gesell ve Leiter'in hiç uyumlu olmadığı kabul edilmiştir.

Kontrol listesindeki 8. kriter ölçeğin uyarlanmış hallerinin denkliğini belirlemek ve uyarlanacağı evrendeki olası problemli yönlerini tanımlamak için uygun istatistiksel teknikleri uygulaması gerektiğini vurgulamaktadır. İlgili kriterle WISC-R, Gesell ve Leiter'in hiç uyumlu olmadığı kabul edilmiştir.

Listedeki 9. kriter ise ölçeğin uyarlanmış hallerinin amaçlanan evrendeki geçerliği ile ilgili bilgi sağlanması gerektiğini belirtmektedir. İlgili kriterle WISC-R'ın tamamen uyumlu olduğu; Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

Listenin 10. kriteri hedef evrende madde eşdeğerliği hakkında istatistiksel kanıt sağlanması gerektiğini vurgulamaktadır. İlgili kriterle WISC-R, Gesell ve Leiter'in hiç uyumlu olmadığı kabul edilmiştir.

Kontrol listesindeki 11. kriter ise ortak puan raporlama ölçeğine "bağlamak" için hedef evrendeki eşdeğer olmayan maddelerin kullanılmaması gerektiğini vurgulamaktadır. Eşdeğer olmama sebebi maddelerin kültürel olarak uygun olmaması ise madde mevcut haliyle kullanılmamalıdır. Eğer bu maddelerin uyarlaması yetersiz ise sorular revize edilmeli veya ölçekten çıkarılmalıdırlar. İlgili kriterle Gesell ve Leiter'in ise hiç uyumlu olmadığı kabul edilmiştir.

İncelenen Ölçeklerin Kontrol Listesindeki Bulunan Standartlara Göre Geliştirilmesi Gereken Özelliklerine İlişkin Bulgular

Araştırmaya alınan ölçekler kontrol listelerine göre incelendiğinde kriterle tamamen uyumlu olanların yanında kriterle kısmen uyumlu olan ve hiç uyumlu olmayan pek çok özellik bulunmuştur. Ölçeklerle kısmen uyumlu olan ve hiç uyumlu olmayan kriterlerde bahsedilen özelliklerin geliştirilmesi önerilmektedir. Her ölçek için; geçerlik, güvenirlik ve ölçek uyarlama kriterlerinden yola çıkarak geliştirilmesi önerilen özellikleri aşağıda sunulmuştur.

Wechsler Çocuklar için Zekâ Ölçeği'nin Gözden Geçirilmiş Hali'nin Geliştirilmesi Gereken Özelliklerine İlişkili Bulgular

Geçerlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Ölçekte yeterli puan aralıkları dışında da puan kriterleri belirtilip nasıl yorumlanacağı açıklanmalıdır.
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplama koşuları detaylandırmalıdır.
- Tipik ölçek koşullarından farklı olan ve bireylerin ölçekteki performansını belirgin şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmelidir.
- Sözel beceri ve performansa dayalı beceri puanların ayrı ayrı yorumlaması ya da gerekçesi sunulmalıdır.
- Katılımcılar belirlenen puan aralıkları dışında puan alması durumuna yönelik olarak farklı kriter aralıkları ve yorumlaması tanımlanmalıdır.

Güvenirlik kriterlerinden yola çıkarak geliştirilmesi gereken özellik aşağıda sunulmuştur:

- Koşullu standart hata birkaç puan seviyesinde raporlanmalıdır.

Ölçek Uyarlama açısından ise geliştirilmesi önerilen kriterler aşağıda sunulmuştur:

- İlgili evrenlerde ölçek ile ölçülen yapıdaki örtüşme miktarı değerlendirilmelidir.
- Uyarlama sürecinde amaçlanan evrende dilsel ve kültürel farklılıkları tam olarak dikkate alınmalıdır.
- Dilsel ve psikolojik eleştirel kanıtlar derlenmelidir.
- Farklı dillere olan tüm uyarlamaların eşdeğerliği ile ilgili kanıtlar derlenmelidir.
- Ölçeğin uyarlanmış hallerinin denkliğini belirlemek ve evrende yetersiz olabilecek problemli bileşenleri tanımlamak için uygun istatistiksel teknikler uygulanmalıdır.
- Hedef evrendeki madde eşdeğerliği hakkında istatistiksel kanıt sağlanmalıdır.

Stanford–Binet Zekâ Ölçeği'nin Geliştirilmesi Gereken Özelliklerine İlişkin Bulgular

Geçerlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Belirlenen standart puanın yorumlamasının nasıl yapılacağı açıklanmalıdır.
- Ölçekte değerlendirilmesi hedeflenen yapı net olarak belirtilmelidir.
- Puanların yorumlanması belli kanıt ya da teoriye dayandırılmalıdır.
- Puanların her kullanım amacı için geçerlikleri değerlendirilmediyse ya da sonuçlarla tutarlı değilse bu durum açıklanmalıdır.
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplama koşuları detaylandırmalıdır.
- Tipik ölçek koşullarından farklı olması olası ve ölçek performansını önemli şekilde etkileyebilecek veri toplama özellikleri de açıklanmalıdır.

Ölçeğin RAM'larda kullanılan versiyonunun güvenirlik çalışması yapılmayıp eski versiyonunun güvenirliğinin devam ettiği belirtilmiştir; fakat ölçek standartları gereği böyle bir durum mümkün değildir. Bu sebeple ölçeğin güvenirliği tekrar yapılması ve bu süreçte ölçek güvenirlik standartlarının göz önünde bulundurulması önerilmektedir.

Ölçeğin Şemin tarafından çevirisinin belirtilmektedir; fakat bu çeviri süreci yazılı bir kaynakta bulunmadığı için ölçek uyarlama standartlara göre incelenememiştir.

Gesell Gelişim Ölçeği'nin Geliştirilmesi Gereken Özelliklerine İlişkin Bulgular

Geçerlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Aynı kişiler farklı alttestlerde farklı kategorilerde performans gösterirse ne yapılacağı belirtilerek puanlarının yorumlaması düzenlenmelidir.
- Değerlendirmeyi amaçladığı yapı ya da yapılar açıkça belirtilmelidir.
- Puanların her kullanım amacı için geçerlikleri değerlendirilmediyse ya da sonuçlarla tutarlı değilse bu durum açıklanmalıdır.
- Alttestlerinin ayırt edilebilirliği, güvenirliği, aralarındaki ilişki ya da yapıyla olan ilişkileri raporlanmalıdır.

Güvenirlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Her amaçlanan puan kullanımının yorumlanmasında uygun güvenilirlik kanıtı sağlanmalıdır.
- Güvenilirlik / hassasiyetin değerlendirildiği tekrarlama aralığı, bu aralığın seçilme gerekçesiyle birlikte açıkça belirtilmelidir.
- Puanların güvenirliği / hassasiyet için sağlanan kanıtlar, ölçek prosedürleriyle ilgili tekrarlar ve ölçek skorlarının kullanımı için planlanmış yorumlarla tutarlı olmalıdır.
- Güvenilirlik tahmin prosedürleri ölçeğin yapısı ile tutarlı olmalıdır.
- Ölçek yayıncıları, ölçeğin önerildiği her ilgili alt grup için 2 yıl içinde güvenilirlik kestirimleri sağlanmalıdır.
- Her yaş seviyesi için güvenilirlik / hassasiyet verileri sağlanmalıdır.
- Genel standart ölçüm hatası, rapor edilen her puanın birimlerinde sağlanmalıdır.
- Koşullu standart ölçüm hataları birkaç puan seviyesinde rapor edilmelidir.

