



## Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Kategoriler Arasındaki Psikometrik Uzaklığının İncelenmesi

### Analyzing of Psychometric Distance Between Categories of Behavioral and Emotional Rating Scale-2

Güneş SALI<sup>a</sup>, Gözde SIRGANCI<sup>a</sup>

<sup>a</sup>Bozok Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Yozgat, Türkiye

#### Öz

Likert türü ölçeklerde tepki kategorileri arasındaki psikometrik uzaklığın eşit olması önemli bir varsayımdır. Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Öğretmen Değerlendirme ve Ebeveyn Değerlendirme Formunun kategoriler arası psikometrik uzaklık varsayımını sağlayıp sağlamadığı incelenmiştir. Çalışma grubunu Yozgat ilinde anasınıfı bulunan ilkokullar ile bağımsız resmi ve özel okulöncesi eğitim kurumlarına devam eden beş yaşındaki çocukların anneleri (n=346) ve öğretmenleri (n=22) oluşturmuştur. Öğretmen Değerlendirme ve Ebeveyn Değerlendirme Formunun beş alt ölçek için kategori değerleri genelleştirilmiş kısmi puanlama modeline göre kestirilmiştir. Her bir alt ölçeğin kategori değerlerinden yararlanılarak ölçek ve dönüştürülmüş ölçek değerleri hesaplanmıştır. Bulgular alt ölçeklerin, kategoriler arası psikometrik uzaklık varsayımını sağladığını göstermektedir.

#### Anahtar Kelimeler

likert ölçeği  
kategoriler arası psikometrik  
uzaklık  
genelleştirilmiş kısmi  
puanlama modeli

#### Keywords

likert scale  
psychometric distances  
between categories  
generalized partial credit  
model

#### Abstract

The most important assumption of a Likert scale is that the psychometric distances between categories is equal. In this study, Teacher Assessment Form and Parents Assessment Form of Behavioral and Emotional Rating Scale-2 (BERS-2) were investigated in terms of psychometric distances between categories. The study group composed of five-year-old children attending to nursery classes of primary schools and independent official and private preschools in Yozgat and their parents (n=346) and teachers (n=22). The estimates of scale values were obtained based on the category parameter by the generalized partial credit model. Scale values and converted scale values were calculated. As a result, both Teacher Assessment Form and Parents Assessment Form provided assumption of psychometric distance between categories.

## Extended Abstract

### Introduction

The “summated rating scale” model as developed by Rensis Likert (1932) is the most frequently used measurement tool in the measurement of psychological traits in social sciences. In Likert scale, the respondents state the degree of their agreement with a particular statement (Tavşancıl, 2010). It is assumed that an accurate measurement of any given psychological trait could be possible only when the psychological distance between the reaction categories is equal (Wakita et al., 2012). Failing to meet the foregoing assumption, would lead to the fact that the unit defined in the scale do not represent the same trait in each area of the measured trait, in other words the unequal intervals would lead to measurements that do not accurately indicate the differences between the objects and individuals. In such cases the scale would fail to indicate the magnitude of inter-individual differences (Tavşancıl, 2010). Therefore, the usability of the results obtained by this scale and the accuracy of decisions made accordingly would become questionable (Wakita et al., 2012).

Recent studies tested the assumption of psychological distance between the reaction categories (Hamby & Levine, 2016; Uyumaz, Sırganıcı & Çokluk, 2016; Wakita, Ueshima, Noguchi, 2012). Wakita et al. (2012) studied how the psychological distances changed in 4-, 5-, and 7-point versions of the neuroticism and extraversion subdomains of the Big Five Scale. They found as a result of the study that inter-category equality was disrupted in 7-point version compared to the 4- and 5-point versions. Furthermore, they suggested that the disruption in psychological distances was higher in items including negative statements. Hamby & Levine (2016) suggested that the individuals tended to less perceive the reaction categories as the number of points increased. Moreover more than required number of points led to difficulty in differentiating the inter-category differences, and less than required number of points increased the error in extreme reaction categories, suggesting that the ideal number of points was five. Testing the assumption of psychological distances is necessary for setting the appropriate type of items and number of points, and it supports the validity of the scales.

### Purpose of the Study

The present study aimed to examine whether the Teacher Rating Scale (TRS) and Parent Rating Scale (PRS) of the Behavioral and Emotional Rating Scale BERS-2 as developed by Epstein & Sharma (1998a) and adapted to Turkish culture by Salı (2017) as Davranışlar ve Duygularla İlgili Değerlendirme Ölçeği-2 met the assumption of equality of psychological distances in practice. Accordingly, an answer to the below study question was sought.

How do the inter-category psychological distances vary in the interpersonal strength, involvement with family, intrapersonal strength, school functioning, and affective strength subdomains that form the teacher rating scale and parent rating scale of BERS-2

### Method

#### Study Group

The study group of the present study was composed of the mothers (n=346) and teachers (n=22) of the five-year-old children attending to the nursery classes of Ministry of national Education primary schools and independent official and private preschool institutions in Yozgat city center.

#### Data Collection Tool

The present study utilized the Teacher Rating Scale (TRS) and Parent Rating Scale (PRS) of the Behavioral and Emotional Rating Scale BERS-2 (Epstein & Sharma, 1998a) as standardized by Epstein (2004) and adapted to Turkish culture by Salı (2017). The scale is composed of 52 items measuring behavioral and emotional strengths and 5 subdomains including the above items. The rating points in the scale are distributed in the range of “Not at all like (0)”, “Not much like (1)”, “Like (2),” and “Very Much Like (3).” The construct validity of the teacher rating scale and parent rating scale was tested by confirmatory factor analysis. Goodness of fit of the five-point model to the data set was examined by  $\chi^2$ /sd, RMSEA, CFI, and TLI goodness of fit indexes and it was concluded that the original five-point model proposed for the teacher rating scale and parent rating scale had an acceptable goodness of fit. The Cronbach’s Alpha values for the teacher rating scale and parent rating scale were calculated as .80 and above.

