



Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi
Ankara University Faculty of Educational Sciences Journal of Special Education

2025, 26(4), 557-577

ARAŞTIRMA MAKALESİ | RESEARCH ARTICLE

Gönderim Tarihi | Recieved Date: 24.08.24

Kabul Tarihi | Accepted Date: 09.04.25

Erken Görünüm | Online First: 12.05.25

**Problem Davranışları Değerlendirme Ölçeğinin Klasik Test Kuramı ve
Madde Tepki Kuramı Geçerlik ve Güvenirlik Kanıtları**

[Türkçe okumak için tıklayınız](#)

**Classical Test Theory and Item Response Theory Validity and Reliability
Evidence of Problem Behavior Evaluation Scale**

[Click here to read in English](#)

Eylem Dayı



Arzu Doğanay-Bilgi



Çetin Toraman



Nevin Güner



Hasan Köse



Etik Onay | Ethical Approve

Gazi Üniversitesi Bilimsel Araştırma Etik Kurulu'nun 17.10.2023 tarih ve E-77082166-604.01.02-777796 sayılı onayı ve MEB'in 23.11.2023 tarih ve E-70297673-605.01-90309581 sayılı araştırma uygulama izni alınmıştır.

Approval dated 17.10.2023 and numbered E-77082166-604.01.02-777796 was obtained from the Scientific Research Ethics Committee of Gazi University, and research implementation permission dated 23.11.2023 and numbered E-70297673-605.01-90309581 was obtained from the Ministry of National Education.

Yazarların Katkı Düzeyleri | Authors' Contributions

Araştırmanın planlama ve uygulamasında yazarlar eşit düzeyde yer almışlardır. Ayrıca birinci yazar araştırmanın raporlaştırılması ve dergiye hazır hale getirilmesinde, ikinci yazar verilerin analizi ve yönteminde, üçüncü yazar giriş, tartışma ve yönteminde, dördüncü yazar makalenin bütününcü incelenmesinde ve beşinci yazar ise İngilizcesinde görev almıştır.

The authors equally contributed to the planning and implementation of the study. Additionally, the first author was responsible for reporting the research and preparing it for journal submission, the second author for data analysis and methodology, the third author for the introduction, discussion, and methodology, the fourth author for reviewing the entire manuscript, and the fifth author for the English language version.

Çıkar Çatışması Beyanı | Conflict of Interest

Bu çalışmada herhangi bir kurum, kuruluş veya kişilerle ekonomik, ticari, yasal ya da mesleki bir herhangi bir çıkar çatışması bulunmamaktadır. There is no economic, commercial, legal or professional conflict of interest with any institution, organization or person in this study.

Destek | Funding

Bu araştırma 223K235 no'lu, "Kaynaştırma/Bütünleştirme Yoluyla Eğitim Uygulamaları Yapılan İlkokul Sınıflarında Bir Davranışsal Müdahale: Classdojo ile Uygulanan İyi Davranış Oyunu" adlı proje kapsamında 1001- TÜBİTAK tarafından desteklenmektedir.

This research is supported by TÜBİTAK under the 1001 Program within the scope of the project numbered 223K235, titled 'A Behavioral Intervention in Primary School Classrooms Implementing Inclusive Education Practices: The Good Behavior Game Implemented via ClassDojo.'

Teşekkür | Acknowledgement

Araştırmayı destekleyen TÜBİTAK'a; ilkokul müdürleri öğretmenlerine, lisans, yüksek lisans ve doktora öğrencilerimize ve çalışma arkadaşlarımıza teşekkür ederiz.

We would like to thank TÜBİTAK for supporting the research; primary school principals and their teachers; our undergraduate, graduate, and doctoral students; and our colleagues.

Dergimizde yayımlanan çalışmaların telif hakları yazarlara ait olup, ticari kullanım hakkı dergimize aittir. Çalışmalar CC-BY-NC-ND lisansı ile açık erişim olarak yayımlanmaktadır.

The copyrights of the studies published in our journal belong to the authors, and the journal reserves the right to commercial use. These studies are published as open access with the CC-BY-NC-ND license.



Problem Davranışları Değerlendirme Ölçeğinin Klasik Test Kuramı ve Madde Tepki Kuramı Geçerlik ve Güvenirlik Kanıtları

Eylem Dayı¹

Çetin Toraman²

Nevin Güner³

Arzu Doğanay-Bilgi⁴

Hasan Köse⁵

Öz

Giriş: Eğitim ortamlarında, okul ve sınıf kurallarına uymamakla ilişkili pek çok davranış kapsayan problem davranışlar, öğrencilerin güvenliğini tehdit eden, öğrenciler arasındaki sosyal etkileşimi olumsuz etkileyen, öğretmeni ve öğrencileri akademik etkinliklerden uzaklaştıran davranışlar olarak tanımlanmaktadır. Eğitim ortamını olumsuz etkileyen problem davranışların belirlenmesi, bu davranışların kontrol edilmesi ve azaltılması açısından önemlidir.

Yöntem: Bu araştırmanın amacı, öğretmenlerin gözlemlerine ve algılarına dayalı olarak öğrencilerin problem davranışlarını değerlendirme amacına hizmet edecek Problem Davranışları Değerlendirme Ölçeğinin (PDDÖ) klasik test kuramı ve madde tepki kuramına göre geçerlik ve güvenirlik kanıtlarını incelemektir. Ölçeğin geliştirilmesi sürecinde öncelikle alanyazın taraması yapılarak ölçekte yer alacak maddeler hazırlanmış, ardından uzman görüşlerinden yararlanılarak ölçeğe uygulama öncesi son hali verilmiştir. Ölçeğin geçerlik güvenirlik kanıtlarını incelemek amacıyla yapılan çalışmaya 590 öğretmen katılmıştır. Toplanan veriler üzerinde açıklayıcı faktör analizi, doğrulayıcı faktör analizi ve madde tepki kuramına göre madde karakteristik eğrileri, madde bilgi fonksiyonları, madde ayırt edicilikleri incelenmiştir. Ölçeğin güvenirlik kanıtları ise Cronbach Alpha McDonald's Omega ve marjinal güvenirlik katsayılarıyla belirlenmiştir.

Bulgular: Yapılan analizlerin sonucunda PDDÖ'nün 18 madde ve iki faktörlü geçerli bir ölçek olduğu belirlenmiştir. PDDÖ'nün güvenirliliği klasik test kuramına göre Cronbach Alpha ve McDonald's Omega katsayıları, madde tepki kuramına göre ise marjinal güvenirlik katsayısı ile belirlenmiştir. Sonuçlar PDDÖ'nün yüksek güvenirlik düzeyinde olduğunu göstermiştir. Analiz sonucunda birinci faktör 10, ikinci faktör sekiz maddeden oluşmuştur.

Tartışma: PDDÖ'nün 18 madde ve iki faktörlü yapısı ile karmaşık olmayan, kolay yanıtlanabilir kullanışlı bir ölçek olduğu söylenebilir. Ölçek, Türkiye'de ilkokullarda öğretmenlerin gözlem ve algılarına dayalı olarak öğrencilerin problem davranışlarının değerlendirilmesi amacıyla kullanılabilecek ve problem davranışlara ilişkin yürütülecek bilimsel çalışmalara destek sağlayabilecektir.

Anahtar sözcükler: Problem davranış, ölçek geliştirme, geçerlik ve güvenirlik klasik test kuramı, madde tepki kuramı, özel eğitim.

Atıf için: Dayı, E., Toraman, Ç., Güner, N., Doğanay-Bilgi, A., & Köse, H. (2025). Problem davranışları değerlendirme ölçeğinin klasik test kuramı ve madde tepki kuramı geçerlik ve güvenirlik kanıtları. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 26(4), 557-577. <https://doi.org/10.21565/ozelegitimdergisi.1538042>

¹**Sorumlu Yazar:** Doç. Dr., Gazi Üniversitesi, E-posta: eylemd@gazi.edu.tr, <https://orcid.org/0000-0003-1020-5846>

²Doç. Dr., Çanakkale Onsekiz Mart Üniversitesi, E-posta: toramanacademic@gmail.com, <https://orcid.org/0000-0001-5319-0731>

³Prof. Dr., Eskişehir Osmangazi Üniversitesi, E-posta: antreh@gmail.com, <https://orcid.org/0000-0002-9135-6429>

⁴Doç. Dr., Gazi Üniversitesi, E-posta: arzudb@yahoo.com, <https://orcid.org/0000-0003-3330-8303>

⁵Dr., Milli Eğitim Bakanlığı, E-posta: hsnkose@hotmail.com, <https://orcid.org/0000-0002-0923-3711>

Giriş

Eğitim ortamlarında, okul ve sınıf kurallarına uymamakla ilişkili pek çok davranışı kapsayan problem davranışlar (Robinson & Ricord-Griesemer, 2006), öğrencilerin güvenliğini tehdit eden, öğrenciler arasındaki sosyal etkileşimi olumsuz etkileyen, öğretmeni ve öğrencileri akademik etkinliklerden uzaklaştıran davranışlar olarak tanımlanmaktadır (Nash vd., 2016; Sucuoğlu & Kargın, 2010). Öğrencilerin büyük bir kısmı genellikle uygun davranışlar sergilese de farklı yıllarda ve farklı eğitim ortamlarında yapılan araştırmalar öğrencilerin yaklaşık %20'sinin orta düzeyde, yaklaşık %6'sının ise ciddi problem davranışlar sergilediğini ortaya koymuştur (Lewis vd., 2006; Pas vd., 2015). Eğitim ortamlarında sıklıkla sergilenen problem davranışlar arasında ders etkinliklerine katılmamak, sınıfta dolaşmak, izinsiz konuşmak, diğer öğrencileri rahatsız etmek, ders dışı iş ve nesnelerle ilgilenmek gibi davranışlar bulunurken sıklığı nispeten az da olsa öfke, saldırganlık ya da zorbalık gibi davranışlarla da karşılaşmaktadır (Robinson & Ricord-Griesemer, 2006).

Öğrencilerin problem davranışlarının öğretmenler, öğrenciler ve eğitim ortamı üzerinde olumsuz etkileri bulunmaktadır. Örneğin öğretmenler problem davranışlarla baş etmek için çok fazla zaman ve enerji harcamak zorunda kalmakta (Clunies-Ross vd., 2008; Simonsen vd., 2008), bu da öğretmenlerin öz yetkinlik algılarında ve iyilik hallerinde düşüşe ve tükenmişlik yaşamalarına yol açabilmektedir (Ford vd., 2019). Öğretmenlik mesleğinin en zor boyutlarından birinin öğrencilerin problem davranışlarıyla baş etmek olduğu alanyazında vurgulanmaktadır (Aloe vd., 2014). Problem davranışlar eğitim ortamı ve öğrencilerle ilgili de önemli sonuçlar ortaya çıkarmaktadır. Problem davranışların yüksek düzeyde sergilendiği eğitim ortamları öğrencilerin kendini güvende hissetmediği, sosyal gelişimin olumsuz etkilendiği ve akademik etkinliklere katılımın ve başarının düşük olduğu ortamlar olarak gösterilmektedir (Charles, 2011; Ødegård & Solberg, 2024). Sürekli ve ciddi problem davranış sergileyen öğrencilerin akademik olarak başarısız öğrenciler oldukları (Kremer vd., 2016; Ross vd., 2023) ve gelecekte suça karışma olasılıklarının da yüksek olduğu bilinmektedir (Lewis vd., 2006). Sonuç olarak öğretmenleri ve öğrencileri akademik etkinliklerden uzaklaştıran ve eğitim ortamını olumsuz etkileyen problem davranışların kontrol edilmesi ve azaltılması için bu davranışların belirlenmesi eğitimin hedeflerine ulaşılabilmesi açısından son derece önemlidir. Çünkü problem davranışların belirlenmesi, bu davranışlar için uygun sağıtım programlarının hazırlanmasının başlangıç noktasını oluşturmaktadır.

Alanyazında eğitim ortamlarında meydana gelen problem davranışların belirlenmesine yönelik ölçek, kontrol listesi ve uyarılama çalışmalarına rastlanmaktadır. Bunlardan biri Kellam ve diğerleri (1975) tarafından geliştirilen sınıfa uyum becerilerinin öğretmen gözlemlerine göre değerlendirildiği Sınıfa Uyum Öğretmen Gözlemi (Teacher Observation of Classroom Adaptation [TOCA]) formudur. Bu ölçme aracında öğretmenler tarafından görüldüğü ve tanımlandığı şekliyle birinci sınıf öğrencilerinin sınıfa uyum davranışlarının değerlendirilmesi amaçlanmıştır. Değerlendirme süreci eğitimli personel tarafından yapılandırılmış görüşmelerle yürütülmüştür. TOCA, Baltimore, Maryland'deki büyük bir müdahale çalışmasında kullanılmak üzere Johns Hopkins Önleme ve Erken Müdahale Merkezi (Johns Hopkins Center for Prevention and Early Intervention) tarafından büyük bir revizyona tabi tutulmuştur (Teacher Observation of Classroom Adaptation-Revised [TOCA-R]-Werthamer-Larsson vd., 1991). Bu versiyon büyük ölçekli araştırmalarda kullanılmıştır (ör. Kellam vd., 1998). Ancak TOCA'nın büyük ölçekli araştırmalarda yapılandırılmış görüşmelerle uygulanması zaman ve maliyet açısından sınırlılıklara neden olduğundan kontrol listesine dayalı yeni versiyonu geliştirilmiştir (Teacher Observation of Classroom Adaptation-Checklist [TOCA-C]- Leaf vd., 2002). Koth ve diğerleri (2009) tarafından geçerlik ve faktör yapısı incelenen TOCA-C'nin okul öncesi eğitim kurumlarından toplanan verilerle Türkçe'ye uyarlaması yapılmıştır (Aydos vd., 2014).

Alanyazında rastlanılan başka bir ölçek olan Çocuk Davranışı Değerlendirme Ölçeği (Child Behavior Checklist [CBCL] Achenbach, 1978; Achenbach & Edelbrock, 1979), ebeveynlerin, öğretmenlerin ve çocukla ilgili diğer kişilerin çocukların davranışsal ve duygusal sorunlarını tanımlamak için doldurduğu standart bir formdur. CBCL'nin 2 ve 3 yaş arası versiyonu (CBCL/2'den 3'e) ve 4 ila 18 yaş arası versiyonu (CBCL/4 ila 18) bulunmaktadır (Achenbach & Ruffle, 2000). Bu ölçek Türkçe'ye çevrilerek birçok çalışmada tipik gelişim gösteren çocuklar için kullanılmıştır (Akkök vd., 1986; Erol vd., 1995; Erol & Şimşek, 1998; Kanay & Girli, 2008). Bir diğer ölçek olan Doğrudan Davranış Derecelendirmesi (Direct Behavior Rating [DBR]) ise sistematik doğrudan gözleme dayalı değerlendirici tarafından algılandığı şekliyle derecelendirme yoluyla verileri kaydetmeyi içermektedir (Chafouleas vd., 2009). DBR kategorisinde psikometrik kanıtların oluşturulduğu Doğrudan Davranış Derecelendirmesi - Tek Maddeli Ölçek (DBR Single-Item Scales [DBR-SIS]) (Chafouleas vd., 2007; Chafouleas

vd., 2010) ve Doğrudan Davranış Derecelendirmesi - Çok Maddeli Ölçek (DBR Multi-Item Scales [DBRMIS]) (Briesch vd., 2016; Daniels vd., 2021; Volpe & Briesch, 2012, 2015, 2016) versiyonları geliştirilmiştir.