Gesell'in dil ve kültürümüze uyarlaması yapılmamıştır. Bütün ölçek uyarlama adımlarına uyularak bir uyarlaması yapılması önerilmektedir.

Leiter Uluslararası Performans Ölçeği'nin Geliştirilmesi Gereken Özelliklerine İlişkin Bulgular

Geçerlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Ölçekte yeterli puan aralıkları dışında da puan kriterleri belirtilip nasıl yorumlanacağı açıklanmalıdır.

- Ölçekte değerlendirilmesi hedeflenen yapı net olarak belirtilmelidir.
- Geçerlik uygulamasındaki katılımcıların yaş ve ırklarının yanı sıra diğer sosyo-demografik ve gelişimsel temel özellikleri de izin verilen ölçüde açıklanmalıdır.
- Geçerlikte motivasyon düzeyi, ölçek puan aralıkları, ölçeğin uygulama şekli gibi verilerin toplama koşuları detaylandırmalıdır.
- Tipik ölçek koşullarından farklı olan ve bireylerin ölçekteki performansını belirgin şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmelidir.
- Alt bölüm puanlarının ayırt edilebilirliği, güvenirliği ve alttest puanları arası ilişkinin yapı ile tutarlı olduğu kanıtlanmalıdır.
- Katılımcılar belirlenen puan aralıkları dışında puan alması durumuna yönelik olarak farklı kriter aralıkları ve yorumlaması tanımlanmalıdır.

Güvenirlik kriterlerinden yola çıkarak geliştirilmesi gereken özellikler aşağıda sunulmuştur:

- Her puan kullanımının yorumlanmasında uygun güvenilirlik kanıtı sağlanmalıdır.
- Alt grup için güvenilirlik / hassasiyet kestirimleri en kısa sürede yapılmalıdır.
- Leiter'de alt gruplar farklı yaş gruplarındaki katılımcılardan oluştuğu için güvenirlik katsayıları bu alt gruplarda raporlanmalıdır.
- Standart ölçüm hatasının her puanın birimlerinde raporlanmalıdır.
- Koşullu standart ölçüm hatası da birkaç puan seviyesinde belirtilmelidir.
- Puanların güvenilirliğini ölçmek için kullanılmış olan iki yarı yöntemi açıkça tanımlanmalıdır.

Türkçe'ye çeviri ya da uyarlaması yapılmadığı için bütün ölçek uyarlama adımlarına uyularak bir uyarlaması yapılması önerilmektedir.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde kontrol listesine göre yapılan inceleme sonuçları alanyazınla ilişkilendirilerek yorumlanmıştır. Sonrasında ise sonuçlar doğrultusunda yapılan önerilere yer verilmiştir. İncelenen ölçekler doğrultusunda elde edilen sonuçlar ve tartışma aşağıda sırasıyla sunulmuştur.

Wechsler Çocuklar İçin Zekâ Ölçeği'nin Standartları Karşılama Düzeyine İlişkin Sonuç ve Tartışma

Wechsler Çocuklar İçin Zekâ Ölçeği'nin Gözden Geçirilmiş Hali'nin incelendiği 9 geçerlik kriterinin 4 tanesi ile tamamen uyumlu, 2 tanesi ile kısmen uyumlu olduğu; 3'ü ile de hiç uyumlu olmadığı görülmüştür.

Ölçeğin tamamen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçeğin uygulanacağı evren ve yapı açıkça tanımlanmıştır.
- Puanların yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur.
- Geçerlik kanıtı elde edilen katılımcı örneklemleri ilgili sosyodemografik ve gelişimsel temel özellikleri içerecek şekilde ayrıntılı olarak tanımlanmıştır.
- Birleşik puan oluşturmak için temel ve gerekçeler verilmiştir.

Ölçeğin kısmen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Yeterli puan kriterleri dışında puanların nasıl yorumlanacağı açıkça belirtilmemiştir.

- Yetersiz kriter performansını seviyeleri hakkında bilgi verilmemiştir.

Ölçeğin hiç uyumlu olmadığı görülen geçerlik kriterleri aşağıda sunulmuştur:

- Verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmamıştır.
- Tipik ölçek koşullarından farklı ve sonuçları önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmemiştir.
- Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında, bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmamıştır.

Wechsler Çocuklar İçin Zekâ Ölçeği'nin Gözden Geçirilmiş Hali'nin incelendiği 9 güvenirlik kriterinin 8'i ile tamamen uyumlu, 1'i ile ise tamamen uyumsuz kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen güvenirlik kriterleri aşağıda sunulmuştur:

- Amaçlanan puan kullanımı yorumlamasına uygun güvenilirlik kanıtı sağlanmıştır.
- Her toplam puan, alt puan veya yorumlanacak puanların kombinasyonu için, uygun güvenilirlik indekslerinin kestirimleri rapor edilmiştir.
- Güvenirlik tahmin prosedürleri ölçeğin yapısı ile tutarlıdır.
- Bir çeşit değişkenliği ele alan bir güvenilirlik veya genellenebilirlik katsayısı diğer değişkenlik türlerine yönelik indekslerle değiştirilebilir olarak yorumlanmamıştır.
- Ölçeğin kullanıldığı her alt grubun 2 yıl içinde güvenilirlik kestirimleri sağlamıştır.
- Her yaş seviyesi için güvenilirlik verileri sağlanmıştır.
- Genel ve koşullu standart ölçüm hatası her puan biriminde raporlanmıştır.
- Puanların güvenilirliğini ölçmek için kullanılan her yöntem açıkça tanımlanmış ve yönteme uygun istatistiklerle ifade edilmiştir.

Ölçeğin hiç uyumlu olmadığı görülen güvenirlik kriteri aşağıda sunulmuştur:

- Koşullu standart ölçüm hataları birkaç puan seviyesinde rapor edilmemiştir.

Tittle (1975) çalışmasında WISC-R'ın el kitabında düzen ve yazımda gelişmeler olduğunu, alttestlerin uygulama talimatlarının iyileştirildiğini ve puanlama sisteminin netleştirildiğini raporlamıştır. Norm çalışmasında katılımcılara ilişkin pek çok temel özelliğin belirtilmesine karşın ebeveynlerin eğitimi ya da geliri gibi sosyoekonomik seviyeyi belirten bir özelliğin açıklanmadığını vurgulamıştır. Ayrıca, norm grubunda ırkı beyaz olmayan katılımcıların yüzdesinin az olması eleştirmiştir. Standart hatanın ise her alttestte belirtildiği raporlanmıştır. Faktör analizi yapılmadığı için ölçeğin en büyük zayıflığının geçerlik

açısından eksikliği oluğu raporlamıştır. El kitabında ölçek sonuçlarının zihinsel gerilik, öğrenme güçlüğü gibi psikolojik tanılama amacıyla kullanılabileceği belirtilmesine rağmen, el kitabının bu konuda ayrıntılı bir araştırma içermediği belirtmiştir.