### Findings

#### Prediction of scale values

Scale values for each category ( $\mu_p$ ) were calculated as based on the category parameters predicted by GPCM. It was seen that the difference between the converted scales values and conventional item values in all the subdomains included in the teacher rating scale and parent rating scale varied between 0.006 and 0.161. While the deviation from the conventional item value was less than 0.01 in the Interpersonal Strength, Involvement with Family, and Intrapersonal Strength subdomains, the highest deviation was in the School Functioning sub-factor of the teacher rating scale with 0.161. Deviations from the category value were very close in each subdomain of the teacher rating scale and the parent rating scale.

### Conclusion and Discussion

The results suggested that the assumption of equal psychological distance between the categories was met in the Interpersonal Strength, Involvement with Family, Intrapersonal Strength, and School Functioning subdomains of the teacher rating scale and the parent rating scale of the Behavioral and Emotional Rating Scale-2. This was an indication of the fact that the scale accurately measured the intended psychological trait (Wakita et al., 2012). Therefore, the findings of the study provided evidence as regards the validity of the teacher rating scale and the parent rating scale of BERS-2 as adapted to the Turkish culture for five-year-old children by Salı (2017).

## 1. Giriş

Sosyal bilimlerde psikolojik özelliklerin ölçülmesinde en yaygın kullanılan ölçme aracı Rensis Likert (1932) tarafından geliştirilen “dereceleme toplamlarıyla ölçme” modelidir (Likert’den aktaran Tavşancıl, 2010). Likert ölçeği, yanıtlayıcıların verilen ifadeye ne ölçüde katılıp katılmadığını derecelere belirttikleri türde bir ölçektir (Tavşancıl, 2010). Likert ölçeğinde bireye bir cümle sunulur ve ondan, cümlenin görüşlerini yansıtıp yansıtmadığına ya da ne derece yansıttığına ilişkin bir tepki kategorisini işaretlemesi beklenir (Tezbaşaran, 2008). Likert türü ölçekler ile yapılan ölçümlerde, psikolojik özelliğin ölçülmesinde karşılaşılan birtakım sorunlar bulunmaktadır. Bunlardan biri bu tür ölçeklerin özellikle bireyin kendisi hakkında bilgi vermesi (self report) esasına dayalı olmasıdır. Burada birey, genel olarak çeşitli özellikleri bakımından kendini gözleyerek, kendisi hakkındaki gözlem sonuçlarını bildirir. Diğer bir ifade ile, bireyden hipotetik olarak ortaya konan durumlarda alacağı tavrın veya göstereceği davranışın ne olacağını belirtmesi istenir (Tezbaşaran, 2008). Bu durumda bireyin kendisi hakkındaki bilgilerin bir kısmını, tasarlanmış olsun ya da olmasın, gizleme veya çarpıtma olasılığı vardır. Dolayısıyla cevaplayıcının bildirimine dayalı olan Likert türü ölçeklerde sosyal kabul edilebilirlikten kaynaklı yanlışlığın etkisi yadsınamaz (Wakita, Ueshima, Noguchi, 2012). Bir psikolojik özelliğin tam olarak ölçülebilmesine neden olan bir diğer sorun ölçeklerde yer alan maddelerle yoklanmak istenen tipik davranışın her cevaplayıcı tarafından aynı şekilde anlaşılabilmesi ve cevaplayıcının genel davranış eğiliminin kestirilememesidir. Bu durumlarda, cevaplayıcı kendisinden beklenen tipik davranışı belirtmekte güçlük çekebilir. Bahsi geçen sorunlar, hem maddenin içeriği yansıtamamasından hem de ölçeğin uygun sayıda tepki kategorisine sahip olmamasından kaynaklanabilir. Böyle durumlarda alınan cevaplara dayalı ölçek puanlarının güvenilirliği ve geçerliği düşük olur (Tezbaşaran, 2008).

Likert ölçeğinde, bir psikolojik özelliğin tam olarak ölçülebilmesinin ancak tepki kategorileri arasındaki psikometrik uzaklığın eşit olduğu durumda gerçekleştiği varsayılmaktadır (Wakita ve diğerleri, 2012). Bu varsayımın sağlanmaması ölçekte tanımlanmış olan birim ölçülen özelliğin her bölgesinde aynı özelliği temsil etmemesi, diğer bir deyişle aralıkların eşit olmaması, obje veya bireyler arasındaki farkların miktarlarını doğru olarak gösteremeyen ölçümlerin elde edilmesine neden olur. Bu durumda, ölçeğin bireyler arası farkların miktarlarını bildirme gücü kaybolur (Tavşancıl, 2010). Böyle bir ölçekle elde edilen sonuçların kullanılabilirliği ve buna dayalı olarak verilen kararların doğruluğu sorgulanır hale gelir (Wakita ve diğerleri, 2012). Bu nedenlerden dolayı Likert türü ölçeklerde kategoriler arasındaki uzaklığın eşitliği test edilmesi gereken önemli bir varsayımdır.

Alanyazın incelendiğinde Likert ölçeğinde bireylerde var olan psikolojik yapıyı ortaya çıkaran en uygun kategori sayısının belirlenmesine yönelik çalışmalar, Garner (1960)’ın çalışmaları ile başlamıştır. Garner (1960), kategori sayısının ancak yirmiden fazla olması halinde bireylerin yanıtlarından gerekli bilginin tam anlamıyla elde edilebileceğini belirtmiştir. Green ve Rao (1970) ise altı ya da yedi kategorili ölçeklerin uygun olduğunu raporlamıştır. Tersine Schuts and Rucker (1975) kategori sayısının katılımcıların tepkilerini etkilemeyeceğini öne sürmüştür. Sonuç olarak bireylerde var olan psikolojik yapıyı ortaya koymak için gerekli kategori sayısının ne olduğuna ilişkin bir fikir birliğine varılamamıştır.