Türkçe uyarlaması yapılan bir ölçek de Davranış Problemleri Kontrol Listesi'nin (Revised Problem Behavior Checklist) gözden geçirilmiş formudur (Quay & Peterson, 1996). Bu çalışmada Kaner ve Çiçekçi (2000) yetersizliği olmayan 7-18 yaş aralığındaki öğrencilerin öğretmenleri tarafından değerlendirilmeleri ile elde edilen verilerle aracın Türkçe formunun psikometrik özelliklerini incelemişlerdir. Kaner ve diğerleri (2011) bir başka çalışmada, Conners ve diğerlerinin 1960'lı yıllarda geliştirip 1997 yılında güncelledikleri Dikkat Eksikliği/Hiperaktivite Bozukluğunun (DEHB) tanılmasında kullanılan Conners Öğretmen Dereceleme Ölçeğini (Conners' Teacher Rating Scale-Revised/Long [CTRS-R/L]) Türk kültürüne uyarlayarak ölçeğin araştırmalarda ve tanılama yapan kliniklerde kullanılabilecek güvenilir ve geçerli bir ölçek olduğunu göstermiştir. Diğer bir çalışmada ise Sucuoğlu (2003) zihin yetersizliği olan ergen ve yetişkinlere uygulanan ilaç tedavisi ya da diğer müdahalelerin etkilerini belirlemek için geliştirilen Sorun Davranışlar Kontrol Listesini (Aberrant Behavior Checklist [ABC], Aman vd., 1985a, 1985b) Türkçe'ye çevirerek psikometrik özelliklerini incelemiş ve aracı kullanarak yaş, cinsiyet ve yerleştirildikleri kurum gibi değişkenlerin zihin yetersizliği olan bireylerin problem davranışları üzerindeki etkilerini belirlemiştir.

Bir diğer çalışmada ise Okul Sosyal Davranış Ölçeklerinin (School Social Behavior Scales [SSBC], Merrell, 1993) Türkçe'ye uyarlama çalışması yapılmıştır (Yukay-Yüksel, 2009). Son olarak okul öncesi ve ilkökul dönemindeki öğrencilerin sosyal becerilerini belirlemek için Sosyal Yeterlilik ve Olumsuz Sosyal Davranışlar alt formlarından oluşan Okul Sosyal Davranış Ölçekleri (OSDÖ) geliştirilmiştir. Bunlardan "Sosyal Yeterlilik Ölçeği", Kişiler Arası İlişkiler, Öz Denetim Becerisi ve Akademik Beceriler içeren üç alt ölçekten oluşmaktadır. "Olumsuz Sosyal Davranışlar Ölçeği" de Saldırgan-Sinirli, Antisosyal-Agresif ve Yıkıcı-Talepkâr olmak üzere problem davranışları da kapsayan üç alt ölçekten oluşmaktadır.

Özetle alanyazında Türkçe'ye çevrilerek uyarlaması yapılan ya da doğrudan geliştirilen ölçekler incelendiğinde bu ölçeklerin ya sadece tipik gelişim gösteren ya da özel gereksinimli öğrenciler için geliştirilmiş ölçekler olduğu anlaşılmaktadır. Bununla birlikte bazı uyarlama çalışmalarının da sadece okul öncesi dönem çocuklar ile yapıldığı görülmektedir. Oysa ilkökul kademesi sayısal olarak büyük bir öğrenci grubunun bulunduğu (MEB, 2023) ve aynı zamanda öğrencilerin önemli bir bölümünün sosyal ve akademik becerileri kazanmak için ilk kez eğitim ortamlarına katıldığı eğitim kademesidir. Bu nedenle ilkökulda öğrenciler okul ve sınıf kurallarını öğrenme ve okula uyum sağlama sürecinde çeşitli problem davranışlar sergileyebilmektedir. Ayrıca ilkökul sınıfları öğrenci özellikleri açısından da büyük bir çeşitliliği barındırmaktadır. Hemen her sınıfta, farklı bireysel özellik ve gereksinimlere sahip tipik gelişen öğrencilerle birlikte akademik ve davranışsal risk altında olan ya da özel gereksinimleri bulunan öğrenciler de bulunmaktadır. Dolayısıyla ilkökul kademesinde doğrudan hem tipik gelişen hem risk altındaki hem de özel gereksinimli öğrencilerin problem davranışlarını öğretmen gözlem ve algısına göre değerlendiren ölçme araçlarının geliştirilmesi ve alanyazına kazandırılmasına ihtiyaç olduğu düşünülmektedir.

Bu araştırmanın amacı, sınıf öğretmenlerinin okul ve sınıf ortamlarındaki gözlemlerine ve algılarına dayalı olarak öğrencilerin problem davranışlarını değerlendirme amacına hizmet edecek Problem Davranışları Değerlendirme Ölçeğinin (PDDÖ-Problem Behavior Evaluation Scale [PBES]) klasik test kuramı (KTT-Classical Test Theory [CTT]) ve madde tepki kuramına (MTK-Item Response Theory [IRT]) göre geçerlik ve güvenirlik kanıtlarını incelemektir.

Yöntem

Araştırma Yöntemi

Bu betimsel araştırma kapsamında sınıf öğretmenlerinin okul ve sınıf ortamlarındaki gözlemlerine ve algılarına dayalı olarak öğrencilerin problem davranışlarını değerlendirme amacına hizmet edecek Problem Davranışları Değerlendirme Ölçeğinin geliştirilmesi amaçlanmıştır. Bu çalışmada PDDÖ'nün geçerlik kanıtları klasik test kuramına göre açıklayıcı faktör analizi ve madde tepki kuramına göre madde karakteristik eğrileri, madde bilgi fonksiyonları, madde ayırt edicilikleri ile incelenmiştir. Güvenirlik kanıtları ise klasik test kuramına göre Cronbach Alpha ve McDonald's Omega katsayıları, madde tepki kuramına göre ise marjinal güvenirlik

katsayısı ile belirlenmiştir. Belirlenen bu özellikler PDDÖ'nün gelecekte başka araştırmalarda kullanılabilirliği açısından önemlidir.

Çalışma Grubu

Bu araştırmada çalışma grubunun belirlenmesinde amaçlı örnekleme yöntemlerinden ölçüt örnekleme kullanılmıştır. Ölçüt örnekleme, araştırmacıların araştırma amaçları doğrultusunda önceden belirledikleri önemli ölçütleri karşılayan birimleri örnekleme seçmesi yoluyla uygulanır (Miles vd., 2014; Toraman & Özdemir, 2023). Ölçüt örnekleme uygulanırken belirlenen ölçütlere uyan ve zengin bilgi verebilecek durumlar çalışmaya alınır (Toraman & Özdemir, 2023). Buna uygun olarak bu araştırmada a) ilkokulda görev yapma, b) araştırmaya katılıma gönüllü olma, c) ulaşılabilir olma ölçütlerini karşılayan 590 sınıf öğretmeninden veri toplanmıştır. Katılımcı öğretmenler için belirlenen ölçütlerle birlikte "İstatistiki Bölge Birimleri Sınıflandırmasındaki-İBBS" ("Nomenclature d'Unités Territoriales Statistiques-NUTS 12 farklı bölgede (Bakanlar Kurulu Kararı, 2002) yer alan illerden (İstanbul, Ankara, Şanlıurfa, Hatay, Eskişehir, Kars, Bolu vb.) katılımcılara ulaşılmaya çalışılmıştır. Katılımcı 590 sınıf öğretmenin demografik bilgileri Tablo 1'de verilmiştir.

Tablo 1

Araştırma Örnekleminde Bulunan Sınıf Öğretmenlerinin Demografik Bilgileri

	Değişken	n (%)
Cinsiyet	Kadın	146(24.7)
	Erkek	444(75.3)
Yaş	20-30 yaş	130(22)
	31-40 yaş	199(33.7)
	41-50 yaş	162(27.5)
	51 yaş ve üstü	99(16.8)
	0-5 yıl	115(19.5)
Kıdem	6-10 yıl	100(16.9)
	11-20 yıl	181(30.7)
	21 Yıl ve üstü	194(32.9)
Öğretmenin halihazırda okuttuğu sınıf düzeyi	1. sınıf	128(21.7)
	2. sınıf	216(36.6)
	3. sınıf	147(24.9)
	4. sınıf	99(16.8)
Öğretmenin halihazırda sınıfı okutma süresi	0-3 ay	28(4.7)
	4-6 ay	85(14.4)
	7-9 ay	100(16.9)
	10-12 ay	16(2.7)
	13 ay ve üstü	361(61.2)
Toplam		590(100)

Veri Toplama Araçları

PDDÖ'nün geliştirilmesindeki motivasyon kaynağı ilkokul öğrencilerinin problem davranışlarının değerlendirilmesidir. Bu amaca hizmet edecek uygun bir ölçek olup olmadığı alanyazından incelenmiştir. Problem davranışlarla ilgili ulusal ve uluslararası kitaplar incelenmiş, DergiPark, Web of Science, Google Akademik ve PubMed veri tabanları üzerinden Türkçe "problem davranışlar", "yıkıcı davranış, istenmeyen davranış", "zorlayıcı davranış", "uygun olmayan davranış", "problem davranışları değerlendirme", "problem davranışları değerlendirme ölçeği" ve İngilizce olarak "problem behaviors", "disruptive behavior", "undesirable behavior", "challenging behavior", "inappropriate behavior", "assessing problem behaviors", "scales for assessing problem behaviors" anahtar kelimeleri ile taramalar yapılmıştır. Ayrıca aynı anahtar kelimelerle Educational Resources Information Center (ERIC), Education Full Text, PsycINFO ve Dissertations and Theses Full Text databases veri tabanları kullanılarak taramalar genişletilmiştir. Bu taramalar sonucunda alanyazında problem davranışlarla ilgili bilgi verebilecek bazı ölçek araçları olduğu belirlenmiştir (ör. Bradshaw & Kush, 2020; Daniels vd., 2021; Kaner & Çiçekçi, 2000; Sucuoğlu, 2003). Bu ölçeklerin ilkokul öğrencilerinin problem davranışlarını öğretmenlerin okul ve sınıf ortamlarındaki gözlemlerine ve algılarına dayalı olarak değerlendirmek için tam anlamıyla uygun olmadığı tespit edilmiştir. Bu nedenle PDDÖ araştırmacılar tarafından geliştirilmiştir.

PDDÖ'nün geliştirilmesinde alanyazının önerdiği ölçme aracı geliştirme adımları izlenmiştir (Cohen & Swerdlik, 2013; DeVellis, 2014; Şeker & Gençdoğan, 2014). Bu adımlar şunlardır: 1) ölçeğin amacını belirleme, 2) ölçekte yer alacak madde türüne karar verme (Likert tipi derecelenmiş madde türü kullanılmıştır), 3) alanyazın

taraması, hedef kitle görüşmeleri, açık uçlu soru uygulamalarından madde yazma ve madde havuzu oluşturma, 4) madde havuzundaki aday maddeler için uzman görüşü alma, 5) Türkçe dil redaksiyonu yapma ve 6) deneme uygulaması yapma, geçerlik ve güvenirlik kanıtları sağlamadır.

Madde havuzu oluşturulması aşamasında iki stratejiden yararlanılmıştır. Bunlardan ilki ilgili alanyazından yararlanmadır. Alanyazın taramasında fikir verebilecek ölçme araçları ile problem davranışlara ait geniş alanyazının sağladığı bilgilerden aday maddeler oluşturulmuştur. İkinci strateji olarak hedef kitleye “açık uçlu soru uygulaması” yapılmıştır. Araştırmacılar tarafından hedef kitleye (ilkokul öğretmenleri) sorulmak üzere “Sınıflarda öğrencilerin sergilediği problem davranışlar nelerdir?” sorusu hazırlanmıştır. Google Forms üzerinden yapılan açık uçlu soru uygulamasına 171 öğretmen katılmıştır. Alınan yanıtlar analiz edilmiş ve aday ölçek maddeleri oluşturulmuştur. Hem ilgili alanyazın hem de öğretmenlerden elde edilen veriler ile 37 maddeden oluşan bir taslak ölçek elde edilmiştir.

37 maddelik bu taslak ölçek formuna uzman görüşü alabilmek için özel eğitim alanında uzman beş akademisyen, ölçme ve değerlendirme alanında uzman üç akademisyen, sınıf öğretmenliği alanında uzman bir akademisyen ve program geliştirme alanında uzman bir akademisyenden oluşan 10 kişilik bir uzman grubu oluşturulmuştur. Uzman grubundan görüş almak için bir form hazırlanmıştır. Forma her madde yazılmış ve maddenin karşısına “madde uygun ve ölçekte yer almalı”, “madde uygun değil, ölçekten çıkarılmalı” ve “madde düzenlenmek şartıyla ölçekte yer alabilir” seçenekleri konulmuş ve uzmanların her maddeye ilişkin değerlendirmelerini sunmaları istenmiştir. Ek olarak uzmanlardan varsa her madde için düzenleme önerilerini de yazmaları istenmiştir. Bu form aracılığıyla alınan uzman görüşleri ile gerekli düzenlemeler yapılmadan önce alınan uzman görüşlerindeki tutarlılık incelenmiştir. Bu inceleme Krippendorff Alpha katsayısı hesaplanarak yapılmıştır (Krippendorff, 2004). Hesaplanan tutarlılık değeri .897 olarak belirlenmiştir. Hesaplanan değer uzman görüşleri arasında tutarlılık gösterdiği için alınan uzman görüşlerine göre maddelerde düzenleme yapılmış ve maddelere son hali verilmiştir. Ayrıca alınan uzman görüşlerine Lawshe tekniği kullanılarak yapılan analiz ile maddelerin her biri için kapsam geçerliği oranı hesaplanmıştır (Lawshe, 1975). Kapsam geçerliği oranı düşük olan altı madde ölçekten çıkarılmış ve deneme uygulamasına çıkarılacak taslak formda 28 madde kalmıştır.

Deneme uygulamasına çıkacak 28 maddeden oluşan taslak ölçek formu açıklık, anlaşılabilirlik, yalınlık vb. bakımlardan değerlendirilmesi için Türk Dili alanında uzman bir akademisyenin görüşüne sunulmuştur. Dil redaksiyonundan geçen maddeler “hiçbir zaman”, “nadiren”, “sıklıkla” ve “her zaman” şeklinde dört dereceli Likert forma dönüştürülmüştür.

Veri Toplama ve Analizi

590 öğretmene “Sınıfınızda oldukça çok, ortalama düzeyde ve çok az problem davranış yaratan üç farklı öğrencinizi düşünün ve bu üç öğrenci için PDDÖ’yü ayrı ayrı doldurun” yönergesi verilmiştir. Bu nedenle 590 öğretmenden 1770 öğrenciye dair veri elde edilmiştir. 1770 öğrencinin problem davranışlarını öğretmen algısını yansıtan veri dosyası ölçek geliştirme alanyazınına uygun olarak (Hair vd., 2014) rastgele yöntemle iki ayrı veri dosyasına bölünmüştür. Bu işlem açıklayıcı faktör analizi (AFA) ve doğrulayıcı faktör analizi (DFA) için iki farklı zamanda veri toplama işleminin yürütülmesi yerine tek seferde yürütülmesi için gerçekleştirilmiştir.

Yapılan rastgele bölme işlemi sonucunda AFA veri dosyasında 915 öğrencinin, DFA veri dosyasında 855 öğrencinin verisi yer almıştır. AFA dosyasında bulunan verilerle PDDÖ’nün yapı geçerliğini keşfetmeye yönelik analizler gerçekleştirildiği gibi güvenirlik kanıtları da bu dosyadaki veri aracılığıyla hesaplanmıştır. DFA dosyasında bulunan verilerle yapısı keşfedilen PDDÖ’nün yapısı doğrulanmaya çalışılmış ve bu doğrultuda analizler gerçekleştirilmiştir. Ek olarak bölünmemiş, öğrencilere ait öğretmenlerin problem davranış algısını gösteren tüm verilerle (1770) MTK doğrultusunda analizler gerçekleştirilmiştir. Tüm veri dosyası, AFA ve DFA için rastgele bölme yöntemiyle oluşturulan veri dosyalarındaki verilerin tanımlayıcı istatistikleri Tablo 2’de gösterilmiştir.

Tüm veri dosyasında öğrenci yaş ortalaması 8.42, standart sapma 1.85’tir. AFA dosyasında öğrenci yaş ortalaması 8.37, standart sapma 1.49’dur. DFA dosyasında ise öğrenci yaş ortalaması 8.34 standart sapma ise 1.34’tür.