Flanagan ve Kaufman (2004) ise WISC-R'ın önceki versiyona göre; yeni bir norm çalışması olduğunu, grafiklerin düzenlendiğini, madde içeriklerinin yenilendiğini, yanlı maddelerin ölçekten çıkarıldığını, puanlama ve uygulama talimatlarını geliştirildiğini raporlamışlardır. WISC-III'te ise WISC-R'a göre norm çalışmasının daha temsili olduğunu, farklı indekslerin eklendiğini, yeni alt test eklendiğini, puanlama ve uygulama talimatlarının geliştirildiğini belirtmişlerdir.

Wechsler Çocuklar İçin Zekâ Ölçeği'nin Gözden Geçirilmiş Hali'nin incelendiği 11 ölçek uyarlama kriterinin 5'i ile tamamen uyumlu olduğu, 2'si ile kısmen uyumlu olduğu ve 4'ü ile hiç uyumlu olmadığı görülmüştür. Ölçeğin tamamen uyumlu olduğu görülen ölçek uyarlama kriterleri aşağıda sunulmuştur:

- Çalışmanın esas amaçları ile ilgisi olmayan kültürel farklılıkların etkileri olabildiğince azaltılmıştır.
- Ölçek talimatlarındaki ve maddelerdeki dilin hedeflenen evren için uygun olmasına dikkat edilmiştir.
- Ölçek tekniklerinin seçimi, madde formatları, ölçek kuralları ve diğer prosedürlerin hedeflenen evrenlere uygun olduğuna dair kanıt sağlanmıştır.
- Madde içeriğinin ve uyarıcı malzemelerinin hedeflenen evrenlere uygun olduğuna dair kanıt sağlanmıştır.
- Ölçeğin uyarlanmış versiyonlarının hedeflenen evrendeki geçerliği hakkında bilgi sağlanmıştır.

Ölçeğin kısmen uyumlu olduğu görülen ölçek uyarlama kriterleri aşağıda sunulmuştur:

- Ölçek ile ölçülen yapıdaki örtüşme miktarı kısmen değerlendirilmiştir.
- Uyarlama sürecinin, amaçlanan evrende dilsel ve kültürel farklılıkları kısmen dikkate alınmıştır.

Ölçeğin hiç uyumlu olmadığı görülen güvenirlik kriterleri aşağıda sunulmuştur:

- Uyarlama sürecinin doğruluğunu artırmak için dilsel ve psikolojik eleştirel kanıtları derlenmemiştir.
- Farklı dillere olan tüm uyarlamaların eşdeğerliği ile ilgili kanıtlar derlenmemiştir.

- Ölçeğin uyarlanmış hallerinin denkliğini belirlemek ve uyarlanacağı evrendeki olası problemli yönlerini tanımlamak için uygun istatistiksel teknikleri uygulanmamıştır.
- Hedef evrendeki madde eşdeğerliği hakkında istatistiksel kanıt sağlanmamıştır.

Dirlik (2013) çalışmasında bu araştırma ile benzer şekilde orijinal form ile hedef dildeki form arasındaki kültürel farklılıkların etkileri olabildiğince azaltıldığını; ölçeğin uygulanacağı evrenin kültürel ve dil yapısına uygun yönerge, puanlama yönergesi ve madde oluşturulduğunu; tekniklerinin seçimi, madde formatı, test düzeni ve diğer işlemlerin hedef evrene uygun olduğunu ve geçerlik çalışması yapıldığını raporlamıştır. Bu araştırmada kısmen uyumlu olarak değerlendirilmesinin aksine Dirlik ölçeğin uyarlama sürecinde diller arası eşitlik sağlanması için dilsel ve psikolojik açıdan sistematik yol izlediğini ve evrendeki ölçülen özelliğin örtüşme düzeyinin belirtildiğini belirtmiştir. Ayrıca, uyarlanan madde ile orijinal maddelerin denk olmasının sağlandığı da raporlanmasına rağmen ilgili madde bu araştırmada hiç uyumlu değil şeklinde değerlendirilmiştir. Bu çalışmada incelenmemiş olmasına rağmen Dirlik'in çalışmasında; uygulama sürecindeki çevresel faktörlerin iki grupta da denk olmasının sağlandığı, hedef evrendeki bireylerin performansları etkileyebilecek sosyokültürel ve ekolojik özelliklerin sonuçlarının yorumlanmasındaki etkilerin belirlendiğini, örneklemler arası puan farklarının ampirik kanıtlarla anlamlılığının kanıtlandığı ve el kitabında ölçek uygulama sürecinde uyulması gereken kesin kurallar bulunduğu raporlanmıştır.

Stanford-Binet Zekâ Ölçeği'nin Standartları Karşılama Düzeyine İlişkin Sonuç ve Tartışma

Stanford-Binet Zekâ Ölçeği'nin incelendiği 7 geçerlik kriterinin 1'i ile tamamen uyumlu, 2'si tanesi ile kısmen uyumlu olduğu; 4'ü ile de hiç uyumlu olmadığı kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen geçerlik kriteri aşağıda sunulmuştur:

- Geçerlik kanıtı elde edilen örneklemlerin bilgileri ayrıntılı olarak tanımlanmıştır.

Ölçeğin kısmen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını kısmen belirtilmiştir.
- Ölçeğin yapısı açıkça tanımlanmamıştır.

Ölçeğin hiç uyumlu olmadığı görülen geçerlik kriterleri aşağıda sunulmuştur:

- Puan yorumlamasının dayandığı kanıt ve teoriler detaylı özetleriyle sunulmamıştır.
- Yorumların geçerliğinin değerlendirilmediği açıklanmamış ve kullanıcılar bu duruma yönelik uyarılmamıştır.
- Veri toplama koşullarının, istatistiksel bulguların yerel koşullarla ilişkisinin karşılaştırılabileceği şekilde detaylıca açıklanmamıştır.
- Veri toplama sürecinde tipik test koşullarından farklı ve performansı önemli şekilde etkileyebilecek durumlara dikkat edilmemiştir.

Stanford- Binet Zekâ Ölçeği'nin kullanılan versiyonunun güvenirlik çalışması yapılmamış olup 1937'de oluşturulan versiyonunun güvenirliğinin geçerli olduğu belirtilmiştir. Bu uygulama güvenirlik ilkelerine aykırı olduğu için Stanford- Binet tüm güvenirlik kriterlerinde kriterle hiç uyumlu değildir şeklinde kabul edilmiştir.

Öner (2008) Stanford- Binet Zekâ Ölçeği'nin Şemin tarafından çevirisinin yapıldığını; fakat uyarlamasının yapılmadığı belirtilmektedir. Belirtilen bu çeviri süreci yazılı bir kaynakta bulunmamaktadır. Bu sebeple Stanford- Binet Zekâ Ölçeği'nin ölçek uyarlama adımlarında göre incelemesi yapılamamıştır.