Bir grup araştırmacı da ölçme hatalarından tesadüfi hatanın etkilediği güvenilirlik ile kategori sayısı arasındaki ilişkiyi incelemiştir (Boote, 1981; Chang, 1994; Cicchetti, Showalter ve Tyrer, 1985; Lissitz and Green, 1975; Oaster, 1989; Preston ve Colman, 2000). Lissitz ve Green (1975), Boote (1981) beş kategorili ölçeklerin güvenilir olduğunu öne sürmüşlerdir. Cicchetti ve diğ. (1985) yapmış oldukları Monte Carlo simülasyon çalışması ile puanlayıcılar arası güvenilirliği incelemişler ve kategori sayısı sekize kadar arttırıldıkça güvenilirliğin arttığını ortaya koymuşlardır. Oaster (1989) yedi kategorili ölçeğin en yüksek test tekrar test güvenilirliğine sahip olduğunu göstermiştir. Preston ve Colman (2000) iki ile dört kategori arasında ölçeğin en düşük; yedi ve üstü kategori sayısında ise en yüksek test tekrar test güvenilirliğine sahip olduğunu gösterirken, iki ile on bir kategori arasında ölçeklerin ölçüt geçerliği ile kategori sayısı arasında bir ilişkinin olmadığını belirtmişlerdir. Chang (1994) ise çalışmasında aynı ölçeğin dört ve altı kategorili durumlarını karşılaştırmış ve kategori sayısındaki artışın her zaman ölçeğin güvenilirliğini arttırmadığını ileri sürmüştür. Bazı çalışmalarda ise güvenilirliğin kategori sayısından bağımsız olduğu gösterilmiştir (Brown, Wilding ve Coulter, 1991; Komorita, 1963; Matell ve Jacoby, 1971).

Preston ve Colman (2000) bir grup katılımcının kategorileri ayırma kolaylığına, yanıtlama için gerekli olan zamana ve katılımcıların görüşlerini açıklamadaki istekliliklerine yönelik görüşlerini almıştır. Bu doğrultuda 5 ile 10 kategori arasında bir derecelendirmenin katılımcılar için kolay olduğunu belirtmişlerdir. Katılımcıların görüşlerini yansıtan kategoriye seçmeleri için gerekli sürenin ise beş kategorili ölçek için yeterli, üç ve dört kategorili ölçek için oldukça yeterli olduğunu ortaya koymuşlardır. Bu bulgu çoğu ölçek için en fazla beş kategorinin yeterli olduğunu desteklemektedir.

Son yıllarda uygun kategori sayısının belirlenmesi için tepki kategorileri arasındaki psikometrik uzaklık varsayı-

mının test edildiği çalışmaların yapıldığı görülmektedir (Hamby ve Levine, 2016; Wakita, Ueshima, Noguchi, 2012; Uyumaz, Sırgancı ve Çokluk, 2016). Bu çalışmalarda uygun kategori sayısının belirlenmesi için sistematik hatanın değerlendirildiği psikometrik uzaklık çalışmalarının Cronbach alfa, sınıf içi korelasyon ve test tekrar test güvenilirliğinden daha uygun olduğu belirtilmektedir (Wakita ve diğ., 2012).

Wakita ve diğ. (2012), Büyük Beşli Kişilik Ölçeği (Big Five Scale)'nin Duygusal Dengesizlik (neuroticism) ve Dışadönüklük (extraversion) alt ölçeklerinin 4-5 ve 7 kategorili durumlarda kategoriler arası psikometrik uzaklıklarının nasıl değiştiğini incelemişlerdir. Çalışma sonucunda tepki kategori sayılarının güvenilirlik kestirimini etkilemediğini ancak madde değerlerini etkilediğini tespit etmişlerdir. Kategori sayısının 7 olduğu durumda 4 ve 5 kategorili duruma göre kategoriler arası eşitliğin bozulduğunu ortaya koymuşlardır. Ayrıca psikometrik uzaklıklardaki bozulmanın olumsuz maddelerde daha fazla olduğu belirtmişlerdir. Özellikle olumsuz maddelerde kullanılan kelimelerin dikkatli bir şekilde belirlenmesi gerektiğini ve sosyal kabul edilebilirliğin yanıtlama davranışına etki ettiğini göstermişlerdir. Hamby ve Levine (2016) kategori sayısı arttıkça bireylerin tepki kategorilerini daha düşük düzeyde algıladıklarını ileri sürmüştür. Kategori sayısının gereğinden fazla oluşunun kategoriler arası ayrımı zorlaştırdığını, gereğinden az oluşunun ise uç tepki kategorilerinde hatayı arttırdığını ortaya koymuşlar ve ideal kategori sayısının beş olduğunu belirtmişlerdir.

Uyumaz ve diğ. (2016), ölçeğin kategori sayısındaki ve bireylerin yetenek dağılımındaki farklılıkların, güvenilirlik kestirimlerini ve kategoriler arası psikometrik uzaklıkları nasıl etkilediğini incelemişlerdir. Çalışmada kategori sayısı arttıkça kategori değerinden sapma miktarının da arttığını görmüşlerdir. Kategori sayısının az olduğu durumlarda, varsayımın sağlanmasına daha yakın sonuçlar elde edildiğini ve varsayımın ihlal derecesinin, kategori sayısına paralel olarak daha da arttığını ortaya koymuşlardır. Ayrıca hem normal, hem de beta dağılımının tüm durumları için güvenilirliğin kategori sayısından bağımsız olduğunu belirtmişlerdir.

Alanyazında Likert ölçeğinde psikolojik özelliğin tam olarak ölçülmesine ilişkin yapılan çalışmaların genellikle tepki kategori sayısı ve buna bağlı ölçeğin güvenilirliği ile ilgili olduğu görülürken kategoriler arasındaki psikometrik uzaklığın etkisinin incelendiği çok az çalışmanın yapıldığı görülmektedir. Oysaki Likert türü ölçeklerin bahsedilen zayıf yönlerinden dolayı ilgili psikolojik yapıyı tam olarak ölçememesi, tepki kategorilerine ilişkin psikometrik uzaklıkların eşitliği varsayımının bozulmasına neden olmaktadır. Böyle bir durumda ölçek puanlarının geçerliği ve güvenilirliği zedelenmekte ve alınan kararların doğruluğu sorgulanabilir hale gelmektedir. Dolayısıyla psikometrik uzaklıkların incelenmesi, Likert türü ölçek geliştirme ve uyarlama çalışmalarında uygun madde türü ve kategori sayısının belirlenmesi bakımından gerekmekte ve ölçeklerin geçerliklerini desteklemektedir.