Araştırmada elde edilen veriler aşağıdaki aşamalardan geçilerek analiz edilmiştir:

1. Araştırmada PDDÖ aracılığıyla 590 sınıf öğretmeninden 1770 öğrenciye dair problem davranış algısını gösteren veri elde edilmiştir.

Tablo 2

Öğrencilere Ait Bazı Değişkenlerin Dağılımı

Değişken		Tüm veri dosyası n (%)	AFA veri dosyası n (%)	DFA veri dosyası n (%)
Öğrencinin cinsiyeti	Erkek	1027(58)	531(58)	481(56.3)
	Kız	743(42)	384(42)	374(43.7)
Öğrencinin uyruğu	TC	1563(88.3)	807(88.2)	758(88.7)
	Diğer	207(11.7)	108(11.8)	97(11.3)
Öğrencinin ailesinin sosyoekonomik düzeyi (SED)	Alt SED	365(20.6)	186(20.3)	169(19.8)
	Orta SED	1275(72)	664(72.6)	623(72.9)
	Üst SED	130(7.3)	65(7.1)	63(7.4)
Öğrencinin resmi tanısı var mı?	Hayır	1636(92.4)	844(92.2)	790(92.4)
	Evet	134(7.6)	71(7.8)	65(7.6)
	Az	790(44.6)	417(45.6)	414(48.4)
Öğrencinin problem davranış sergileme düzeyi	Orta	638(36)	328(35.8)	304(35.6)
	Çok	342(19.3)	170(18.6)	137(16)
Toplam		1770(100)	915(100)	855(100)

Not: TC = Türkiye Cumhuriyeti.

2. Elde edilen verilerde yanıt verilmemiş madde olmadığı tespit edilmiştir.
3. Veriler toplam veri (1770 öğrenciye ait), rastgele bölme yöntemiyle AFA verisi (915 öğrenciye ait veri) ve DFA verisi (855 öğrenciye ait veri) olmak üzere üç dosya halinde kaydedilmiştir.
4. Veriler JAMOVİ ve R istatistiksel yazılımlarına aktarılmıştır.
5. Her üç veri dosyasında bulunan veri setinin çok değişkenli normal dağılım gösterip göstermediği incelenmiştir. Bu inceleme R Paket Programı, “mvn” scripti (Korkmaz vd., 2014) ile gerçekleştirilmiştir.
6. AFA dosyasında Kaiser-Meyer-Olkin testi (KMO) gerçekleştirilmiştir. KMO, hesaplanan korelasyon katsayılarının büyüklüklerini kısmi korelasyon katsayılarının büyüklükleriyle karşılaştırır, bu yolla örneklem yeterliliği hakkında bilgi sunar (Pett vd., 2003). KMO değerinin yorumlanmasında .90 üstü “mükemmel”, .80-.90 “iyi”, .70-.80 “kabul edilebilir”, .60-.70 “orta düzeyde”, .50-.60 “düşük düzey” ve .50 altı “kabul edilemez” ölçütleri kullanılmıştır (Kaiser & Rice, 1974).
7. Bartlett’s Test of Sphericity aracılığıyla veri setinin birim matrisi (identity matrix) (köşegenleri 1, diğer elemanları 0 olan matris) özelliği gösterip göstermediği incelenmiştir. Birim matris özelliği gösteren bir korelasyon matrisi, faktör analizinde tercih edilmemektedir. Birim matriste maddeler arasında karşılıklı ilişki bulunmaz. Bartlett test sonucunun anlamlı olması veri setinin birim matris olmadığını göstermektedir ve faktör analizi için istenen durum budur (Field, 2018; Pett vd., 2003).
8. Principal Axis Factoring (PAF) yöntemiyle faktör analizi (AFA) gerçekleştirilmiştir. Sosyal ve davranış bilimleri araştırma raporlarında (makalelerinde) daha çok kullanılan PAF, bir dizi X ölçümündeki paylaşılan varyansı modellemektedir (Warner, 2013).
9. Faktör sayısını belirlemek için “öz değer (Eigenvalue)” incelemesi yapılmıştır. Öz değer referansı olarak “1” değeri alınmıştır. Öz değerleri 1’den büyük olan faktörler birer faktör adayı olarak anlamlı kabul edilmektedir. Değişken sayısı (madde sayısı) 20 ile 50 arasında olduğunda referans olarak öz değer kullanılması oldukça güvenilirdir.
10. Açıklanan toplam varyans miktarı incelenmiştir. Her ölçekte bir özellik ölçülmek istenmektedir. PDDÖ’de ise öğretmenlerin algısından öğrencilerinin problem davranış düzeyleri ölçülmek istenmektedir. PDDÖ bu özelliği ölçerken problem davranışlardaki varyansın (değişkenliğin) en az ne kadarı hakkında bilgi vermektedir? Açıklanan varyans düzeyi bu sorunun cevabıdır. Hair ve diğerlerine (2014) göre bilginin genellikle daha az kesin olduğu sosyal bilimlerde, toplam varyansın %60’ını (ve bazı durumlarda daha da azını) karşılayan bir çözüm tatmin edicidir. Warner’a (2013) göre yeterli olarak kabul edilebilir sınırlar %40 ile %70 arasındadır.
11. Faktörler altında gruplanan maddelerin elde ettikleri yük değerleri (korelasyon değerleri) incelenmiştir. Faktör yapısının yorumlanmasında faktör yük değerleri için aranması gereken minimum değerler $\pm .30$

ile $\pm .40$ aralığıdır. $\pm .50$ önemli kabul edilmesi gereken bir faktör yük değeridir. $\pm .70$ ve üzeri ise iyi tanımlanmış bir yapının göstergesi olarak kabul edilmektedir (Hair vd., 2014).

12. DFA analizinde, kurulan modelin uygunluk derecesi incelenmiştir. DFA için belirlenen uyum indekslerinde (fit-index) alanyazına göre referans değerler olarak RMSEA için $0.05 < RMSEA \leq 0.08$ kabul edilebilir, $0 > RMSEA \leq 0.05$ mükemmel, TLI ve CFI için .95 ve üzeri mükemmel, X^2/sd için $2 < X^2/sd \leq 5$ kabul edilebilir, $0 > X^2/sd \leq 2$ mükemmel aralık olarak alınmıştır (Anderson & Gerbing, 1984; Bentler, 1990; Hooper vd., 2008; Hu & Bentler, 1999; Kline, 2005; Marsh vd., 1988; Tabachnick & Fidell, 2013; Vieira, 2011).
13. Ölçeğin güvenirlik düzeyi belirlenmiştir. Güvenirlik düzeyi, Cronbach Alpha, McDonald Omega ve Marjinal Güvenirlik katsayısı ile hesaplanmıştır. Nunnally ve Bernstein'a (1994) göre yeterli güvenirlik en az .70 ve üzerinde olmalıdır.
14. MTK ile geçerlik ve güvenirlik incelemeleri yapılmıştır. İlk olarak tek boyutluluk (unidimensionality) ve yerel bağımsızlık (local independence) varsayımları incelenmiştir (Zhao, 2008). Tek boyutluluk bireylerin ölçme aracındaki performansını etkileyen bir özelliğin (ölçme aracının ilgili maddelerinin yalnızca bir özelliğe yönelik olması) gerektirmektedir (Hambleton vd., 1991). Tek boyutluluk AFA ile değerlendirilmiştir. Yerel bağımsızlık varsayımı Q3 istatistiği (Yen, 1993) kullanılarak test edilmiş ve MTK kalibrasyonları R v.4.1.2 programı içinde yer alan "Mirt v.1.30" (Chalmers, 2012) paket programı ile yapılmıştır.
15. Madde ayırt edicilik ve güçlük düzeyleri incelenmiştir. MTK'da ideal bir ölçek maddesinin ayırt edicilik değeri (yani "a" parametresi) .5 ile 2 arasında olmalıdır. Alanyazında bu parametrenin .75 ile 2.50 aralığında olması kabul edilir aralıkta olduğunu gösterir (Flannery vd., 1995). Madde güçlük düzeyleri (yani "b" parametresi) için ideal (orta güçlük düzeyi) sınırlar -1.00 ile 1.00 arasında kabul edilmektedir (Hambleton, 1994). Yetenek ya da başarı testlerinde güçlük düzeyi -1.00'den düşük maddeler kolay, 1.00'dan büyük maddeler ise zor olarak nitelendirilmektedir.
16. Madde bilgi fonksiyonları ve madde karakteristik eğrileri incelenmiştir.

Bulgular

PDDÖ'nün KTT Geçerlik ve Güvenirlik Kanıtları

915 öğrenciye ait AFA verisinin çok değişkenli normal dağılımı incelenmiş ve veri setinin çok değişkenli normal dağılım göstermediği belirlenmiştir. KMO değeri .970 olarak belirlenmiştir. Bu durumda veri dosyasının kısmi korelasyonlar bağlamındaki incelemesinin faktör analizi için uygun olduğu söylenebilir. Bartlett test sonucu 15671 olarak ve test sonucu anlamlı ($p < .05$) belirlenmiştir. Bu sonuç veri matrisinin birim matris olmadığını göstermektedir ve faktör analizi için uygun bir matris yapısı söz konusudur.

Veri analizi bölümünde açıklandığı üzere faktör belirleme yöntemi olarak PAF yöntemi tercih edilmiştir. Analize PDDÖ taslak formunda bulunan 28 madde ile girilmiştir. Bazı maddeler analizde alanyazının önerdiği faktör korelasyon yük değerlerine ulaşamamıştır. Bu maddeler; 1, 3, 4, 5, 6, 10, 16, 18, 26 ve 27 numaralı maddelerdir. Ölçekten dışlanan bu maddelerden sonra ölçekte 18 madde kalmıştır. Kalan 18 madde iki faktör altında toplanmıştır. Faktörleşmeyi netleştirmek, faktörler altında yer alan maddeleri tespit edebilmek için öz değer, yamaç grafiği (scree plot) incelemesi ve oblimin döndürme yönteminden yararlanılmıştır. Oblimin döndürme yöntemi oluşan iki faktörün birbiri ile yüksek ilişkili olmasından dolayı seçilmiştir. AFA kapsamında yapılan analiz sonuçları Tablo 3'te gösterilmiştir.

Tablo 3

AFA Analiz Sonuçları

	Maddeler	Faktörler				Öz değer (Eigenvalue)	Açıklanan varyans düzeyi	Faktör adı	Tersten kodlanan maddeler	Alınabilecek en düşük puan	Alınabilecek en yüksek puan
		F1 FLVR	ITC	F2 FLVR	ITC						
I28	Arkadaşlarına zorbalık yapar.	.978	.852								
I14	Arkadaşlarıyla kavga eder.	.878	.877								
I8	Arkadaşlarıyla alay eder.	.863	.809								
I19	Arkadaşlarına zarar verir.	.850	.854								
I24	Haksız suçlamalarda bulunur.	.817	.817								
I25	Küfürlü/argo sözcükler kullanır.	.806	.817			6.72	37.3	Sınıf iklimini etkileyen problem davranışlar	Yok	10	40
I21	Arkadaşlarını sürekli şikâyet eder.	.744	.766								
I11	Arkadaşlarının dersi dinlemesini engelleyici davranışlarda bulunur.	.623	.826								
I12	Ders sırasında sınıfta gezinir.	.528	.709								
I22	Dersten çıkmak için bahaneler üretir.	.492	.741								
I20	Ev ödevlerini yapmaz.			.938	.792						
I7	Sınıf içinde verilen görevleri yerine getirmez.			.851	.848						
I2	Ders araç-gereçlerini yanında getirmez.			.837	.783						
I9	Grup etkinliklerine katılmaz.			.785	.790						
I17	Dersi dinlemez.			.765	.843	5.59	31	Kendi öğrenmesini etkileyen problem davranışlar	Yok	8	32
I23	Ders araç-gereçlerini dersin amaçlarına uygun olarak kullanmaz.			.765	.820						
I13	Grup etkinliklerinin kurallarına uymaz.			.606	.817						
I15	Gereksinim duysa da yardım istemez.			.522	.555						

Not: F1 = faktör 1; F2 = faktör 2; FLVR = dönüşüm sonrası faktör yük değeri; ITC = madde toplam korelasyonu.

Faktör 1'deki maddeler anlamsal olarak okunduğunda maddelerin, sınıf iklimini etkileyen problem davranışlarla ilgili olduğu belirlenmiştir. Bu faktörde 10 madde yer almıştır. Tersten kodlanan (puanlanan) madde bulunmamaktadır. Faktör altındaki maddeler .492 ile .978 arası korelasyon yükü, .709 ile .877 madde toplam korelasyonu göstermiştir. Bu faktör 6.72 öz değere sahip ve tek başına öğrenci problem davranışlarındaki varyansın %37'sini açıklamaktadır.

Faktör 2'deki maddeler anlamsal olarak okunduğunda maddelerin, öğrencinin kendi öğrenmesini etkileyen problem davranışlarıyla ilgili olduğu belirlenmiştir. Bu faktörde 8 madde yer almıştır. Tersten kodlanan (puanlanan) madde bulunmamaktadır. Faktör altındaki maddeler .522 ile .938 arası korelasyon yükü, .555 ile .848 madde toplam korelasyonu göstermiştir. Bu faktör 5.59 öz değere sahip ve tek başına öğrenci problem davranışlarındaki varyansın %31'ini açıklamaktadır. PDDÖ güvenirlilik düzeyi CTT'ye göre Cronbach Alpha ve McDonald's Omega güvenirlilik analizleri ile ölçek genel toplam puanları ve ölçek faktörleri için incelenmiştir. Tablo 4'te güvenirlilik analizi sonuçları verilmiştir.

Tablo 4

PDDÖ Güvenirlilik Analizi Sonuçları

Ölçek	Cronbach Alpha	McDonald's Omega
F1 sınıf iklimini etkileyen problem davranışlar	.956	.956
F2 kendi öğrenmesini etkileyen problem davranışlar	.938	.939
PDDÖ toplam puanları	.966	.967

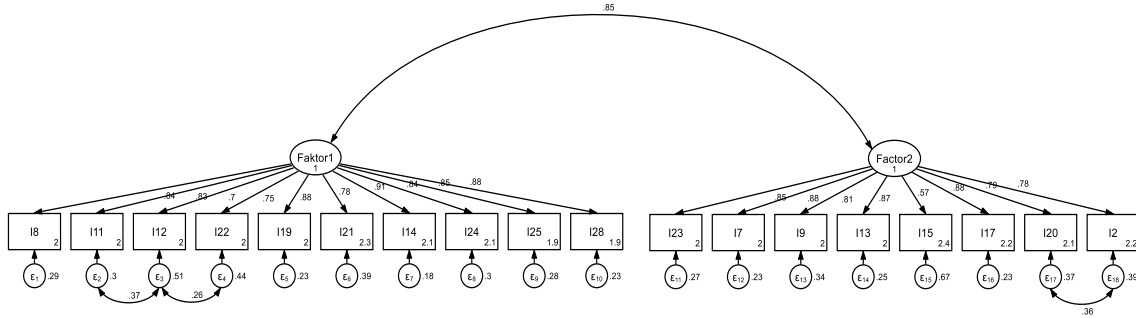
Not: F1 = faktör 1, F2 = faktör 2, PDDÖ = problem davranışları değerlendirme ölçeği.

PDDÖ ve faktörleri için hesaplanan güvenirlilik değerleri .900 üzerindedir. Veri analizi bölümünde açıklandığı üzere .700 ve üzeri güvenirlilik değeri yüksek güvenirlilik düzeyi olarak kabul edilmektedir. Ölçek ve faktörleri için yüksek güvenirlilik düzeyi elde edilmiştir.

CTT kapsamında uygulanan AFA sonucunda PDDÖ'nün 18 madde ve iki faktörden oluşan yapısı keşfedilmiştir. Keşfedilen bu yapı genel veri dosyasından rastgele bölme yöntemi ile oluşturulmuş 855 öğrenciye ait DFA veri dosyasında doğrulanmıştır. Gerçekleştirilen DFA modellemesi sonucu oluşan diyagram Şekil 1'de gösterilmiştir.