Bu araştırma ile benzer şekilde Waddell (1980) ölçekte standartlaştırılmanın yapıldığı örneklemin sosyoekonomik düzey, etnik yapı, cinsiyet ve Zekâ seviyeleri açısından yeterince detaylandırılmadığını ve anadili İngilizce olmayan katılımcıların örnekleme alınmadığını raporlamıştır. Ayrıca, ilgili versiyon için IQ puanlarının dönüşümünü içeren bir tablo olmadığı ve güvenirlik çalışmasının yapılmadığı da raporlanmıştır. Tyler-Wood vd. (2014) araştırmalarında Stanford-Binet Zekâ Ölçeği'nin 1973, 1986 ve 2003 versiyonlarını karşılaştırmışlardır. Bu karşılaştırma sonucunda 1973 versiyonundaki kültürel yanlılık, puanlamada zorluk, sonuçların yorumlanmasında yanlılık, puanlama prosedüründe alttestlerin ayrı puanlamasının olmaması, kelime alttesindeki madde türlerinin her yaş seviyesinde yeterli oranda temsil edilmemesi, puanlamada maddelerin konuları göz ardı edilerek yaş seviyelerine göre gruplanması ve ölçek bırakma prosedürleri konularını eleştirmişlerdir.

Bu çalışma sonuçlarının aksine Neale (1962) ölçeğin el kitapçığının puanlama ve uygulama talimatları açısından açık ve ayrıntılı olduğu raporlamıştır. Becker (2003) ise ölçek el kitapçığında açık puanlama talimatları içerdiğini ve ölçeğin temellendiği ayrıntılı alanyazın içerdiğini raporlamıştır.

Gesell Gelişim Ölçeği'nin Standartları Karşılama Düzeyine İlişkin Sonuç ve Tartışma

Gesell Gelişim Ölçeği'nin incelendiği 6 geçerlik kriterinin 2'si ile tamamen uyumlu, 2'si tanesi ile kısmen uyumlu olduğu; 2'si ile de hiç uyumlu olmadığı kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Geçerlik kanıtı elde edilen örneklemlerin bilgileri pratik ve izin verilen ölçüde ayrıntılı olarak tanımlanmıştır.
- Birleşik puan oluşturmak için temel ve gerekçeler verilmiştir.

Ölçeğin kısmen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçek puanlarının nasıl yorumlanacağını ve kullanılacağını kısmen belirtilmiştir.
- Ölçeğin yapısı açıkça tanımlanmamıştır.

Ölçeğin hiç uyumlu olmadığı görülen geçerlik kriterleri aşağıda sunulmuştur:

- Yorumların geçerliği değerlendirilmediği açıklanmamış ve kullanıcılar uyarılmamıştır.
- Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlaması sürecinde yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmamıştır.

Gesell Gelişim Ölçeği'nin toplamda incelendiği 11 güvenirlik kriterinin 3'ü ile tamamen uyumlu, 1'i ile kısmen uyumlu olduğu; 7'si ile de hiç uyumlu olmadığı kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen kriterler aşağıda sunulmuştur:

- Her bir toplam puan, alt puan veya yorumlanacak puanların kombinasyonu için, uygun güvenilirlik indekslerinin kestirimleri rapor edilmiştir.
- Bir çeşit değişkenliği ele alan bir güvenilirlik veya genellenebilirlik katsayısı diğer değişkenlik türlerini ele alan indekslerle değiştirilebilir olarak yorumlanmamıştır.
- Puanların güvenilirliğini ölçmek için kullanılan her yöntem açıkça tanımlanmış ve yönteme uygun istatistiklerle ifade edilmiştir.

Ölçeğin kısmen uyumlu olduğu görülen güvenirlik kriteri aşağıda sunulmuştur:

- Güvenirlik tahmin prosedürleri ölçeğin yapısı ile tutarlıdır.

Ölçeğin hiç uyumlu olmadığı görülen güvenirlik kriterleri aşağıda sunulmuştur:

- Tüm amaçlanan puan kullanımı yorumlamasında güvenilirlik kanıtı sağlanmamıştır.
- Güvenilirlik tekrarlama aralığı ve aralığın seçilme gerekçesi belirtilmemiştir.

- Puanların güvenirlik kanıtları, uygulama prosedürleri ve puanların yorumlarıyla tutarlı değildir.
- Ölçeğin önerildiği alt gruplar için güvenilirlik kestirimleri sağlanmamıştır.
- Her yaş veya sınıf seviyesi alt grubu için güvenilirlik verileri sağlanmamıştır.
- Genel standart ölçüm hatası raporlanmamıştır.
- Koşullu standart ölçüm hatası raporlanmamıştır.

Gesell Gelişim Ölçeği'nin Türkiye'deki alanyazında herhangi bir çeviri ya da uyarlaması bulunmadığı için 12 ölçek uyarlama kriteri ile hiç uyumlu değildir şeklinde kabul edilmiştir.

Bu sonuçlarla benzer şekilde Kaplan ve Saccuzzo (2013) da ölçekteki puanlama sisteminin yetersiz olduğunu, yönergelerin net olmadığını ve ölçeğin psikometrik açıdan istenilen seviyede olmadığını raporlamışlardır. Huelsman ise (1965) ölçeğin yaş ya da sınıf seviyesine göre normlarının bulunmadığını, toplam zekâ puanı belirtilmediğini, ölçeğin kullanım yararlarının açıklanmamış olduğunu, standart sapmanın belirtilmemesini, geçerlik-güvenirlik çalışmalarının yapılmamış olmasını, ölçeğin kullanım yararları hakkındaki açıklamaların yetersizliğini ve ölçekle ilgili alanyazın taramasında yazarların çoğunlukla kendi çalışmalarına atıf yaptıklarını raporlamıştır. Lichtenstein (1990) Gesell Gelişim Ölçeği sonuçlarını incelediğinde puanlayıcılar arası farklılıklar olduğunu raporlamış ve bu durumun ölçekte net bir puanlama sistemi olmamasından kaynaklandığını belirtmiştir. Shepard (1997) ise çalışmasında ölçekte test- tekrar test ve puanlayıcılar arası güvenirlik katsayısı gibi önemli psikometrik özelliklerin bulunmadığını raporlamıştır.

Leiter Uluslararası Performans Ölçeği'nin Standartları Karşılama Düzeyine İlişkin Sonuç ve Tartışma

Çalışma sonucunda Leiter Uluslararası Performans Ölçeği'nin incelendiği 9 geçerlik kriterinin 2'si ile tamamen uyumlu olduğu, 4'ü ile kısmen uyumlu olduğu; 3'ü ile de hiç uyumlu olmadığı kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçek puanlarının yorumlanmasında dayandığı kanıtlar ve teoriler ayrıntılı özetleriyle sunulmuştur.
- Birleşik puanların temel ve gerekçeleri verilmiştir.