### **Araştırmanın Amacı**

Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin öğretmen ve ebeveyn formunun kategoriler arası psikometrik uzaklıkların eşitliği varsayımını uygulamada sağlayıp sağlamadığının incelenmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıda verilen araştırma sorusuna yanıt aranmıştır.

Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan Kişilerarası Güçlü Yön, Aile Bağları, Kendine Dönük (içsel) Güçlü Yön, Okulun İşlevi, Duyuşsal Güçlü Yön alt ölçeklerinin kategoriler arası psikometrik uzaklıkları nasıl değişmektedir?

## **2. Yöntem**

### **Çalışma Grubu**

Bu araştırmanın çalışma grubunu Milli Eğitim Bakanlığına bağlı eğitim veren, Yozgat il merkezi bünyesinde anasınıfı bulunan ilkokullar ile bağımsız resmi ve özel okulöncesi eğitim kurumlarına devam eden beş yaşındaki çocukların anneleri (n=346) ve öğretmenleri (n=22) oluşturmuştur.

### **Veri Toplama Aracı**

Bu çalışmada Epstein (2004) tarafından standardize edilen (Epstein and Sharma, 1998a) ve Salı (2017) tarafından Türk kültürüne uyarlanan Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin (Behavioral and Emotional Rating Scale BERS-2) Öğretmen Değerlendirme Formu (ÖDF) ve Ebeveyn Değerlendirme Formu (EDF) kullanılmıştır. Öğretmen Değerlendirme Formu ve Ebeveyn Değerlendirme Formu 5 alt ölçek ve toplamda 52 maddeden oluşmaktadır. Her iki formda aynı ölçek maddeleri yer almaktadır. Dört dereceli likert türünde düzenlenmiş olan ölçeğin dereceleme maddeleri; Hiç uygun değil (0), Uygun değil (1), Uygun (2) ve Çok Uygun (3) arasında dağılmaktadır. Tablo 1'de ölçeğin alt ölçek, madde ve içeriğine dair bilgilere yer verilmiştir.

**Tablo1. Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Alt Ölçek, Madde ve İçeriği**

Alt Ölçek	Madde	Ölçülen özellik
Kişilerarası güçlü yön (Interpersonal Strength)	10,12,16,17,18,28,30,33,35,37,43,44,46,49,50	Sosyal durumlar içindeki duygusal ve davranışsal kontrol yetenekleri; olumsuz ve hayal kırıklığı yaratan durumlarda gösterilen duygusal ve davranışsal olumlu tepkiler
Aile bağları (Family Involvement)	1,2,4,7,11,15,19,29,36,45	Aile bireyleriyle olan olumlu ilişkiler, aile etkinliklerine katılım, yaşamındaki önemli kişilere güven
Kendine dönük güçlü yön (Intrapersonal Strength)	5,8,20,21,22,26,27,32,38,42,48	Kendine olan güven, mizah duygusu, yaşına uygun alışkanlıklar, akran ilişkileri, duyguları ifade edebilmede güçlük, yaşam sevinci ölçülmektedir.
Okulun işlevi (School Functioning)	14,24,31,39,40,41,47,51,52	Okul etkinliklerine düzenli olarak katılma, ev ödevlerini yapma alışkanlığı, sınıfa uyum sağlama, okula düzenli olarak devam etme, sınıf içi görevleri yerine getirme, okul görevlerini zamanında bitirme
Duyuşsal güçlü yön (Affective Strength)	3,6,9,13,23,25,34	Duyularını paylaşabilme, insanları etkileyebilme ve onlardan etkilenme, başkalarına ilgi gösterme, acı veren durumları kabullenebilme, yaşama olan esnek, diğer insanlarla duygularını paylaşabilme yeterliği

Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Öğretmen Değerlendirme Formu ve Ebeveyn Değerlendirme Formunun yapı geçerliliği doğrulayıcı faktör analizi ile test edilmiştir. Beş faktörlü modelinin veri setine uyumu,  $\chi^2$  /sd, RMSEA, CFI, TLI uyum iyiliği indeksleri aracılığıyla incelenmiştir. Literatürde, Ki-karenin serbestlik derecesine oranının ikiden büyük beşten küçük ya da eşit olması ( $3 < \chi^2 /sd \leq 5$ ), RMSEA değerinin .05 ila .10 aralığında olması ( $0.05 \leq RMSEA \leq 0.10$ ), ve CFI, TLI değerlerinin ise .90 ve üzerinde ancak .95' ten küçük olması ( $0.90 \leq CFI/TLI \leq 0.95$ ) kabul edilebilir uyuma işaret etmektedir. Diğer taraftan,  $\chi^2$  /sd oranının 3' ten küçük olması, RMSEA' nın .05' in altında olması, TLI ve CFI' nin .95 ve üzerinde olması ise mükemmel uyuma işaret etmektedir (Byrne, 2013; Hooper, Coughlan ve Mullen, 2008; Hu ve Bentler, 1999; Marsh, Balla ve McDonald, 1988; McDonald ve Marsh, 1990; Schreiber, Nora, Stage, Barlow ve King, 2006). Öğretmen Değerlendirme Formunun uyum iyiliği indeksleri incelendiğinde,  $\chi^2$  /sd: 1.30 değeri ile mükemmel, CFI: .92 değeri ile kabul edilebilir, TLI: .91 değeri ile kabul edilebilir ve RMSEA: .03 değeri ile mükemmel uyuma işaret etmektedir. Ebeveyn Değerlendirme Formuna ilişkin uyum iyiliği indeksleri de benzer şekilde  $\chi^2$  /sd: 1.28 değeri ile mükemmel, CFI: .92 değeri ile kabul edilebilir, TLI: .92 değeri ile kabul edilebilir ve RMSEA: .03 değeri ile mükemmel uyuma işaret etmektedir. Sonuçlar bir bütün olarak değerlendirildiğinde, öğretmen formu ve ebeveyn formu için önerilen orijinal beş faktörlü modelin kabul edilebilir bir uyuma sahip olduğu belirlenmiştir.

Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin güvenilirliği için madde toplam korelasyonları ve Cronbach alpha iç tutarlılık katsayıları hesaplanmıştır. Öğretmen Değerlendirme Formuna Cronbach alpha iç tutarlılık katsayısı ise .97' dir. Öğretmen Değerlendirme Formunda Kişiler Arası Güçlü Yön Alt Ölçeği .95, Aile Bağları Alt Ölçeği .87, İçsel (Kendine Dönük) Güçlü Yön Alt Ölçeği .87, Okulun İşlevi Alt Ölçeği .91, Duyuşsal Güçlü Yön Alt Ölçeği .85 Cronbach alpha iç tutarlılık katsayılarına sahiptir. Ölçeğin Öğretmen Değerlendirme Formunun madde toplam korelasyonları .26 ila .75 arasında değişmektedir. Ebeveyn Değerlendirme Formunun Cronbach alpha iç tutarlılık katsayısı ise .97' dir. Ebeveyn Değerlendirme Formu alt ölçekleri için Cronbach alpha katsayıları ise Kişiler Arası Güçlü Yön Alt Ölçeği İçin .95, Aile Bağları Alt Ölçeği İçin .86, İçsel (Kendine Dönük) Güçlü Yön Alt Ölçeği İçin .87, Okulun İşlevi Alt Ölçeği İçin .91, Duyuşsal Güçlü Yön Alt Ölçeği için .83 olarak hesaplanmıştır. Ebeveyn Değerlendirme Formu madde toplam korelasyonları .25 ila .75 arasında değişmektedir. Güvenirliğe ilişkin bulgular değerlendirildiğinde ölçek madde toplam korelasyonlarının ilgili kriterleri karşıladığı ve .80 ve üstü Cronbach Alpha değerleri ile Öğretmen ve Ebeveyn Değerlendirme Formlarının tamamı ve alt ölçeklerinin yüksek derecede güvenilir olduğu söylenebilir (Kalaycı, 2010).

### Verilerin Analizi

Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan alt ölçeklerin ölçek değerlerinin hesaplanması için öncelikle genelleştirilmiş kısmi puanlama modeli (GPCM) ile kategori parametreleri kestirilmiştir. MTK varsayımları incelenmiştir. Analiz PARSCALE 4.1 programı ile yapılmıştır (Muraki ve Bock, 2003).

Kategori parametrelerinden yararlanılarak hesaplanan ölçek değerleri (scale value) ve dönüştürülmüş ölçek değerleri Wakita ve diğ. (2012) tarafından Excel' de geliştirilen formül kullanılarak hesaplanmıştır.

### 3. Bulgular

Bu bölümde her bir alt ölçeğin MTK'ya dayalı genelleştirilmiş kısmi puanlama modeli ile kestirilen kategori değerlerine, kategoriler arası psikometrik uzaklıklara dair ölçek değerlerine ve dönüştürülmüş ölçek değerlerine ait bulgulara yer verilmiştir.

#### Ölçek Değerlerinin Kestirilmesi

Madde tepki kuramının temel varsayımı olan tek boyutluluğu (Hambleton ve Swaminathan, 1986) doğrulamak için tüm alt ölçeklerin öz değerleri hesaplanmıştır. Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan her bir alt ölçeğin birinci ve ikinci öz değerleri ve bunların oranları Tablo 2'de verilmiştir.

**Tablo 2. Alt Ölçeklerin Öz Değerleri**

	Özdeğer ( $\lambda$ )	Kişiler Arası Güçlü Yön	Aile Bağları	Kendine Dönük Güçlü Yön	Okulun İşlevi	Duyuşsal Güçlü Yön
Öğretmen Değerlendirme Formu	$\lambda_1$	8.636	4.715	5.248	5.238	3.677
	$\lambda_2$	0.902	0.962	1.003	0.900	0.868
	$\lambda_1/\lambda_2$	9.574	4.901	5.232	5.820	4.236
Ebeveyn Değerlendirme Formu	$\lambda_1$	8.510	4.524	5.104	5.164	3.583
	$\lambda_2$	0.900	0.959	1.047	0.892	0.888
	$\lambda_1/\lambda_2$	9.455	4.717	4.875	5.789	4.035

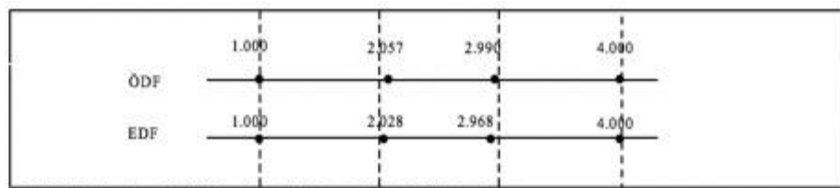
Her kategori için ölçek değerleri ( $\lambda$ ) GPCM ile kestirilen kategori parametrelerine dayalı olarak hesaplanmıştır. Aile Bağları alt ölçeğinin hem Öğretmen hem de Ebeveyn Değerlendirme Formundaki iki maddesinin (1. ve 36. madde), "hiç uygun değil" kategorisine hiçbir katılımcı yönelmediği için aile bağları alt ölçeğinin kategori parametreleri ve ölçek değerleri bu iki madde hariç tutularak hesaplanmıştır. Tüm alt ölçekler için hesaplanan ölçek değerleri, ölçek 4 kategorili olduğu için, 1-4 puan ranjına dönüştürülmüştür. Öğretmen ve Ebeveyn Değerlendirme Formlarına ait her bir alt ölçeğin genelleştirilmiş kısmi puanlama modeli ile kestirilen kategori parametreleri, ölçek değerleri ve dönüştürülmüş ölçek değerleri tablo 3'de verilmiştir. Şekil 1-5'de ise Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan alt ölçeklerin dönüştürülmüş ölçek değerleri geleneksel kategori değerlerine göre konumları ile gösterilmiştir.