Şekil 1

PDDÖ Doğrulayıcı Faktör Analizi Diyagramı ve Standartlaştırılmış Korelasyon Değerleri



DFA'da elde edilen uyum indeksleri (fit-indexes); CFI = .953, TLI = .945, RMSEA = .079, $\chi^2/df = 4.218$ olarak hesaplanmıştır. DFA analizinde elde edilen bu uyum indeksleri alanyazına göre kabul edilebilir sınırlar içerisinde. Bu sonuçlara göre PDDÖ'nin iki faktörlü yapısının doğrulandığı sonucuna ulaşılmıştır.

PDDÖ'nün MTK Geçerlik ve Güvenirlilik Kanıtları

590 sınıf öğretmeni aracılığıyla 1770 öğrenciye dair toplanan veriler gerçekleştirilen MTK analizlerinde veri seti olarak kullanılmıştır. CTT kanıtlarına göre PDDÖ iki faktörden oluşmuştur. Her iki faktör bağımsız faktör olarak kabul edilmiş ve IRT analizleri gerçekleştirilmiştir. Faktör 1 yani Sınıf İklimini Etkileyen Problem Davranışlar olarak isimlendirilebilecek faktör için yerel bağımsızlık test edilmiştir ve yerel bağımsızlığı bozan herhangi bir madde tespit edilememiştir. MTK'de madde kalibrasyonları Generalized Partial Credit Model

(GPCM) ile gerçekleştirilmiştir. GPCM'ye göre Faktör 1'de yer alan maddelere dair hesaplanan istatistikler Tablo 5'te verilmiştir.

Tablo 5

PDDÖ Faktör 1 (Sınıf İklimini Etkileyen Problem Davranışlar) İçin Madde Uyum İndeksleri (Fit-Indexes)

Madde	$S \chi^2$	GPCM df	RMSEA
I8	83.707	39	0.025
I11	83.823	42	0.024
I12	73.596	53	0.015
I14	54.639	32	0.020
I19	29.409	31	0.000
I21	66.909	45	0.017
I22	107.189	50	0.025
I24	59.274	39	0.017
I25	76.360	38	0.024
I28	36.914	33	0.008

MTK kapsamında madde uyumuna gösteren RMSEA değerlerinin incelenmesi oldukça önemlidir. RMSEA için uyum gösteren ölçüt düzey .080 ve altındır (Stout, 1990). PDDÖ Faktör 1 içinde yer alan 10 madde için RMSEA değerleri .025 ve altındadır. Maddelere ilişkin gerekli uyum düzeyi elde edilmiştir. Faktör 1'deki 10 madde için "a" (madde ayırt edicilik) ve "b" (madde güçlüğü) parametreleri ile standart hataları (standart error) belirlenmiştir. Sonuçlar Tablo 6'da sunulmuştur.

Tablo 6

PDDÖ Faktör 1 (Sınıf İklimini Etkileyen Problem Davranışlar) İçin Madde Parametreleri

Madde	$a(SE)$	$b1(SE)$	$b2(SE)$	$b3(SE)$
I8	3.113 (0.156)	-0.023 (0.035)	0.937 (0.041)	1.669 (0.061)
I11	2.993 (0.148)	-0.208 (0.037)	0.575 (0.037)	1.433 (0.052)
I12	1.828 (0.092)	0.126 (0.045)	0.975 (0.053)	1.809 (0.084)
I14	5.174 (0.278)	-0.320 (0.033)	0.613 (0.032)	1.416 (0.044)
I19	5.278 (0.302)	0.071 (0.031)	0.860 (0.034)	1.652 (0.053)
I21	2.381 (0.114)	-0.618 (0.042)	0.598 (0.040)	1.572 (0.061)
I22	1.934 (0.098)	0.268 (0.044)	0.976 (0.051)	2.116 (0.098)
I24	3.201 (0.160)	-0.041 (0.035)	0.876 (0.039)	1.775 (0.064)
I25	3.450 (0.181)	0.147 (0.034)	0.915 (0.039)	1.608 (0.057)
I28	4.821 (0.273)	0.132 (0.031)	0.899 (0.036)	1.564 (0.051)

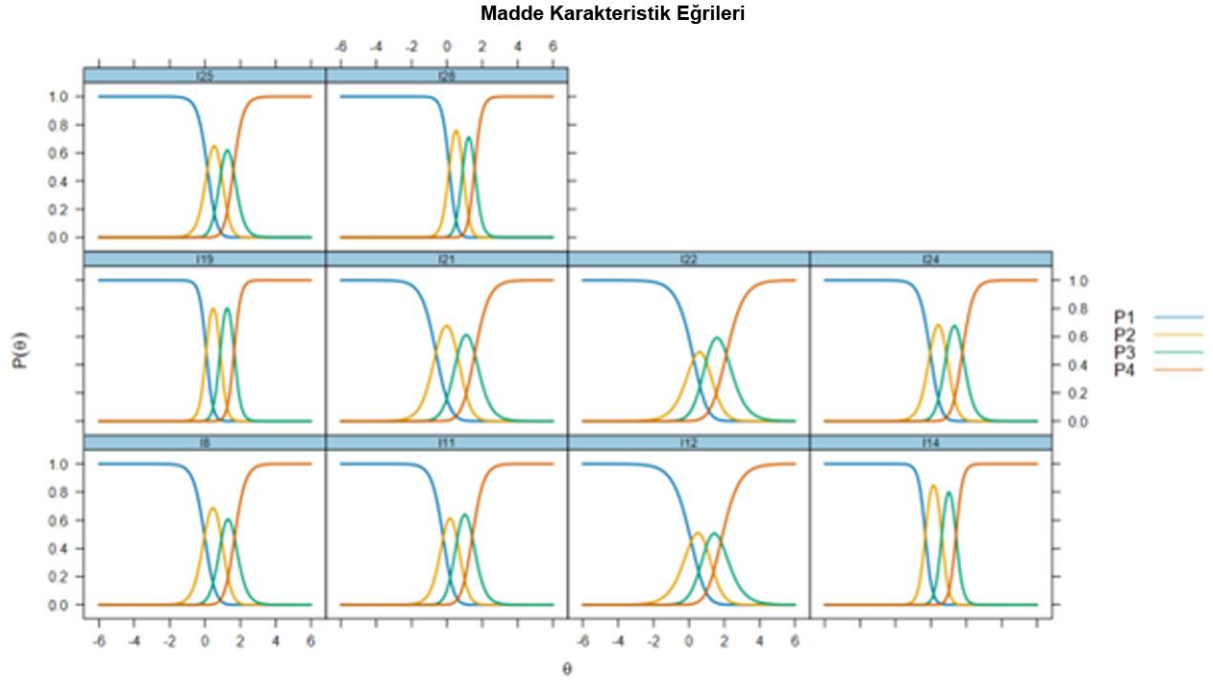
Not: Iteration = 163; LogLikelihood: - 13840.758; $p < .05$.

Ölçek maddelerinde her bir madde için ideal ayırt edicilik .5 ile 2 arasında kabul edilmekle birlikte "a" parametresi için .75 ile 2.50 aralığı da kabul edilir değerlerdir. PDDÖ aktör 1 altında yer alan 10 madde içinde I12, I21 ve I22 ideal ayırt edicilik düzeyini yakalamıştır. GPCM'ye göre yapılan kestirimlere (LogLikelihood, $p < .05$) göre madde uyumu sağlanmıştır. Madde karakteristik eğrileri Şekil 2'de gösterilmiştir.

Madde karakteristik eğrileri Faktör 1 altında yer alan 10 maddenin hemen hemen tamamının yanıt setindeki seçenekleriyle birlikte iyi çalıştığını göstermektedir. Madde bilgi fonksiyonları Şekil 3'te gösterilmiştir. Madde bilgi fonksiyonu maddenin ölçme aracını alan bireyleri en iyi ayırt ettiği özellik (ölçekte ölçülmeye çalışılan özellik) aralığını gösteren bir grafiksel gösterimdir (Edelen & Reeve, 2007). Madde bilgi fonksiyonunda eğrinin tepe noktası ne kadar yükselirse madde o kadar çok bilgi vermektedir anlamı çıkarılır. PDDÖ Faktör 1 maddelerinin, madde bilgi fonksiyonlarına bakıldığında en çok bilgi elde edilen maddeler 14, 19, 25 ve 28'dir. 12, 21 ve 22. maddeler diğer maddelere göre daha az bilgi elde edilen maddelerdir.

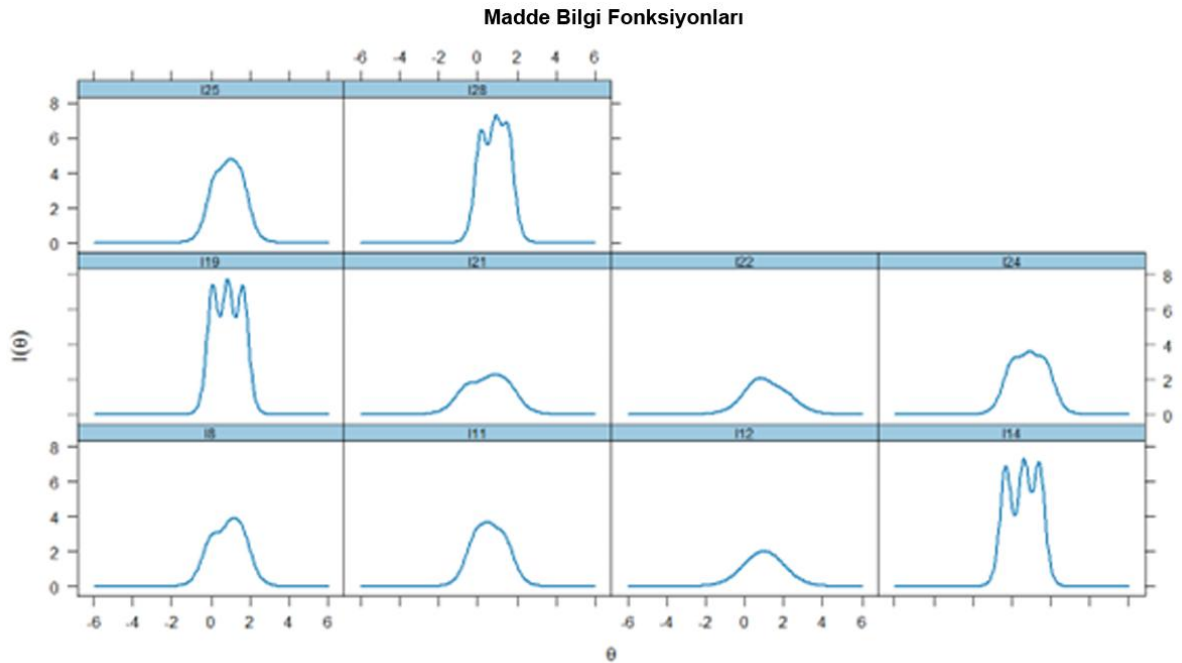
Şekil 2

PDDÖ Faktör 1 Maddelerine Ait Madde Karakteristik Eğrileri



Şekil 3

PDDÖ Faktör 1 Maddelerine Ait Madde Bilgi Fonksiyonları



Faktör 2 yani Kendi Öğrenmesini Etkileyen Problem Davranışlar olarak isimlendirilebilecek faktör için yerel bağımsızlık test edilmiştir. Bu test sonucunda yerel bağımsızlığı bozan herhangi bir madde tespit edilememiştir. Faktör 2’de yer alan maddelere dair hesaplanan istatistikler Tablo 7’de verilmiştir.

Tablo 7

PDDÖ Faktör 2 (Kendi Öğrenmesini Etkileyen Problem Davranışlar) için Madde Uyum İndeksleri (Fit-Indexes)

Madde	$S \chi^2$	GPCM df	RMSEA
I2	45.495	31	.016
I7	40.743	26	.018
I9	62.533	29	.026
I13	41.415	28	.016
I15	96.480	43	.027
I17	48.369	27	.021
I20	76.344	32	.028
I23	52.812	28	.022

PDDÖ faktör 2 içinde yer alan sekiz madde için RMSEA değerleri .028 ve altındadır. Maddelere ilişkin gerekli uyum düzeyi elde edilmiştir. Faktör 2'deki sekiz madde için “a” (madde ayırt edicilik) ve “b” (madde güçlüğü) parametreleri ile standart hataları (standart error) belirlenmiştir. Sonuçlar Tablo 8’de sunulmuştur.

Tablo 8

PDDÖ Faktör 2 (Kendi Öğrenmesini Etkileyen Problem Davranışlar) için Madde Parametreleri

Madde	$a(SE)$	$b1(SE)$	$b2(SE)$	$b3(SE)$
I2	2.733 (0.135)	-.370 (0.038)	.902 (0.042)	1.789 (0.069)
I7	4.233 (0.229)	-.151 (0.033)	.725 (0.035)	1.609 (0.053)
I9	3.091 (0.162)	.025 (0.035)	1.047 (0.043)	1.743 (0.066)
I13	3.504 (0.185)	-.069 (0.034)	.728 (0.037)	1.680 (0.059)
I15	1.169 (0.063)	-.598 (0.059)	1.387 (0.079)	2.227 (0.133)
I17	4.109 (0.216)	-.432 (0.035)	.600 (0.034)	1.448 (0.048)
I20	2.699 (0.134)	-.224 (0.037)	.890 (0.042)	1.647 (0.063)
I23	3.572 (0.189)	-.033 (0.034)	.840 (0.038)	1.743 (0.061)

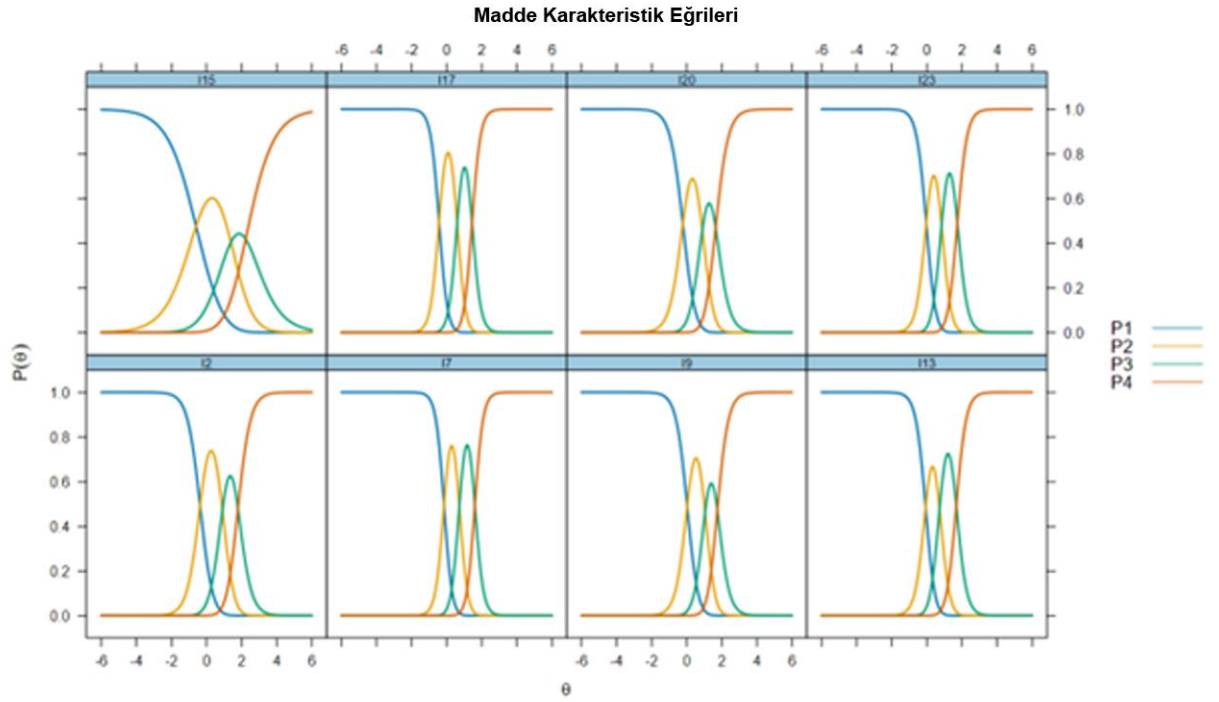
Not: Iteration = 124; LogLikelihood = - 11707.660; $p < .05$.