Ölçeğin kısmen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçek yorumlamasında sadece “yeterli puan” aralığı belirtilmiştir.
- Ölçeğin uygulanacağı evren tanımlanmasına rağmen yapı açıkça tanımlanmamıştır.
- Geçerlik kanıtı elde edilen örneklemde katılımcıların yaş ve ırkları dışında sosyodemografik ve gelişimsel temel özellikleri ayrıntılı olarak tanımlanmamıştır.
- Yetersiz kriter performansını seviyeleri hakkında bilgi verilmemiştir.

Ölçeğin hiç uyumlu olmadığı görülen geçerlik kriterleri aşağıda sunulmuştur:

- Verilerin toplandığı koşullar, ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanmamıştır.
- Tipik ölçek koşullarından farklı olması olası ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmemiştir.
- Alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlaması sürecinde yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanmamıştır.

Leiter Uluslararası Performans Ölçeği toplamda incelendiği 8 güvenirlik standartının 2’si ile tamamen uyumlu, 2’si ile kısmen uyumlu olduğu; 4’ü ile de hiç uyumlu olmadığı kabul edilmiştir. Ölçeğin tamamen uyumlu olduğu görülen kriterler aşağıda sunulmuştur:

- Güvenilirlik tahmin prosedürleri ölçeğin yapısı ile tutarlıdır.
- Bir çeşit değişkenliği ele alan bir güvenilirlik veya genellenebilirlik katsayısı diğer değişkenlik türlerini ele alan indekslerle değiştirilebilir olarak yorumlanmamıştır.

Ölçeğin kısmen uyumlu olduğu görülen geçerlik kriterleri aşağıda sunulmuştur:

- Her amaçlanan puan kullanımı yorumlamasına uygun güvenilirlik kanıtı sağlanmamıştır.
- Puanların güvenilirliğini ölçmek için kullanılan her yöntem açıkça tanımlanmamış ve yönteme uygun istatistiklerle ifade edilmemiştir.

Ölçeğin hiç uyumlu olmadığı görülen geçerlik kriterleri aşağıda sunulmuştur:

- Ölçeğin önerildiği her alt grup için güvenilirlik kestirimleri sağlanmamıştır
- Her yaş aralığı için güvenilirlik verileri sağlanmamıştır.
- Genel ve koşullu standart ölçüm hatası, her puanın birimlerinde raporlanmamıştır.
- Koşullu standart ölçüm hataları birkaç puan seviyesinde rapor edilmemiştir.

Arthur (1949), ise ölçekteki norm seviyelerinin ortalama bir çocuk için çok yüksek olduğunu belirtmiştir. Her yaş seviyesi için 4 test verilmesi sonucunda bir hatanın bile toplam puan üzerinde çok etkili olabileceğine bunun da ölçeğin geçerliğini düşüreceği de raporlanmıştır.

Araştırma sonuçlarına genel olarak bakıldığında ölçeklerin geçerlik alanında eksik olan pek çok alan görülmektedir. Bunların en önemli nedenlerinden birisi olarak ölçeklerin uluslararası alanda kullanılan güncel hallerine göre revizyonlarının yapılmaması olduğu söylenebilir. Güvenirlik açısından ise ölçeklerde yine eski revizyonların kullanımından kaynaklı eksiklikler olduğu görülmektedir. Gesell ve Leiter'in Türkiye'de norm değerleri çalışmalarının yapılmamış olması da bu süreçte göze çarpmaktadır. Ölçek uyarlama açısından ise WISC-R haricinde hiçbir ölçeğin uyarlamasının yapılmadığı göze çarpmaktadır. Araştırma sonuçları; Gülgöz (1994), Pişkin (2006) ve Yurtsever (2013)'in önceki bulgularını destekler niteliktedir.

Öneriler

Bu bölümde araştırma sonucunda geliştirilen öneriler uygulamaya ve yapılacak araştırmalara yönelik olmak üzere iki başlıkta ele alınmıştır.

Uygulamaya Yönelik Öneriler

- Gesell Gelişim Ölçeği'nin güvenirliğe dair herhangi bir çalışma yapılmadığı için güvenirlik standartlarına uygun bir güvenirlik çalışması yapılmalıdır. Benzer şekilde ölçeğin dil ve kültürümüze uyarlaması da bulunmadığı için bu uyarlama çalışmasının da yapılması önerilmektedir.
- Rehberlik ve Araştırma Merkezlerinde ya da okullarda görev yapan psikolojik danışmanların, kullanılan psikolojik ölçekleri araştırma kapsamında oluşturulan kontrol listesine göre inceleyip kullanacakları ölçeklerin seçiminde bu sonuçlardan yararlanmaları önerilmektedir.
- Ölçek uyarlama sürecinde görev alacak olan kişilerin hem kontrol listesindeki maddelere dikkat etmesi hem de süreçte ölçme ve değerlendirme uzmanından yardım alması önerilmektedir.

- Stanford- Binet Zekâ Ölçeği'nin geçerlik, güvenirlik ve ölçek uyarlaması açısından önemli eksiklikleri olduğu için bu ölçek yerine geçerlik, güvenirlik ve test uyarlama kriterlerini büyük ölçüde karşılayan WISC-R'ın uygulanması önerilmektedir.
- Gesell Gelişim Ölçeği ve Leiter Uluslararası Performans Ölçeği'nin ilgili revizyonlarının uyarlamalarının yapılmadığı ve el kitaplarının bulunmadığı göz önüne alınarak ölçek seçiminde bu özellikler göz önünde bulundurulmalıdır.

Araştırmacılara Yönelik Öneriler

- WISC-R'da geçerlik açısından verilerin toplandığı koşulların ölçek kullanıcıları istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanması; tipik ölçek koşullarından farklı ve sonuçları önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmesi ve alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlamasında bu yorumlamayı destekleyen gerekçe ile ilgili kanıtlar sağlanması önerilmektedir.
- WISC-R'da güvenirlik açısından ise koşullu standart ölçüm hatalarının farklı puan seviyelerinde rapor edilmesi önerilmektedir.
- WISC-R'da test uyarlama açısından ise dilsel ve psikolojik eleştirel kanıtları derlenmesi, farklı dillere olan tüm uyarlamaların eşdeğerliği ile ilgili kanıtlar derlenmesi, ölçeğin uyarlanmış hallerinin denkliğinin belirlenmesi, uyarlanacağı evrendeki olası problemli yönlerini tanımlamak için uygun istatistiksel tekniklerinin uygulanması ve hedef evrendeki madde eşdeğerliği hakkında istatistiksel kanıt sağlanması önerilmektedir.
- Stanford – Binet Zekâ Ölçeği'nde geçerlik açısından puan yorumlamasının dayandığı kanıt ve teorilerin detaylı özetleriyle sunulması, yorumların geçerliğinin değerlendirilmediğinin açıklanması, veri toplama koşullarının istatistiksel bulguların yerel koşullarla ilişkisinin karşılaştırılabileceği şekilde detaylıca açıklanması ve veri toplama sürecinde tipik test koşullarından farklı olan performansı önemli şekilde etkileyebilecek durumlara dikkat edilmesi önerilmektedir.
- Stanford – Binet Zekâ Ölçeği'nde RAM'larda kullanılan versiyonunun güvenirlik çalışması yapılmayıp eski versiyonunun güvenirliğinin varsayıldığı belirtilmiştir;

fakat ölçek standartları gereği böyle bir durum söz konusu olamaz. Bu sebeple bütün güvenirlik standartlarının geliştirilmesi önerilmektedir.