**Tablo 3. MTK'ya Göre Kestirilen Kategori Parametreleri ve Ölçek Değerleri**

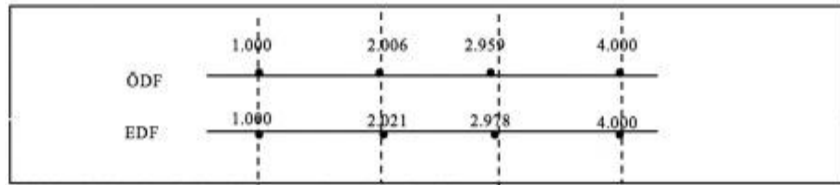
Alt Ölçekler	Öğretmen Değerlendirme Formu			Ebeveyn Değerlendirme Formu		
	Kategori Parametresi	Ölçek Değeri	Dönüştürülmüş Ölçek Değeri	Kategori Parametresi	Ölçek Değeri	Dönüştürülmüş Ölçek Değeri
Kişiler Arası Güçlü Yön	1.143	-1.641	1.000	1.159	-1.654	1.000
	-0.033	-0.494	2.057	0.003	-0.520	2.028
	-1.110	0.519	2.990	-1.162	0.517	2.968
Aile Bağları		1.614	4.000		1.656	4.000
	1.252	-1.730	1.000	1.302	-1.771	1.000
	0.025	-0.536	2.006	0.001	-0.565	2.021
	-1.277	0.543	2.959	-1.303	0.564	2.978
Kendine Dönük Güçlü Yön		1.751	4.000		1.772	4.000
	1.319	-1.786	1.000	1.345	-1.808	1.000
	-0.073	-0.530	2.073	-0.088	-0.529	2.083
	-1.247	0.588	3.028	-1.257	0.600	3.039
Okulun İşlevi		1.726	4.000		1.734	4.000
	1.295	-1.766	1.000	1.309	-1.778	1.000
	-0.189	-0.460	2.161	-0.160	-0.479	2.138
	-1.106	0.604	3.105	-1.149	0.603	3.086
Duyuşsal Güçlü Yön		1.611	4.000		1.646	4.000
	1.183	-1.673	1.000	1.287	-1.760	1.000
	-0.124	-0.459	2.123	-0.163	-0.471	2.142
	-1.059	0.550	3.055	-1.125	0.596	3.087
	1.572	4.000		1.626	4.000	

Ölçeğin Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan alt ölçeklerin tümünde dönüştürülmüş ölçek değerleri ile geleneksel madde değerleri (conventional item values) arasındaki farklar hesaplanmıştır. Bu farklar dört kategoriden oluşan ölçeğin geleneksel kategori değerinden Wakita ve diğ. (2012), tarafından yazılan algoritma ile hesaplanan dönüştürülmüş ölçek değerleri çıkarılarak hesaplanmıştır. Örneğin Öğretmen Değerlendirme Formu'nun Kişiler Arası Güçlü Yön alt boyutunun geleneksel kategori değerinden sapma miktarı, geleneksel kategori değeri 2 için, bu değerden dönüştürülmüş ölçek değeri 2.057'nin çıkarılması yoluyla  $|2-2.057|=0.057$  hesaplanmıştır. Buna göre geleneksel kategori değerlerinden sapma miktarlarının her bir alt ölçek için 0.006 ile 0.161 arasında değişmekte olduğu görülmektedir. Kişiler Arası Güçlü Yön, Aile Bağları ve Kendine Dönük Güçlü Yön alt ölçeklerinde bilinen kategori değerinden sapma 0.1'den küçükken en büyük sapmanın 0.161 değeri ile öğretmen formunda okulun işlevi alt ölçeğinde olduğu görülmüştür. Okulun işlevi alt ölçeğindeki bu kaymaların "uygun değil" kategorisinden "uygun" ve "uygun" kategorisinden "çok uygun" kategorisine doğru olduğu görülmüştür. Duygusal güçlü yön alt ölçeğinde ise öğretmen değerlendirme formunda 0.123; ebeveyn değerlendirme formunda 0.142 değerinde uygun değil" kategorisinden "uygun" kategorisine doğru bir kayma olmuştur.

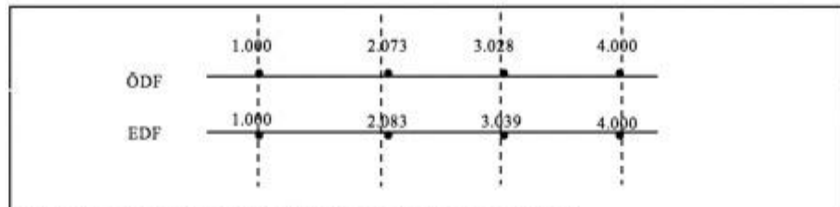
Öğretmen Değerlendirme Formu ve Ebeveyn Değerlendirme Formunun her bir alt ölçeğinde kategori değerinden sapmalar birbirine çok yakındır. Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin Öğretmen ve Ebeveyn Değerlendirme Formlarının MTK'ya dayalı hesaplanan kategoriler arası uzaklıkları incelendiğinde Likert ölçeğinin önemli bir varsayımı olan tepki kategorileri arasındaki uzaklığın sağlandığı görülmektedir.



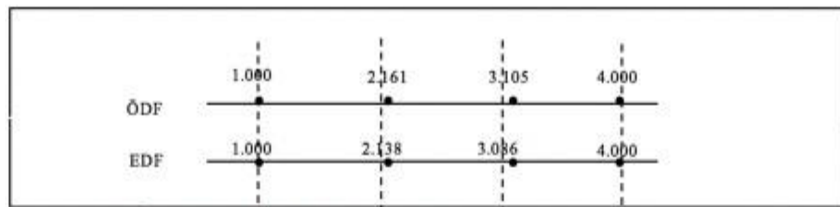
Şekil 1. Kişiler Arası Güçlü Yön alt ölçeği dönüştürülmüş ölçek değerleri



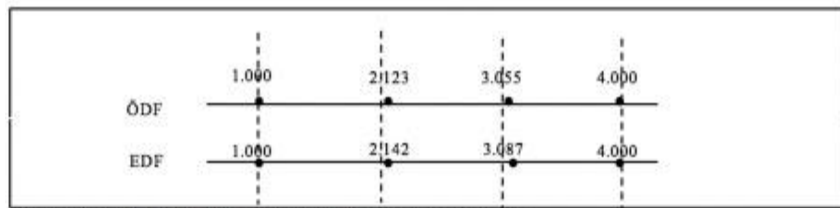
Şekil 2. Aile Bağları alt ölçeği dönüştürülmüş ölçek değerleri