PDDÖ faktör 2 altında yer alan sekiz madde içinde M2, M15 ve M20 ideal ayırt edicilik düzeyini yakalamıştır. GPCM’ye göre yapılan kestirimlere (LogLikelihood, $p < .05$) göre madde uyumu sağlanmıştır. Madde karakteristik eğrileri Şekil 4’te gösterilmiştir.

Madde karakteristik eğrileri faktör 2 altında yer alan sekiz maddenin hemen hemen tamamının yanıt setindeki seçenekleriyle birlikte iyi çalıştığını göstermektedir. Madde bilgi fonksiyonları Şekil 5’te gösterilmiştir. PDDÖ Faktör 2 maddelerinin, madde bilgi fonksiyonlarına bakıldığında en çok bilgi veren maddeler 2, 7, 9, 13, 17, 20 ve 23’tür. 15. madde diğer maddelere göre daha az bilgi vermektedir. Hesaplanan marjinal güvenirlik katsayı Faktör 1 için 893, Faktör 2 için .887’dir. Veri analizi bölümünde açıklandığı üzere, alanyazına göre bu değer yüksek güvenirlik düzeyi göstergesidir.

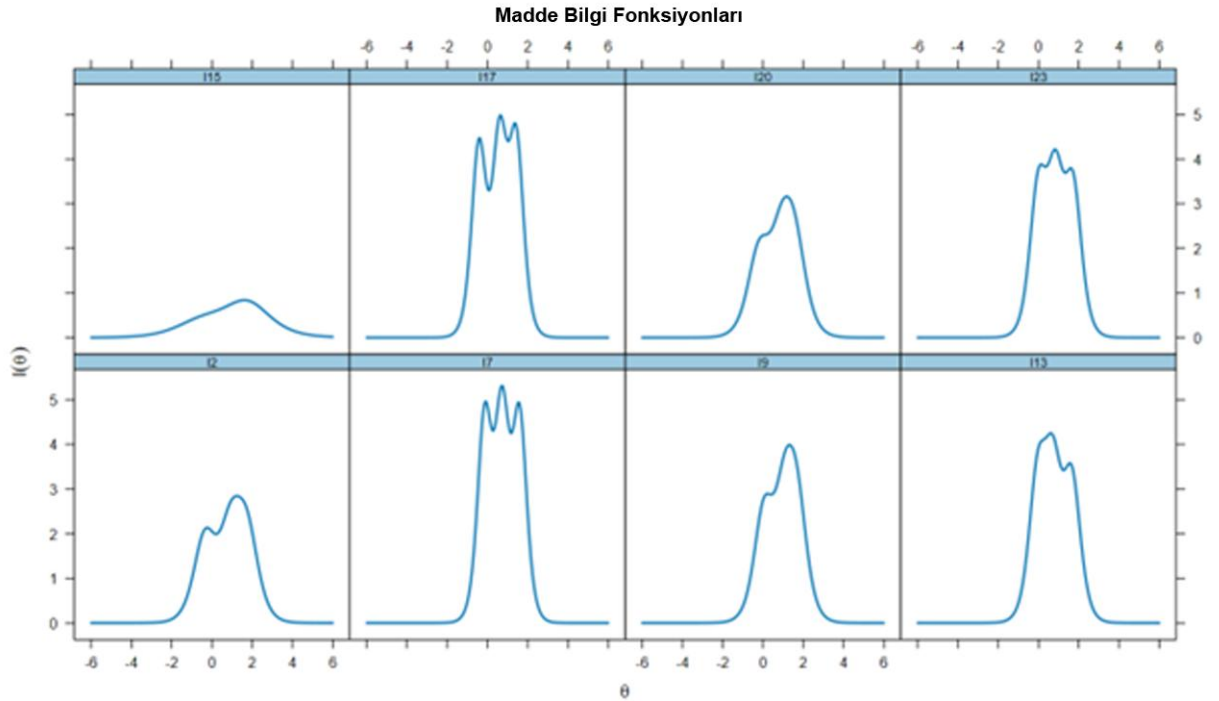
Şekil 4

PDDÖ Faktör 2 Maddelerine Ait Madde Karakteristik Eğrileri



Şekil 5

PDDÖ Faktör 2 Maddelerine Ait Madde Bilgi Fonksiyonları



Tartışma

Bu çalışma, öğretmenlerin okul ve sınıf ortamlarındaki gözlemlerine ve algılarına dayalı olarak öğrencilerin problem davranışlarını değerlendirme amacına hizmet edecek Problem Davranışları Değerlendirme

Ölçeğin klasik test kuramı ve madde tepki kuramına göre geçerlik ve güvenirlik kanıtlarını incelemek amacıyla gerçekleştirilmiştir. Bu amaç doğrultusunda planlanan araştırmaya ilkokullarda görev yapan toplam 590 sınıf öğretmeni katılmıştır. İlk aşamada, ölçeğin amacı ve konusuna uygun olarak alanyazın taraması yapılmıştır. Alanyazın taraması ve uzman görüşleri sonucunda ölçeğe ilişkin madde havuzu oluşturulmuştur. PDDÖ'nün geçerlik kanıtları klasik test kuramına göre açıklayıcı faktör analizi ve madde tepki kuramına göre madde karakteristik eğrileri, madde bilgi fonksiyonları, madde ayırt edicilikleri ile incelenmiştir. Sonuç olarak PDDÖ'nün 18 madde ve iki faktörlü geçerli bir ölçek olduğu belirlenmiştir. PDDÖ'nün güvenirliği klasik test kuramına göre Cronbach Alpha ve McDonald's Omega katsayıları, madde tepki kuramına göre ise marjinal güvenirlik katsayısı ile belirlenmiştir. Sonuçlar PDDÖ'nün yüksek güvenirlik düzeyinde olduğunu göstermiştir. PDDÖ'nün 18 madde ve iki faktörlü yapısı ile karmaşık olmayan, kolay yanıtlanabilir ve kullanışlı bir ölçek olduğu söylenebilir. Analiz sonucunda birinci faktör 10, ikinci faktör sekiz maddeden oluşmuştur.

Maddelerin özellikleri ve kapsamına uygun olarak birinci faktör “sınıf iklimini etkileyen problem davranışlar” ve ikinci faktör “kendi öğrenmesini etkileyen problem davranışlar” olarak adlandırılmıştır. Alanyazın incelendiğinde geliştirilen ya da uyarlaması yapılan ölçeklerde farklı maddelerin farklı faktör sayısı ve isimleriyle yer aldığı görülmektedir. Örneğin, öğretmenlerinin gözlemlerine dayalı olarak öğrencilerin sınıfa uyum davranışlarının değerlendirilmesini amaçlayan ve Werthamer-Larsson ve diğerleri (1991) tarafından geliştirilen TOCA-R, 43 maddelik üç faktörden (konsantrasyon problemleri [concentration problems], yıkıcı/saldırgan davranış [disruptive/aggressive behavior] ve utangaç davranış [shy behavior]) oluşmaktadır. Leaf ve diğerleri (2002) tarafından geliştirilip Koth ve diğerleri (2009) tarafından geçerlik ve faktör yapısı incelenen TOCA-C 25 maddelik üç faktörden (konsantrasyon problemleri [concentration problems], yıkıcı davranış [disruptive behavior] ve olumlu sosyal davranış [prosocial behavior]) oluşmaktadır. TOCA-C'nin okul öncesi eğitim kurumlarından elde edilen verilerle Türkçe'ye uyarlaması ise TOCA-C ile aynı faktörlere sahip toplam 21 maddeden oluşmaktadır (Aydos vd., 2014). Alanyazın incelendiğinde geliştirilen ya da uyarlaması yapılan ölçeklerin genellikle doğrudan öğrencilerin sınıfa uyum davranışlarını (uygun ve uygun olmayan davranışlar) ölçtüğü görülmekte ve dolayısıyla üçüncü bir faktör olan olumlu sosyal davranışlar ya da benzeri adlı bir faktör bulunmaktadır. Ancak PDDÖ doğrudan problem davranışları değerlendirmeyi amaçladığından faktörler arasında olumlu sosyal davranışlar yer almamaktadır. Bu araştırmada geliştirilen PDDÖ bu yönüyle daha önce geliştirilmiş ölçeklerden farklılık göstermektedir. Diğer yandan Quay ve Peterson, (1996) tarafından geliştirilen ve Kaner ve Çiçekçi, (2000) tarafından Türkçe formunun psikometrik özellikleri incelenen Gözden Geçirilmiş Davranış Problemleri Kontrol Listesi (Revised Problem Behavior Checklist) her ne kadar doğrudan problem davranışları ölçmeyi amaçlasa da madde sayısı uygulamada güçlükler yaratacak miktardadır. Ölçekler incelendiğinde asıl ölçekte 89 maddelik altı faktör (Davranım Bozukluğu, Toplumsallaşmış Saldırganlık, Dikkat Problemleri-Toyluk, Kaygı- Gerçekkilme, Psikotik Davranış, Motor Gerilim) yer alırken Türkçe uyarlamasında 56 maddelik üç faktörün (Davranım Bozukluğu-Hiperaktivite ve İmpulsivite, Depresyon - Dikkat Eksikliği ve Toplumsallaşmış Saldırganlık) yer aldığı görülmektedir. Bu bakımdan PDDÖ'nün 18 maddelik iki faktörlü yapısıyla kullanımı nispeten kolay bir ölçek olduğu söylenebilir. Bunun yanında öğretmenlerin sınıftaki problem davranışları belirlemede ek bir değerlendirmeye gereksinim duymadan kullanabilecekleri bir ölçek olması açısından da uygulama ortamları için PDDÖ'nün oldukça kullanışlı bir ölçek olduğu düşünülmektedir. Alanyazında bir davranış değerlendirme yöntemi olarak doğrudan gözlemin olumlu yanlarıyla birlikte ek iş gücü ve zaman gerektirmesi açısından kullanım zorlukları vurgulanırken öğretmenlerin kolayca kullanabilecekleri davranış değerlendirme ölçeklerinin ise verimlilikle ilişkilendirildiği görülmektedir (Chafouleas vd., 2009). Bu açılarından PDDÖ'nün uygulama ortamlarında öğrencilerin problem davranışlarını belirlemede kullanılabilecek pratik ve kullanışlı bir ölçek olduğu düşünülmektedir.

Mevcut araştırmada geliştirilen PDDÖ'nün iki faktöründen birisi “Sınıf İklimini Etkileyen Problem Davranışlar” olarak adlandırılmıştır. Bu faktör isminde geçen sınıf iklimi ifadesi, alanyazında sınıf atmosferi, sınıf ortamı, öğrenme çevresi gibi farklı adlandırmalarla kullanılabilmektedir (Adelman & Taylor, 2005). Sınıf iklimi, sınıftaki öğretmen-öğrenci ile öğrenci-öğrenci arasındaki ilişkiler, sınıf kuralları, sınıfın fiziksel yapısının sınıfta yarattığı psikolojik, sosyal ve fiziksel etkiler şeklinde tanımlanmaktadır (Açıkgöz, 1998). Uluslararası ve ulusal alanyazında doğrudan sınıf iklimini değerlendirmeyi amaçlayan farklı okul kademeleri için hazırlanmış çeşitli ölçek ve ölçek uyarlamaları bulunmaktadır (ör. Çengel & Türkoğlu, 2015; Dwyer vd., 2004; Özbay & Şahin, 2000; Sağkal vd., 2015). Alanyazın incelendiğinde genellikle öğretmenin sınıf yönetimi ve öğretim stili ile öğretmen-öğrenci etkileşimini kapsayan boyutların sınıf iklimini oluşturduğu görülmektedir. (Evans vd., 2009). Yapılan araştırmalarda sınıf iklimi, derse katılım ve motivasyon (Ryan & Patrick, 2001), ders başarısı (Pierce, 1994), suç işleme durumu (Sprott, 2004), derse ilişkin tutum ve özsaygı (Chionh & Fraser, 2009), kişinin akıl sağlığı (Somersalo vd., 2002) gibi değişkenlerle ilişkilendirilmektedir. Bu araştırmada geliştirilen PDDÖ'nün

“Sınıf İklimini Etkileyen Problem Davranışlar” faktörünü oluşturan maddeler de alanyazında belirtilen içeriği desteklemektedir.

PDDÖ'nin ikinci faktörü öğrencilerin kendi öğrenmelerini olumsuz etkileyen problem davranışlarıyla ilgilidir. Bu faktörde yer alan maddeler incelendiğinde öğrencilerin başarılı olmaları için gerekli derse katılım ve görevi yerine getirme davranışlarına ilişkin problem davranışların yer aldığı görülmektedir. Alanyazında öğrencilerin okulda öğrendiklerinin ders etkinliklerine katılmalarının düzeyiyle doğrudan ilişkili olduğu vurgulanmaktadır (Brophy,1979; Greenwood vd., 2002). Aynı zamanda araştırmalar yüksek düzeyde katılımın başarıyla, düşük düzeyde katılımın ise problem davranışlarla ilişkili olduğunu ortaya koymaktadır (Baker vd., 2008; Noltemeyer vd., 2015). Ders etkinliklerine düşük düzeyde katılım sadece öğrencinin kendisi için değil, sınıftaki diğer öğrenciler ve öğretmen için de çeşitli zorluklar ortaya çıkarmaktadır (Alter vd., 2013). Ek olarak düşük düzeyde derse katılım öğrencilerin uzun dönemli başarıları açısından da önemli bir gösterge olarak kabul edilmektedir (Skinner vd., 1998). Öğretmen ve öğrenciler açısından olumsuz sonuçlara yol açan bu davranışlar PDDÖ'nün ikinci faktörüyle ölçülebilecektir. Sonuç olarak PDDÖ'nün iki faktörü ile sınıf iklimini, öğretmen etkililiğini ve öğrencilerin yakın ve uzak vadeli başarılarını doğrudan etkileyebilecek problem davranışların doğru ve pratik bir şekilde değerlendirilebileceği düşünülmektedir.

Geliştirilen bu ölçek Türkiye’de ilkokullarda öğretmenlerin gözlem ve algılarına dayalı olarak öğrencilerin problem davranışlarının değerlendirilmesi amacıyla kullanılabilecektir. Ölçeğin problem davranışlara ilişkin yürütülecek bilimsel çalışmalara destek sağlayabileceği düşünülmektedir. Bunun yanında PDDÖ'nün ilkokula devam eden öğrencilerin problem davranışlarının hızlı bir şekilde belirlenerek gerekli düzenlemeler yapılmasına katkı sağlayacağı da umulmaktadır.

Sonuç olarak, bu araştırmada bir ölçek geliştirme çalışması yapılmıştır. Araştırmanın amacı doğrultusunda ilkokullarda görev yapan sınıf öğretmenleri ile çalışılmıştır. Farklı okul türleri ve kademelerinde, branş öğretmenleri ve özel eğitim öğretmenleri ile çalışma genişletilebilir. Bu ölçek yalnızca öğretmenlerin yanıtlarına dayalı olarak sonuçlandırılmıştır. Okullarda gözlemler, nitel görüşmeler yapılabilir, öğretmen tutanakları, öğrenci çıktıları vb. incelenerek çalışmanın kapsamı genişletilebilir. Ölçeğin veri toplama sürecinde ulaşılan katılımcı sayısı analizler için yeterli görünse de ileriki araştırmalarda ölçeğin farklı katılımcılar ile geçerlik ve güvenirlik çalışması tekrar test edilebilir.