- Stanford – Binet Zekâ Ölçeği'nin 1972 yılında Şemin tarafından çevirisi yapıldığı belirtilmektedir; fakat hem bu çeviri süreci yazılı bir kaynakta bulunmadığı için hem de uyarlamasına ilişkin bir uygulama yapıldığı belirtilmediği için ölçek uyarlama standartları göz önüne alınarak uyarlamasının tekrar yapılması önerilmektedir.
- Gesell Gelişim Ölçeği'nin geçerlik açısından; yorumların geçerliğinin değerlendirilmediğinin açıklanması ve alt bölüm puanlarının, puan farklılıkları veya profillerin yorumlaması sürecinde yorumlamayı destekleyen gerekçe ile ilgili kanıtların sağlanması önerilmektedir.
- Gesell Gelişim Ölçeği güvenirlik açısından tüm amaçlanan puan kullanımı yorumlamasında güvenilirlik kanıtı sağlanması, güvenilirlik tekrarlama aralığı ile aralığın seçilme gerekçesinin belirtilmesi, puanların güvenirlik kanıtları ile uygulama prosedürleri ve puanların yorumlarıyla tutarlı hale getirilmesi, ölçeğin önerildiği alt gruplar için güvenilirlik kestirimlerinin sağlanması, her yaş veya sınıf seviyesi alt grubu için güvenilirlik verilerinin sağlanması, genel standart ölçüm hatasının ve koşullu standart ölçüm hatasının raporlanması önerilmektedir.
- Gesell Gelişim Ölçeği'nin Türkçe'ye çevirisi ya da uyarlaması yapılmadığı için bütün ölçek uyarlama adımlarına uyularak uyarlamasının yapılması önerilmektedir.
- Leiter Uluslararası Performans Ölçeği'nin geçerlik açısından verilerin toplandığı koşullar, ölçek kullanıcılarının istatistiksel bulguların yerel koşullarla ilişkisini yargılayabilecekleri kadar ayrıntılı olarak açıklanması, tipik ölçek koşullarından farklı olması olası ve ölçek performansını önemli şekilde etkileyebilecek olan veri toplama özelliklerine dikkat edilmesi ve alt bölüm puanlarının puan farklılıkları veya profillerin yorumlaması sürecinde yorumlamayı destekleyen gerekçe ile ilgili kanıtların sağlanması önerilmektedir.
- Leiter Uluslararası Performans Ölçeği'nin güvenirlik açısından ölçeğin önerildiği her alt grup için güvenilirlik kestirimlerinin sağlanması, her yaş aralığı için güvenilirlik verilerinin sağlanması, genel ve koşullu standart ölçüm hatasının her puan birimlerinde raporlanması ve koşullu standart ölçüm hatalarının birkaç puan seviyesinde rapor edilmesi önerilmektedir.

- Leiter Uluslararası Performans Ölçeği'nin Türkçe'ye çevirisi ya da uyarlaması yapılmadığı için bütün ölçek uyarlama adımlarına uyularak uyarlamasının yapılması önerilmektedir.
- Araştırmacılar, Türkiye'de kullanılan farklı psikolojik ölçekleri araştırmada hazırlanan kontrol listesine göre inceleyebilirler.
- Araştırma kapsamında oluşturulan kontrol listesi farklı alanlar da eklenerek standartlaştırılabilir. Böylece, ölçeklerin geliştirilme ya da uyarlama açısından değerlendirilme süreçleri sistematik hale getirilebilir.

KAYNAKLAR

Acar Güvendir, M., ve Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi, 14*(52), 23-33. doi:10.17755/esosder.54872

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

Becker, K. A. (2003). History of the Stanford-Binet intelligence scales: Content and psychometrics. Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No 1. Itasca, IL: Riverside Publishing.

Boztunç Öztürk, N., Eroğlu, M. G., & Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim, 40*(178), 123-137. doi:10.15390/EB.2015.4091

Büyüköztürk, Ş., Kılıç, E., Akgün, E., Karadeniz, Ş., & Demirel, F. (2014). *Bilimsel araştırma yöntemleri*. Ankara: Pegem.

Brennan, R. L. (2009). *Generalizability theory*. New York: Springer-Verlag

Crocker, L., ve Algina, J. (1986). *Introduction to classical and modern test theory.* New York: CBS College Publishing Company.

Cronbach, L. (1990). *Essentials of Psychological Testing.* New York: Harper ve Row.

Çüm, S. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. (Yüksek Lisans Tezi). https://tez.yok.gov.tr/UlusalTezMerkezi/ sayfasından erişilmiştir.

Çüm, S., ve Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama Dergisi, 12*(24), 128-131.

Dirlik; E. (2013). *Eğitim Kurumlarında Kullanılan Psikolojik Testlerin Ölçme Standartlarına Göre İncelenmesi* (Yüksek Lisans Tezi). https://tez.yok.gov.tr/UlusalTezMerkezi/ sayfasından erişilmiştir.

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni, 13*(40), 17-25.

Evirgen, N., Kayhan, E., & Erden, G. (2015, Mayıs). *Gesell Gelişim Figürleri'nin anasınıfı çocuklarında güvenirliğine yönelik bir ön çalışma.* Uluslararası Katılımlı 3. Çocuk Gelişimi ve Eğitimi Kongresi'nde sunulan bildiri, Ankara.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: the problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543-549.

Flanagan, D, & Kaufman, A. (2004). Essentials of WISC-IV assessment, N.J.: John Wiley & Sons.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378-382. doi: 10.1037/h0031619

Gesell Institute (2020). *What is the Gesell Developmental Observation?* https://gesellinstitute.org/pages/what-is-the-gesell-developmental-observation adresinden erişilmiştir.

Gülgöz, S. (1994). Test kullanımında temel konular. *Türk Psikoloji Dergisi*, *9*(33), s. 1-8.

Hambleton, R., & Patsula, L. (1999). Increasing the validity of adapted tests: myths to be avoided and guidelines for ımproving test adaptation practices. *Journal of Applied Testing Technology*, *1*(1), 1-13.

Hambleton, R., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment.* New Jersey: Lawrence Erlbaum Associates.

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.

International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf adresinden erişilmiştir.

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. *International Journal of Academic Research* 4(1), 16-33.

Kaufman, A. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.

Koç, N. (1993). Eğitim sistemimizde ölçme ve değerlendirme alanındaki gelişmeler. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 25*(2), 387- 407. doi:10.1501/Egifak_0000000576

Koç, N. (2008). Eğitimimizde Psikolojik Testlerin (Standart Testlerin) Kullanım Durumu, Sorunlar ve Öneriler. *1. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi*, 85-103.

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. R., & Koch, G. 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.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores.* Massachusetts: Addison-Wesley.