Şekil 3. Kendine Dönük Güçlü Yön alt ölçeği dönüştürülmüş ölçek değerleri



Şekil 4. Okulun İşlevi alt ölçeği dönüştürülmüş ölçek değerleri



Şekil 5. Duygusal Güçlü Yön alt ölçeği dönüştürülmüş ölçek değerleri

#### 4. Sonuç, Tartışma ve Öneriler

Bu çalışmada Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin hem Öğretmen hem de Ebeveyn Değerlendirme Formunu oluşturan Kişiler Arası Güçlü Yön, Aile Bağları, Kendine Dönük Güçlü Yön, Okulun İşlevi ve Duygusal Güçlü Yön alt ölçeklerinin tepki kategorileri arasındaki psikometrik uzaklıkların eşitliği incelenmiştir. Ölçeğin her iki formunun doğrulayıcı faktör analizi sonuçları yapı geçerliğinin sağlandığını; madde toplam korelasyonları ve alpha güvenirlik sonuçları ise güvenilir olduğunu göstermektedir. Her bir alt ölçeğe ilişkin kategori parametreleri MTK'ya dayalı genelleştirilmiş kısmi puanlama modeli ile kestirilmiştir. Aile Bağları alt ölçeğinin hem ebeveyn hem de öğretmen formunda iki maddenin "hiç uygun değil" tepki kategorisi hiçbir yanıtlayıcı tarafından işaretlenmediğinden bu iki madde hariç tutularak ölçek değeri hesaplanmıştır.

Kategoriler arasındaki psikometrik uzaklığın geleneksel kategori değerinden en fazla sapmanın öğretmen ve ebeveyn formunda okulun işlevi alt ölçeğinde olduğu görülmüştür. Tepki kategorilerindeki bu sapma hiç uygun değil tepki kategorisinden çok uygun tepki kategorisine doğru olmuştur. Bu durumun nedeni olarak, çalışma verilerinin eğitim öğretim yılı ikinci dönemin sonuna doğru toplanmış olması nedeniyle örneklem grubundaki çocukların bir öğretim yılı boyunca aldıkları okulöncesi eğitimden yeterince yararlanabilmiş olmaları düşünülebilir. Bu durumda tepki kategorilerinde olumsuzdan olumlu ifadeye yönelik olan sapmanın sosyal kabul edilebilirlik gibi yanıtlayıcı davranışından; maddenin içeriği yansıtamaması, yanıtlayıcılar tarafından anlaşılır olmaması gibi maddenin yapısından ya da ölçülecek özelliğe uygun tepki kategori sayısının belirlenememesinden kaynaklanmadığı düşünülmektedir. Aksine çocukların okul öncesi eğitimin hedeflerini kazanmış olmaları ve bunu davranış ve duygularına yansıtmaları sonucunda öğretmen ve ebeveynlerin okulun işlevine yönelik algılarının olumlu yönde gelişmiş olmasından kaynaklanmış olabileceği düşünülmektedir. Hem Öğretmen hem de Ebeveyn Değerlendirme Formunda gözlenen sapmalardaki tutarlılık, her iki cevaplayıcı grubunun, benzer yanıtlama davranışına sahip olduğunu göstermektedir. Bunun nedeni de yukarıda açıklandığı gibi çocukların aldığı okulöncesi eğitimin hedefine ulaşmış olmasından kaynaklanan bir sapmanın olduğunun kanıtı olabilir.

Okulun İşlevi alt ölçeğinden sonra geleneksel kategori değerlerinden en büyük sapma Duygusal Güçlü Yön alt ölçeğinde bulunmuştur. Hem öğretmen hem de ebeveyn formunda tepki kategorilerindeki bu sapmanın uygun değil kategorisinden uygun kategorisine doğru olduğu bulunmuştur. Bunun nedeninin de yine yukarıda açıklandığı gibi çocukların aldıkları okulöncesi eğitimin, onların duygusal ve sosyal gelişimlerine olan olumlu katkısının cevaplayıcı grupların her ikisinin de ölçekteki maddelere verdikleri tepkiyi etkilemiş olabileceği düşünülmektedir. Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin okulun işlevi alt ölçeğinin tepki kategorilerinde en büyük sapmanın olduğu görülmekle birlikte bu alt ölçeğin güvenirliği hem öğretmen hem de ebeveyn formu için .91'dir. Benzer şekilde tepki kategorilerindeki ikinci en büyük sapma Duygusal güçlü yön alt ölçeğinde olup bu ölçeğin güvenirliği ise öğretmen formu için .85, ebeveyn formu için ise .83 olarak hesaplanmıştır. Bulgular birlikte değerlendirildiğinde alt ölçeklerin hem öğretmen hem de ebeveyn formu için güvenirliklerinin tepki kategorilerindeki sapmadan bağımsız olduğu görülmektedir. Bu bulgu alanyazındaki bazı araştırmalar ile tutarlık göstermektedir. Brown, Wilding ve Coulter, 1991; Komorita, 1963; Matell ve Jacoby, 1971, Wakita ve diğ., 2012).

Elde edilen sonuçlara göre Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin hem Öğretmen hem de Ebeveyn Değerlendirme Formunu oluşturan Kişiler Arası Güçlü Yön, Aile Bağları ve Kendine Dönük Güçlü Yön, Okulun İşlevi ve Duygusal Güçlü Yön alt ölçeklerinde tepki kategorilerinin geleneksel değerlerinden sapmanın en fazla .161 değerinde olduğu bulunmuştur. Hamby ve Levine (2016) ve Wakita ve diğ. (2012)'nin çalışmaları incelendiğinde geleneksel kategori değerlerinden bu miktarda bir sapmanın düşük düzeyde bir sapma olduğu düşünülmektedir. Dolayısıyla, Davranış ve Duyguları Değerlendirme Ölçeği-2 (BERS-2)'nin hem Öğretmen hem de Ebeveyn Değerlendirme Formunu oluşturan Kişiler Arası Güçlü Yön, Aile Bağları ve Kendine Dönük Güçlü Yön, Okulun İşlevi ve Duygusal Güçlü Yön alt ölçeklerinde tepki kategorileri arasındaki psikometrik uzaklık varsayımının sağlandığı söylenebilir. Bu bulgu ölçek ile ölçülmek istenen psikolojik özelliğin tam olarak ölçülebildiğinin bir göstergesidir (Wakita ve diğ., 2012). Bu yönüyle çalışma sonucunda, Salı (2017) tarafından beş yaş grubu çocukları için Türk kültürüne uyarlanan ölçeğin Öğretmen ve Ebeveyn Değerlendirme Formlarının geçerliğini destekler yönde bir kanıt sunulmuştur.