Kaynaklar

- Achenbach, T. M. (1978). The child behavior profile: I. boys aged 6-11. *Journal of Consulting and Clinical Psychology*, 46(3), 478-488. <https://doi.org/10.1037/0022-006X.46.3.478>
- Achenbach, T. M., & Edelbrock, C. S. (1979). The Child Behavior Profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology*, 47(2), 223-233. <https://doi.org/10.1037/0022-006X.47.2.223>
- Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, 21(8), 265-271. <https://doi.org/10.1542/pir.21-8-265>
- Açıkgöz, K. Ü. (1998). *Etkili öğrenme ve öğretme* (2. baskı). Kanyılmaz Matbaası.
- Adelman, H. S., & Taylor, L. (2005). *Classroom climate*. In S. W. Lee (Ed.), *Encyclopedia of school psychology* (pp. 88-90). Sage Publications.
- Akkök, F., Aşkar, P., & Sucuoğlu, B. (1986). *Öğretmen gözlem formu el kitabı*. Türk Eğitim Derneği Yayınları.
- Aloe, A. M., Shisler, S. M., Norris, B. D., Nickerson, A. B., & Rinker, T. W. (2014). A multivariate meta-analysis of student misbehavior and teacher burnout. *Educational Research Review*, 12, 30-44. <https://doi.org/10.1016/j.edurev.2014.05.003>
- Alter, P., Walker, J. N., & Landers, E. (2013). Teachers' perceptions of students' challenging behavior and the impact of teacher demographics. *Education and Treatment of Children*, 36(4), 51-69. <https://www.jstor.org/stable/42900226>
- Aman, G. M., Singh, N. N., Stewart, A. W., & Field, C. J. (1985a). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, 89(5), 485-491.
- Aman, G. M., Singh, N. N., Stewart, A. W., & Field, C. J. (1985b). Psychometric characteristics of the aberrant behavior checklist. *American Journal of Mental Deficiency*, 89(5), 492-502.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173. <https://doi.org/10.1007/BF02294170>
- Aydos, E. H., Uysal, H., & Akman, B. (2014). Sınıfa Uyum Öğretmen Gözlemi-Kontrol Listesi'nin Türkçeye uyarılama çalışması. *Kastamonu Eğitim Dergisi*, 23(2), 809-824. <https://dergipark.org.tr/en/pub/kefdergi/issue/22599/241434>
- Bakanlar Kurulu Kararı, (2002). *İstatistiki bölge birimleri sınıflandırması*. 28/08/2002 Tarihli ve 2002/4720 Sayılı Kararname. <https://www.resmigazete.gov.tr/eskiler/2002/09/20020922.htm>
- Baker, J. A., Clark, T. P., Maier, K. S., & Viger, S. (2008). The differential influence of instructional context on the academic engagement of students with behavior problems. *Teaching and Teacher Education*, 24, 1876-1883. <https://doi.org/10.1016/j.tate.2008.02.019>
- Bentler P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bradshaw, C. P., & Kush, J. M. (2020). Teacher observation of classroom adaptation-checklist: Measuring children's social, emotional, and behavioral functioning. *Children & Schools*, 42(1), 29-40. <https://doi.org/10.1093/cs/cdz022>
- Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2016). *Direct behavior rating (DBR): Linking assessment, communication, and intervention*. Guilford Press.
- Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71(6), 733-750. <https://doi.org/10.1037/0022-0663.71.6.733>

- Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. A. (2007). Generalizability and dependability of direct behavior ratings to assess social behavior of preschoolers. *School Psychology Review*, 36(1), 63-79. <https://doi.org/10.1080/02796015.2007.12087952>
- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct behavior rating (DBR) an emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, 34(4), 195-200. <https://doi.org/10.1177/1534508409340391>
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, 48(3), 219-246. <https://doi.org/10.1016/j.jsp.2010.02.001>
- Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. <https://doi.org/10.18637/jss.v048.i06>
- Charles, C. M. (2011). *Building classroom discipline*. Pearson.
- Chionh, Y. H., & Fraser, B. J. (2009). Classroom environment, achievement, attitudes and self-esteem in geography and mathematics in Singapore. *International Research in Geographical and Environmental Education*, 18(1), 29-44. <https://doi.org/10.1080/10382040802591530>
- Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28(6), 693-710. <https://doi.org/10.1080/01443410802206700>
- Cohen, R. J., & Swerdlik, M. E. (2013). Test geliştirme (A. Kan, Çev.). E. Tavşancıl (Ed.), *Psikolojik test ve değerlendirme, testler ve ölçmeye giriş* içinde (ss. 233-277). Nobel Yayıncılık. (Orijinal kitabın yayın tarihi 2010)
- Çengel, M., & Türkoğlu, A. (2015). Öğrenciler tarafından algılanan sınıf iklimi ölçeği: Geçerlik ve güvenirlik. *Eğitimde Kuram ve Uygulama*, 11(4), 1240-1257. <https://dergipark.org.tr/en/pub/eku/issue/5467/74233>
- Daniels, B., Briesch, A. M., Volpe, R. J., & Owens, J. S. (2021). Content validation of direct behavior rating multi-item scales for assessing problem behaviors. *Journal of Emotional and Behavioral Disorders*, 29(2), 71-82. <https://doi.org/10.1177/1063426619882345>
- DeVellis, R. F. (2014). *Ölçek geliştirme, kuram ve uygulamalar* (T. Totan Çev.; 3. baskı). Nobel Yayıncılık. (Orijinal kitabın yayın tarihi 2011).
- Dwyer, K. K., Bingham, S. G., Carlson, R. E., Prisbell, M., Cruz, A. M., & Fus, D. A. (2004). Communication and connectedness in the classroom: Development of the connected classroom climate inventory. *Communication Research Reports*, 21(3), 264-272. <https://doi.org/10.1080/08824090409359988>
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16, 5-18. <https://doi.org/10.1007/s11136-007-9198-0>
- Erol, N., Arslan, B. L., & Akçakin, M. (1995). The adaptation and standardization of the Child Behavior Checklist among 6-18 year old Turkish children. In J. Sergeant (Ed.) Eunethydis: European Approaches to Hyperkinetic Disorder. Zurich: Fotoratar, (pp. 97-113).
- Erol, N., & Şimşek, Z. (1998). Türkiye ruh sağlığı profili: Çocuk ve gençlerde ruh sağlığı (Erol, N., Kılıç, C., Ulusoy, M., Keçeci, M., Şimşek, Z.) Türkiye ruh sağlığı profili [Türkiye mental health profile]. T.C. Sağlık Bakanlığı Sağlık Projesi Genel Müdürlüğü.
- Evans, I. M., Harvey, S. T., Buckley, L., & Yan, E. (2009). Differentiating classroom climate concepts: Academic, management, and emotional environments. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 4(2), 131-146. <https://doi.org/10.1080/1177083X.2009.9522449>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

- Flannery, W. P., Reise, S. P., & Widaman, K. F. (1995). An item response theory analysis of the general and academic scales of the self-description questionnaire II. *Research in Personality*, 29(2), 168-188. <https://doi.org/10.1006/jrpe.1995.1010>
- Ford, T., Hayes, R., Byford, S., Edwards, V., Fletcher, M., Logan, S., Norwich, B., Pritchard, W., Allen, K., Allwood, M., Ganguli, P., Grimes, K., Hansford, L., Longdon, B., Norman, S., Price, A., Russell, A. E., & Ukoumunne, O. C. (2019). Training teachers in classroom management to improve mental health in primary school children: The STARS cluster RCT. *Public Health Research*, 7(6), 1-150. <https://doi.org/10.3310/phr07060>
- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, 31(3), 328-349. <https://doi.org/10.1080/02796015.2002.12086159>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. Pearson Education Limited.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publication.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological test: A progress report. *European Journal of Psychological Assessment*, 10(3), 229-244. <https://eric.ed.gov/?id=ED399291>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60. <https://academic-publishing.org/index.php/ejbrm/article/view/1224>
- Hu, L. T., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, 34, 111-117. <https://doi.org/10.1177/001316447403400115>
- Kanay, A., & Girli, A. (2008). Analyzing the relations between adaptive behaviors, self concept and academic achievements of elementary school students aged between 9-13 years old with attention deficit hyperactivity disorder. *Journal of Buca Faculty of Education*, 23, 184-191. <https://dergipark.org.tr/en/pub/deubefd/issue/25428/268284>
- Kaner, S., & Çiçekçi, A. U. (2000). Gözden geçirilmiş davranış problemleri kontrol listesinin Türkçeye uyarlanması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 2(4), 23-34. https://doi.org/10.1501/Ozlegt_0000000048
- Kaner, S., Büyüköztürk, Ş., İşeri, E., Ak, A., & Özyayın, L. (2011). Conners öğretmen dereceleme ölçeği-yenilenmiş/uzun: Türk çocukları için psikometrik özelliklerinin değerlendirilmesi. *Klinik Psikiyatri Dergisi*, 14(3), 150-163.
- Kellam, S. G., Branch, J. D., Agrawal, K. C., & Ensminger, M. E. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention and evaluation*. University of Chicago Press.
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, 10, 165-185. <https://doi.org/10.1017/S0954579498001564>
- Kline, T. J. B. (2005). *Psychological testing, a practical approach to design and evaluation*. Sage.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151-162. <https://doi.org/10.32614/RJ-2014-031>
- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation-checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*, 42(1), 15-30. <https://doi.org/10.1177/0748175609333560>

- Kremer, K. P., Flower, A., Huang, J., & Vaughn, M. G. (2016). Behavior problems and children's academic achievement: A test of growth-curve models with gender and racial differences. *Children and Youth Services Review*, 67, 95-104. <https://doi.org/10.1016/j.childyouth.2016.06.003>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage Publications.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Leaf, P. J., Schultz, D., Keys, S., & Jalongo, N. (2002). *The Teacher Observation of Classroom Adaptation-Checklist (TOCA-C)*. Johns Hopkins Center for the Prevention of Youth Violence.
- Lewis, T. J., Newcomer, L. L., Trussell, R., & Richter, M. (2006). Schoolwide positive behavior support: Building systems to develop and maintain appropriate social behavior. In C. Evertson, & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73-95). Lawrence Erlbaum Associates.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410. <https://doi.org/10.1037/0033-2909.103.3.391>
- Merrell, K. W. (1993). Using behavior rating scales to assess social skills and antisocial behavior in school settings: Development of the school social behavior scales. *School Psychology Review*, 22(1), 115-133. <https://doi.org/10.1080/02796015.1993.12085641>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications, Inc.
- Milli Eğitim Bakanlığı (2023). Millî Eğitim İstatistikleri Örgün Eğitim 2022/2023 https://sgb.meb.gov.tr/www/icerik_goruntule.php?KNO=508
- Nash, P., Schlosser, A., & Scarr, T. (2016). Teachers' perceptions of disruptive behaviour in schools: A psychological perspective. *Emotional & Behavioural Difficulties*, 21(2), 167-180. <https://doi.org/10.1080/13632752.2015.1054670>
- Noltemeyer, A. L., Ward, R. M., & McLoughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2), 224-240. <https://doi.org/10.17105/spr-14-0008.1>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.
- Ødegård, M., & Solberg, S. (2024). Identifying teachers' reactive strategies towards disruptive behavior in classrooms. *Teaching and Teacher Education*, 145, 1-9. <https://doi.org/10.1016/j.tate.2024.104627>
- Özbay, Y., & Şahin, M. (2000). Empatik sınıf atmosferi ölçeği (ESATÖ): Geçerlik ve güvenirlik çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 19(19), 104-113. https://dergipark.org.tr/tr/pub/hunefd/issue/7819/102747#article_cite
- Pas, E., Cash, A., O'Brennan, L., Debnam, K., & Bradshaw, C. (2015). Profiles of classroom behavior in high schools: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology*, 53(2), 137-148. <https://doi.org/10.1016/j.jsp.2014.12.005>
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis*. Sage Publications.
- Pierce, C. (1994). Importance of classroom climate for at-risk learners. *The Journal of Educational Research*, 88(1), 37-42. <https://www.jstor.org/stable/27541952>
- Quay, H. C., & Peterson, D. R. (1996). *Revised behavior problem checklist. Professional manual*.
- Robinson, S. L., & Ricord-Griesemer, S. M. (2006). Helping individual students with problem behavior. In C. Evertson, & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73-95). Lawrence Erlbaum Associates.
- Ross, R., Starrett, A., & Irvin, M. J. (2023). Examining the relationships between kindergarteners' demographic characteristics and behavior problems across geographic locale. *Early Childhood Education Journal*, 51(8), 1427-1439. <https://doi.org/10.1007/s10643-022-01391-6>

- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2), 437-460. <https://doi.org/10.3102/00028312038002437>
- Sağkal, A. S., Kabasakal, Z. T., & Tümmüklü, A. (2015). Sınıf İklimi Envanteri'nin (SİE) Türkçe'ye uyarlanması. *İlköğretim Online*, 14(4), 1179-1192. <http://dx.doi.org/10.17051/io.2015.30422>
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351-380. <https://www.jstor.org/stable/42899983>
- Skinner, E. A., Zimmer-Gembeck, M. J., & Connell, J. P. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, 63(2-3), 220-221. <https://doi.org/10.2307/1166220>
- Somersalo, H., Solantaus, T., & Almqvist, F. (2002). Classroom climate and the mental health of primary school children. *Nordic Journal of Psychiatry*, 56(4), 285-290. <https://doi.org/10.1080/08039480260242787>
- Sprott, J. B. (2004). The development of early delinquency: Can classroom and school climates make a difference? *Canadian Journal of Criminology and Criminal Justice*, 46(5), 553-572. <https://doi.org/10.3138/cjccj.46.5.553>
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325. <https://link.springer.com/article/10.1007/BF02295289>
- Sucuoğlu, N. (2003). Sorun davranışlar kontrol listesinin Türkçe formunun psikometrik özelliklerinin incelenmesi. *Türk Psikoloji Dergisi*, 18(52), 77-91. <https://psikiyatridizini.net/viewarticle.aspx?articleid=21399&tammetinvar=yes>
- Sucuoğlu, B., & Kargın, T. (2010). *İlköğretimde kaynaştırma uygulamaları: Yaklaşımlar, yöntemler, teknikler*. Kök Yayıncılık.
- Şeker, H., & Gençdoğan, B. (2014). *Psikolojide ve eğitimde ölçme aracı geliştirme*. Nobel Yayıncılık.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson Education.
- Toraman, Ç., & Özdemir H. F. (2023). Evren ve örneklem. A. İlğan, & A. Ekinci (Eds.), *Eğitimde araştırma yöntemleri içinde* (ss. 135-167). Nobel Akademik Yayıncılık.
- Vieira, A. L. (2011). *Interactive LISREL in practice, getting started with a SIMPLIS Approach*. Springer. <https://doi.org/10.1007/978-3-642-18044-6>
- Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review*, 41(3), 246-261. <https://doi.org/10.1080/02796015.2012.12087506>
- Volpe, R. J., & Briesch, A. M. (2015). Multi-item direct behavior ratings: Dependability of two levels of assessment specificity. *School Psychology Quarterly*, 30(3), 431-442. <http://dx.doi.org/10.1037/spq0000115>
- Volpe, R. J., & Briesch, A. M. (2016). Dependability of two scaling approaches to direct behavior rating multi-item scales assessing disruptive classroom behavior. *School Psychology Review*, 45(1), 39-52. <https://doi.org/10.17105/SPR45-1.39-52>
- Warner, R. M. (2013). *Applied statistics, from bivariate through multivariate techniques*. Sage Publications.
- Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). *Teacher observation of classroom adaptation-revised (TOCA-R) [Database record]*. APA PsycTests. <https://doi.org/10.1037/t31163-000>
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187-213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>

- Yukay-Yüksel, M. (2009). Okul sosyal davranış ölçeklerinin (OSDÖ) Türkçeye uyarlanması. *Kuram ve Uygulamada Eğitim Bilimleri*, 9(3), 1605-1650. <https://tr-scales.arabpsychology.com/wp-content/uploads/pdf/okul-sosyal-davranis-olcekleri-b-formu-olumsuz-sosyal-davranislar-toad.pdf>
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data*. [Doctoral dissertation, University of Massachusetts Amherst]. <https://www.proquest.com/docview/220010848?fromopenview=true&pqorigsite=gscholar&sourcetype=Dissertations%20&%20Theses>



Classical Test Theory and Item Response Theory Validity and Reliability Evidence of Problem Behavior Evaluation Scale

Eylem Dayı ¹

Çetin Toraman ²

Nevin Güner ³

Arzu Doğanay-Bilgi ⁴

Hasan Köse ⁵

Abstract

Introduction: In educational environments, problem behaviors, including many related to non-cooperation with school and classroom rules, are defined as those that threaten students' safety, negatively affect their social interaction, and distract teachers and students from academic activities. Identifying problem behaviors that negatively affect the educational environment is important in controlling and reducing these behaviors.