Millî Eğitim Bakanlığı Rehberlik ve Psikolojik Danışmanlık Hizmetleri Yönetmeliği. (2017, 10 Kasım). Resmi Gazete. Sayı: 30236. https://resmigazete.gov.tr/eskiler/2017/11/20171110-2.htm adresinden erişilmiştir.

Milli Eğitim Temel Kanunu. (1973, 24 Haziran). Resmi Gazete. Sayı: 14574. https://resmigazete.gov.tr/arsiv/14574.pdf adresinden erişilmiştir.

Murphy, K., & Davidshofer, C. O. (2005). *Psychological testing: principles, and applications.* New Jersey: Pearson/Prentice Hall.

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 testlerden örnekler: bir başvuru kaynağı.* İstanbul: Boğaziçi Üniversitesi.

Özçelik, D. A. (2013). *Okullarda ölçme ve değerlendirme: öğretmen el kitabı.* Ankara: Pegem.

Özgüven, İ. E. (2014). *Psikolojik testler.* Ankara: Nobel.

Pişkin, M. (2006). Türkiye'de psikolojik danışma ve rehberlik hizmetlerinin dünü, bugünü ve yarını. M. Hesapçıoğlu & A. Durmuş (Ed.) *Türkiye'de Eğitim Bilimleri: Bir Bilanço Denemesi* içinde (s. 457–501). *Ankara: Nobel*.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297. doi: 10.1037/10403590.12.3.287

Roid, G. H., & Barram, R. A. (2004). Essentials of Stanford-Binet intelligence scales (SB5) assessment. New Jersey: John Wiley & Sons.

Roid, G. H., & Koch, C. (2017). Leiter-3: Nonverbal cognitive and neuropsychological assessment. In Handbook of nonverbal assessment (pp. 127-150). Springer, Cham.

Sattler, J. (1965). Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, form L-M. *Journal of Clinical Psychology, 21*(2), 173–179. doi:10.1002/1097-4679(196504)21:2<173::AID-JCLP2270210211>3.0.CO;2-F

Sattler, J., & Hoge, R. (2006). *Assessment of children: behavioral, social, and clinical foundations.* San Diego: Jerome M. Sattler Publisher Inc.

Savaşır, I. (1994). Ölçek uyarlamasındaki sorunlar ve bazı çözüm yolları. *Türk Psikoloji Dergisi, 9*(33), 27-32.

Savaşır, I., ve Öktem, F. (1980) WISC uyarlama çalışmaları ön rapor – II. *Türk Psikoloji Dergisi, 3*(7), 16-22.

Savaşır, I., ve Şahin, N. (1978). WISC uyarlama çalışmaları ön rapor – I. *Türk Psikoloji Dergisi, 1*(1), 33-37.

Savaşır, I., ve Şahin, N. (1995). *Wechsler çocuklar için zekâ ölçeği (WISC-R) el kitabı*. Ankara: Türk Psikologlar Derneği.

Seçer, İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci: Spss ve Lisrel uygulamaları.* Ankara: Anı.

Shepard, L. A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools, 34*(2), 85–97. doi:10.1002/(SICI)1520-6807(199704)34:2<85::AID-PITS2>3.0.CO;2-R

Şemin, R. (1972). *Zekânın değerlendirilmesi: Binet Stanford testlerinin İstanbul çocuklarına standartlanması.* İstanbul: Edebiyat Fakültesi Matbaası.

Tan, H. (1992). *Psikolojik danışma ve rehberlik.* İstanbul: Milli Eğitim Bakanlığı.

Tekin, H. (1979). *Eğitimde ölçme ve değerlendirme.* Ankara: Mars.

Tekindal, S. (2008). *Eğitimde ölçme ve değerlendirme.* Ankara: Pegem.

Tittle, C. (1975). Wechsler intelligence scales for children-revised. *Journal of Educational Measurement, 12*(2), 140-144.

Turgut, F. (1977). *Eğitimde ölçme ve değerlendirme metotları.* Ankara: Nove.

Türk Psikologlar Derneği. (2020). *Türk Psikologlar Derneği tüzüğü.* https://www.psikolog.org.tr/TPD-Tuzuk.pdf adresinden erişilmiştir.

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.

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.

Waddell, D. D. (1980). The Stanford-Binet: An evaluation of the technical data available since the 1972 restandardization. *Journal of School Psychology*, *18*(3), 203-209.

Wechsler, D. (1974). *Wechsler intelligence scale for children-revised edition*. San Antonio, TX: The Psychological Corporation.

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.

EKLER

EK 1. **APA, AERA ve NCME'ya Dayalı Geliştirilen 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 sosyodemografik 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 ulaşıp ulaşmadıklarını göstermiş ve ulaşılan anlaşma seviyesini belirtmiştir.
8D	Katılımcılar birbirleriyle etkileşime girdiyse veya bilgi alışverişinde bulunduysa, 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 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	Ranjda 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 araştırmalar, ölçek uygulamasının diğer 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 güvenirliğinin olmadığı ve ranjda daralma gibi yapı düzeltilmesi için yapılan varsayımlar sunulup bu varsayımların karşılandığı gösterilmiştir.
22	Eğer ölçek bireyleri alternatif yöntemlere yönlendirmede kullanılıyorsa ve bu yöntemlerin sonuçları ortak bir kritere göre karşılaştırılabiliyorsa, farklı sonuçların destekleyen kanıtlar sunulmalıdır.

EK 2. **APA, AERA ve NCME'ya Dayalı Geliştirilen Kontrol Listesindeki Güvenirlik Kriterleri**

1	Her bir amaçlanan puan kullanımının yorumlanmasında uygun güvenilirlik / hassasiyet kanıtı sağlanmıştır.
2	Güvenilirlik / hassasiyetin değerlendirildiği tekrarlama aralığı, bu aralığın seçilme gerekçesiyle birlikte açıkça belirtilmiştir.
3	Puanların güvenirliği / hassasiyet için sağlanan kanıtlar, ölçek prosedürleriyle ilgili tekrarlar ve ölçek skorlarının kullanımı için planlanmış yorumlarla tutarlıdır
4	Her bir toplam puan, alt puan veya yorumlanacak puanların kombinasyonu için, uygun güvenilirlik / hassasiyet indekslerinin kestirimleri rapor edilmiştir.
5	Bir ölçek puanı yorumlaması bir kişinin gözlemlenen iki puanı ya da bir grubun iki ortalamasının arasındaki farkları vurguladığında, bu gibi farklılıklar için standart hatalar da dahil olmak üzere güvenilirlik / hassasiyet verileri sağlanmıştır.
6	Güvenilirlik tahmin prosedürleri ölçeğin yapısı ile tutarlıdır.
7	Bir çeşit değişkenliği ele alan bir güvenilirlik veya genellenebilirlik katsayısı (veya standart hata), ölçüm hatası tanımları eşdeğer olarak kabul edilemezse, diğer değişkenlik türlerini ele alan indekslerle değiştirilebilir olarak yorumlanmamıştır.
8A	Ölçek puanlamasında puanlayıcılar öznel yargılar kullanıldığında, puanlayıcılar arası tutarlılık hem puanlama açısından hem de bireyler bazında puanlama şeklinde tekrarlanan ölçümlerle sunulmuştur.
8B	(i) Aynı performansları veya ürünleri puanlayan bağımsız puanlayıcı panelleri, (ii)ardışık performansları veya yeni ürünleri puanlayan tek bir panel ve (iii) ardışık performansları veya yeni ürünleri puanlayan bağımsız paneller olarak 3 kategoride güvenilirlik verileri arasında net bir ayrım yapılmıştır.
9	Açık uçlu sorular içeren ölçekler yerel olarak puanlandığında, yeterli büyüklükteki örneklem mevcut olduğunda yerel puanlama için güvenilirlik / hassasiyet verileri toplanmış ve raporlanmıştır.