Ölçeğin Öğretmen ve Ebeveyn Değerlendirme Formlarını oluşturan alt ölçekler 4'lü likert tipindedir. Wakita ve diğerleri (2012) tepki kategori sayısının dört olduğu durumda psikometrik uzaklıklardaki bozulmanın en az olduğunu ortaya koymuştur. Uyumaz ve diğerleri (2016) kategoriler arası psikometrik uzaklık varsayımındaki bozulmanın normal dağılım ve beta dağılımının her durumu için, en az dört kategorili ölçeklerde olduğunu göstermiştir. Hamby ve Levine (2016)'de tepki kategori sayısındaki artışın cevaplayıcıların kategoriler arasındaki ayrımları yapmakta zorlanmalarına neden olduğunu belirtmiştir. Bu çerçevede araştırmaya konu olan ölçeğin kategori sayısının uygun olduğu belirtilen çalışmalar ile desteklenmektedir.



Bu çalışmanın ölçek geliştirme ve uyarlama aşamalarında ölçülecek psikolojik yapıya uygun tepki kategori sayısının belirlenmesinde ve dolayısıyla ölçeklerin geçerliğini artırıcı yönde önlemlerin alınmasında alana katkı sağlayacağı düşünülmektedir. Araştırmacılar geliştirdikleri ölçeklerin farklı sayıda tepki kategorilerinde kategoriler arası psikometrik uzaklığın nasıl değiştiğine yönelik analizler yaparak uygun kategori sayısını belirleyebilirler. Bunun yanı sıra kategoriler arası psikometrik uzaklık varsayımının maddenin yönüne (olumlu-olumsuz), dağılımın tipine ya da maddenin içeriğine göre nasıl değiştiğine yönelik çalışmalar yapılabilir.

## 5. Kaynakça

- Boote, A. S. (1981). Reliability testing of psychographic scales: Five-point or seven-point? Anchored or labeled? *Journal of Advertising Research*, 21, 53-60.
- Brown, G., Wilding, R. E., & Coulter, R. L. (1991). Customer evaluation of retail salespeople using the SOCO scale: A replication extension and application. *Journal of the Academy of Marketing Science*, 9, 347-351.
- Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic Concepts, Applications, and Programming*, Second Edition. Routledge.
- Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18, 205-215.
- Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of inter-rater reliability: A Monte-Carlo investigation. *Applied Psychological Measurement*, 9, 31-36.
- Epstein, M. H., & Sharma, J. (1998a). *Behavioral and emotional rating scale: A strengths-based approach to assessment*. Austin, TX: PRO-ED.
- Epstein, M. H. (2004). *Behavioral and emotional rating scale: A strength-based approach to assessment: Examiner's manual*. Austin, Texas: Pro-Ed.
- Garner, W. R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, 67, 343-352.
- Green, P. E., & Rao, V. R. (1970). Rating scales and information recovery: How many scales and response categories to use? *Journal of Marketing*, 34, 33-39.
- Hambleton, R. K. ve Swaminathan, H. (1986). *Item response theory*. Boston: Kluwer
- Hamby, T. & Levine, D. (2016). Response- scale formats and psychological distances between categories. *Applied Psychological Measurement*, 40(1) 73-75.
- Hooper, D., Coughlan, J. ve Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-59.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A multidisciplinary journal*, 6(1), 1-55.
- Kalaycı, Ş. (2010). *SPSS uygulamalı çok değişkenli istatistik teknikleri* (Beşinci Baskı.). Ankara: Asil Yayın Dağıtım.
- Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *Journal of Social Psychology*, 61, 327-334.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte-Carlo approach. *Journal of Applied Psychology*, 60, 10-13.
- Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin*, 103(3), 391.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: Reliability and validity. *Educational and Psychological Measurement*, 31, 657-674.
- McDonald, R. P. & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological bulletin*, 107(2), 247.
- Muraki, E., & Bock, R. D. (2003). *PARSCALE: Parameter scaling of rating data [Computer program]*. Chicago, IL: Scientific Software.
- Oaster, T. R. F. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills*, 68, 549-550.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.
- Salı, G. (2017). Davranış ve duyguları değerlendirme ölçeği-2 (BERS-2)'nin beş yaş grubu için Türk kültürüne uyarlama çalışması. *Uluslararası Avrasya Sosyal Bilimler Dergisi* 23, 43-61.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338. doi:10.3200/JOER.99.6.323-338

- Schuts, H. G., & Rucker, M. H. (1975). A comparison of variables configurations across scale lengths: An empirical study. *Educational and Psychological Measurement*, 35, 319-324.
- Tavşancıl, E. (2010). *Tutumların ölçülmesi ve SPSS ile veri analizi* (4. baskı). Ankara: Nobel.
- Tezbaşaran, A. (1997). *Likert tipi ölçek geliştirme kılavuzu*. Ankara: Türk Psikologlar Derneği.
- Uyumaz, G., Sırgancı, G. ve Çokluk, Ö. (2016). Farklı kategori sayısı ve yetenek dağılımı durumlarında kategoriler arası psikolojik uzaklıkların incelenmesi. V. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Bildiri olarak sunuldu) 1-3 Eylül 2016 Akdeniz Üniversitesi.
- Wakita, T., Ueshima, N., & Noguchi, H. (2012) Psychological distance between categories in the likert scale: Comparing Different Numbers of Options. *Educational and Psychological Measurement*, 72(4) 533–546.