Method: This study aims to examine the validity and reliability of the Problem Behavior Evaluation Scale (PBES) evidence, which will evaluate students' problem behaviors based on teachers' observations and perceptions, according to classical test theory and item response theory. We prepared the scale items by conducting a literature review and finalizing them by assessing expert opinions. A total of 590 teachers participated in the study to examine the scale's reliability and validity. Based on the collected data, we examined exploratory factor analysis, confirmatory factor analysis, item characteristic curves, item information functions, and item discriminations according to item response theory. We determined the scale's reliability evidence with Cronbach's Alpha, McDonald's Omega, and marginal reliability coefficients.

Findings: As a result of the analyses, we determined that the PCES is a valid scale with 18 items and two factors. We determined the reliability of PBES by Cronbach's Alpha and McDonald's Omega coefficients according to classical test theory and the marginal reliability coefficient according to item response theory. The results showed that the PBES had a high level of reliability. As a result of the analysis, the first factor consisted of 10 items, and the second factor consisted of eight items.

Discussion: With its 18-item and two-factor structure, the PBES is a proper, simple, easy-to-respond scale. The scale can be used in primary schools in Türkiye to assess students' problem behaviors based on teachers' observations and perceptions. It can also provide support for scientific studies on problem behaviors.

Keywords: Problem behavior, scale development, validity and reliability, classical test theory, item response theory, special education.

To cite: Dayı, E., Toraman, Ç., Güner, N., Doğanay-Bilgi, A., & Köse, H. (2025). Classical test theory and item response theory validity and reliability evidence of problem behavior evaluation scale. *Ankara University Faculty of Educational Sciences Journal of Special Education*, 26(4), 557-577. <https://doi.org/10.21565/ozelegitimdergisi.1538042>

¹**Corresponding Authors:** Assoc. Prof., Gazi University, E-mail: eylemd@gazi.edu.tr, <https://orcid.org/0000-0003-1020-5846>

²Assoc. Prof., Çanakkale Onsekiz Mart University, E-mail: toramanacademic@gmail.com, <https://orcid.org/0000-0001-5319-0731>

³Prof., Eskişehir Osmangazi University, E-mail: antreh@gmail.com, <https://orcid.org/0000-0002-9135-6429>

⁴Assoc. Prof., Gazi University, E-mail: arzudb@yahoo.com, <https://orcid.org/0000-0003-3330-8303>

⁵PhD., Ministry of National Education, E-mail: hsnkose@hotmail.com, <https://orcid.org/0000-0002-0923-3711>

Introduction

In educational environments, problem behaviors, including many related to non-cooperation with school and classroom rules (Robinson & Ricord-Griesemer, 2006), are defined as those that threaten students' safety, negatively affect social interaction, and distract teachers and students from academic activities (Nash et al., 2016; Sucuoğlu & Kargin, 2010). Although most students generally exhibit appropriate behaviors, studies conducted in different years and educational settings have revealed that approximately 20% of students exhibit moderate problem behaviors, and approximately 6% exhibit severe problem behaviors (Lewis et al., 2006; Pas et al., 2015). Among the problem behaviors frequently exhibited in educational settings are non-involvement in classroom activities, wandering around the classroom, talking without permission, disturbing other students, and being interested in extracurricular activities and objects. Anger, aggression, or bullying are also less frequently encountered (Robinson & Ricord-Griesemer, 2006).

Students' problem behaviors negatively affect teachers, students, and the educational environment. For example, teachers have to spend a lot of time and energy dealing with problem behaviors (Clunies-Ross et al., 2008; Simonsen et al., 2008), which can decrease their perceptions of self-efficacy, well-being, and burnout (Ford et al., 2019). Researchers emphasize that one of the most challenging aspects of the teaching profession is dealing with students' problem behaviors (Aloe et al., 2014). Problem behaviors also have important consequences for the educational environment and students. Educational environments with high problem behaviors are shown as environments where students do not feel safe, social development is negatively affected, and participation and success in academic activities are low (Charles, 2011; Ødegård & Solberg, 2024). Frequently, students who exhibit persistent and severe problem behaviors are academically unsuccessful (Kremer et al., 2016; Ross et al., 2023) and are more likely to be involved in crime in the future (Lewis et al., 2006). As a result, to control and reduce problem behaviors that distract teachers and students from academic activities and negatively affect the learning setting, it is essential to identify these behaviors to achieve educational goals because the identification of problem behaviors constitutes the starting point for the preparation of appropriate treatment programs for these behaviors.

There are scales, checklists, and adaptation studies to help determine problem behaviors in educational environments. One is the Teacher Observation of Classroom Adaptation (TOCA) form developed by Kellam et al. (1975), in which classroom adaptation skills are evaluated according to teacher observations. This measurement tool aimed to assess the classroom adaptation behaviors of first-grade students as seen and defined by teachers. Trained personnel conducted the assessment process through structured interviews. The TOCA underwent a significant revision by the Johns Hopkins Center for Prevention and Early Intervention (Johns Hopkins Center for Prevention and Early Intervention) for use in a large intervention study in Baltimore, Maryland (Teacher Observation of Classroom Adaptation- Revised [TOCA-R]-Werthamer-Larsson et al., 1991). Researchers have utilized this version in extensive studies (e.g., Kellam et al., 1998). However, due to the time and cost limitations of applying TOCA with structured interviews in large-scale studies, researchers designed a new checklist-based version (Teacher Observation of Classroom Adaptation-Checklist [TOCA-C] - Leaf et al., 2002). Koth et al. (2009) investigated TOCA-C's validity and factor structure, while Aydos et al. (2014) adapted it into Turkish using data collected from preschoolers.

Another scale is the Child Behavior Checklist (CBCL) (Achenbach, 1978; Achenbach & Edelbrock, 1979), is a standardized form that parents, teachers, and others involved with children fill out to identify children's behavioral and emotional problems. The CBCL has a version for ages 2 and 3 (CBCL/2 to 3) and a version for ages 4 to 18 (CBCL/4 to 18) (Achenbach & Ruffle, 2000). This scale was translated into Turkish and used in many studies for children with typical development (Akkök et al., 1986; Erol et al., 1995; Erol & Şimşek, 1998; Kanay & Girli, 2008). Another scale, the Direct Behavior Rating (DBR), involves recording data by rating them as perceived by the rater based on systematic direct observation (Chafouleas et al., 2009). In the DBR category, Direct Behavior Rating - Single-Item Scales (DBR Single-Item Scales [DBR-SIS]) (Chafouleas et al., 2007; Chafouleas et al., 2010) and Direct Behavior Rating - Multi - Item Scales (DBR Multi - Item Scales [DBRMIS]) (Briesch et al., 2016; Daniels et al., 2021; Volpe & Briesch, 2012, 2015, 2016) versions have been developed for which psychometric evidence has been established.

Another scale adapted into Turkish is the Revised Problem Behavior Checklist (Quay & Peterson, 1996). In the study, Kaner and Çiçekçi (2000) examined the psychometric properties of the Turkish form of the instrument with the data obtained from the evaluation of 7 - 18-year-old students without disabilities by their teachers. In another study, Kaner et al. (2011) adapted to Turkish culture the Conners' Teacher Rating Scale-Revised/Long (CTRS-R/L), which was developed by Conners et al. in the 1960s and updated in 1997 and used in the diagnosis

of Attention Deficit/Hyperactivity Disorder (ADHD). They showed that the scale was reliable and valid for use in research and diagnostic clinics. In another study, Sucuoğlu (2003) translated into Turkish the Problem Behavior Checklist (Aberrant Behavior Checklist [ABC], Aman et al., 1985a, 1985b), which was developed to determine the effects of drug treatment or other interventions applied to adolescents and adults with intellectual disabilities. The author also examined its psychometric properties.

In another study, Yukay-Yüksel (2009) adapted the School Social Behavior Scales (SSBSs) (Merrell, 1993) to Turkish. Merrell (1993) developed the SSBSs, which consist of Social Competence and Negative Social Behavior sub-forms, to assess the social skills of preschool and primary school students. The "Social Competence Scale" consists of three subscales: Interpersonal Relationships, Self-Control Skills, and Academic Skills. The "Negative Social Behavior Scale" consists of three subscales, including problem behaviors, namely Aggressive-Nervous, Antisocial-Aggressive, and Destructive-Demanding.

In summary, considering the scales translated into Turkish and adapted or directly developed in the literature, they are aimed at either students with typical development or students with special needs. In addition, some adaptation studies were conducted only with preschool children. However, primary school is an educational level with a numerically large group of students (Ministry of National Education, 2023). At the same time, many students participate in educational environments for the first time to acquire social and academic skills. For this reason, students in primary school may exhibit various problem behaviors when learning school and classroom rules and adapting to school. In addition, primary school classes contain great diversity in student characteristics. Almost every preschool classroom has typically developing students with different individual characteristics and needs, students at academic and behavioral risk, or with special needs. Therefore, measurement tools need to be used to directly assess the problem behaviors of typically developing, at-risk, and special needs students according to teacher observation and perception.

Taking these as a starting point, the purpose of this study is to examine the validity and reliability evidence of the PBES, which will serve the purpose of assessing students' problem behaviors based on classroom teachers' observations and perceptions in school and classroom environments, according to the Classical Test Theory (CTT) and Item Response Theory (IRT).

Method

Research Design

This descriptive study aims to develop a PBES that would assess students' problem behaviors based on classroom teachers' observations and perceptions in school and classroom environments. In this study, we investigated the validity of the PBES evidence using exploratory factor analysis based on classical test theory and item characteristic curves, item information functions, and item discriminations from item response theory. We assessed reliability evidence through Cronbach's Alpha and McDonald's Omega coefficients per classical test theory, alongside the marginal reliability coefficient from item response theory. These features are important in terms of the usability of the PBES in future studies.

Study Group

To identify the study group, we utilized criterion sampling, a purposeful sampling method. This involved selecting units that fulfilled key criteria established by our research objectives (Miles et al., 2014; Toraman & Özdemir, 2023). While applying criterion sampling, cases that meet the criteria and can provide rich information are included in the study (Toraman & Özdemir, 2023). Accordingly, this study collected data from 590 classroom teachers who met the criteria of a) working in a primary school, b) volunteering to participate in the study, and c) being accessible. By these criteria, the researchers tried to reach the participants from the provinces (Istanbul, Ankara, Şanlıurfa, Hatay, Eskişehir, Kars, Bolu, etc.) located in 12 different regions (Council of Ministers Decree, 2002) in the "Nomenclature of Territorial Units for Statistics". Demographic information of 590 classroom teachers is given in Table 1.

Table 1

Demographic Information of Classroom Teachers in the Research

Variable		n(%)
Gender	Female	146 (24.7)
	Male	444 (75.3)
Age	20-30 years	130 (22)
	31-40 years	199 (33.7)
	41-50 years	162 (27.5)
	51 years and over	99 (16.8)
	0-5 years	115 (19.5)
Seniority	6-10 years	100 (16.9)
	11-20 years	181 (30.7)
	21 years and over	194 (32.9)
Current grade level of the teacher	1st grade	128 (21.7)
	2nd grade	216 (36.6)
	3rd grade	147 (24.9)
	4th grade	99 (16.8)
	0-3 months	28 (4.7)
Length of time the teacher has taught the current class	4-6 months	85 (14.4)
	7-9 months	100 (16.9)
	10-12 months	16 (2.7)
Total	13 months and over	361 (61.2)
		590 (100)

Data Collection Tools

The motivation for developing the PBES was to assess the problem behaviors of primary school students. First, we scanned the literature to determine whether there was a suitable scale for this purpose. We also examined national and international books on problem behaviors. We searched the terms "problem behaviors," "disruptive behavior, unwanted behavior," "challenging behavior," "inappropriate behavior," and "evaluating problem behaviors" through DergiPark, Web of Science, Google Scholar, and PubMed databases. In addition, we expanded the searches using the same keywords in the Educational Resources Information Center (ERIC), Education Full Text, PsycINFO, Dissertations, and Theses Full Text databases. These searches revealed that some measurement tools in the literature could provide information about problem behaviors (e.g., Bradshaw & Kush, 2020; Daniels et al., 2021; Kaner & Çiçekçi, 2000; Sucuoğlu, 2003). However, we found that these scales were not entirely suitable for assessing the problem behaviors of primary school students based on teachers' observations and perceptions in school and classroom environments. Therefore, we developed the PBES.

While developing the PBES, we followed the measurement tool development steps recommended by the literature (Cohen & Swerdlik, 2013; DeVellis, 2014; Şeker & Gençdoğan, 2014). These steps are as follows: 1) determining the purpose of the scale, 2) deciding on the type of item to be included in the scale (Likert-type items were used), 3) writing items and creating an item pool from literature review, target group interviews, open-ended question applications, 4) getting expert opinion for candidate items in the item pool, 5) Turkish language editing, and 6) conducting a pilot study and providing evidence of validity and reliability.

We employed two strategies to develop the item pool. The first strategy involved reviewing relevant literature, from which we generated potential items using insights from measurement tools that reflect the existing research on problem behaviors. The second strategy involved using open-ended questions directed at the target group. The researchers developed the question, "What are the problem behaviors exhibited by students in classrooms?" for the target audience (primary school teachers). One hundred seventy-one teachers participated in the open-ended question application via Google Forms. We analyzed the responses, created candidate scale items, and obtained a draft scale consisting of 37 items with the relevant literature and teacher data.

To get an expert opinion on this 37-item draft scale form, we formed a 10-person expert group consisting of five academicians specializing in special education, three specializing in educational assessment, one specializing in classroom teaching, and one specializing in curriculum development. We prepared a form to get opinions from the expert group. Each item was written on the form, and the options "the item is appropriate and should be included in the scale," "the item is not appropriate and should be removed from the scale," and "the item can be included in the scale provided that it is edited" were placed opposite the item. We asked the experts to

present their evaluations of each item. In addition, we wanted the experts to write their suggestions for editing each item, if any. We examined the consistency between the expert opinions received through this form and those received before the necessary arrangements were made. We examined the data by calculating the Krippendorff Alpha coefficient (Krippendorff, 2004) and determined the calculated consistency value as 0.897, which showed consistency between the expert opinions. Based on the expert feedback, we edited and finalized the items. Additionally, we calculated the content validity ratio for each item by analyzing the expert opinions using the Lawshe technique (Lawshe, 1975). We removed six items with low content validity from the scale, leaving 28 items in the draft form for the trial application.

We submitted the draft scale form, consisting of 28 items, to an academician specializing in the Turkish Language field for evaluation in terms of clarity, comprehensibility, simplicity, and other factors. After language editing, we converted the items into a four-point Likert scale: "never," "rarely," "frequently," and "always."

Data Collection and Analysis

We instructed five hundred ninety teachers to "Think of three different students in your classroom who have many, average, and few problem behaviors and fill out the PBES separately for these three students." This approach allowed us to gather data on 1,770 students from 590 teachers. We then randomly divided the data file reflecting the teachers' perception of 1,770 students' problem behaviors into two separate data files, following the scale development literature (Hair et al., 2014). We did this to simultaneously conduct both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) instead of collecting data at two different times.

As a result of the random division process, we included data from 915 students in the EFA data file and data from 855 students in the CFA data file. Using the data in the EFA file, we conducted analyses to explore the construct validity of the PBES and calculated reliability evidence. With the data in the CFA file, we attempted to verify the structure of the PBES, which had been discovered and conducted analyses in this direction. Additionally, we carried out analyses with all undivided data (1,770), reflecting teachers' perceptions of problem behaviors in line with the IRT. Descriptive statistics of the data in the data files created by the random division method for the whole data file, EFA, and CFA are shown in Table 2.