10	Bir ölçeğin hem uzun hem de kısa versiyonları mevcutsa her versiyonun puanları için, tercihen her ölçeğin bağımsız uygulamalarına göre bağımsız örnekleme dayanarak, güvenilirlik / hassasiyet kanıtı rapor edilmiştir.
11	Ölçeklerde veya ölçek uygulama prosedürlerinde önemli değişikliklere izin verildiğinde; yeterli örneklem sayısı varsa, her büyük değişiklik için oluşturulan puanlar için ayrı güvenilirlik / hassasiyet analizleri sağlanmıştır.
12	Ölçeğin önerildiği her ilgili alt grup için mümkün olan en kısa sürede (2 yıl içinde) güvenilirlik / hassasiyet kestirimleri sağlamıştır.
13	Bir ölçeğin birkaç sınıf seviyesinde veya belli yaş aralığında kullanılması öneriliyorsa ve her sınıf seviyesi veya her yaş aralığı için ayrı normlar sağlanmışsa, her yaş veya sınıf seviyesi alt grubu için güvenilirlik / hassasiyet verileri sağlanmıştır.
14	Hem genel hem de koşullu (eğer varsa) standart ölçüm hatası, rapor edilen her puanın birimlerinde sağlanmıştır.
15A	Uygun olduğunda ve standart hatanın puan seviyeleri arasında sabit olduğuna dair kanıt olmadığı sürece, koşullu standart ölçüm hataları birkaç puan seviyesinde rapor edilmelidir.
15B	Seçim veya sınıflandırma için kesme puanlarının belirtildiği durumlarda, standart kesim hataları, her kesme puanının yakınında rapor edilmiştir.
16	Koşullu standart ölçüm hatalarının veya ölçek bilgi fonksiyonlarının hatalarının çeşitli alt gruplar için büyük ölçüde farklı olacağına dair inandırıcı bir kanıt olduğunda, bu farklılıkların kapsamının ve etkisinin araştırılması mümkün olan en kısa sürede (2 yıl içinde) yapılmış ve raporlanmıştır.
17	Sınıflandırma yapmak için bir ölçek veya ölçümler kombinasyonu kullanıldığında, prosedürün iki tekrarında da aynı şekilde sınıflandırılan ölçek katılımcılarının yüzdesi hakkında tahminler yapılmıştır.

18	Grupların ortalama ölçek puanları, ölçek sonuçlarının yorumunun odağını oluşturduğunda ve ölçek edilen gruplar genellikle daha geniş bir evrenden alınan bir örneklem olarak kabul edildiyse örneklemin ortalamasının standart hatası rapor edilmiştir; çünkü bu sonuç örneklem seçimi ve bireysel ölçüm hatasından kaynaklanan değişkenliği yansıtır.
19A	Ölçeğin amacı bireyler yerine grupların performansını ölçmek olduğunda, maddelerin alt kümeleri farklı inceleme alt örneklemlere rastgele atanabilir.
19B	Veriler, grup performansının bir ölçüsünü elde etmek için alt örneklem ve madde alt kümeleri arasında toplanır.
19C	Bu tür prosedürler program değerlendirmesi veya popülasyon betimlemeleri için kullanıldığında, güvenilirlik / hassasiyet analizleri örneklem seçimini dikkate alınmıştır.
20A	Puanların güvenilirliğini / hassasiyeti ölçmek için kullanılan her yöntem açıkça tanımlanmış ve yönteme uygun istatistiklerle ifade edilmiştir.
20B	Güvenilirlik / hassasiyet analizleri için kullanılan çalışmalardaki örneklem seçme prosedürleri ve bu örneklemlerdeki betimleyici istatistikler rapor edilmiştir.
21A	Güvenirlik katsayıları, değişkenliğin veya ranjda daralma için düzeltme yapılmışsa; düzeltme işlemi ile düzeltilmiş ve düzeltilmemiş katsayılar rapor edilmiştir.
21B	Gerçekte ölçülen grubun ve hedef evrenin standart sapmalarının yanı sıra düzenleme gerekçesinin de sunulmuştur.

EK 3. **ITC'ye Dayalı Geliştirilen Kontrol Listesindeki Ölçek Uyarlama Kriterleri**

1	Çalışmanın ana amaçları için önemli olmayan kültürel farklılıkların etkileri, mümkün olduğu kadar en aza indirilmiştir.
2	İlgili evrenlerde ölçek ile ölçülen yapıdaki örtüşme miktarı değerlendirilmiştir.
3	Uyarlama sürecinin, amaçlanan evrende dilsel ve kültürel farklılıkları tam olarak dikkate alması sağlanmıştır.
4	Ölçek talimatlarında, rubriklerde ve maddelerde kullanılan dilin ölçeğin kültürel ve dil açısından hedef kişiler için uygun olduğuna dair kanıt sağlanmıştır.
5	Ölçek tekniklerinin seçimi, madde formatları, ölçek kuralları ve diğer prosedürlerin amaçlanan tüm evrenlere uygun olduğuna dair kanıt sağlanmıştır.
6	Madde içeriğinin ve uyarıcı malzemelerinin amaçlanan tüm evrenlere uygun olduğuna dair kanıt sağlanmıştır.
7A	Uyarlama sürecinin doğruluğunu artırmak için dilsel ve psikolojik eleştirel kanıtlar derlenmiştir.
7B	Farklı dillere olan tüm uyarlamaların eşdeğerliği ile ilgili kanıtlar derlemiştir.
8	Ölçeğin dillere uyarlanmış hallerinin denkliğini belirlemek ve ölçeğin uygulanması amaçlanan bir ya da daha fazla evrende yetersiz olabilecek problemli bileşenleri veya yönlerini tanımlamak için uygun istatistiksel teknikler uygulanmıştır.
9	Ölçeğin uyarlanmış versiyonlarının hedeflenen evrendeki geçerliği hakkında bilgi sağlanmıştır.
10	Hedef evrendeki madde eşdeğerliği hakkında istatistiksel kanıt sağlamıştır.
11	Hedef evrendeki eşdeğer olmayan maddeler, ölçeğin uyarlanmış versiyonlarını ortak bir puan raporlama ölçeğine "bağlamak" için kullanılmamıştır.

GAZİLİ OLMAK AYRICALIKTIR…