Table 2

Distribution of Some Variables of Students

Variable		All data file	EFA data file	CFA data file
		<i>n (%)</i>	<i>n (%)</i>	<i>n (%)</i>
Gender	Male	1.027(58)	531(58)	481(56.3)
	Female	743(42)	384(42)	374(43.7)
Nationality	RT	1.563(88.3)	807(88.2)	758(88.7)
	Others	207(11.7)	108(11.8)	97(11.3)
Socioeconomic level (SES) of the student's family	Lower SES	365(20.6)	186(20.3)	169(19.8)
	Middle SES	1.275(72)	664(72.6)	623(72.9)
	Upper SES	130(7.3)	65(7.1)	63(7.4)
Does the student have a formal diagnosis?	No	1.636(92.4)	844(92.2)	790(92.4)
	Yes	134(7.6)	71(7.8)	65(7.6)
Student's level of problem behavior	Mild	790(44.6)	417(45.6)	414(48.4)
	Moderate	638(36)	328(35.8)	304(35.6)
	Severe	342(19.3)	170(18.6)	137(16)
Total		1770(100)	915(100)	855(100)

Note: RT = Republic of Türkiye.

In the whole data file, the mean age of the students was 8.42, with a standard deviation of 1.85. In the EFA file, the mean age of the students was 8.37, with a standard deviation of 1.49. In the CFA file, the mean age of the students was 8.34, and the standard deviation was 1.34.

We analyzed the data obtained in the research by going through the following stages:

1. In the study, data showing the perception of problem behaviors of 1,770 students were obtained from 590 classroom teachers through PBES.
2. There were no unanswered items in the data obtained.
3. We saved the data in three files: total data (1.770 students), EFA data (915 students), and CFA data (855 students) using the random division method.

4. We transferred the data to JAMOVİ and R statistical software.
5. We examined the data set in all three data files to determine whether it showed a multivariate normal distribution. We performed this examination using the "mvn" script from the R Package Program (Korkmaz et al., 2014).
6. We performed the Kaiser-Meyer-Olkin (KMO) test on the EFA file. The KMO test compares the magnitudes of the calculated correlation coefficients with the magnitudes of the partial correlation coefficients, providing information about sampling adequacy (Pett et al., 2003). We used the following criteria to interpret the KMO value: above .90 "excellent," .80-.90 "good," .70-.80 "acceptable," .60-.70 "moderate," .50-.60 "low level," and below .50 "unacceptable" (Kaiser & Rice, 1974).
7. We examined Bartlett's Test of Sphericity to determine whether the data set showed the characteristics of a unit matrix (identity matrix) with diagonals of 1 and other elements of 0. A correlation matrix with a unit matrix is not preferred in factor analysis. There is no reciprocal relationship between items in a unit matrix. A significant Bartlett test result indicates that the data set is not a unit matrix, the desired condition for factor analysis (Field, 2018; Pett et al., 2003).
8. We conducted factor analysis (EFA) using the Principal Axis Factoring (PAF) method. PAF, commonly used in social and behavioral science research, models the shared variance in a set of X measures (Warner, 2013).
9. We examined the "Eigenvalue" to determine the number of factors. We used the value "1" as the eigenvalue reference. Factors with eigenvalues greater than one are considered significant factor candidates. When the number of variables (number of items) is between 20 and 50, using the eigenvalue as a reference is quite reliable.
10. We analyzed the total variance explained. Each scale aims to measure one characteristic. PBES measures the student's problem behavior levels based on the teachers' perception. While measuring this characteristic, how much of the variance (variability) in problem behaviors does the PBES provide information about? The explained variance level is the answer to this question. According to Hair et al. (2014), in social sciences, where information is generally less precise, a solution that accounts for 60% (and sometimes even less) of the total variance is satisfactory. According to Warner (2013), the acceptable limits of adequacy are between 40% and 70%.
11. We analyzed the loading values (correlation values) obtained by the items grouped under the factors. The minimum values that should be sought for factor loading in interpreting the factor structure are between $\pm .30$ and $\pm .40$. On the other hand, $\pm .50$ is a factor loading value that should be considered important. $\pm .70$ and above is accepted as an indicator of a well-defined structure (Hair et al., 2014).
12. In the CFA analysis, we examined the degree of fit of the model. In the fit-indexes determined for CFA, according to the literature, $0.05 < RMSEA \leq 0.08$ is acceptable and $0 > RMSEA \leq 0.05$ excellent for RMSEA, 0.95 and above excellent for TLI and CFI, $2 < X^2/sd \leq 5$ acceptable and $0 > X^2/sd \leq 2$ excellent for X^2/sd (Anderson & Gerbing, 1984; Bentler, 1990; Hooper et al, 2008; Hu & Bentler, 1999; Kline, 2005; Marsh et al., 1988; Tabachnick & Fidell, 2013; Vieira, 2011).
13. We determined the scale's reliability level using Cronbach's Alpha, McDonald's Omega, and the Marginal Reliability coefficient. According to Nunnally and Bernstein (1994), adequate reliability should be at least .70 and above.
14. We examined validity and reliability using the ITS. First, we analyzed the assumptions of unidimensionality and local independence (Zhao, 2008). Unidimensionality requires only one characteristic (only one characteristic of the relevant items of the measurement tool) that affects the performance of individuals in the measurement tool (Hambleton et al., 1991). We evaluated unidimensionality with EFA. We tested the assumption of local independence using the Q3 statistic (Yen, 1993) and performed ITS calibrations with the "Mirt v.1.30" (Chalmers, 2012) package in the R v.4.1.2 program.
15. We analyzed item discrimination and difficulty levels. The discrimination value of an ideal scale item (i.e., the "a" parameter) should be between 0.5 and 2. This parameter between 0.75 and 2.50 in the literature indicates that it is in the acceptable range (Flannery et al., 1995). We accepted the ideal (medium

difficulty level) limits for item difficulty levels (i.e., the “b” parameter) between -1.00 and 1.00 (Hambleton, 1994). In aptitude or achievement tests, we considered items with difficulty levels lower than -1.00 as manageable and those higher than 1.00 as problematic.

16. We analyzed item information functions and item characteristic curves.

Findings

CTT Validity and Reliability Evidence of PBES

We examined the multivariate normal distribution of the EFA data from 915 students and determined that the data set did not follow a multivariate normal distribution. The KMO value was .970. In this case, examining the data file in terms of partial correlations is suitable for factor analysis. Bartlett's test result was 15671, which was significant ($p < .05$). This result shows that the data matrix is not a unit matrix and has a suitable matrix structure for factor analysis.

As the data analysis section explained, we preferred the PAF method for factor identification. We analyzed with 28 items in the PBES draft form. Some items did not reach the factor correlation loading values recommended by the literature. These items are numbered 1, 3, 4, 5, 6, 10, 16, 18, 26, and 27. After these items were excluded from the scale, 18 items remained and were grouped under two factors. To clarify the factor structure and identify the items under each factor, we used eigenvalue analysis, scree plot analysis, and the oblimin rotation method. We chose the oblimin rotation method because the two factors were highly correlated. Table 3 displays the analysis results conducted within the scope of EFA.

When we read the items in Factor 1 semantically, we determined they were related to problem behaviors affecting the classroom climate. There were ten items in this factor and no reverse-coded (scored) items. The items under the factor showed correlation loadings between .492 and .978 and item-total correlations between .709 and .877. This factor has an eigenvalue of 6.72 and alone explains 37% of the variance in student problem behaviors.

Analyzing the items in Factor 2 semantically, we found that they were related to problem behaviors that affect students' learning. There were eight items in this factor. There were no reverse-coded (scored) items. The items under the factor showed correlation loadings between .522 and .938 and item-total correlations between 0.555 and 0.848. This factor had an eigenvalue of 5.59 and alone explained 31% of the variance in student problem behaviors. We examined the reliability level of the PBES for the total scale scores and scale factors using Cronbach's Alpha and McDonald's Omega reliability analyses according to the CTT. Table 4 shows the reliability analysis results.

Table 3*EFA Analysis Results*

Items		Factors				Eigenvalue	Explained variance	Factor name	Items coded reversed	Lowest possible score	Highest possible score
		F1		F2							
		FLVR	ITC	FLVR	ITC						
I28	Bullies his/her friends.	.978	.852								
I14	Fights with friends.	.878	.877								
I8	Mocks his/her friends.	.863	.809								
I19	Harms his/her friends.	.850	.854								
I24	Makes unfair accusations.	.817	.817								
I25	Uses abusive/slang words.	.806	.817								
I21	Constantly complains about his/her friends.	.744	.766			6.72	37.3	Problem behaviors affecting whole classroom	None	10	40
I11	Behaves in a way that prevents his/her friends from listening to the teacher.	.623	.826								
I12	Walks around the classroom during class.	.528	.709								
I22	Makes excuses to leave the class.	.492	.741								
I20	Does not do homework.			.938	.792						
I7	Does not fulfill the tasks in class.			.851	.848						
I2	Does not bring the necessary course materials with him/her.			.837	.783						
I9	Does not participate in group activities.			.785	.790						
I17	Does not listen to the teacher.			.765	.843	5.59	31	Problem behaviors affecting own learning	None	8	32
I23	Does not use the course materials in accordance with the goals of the lesson.			.765	.820						
I13	Does not follow the rules of group activities.			.606	.817						
I15	Does not ask for help even if needed.			.522	.555						

Note: F1 = factor 1; F2 = factor 2; FLVR = factor load value after rotation; ITC = item total correlation.

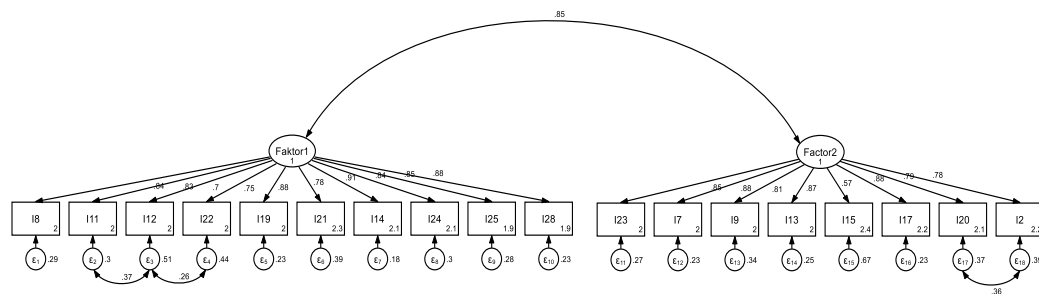
Table 4*Reliability Analysis Results of the PBES*

Scale	Cronbach's Alpha	McDonald's Omega
F1 problem behaviors affecting whole classroom	.956	.956
F2 problem behaviors affecting own learning	.938	.939
PBES total scores	.966	.967

Note: F1 = factor 1; F2 = factor 2; PBES = problem behavior evaluation scale.

The reliability values calculated for the PBES and its factors are above .900. The data analysis section explains that a reliability value of .700 and above is accepted as a high level of reliability. Hence, a high level of reliability was obtained for the scale and its factors.

The EFA applied within the scope of CTT revealed the structure of the PBES, consisting of 18 items and two factors. This structure was confirmed in the CFA data file of 855 students created by the random division method from the general data file. The diagram resulting from the CFA modeling is shown in Figure 1.

Figure 1*PBES Confirmatory Factor Analysis Diagram and Standardized Correlation Values*

The fit indexes obtained in CFA were calculated as CFI=.953, TLI=.945, RMSEA=.079, and $X^2/df = 4.218$. According to the literature, these fit indices are within acceptable limits. These results confirmed the two-factor structure of the PBES.

IRT Validity and Reliability Evidence of PBES

We used data collected from 1.770 students through 590 classroom teachers as the data set in the CTT analyses. According to the CTT evidence, the PBES consisted of two factors. We accepted both factors as independent and conducted IRT analyses. For Factor 1, Problem Behaviors Affecting Classroom Climate, we tested local independence and found no item violating local independence. We performed item calibrations in the IRT using the Generalized Partial Credit Model (GPCM). According to the GPCM, the statistics calculated for the items in Factor 1 are given in Table 5.

Table 5*Item Fit-Indexes for PCBS Factor 1 (Problem Behaviors Affecting Whole Classroom Climate)*

Item	$S \chi^2$	GPCM df	RMSEA
I8	83.707	39	.025
I11	83.823	42	.024
I12	73.596	53	.015
I14	54.639	32	.020
I19	29.409	31	.000
I21	66.909	45	.017
I22	107.189	50	.025
I24	59.274	39	.017
I25	76.360	38	.024
I28	36.914	33	.008

Examining the RMSEA values, indicating that items fit within the scope of IRT, is essential. The criterion level for RMSEA is 0.080 and below (Stout, 1990). The RMSEA values for the ten items in Factor 1 of the PBES

are 0.025 and below. The required level of fit for the items was obtained. The “a” (item discrimination) and “b” (item difficulty) parameters and standard errors were determined for the ten items in Factor 1. Table 6 presents the results.

Table 6

Item Parameters for PCRS Factor 1 (Problem Behaviors Affecting Whole Classroom Climate)

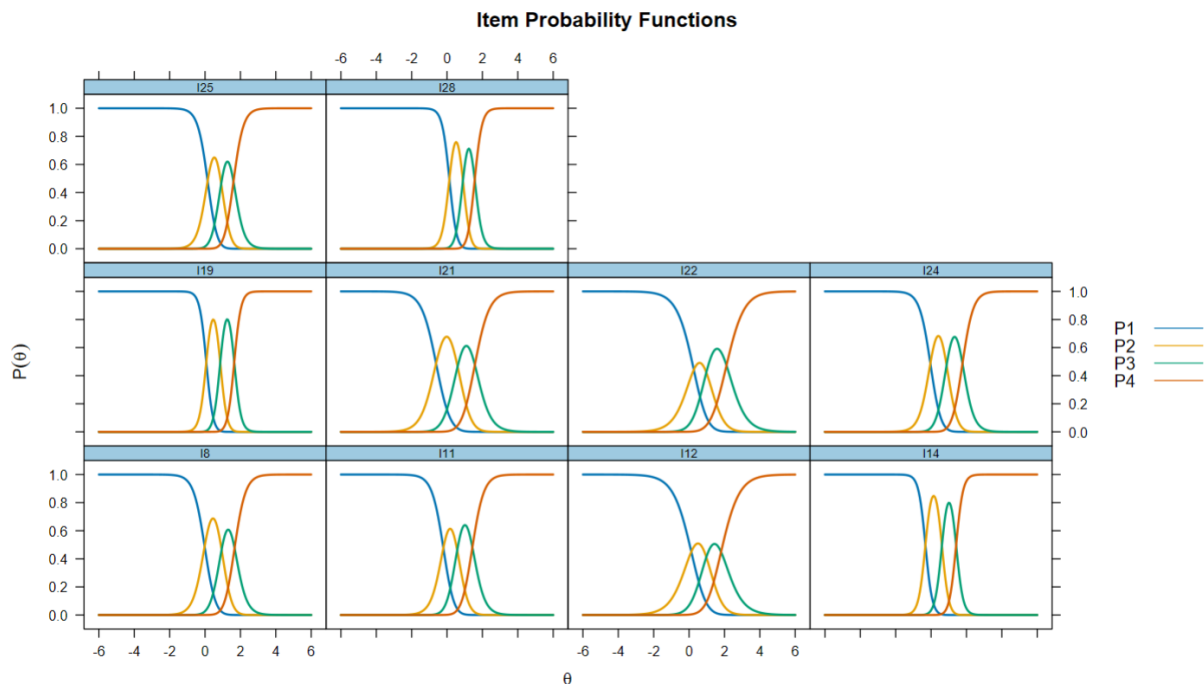
Item	<i>a</i> (SE)	<i>b</i> ₁ (SE)	<i>b</i> ₂ (SE)	<i>b</i> ₃ (SE)
I8	3.113 (0.156)	-.023 (0.035)	.937 (0.041)	1.669 (0.061)
I11	2.993 (0.148)	-.208 (0.037)	.575 (0.037)	1.433 (0.052)
I12	1.828 (0.092)	.126 (0.045)	.975 (0.053)	1.809 (0.084)
I14	5.174 (0.278)	-.320 (0.033)	.613 (0.032)	1.416 (0.044)
I19	5.278 (0.302)	.071 (0.031)	.860 (0.034)	1.652 (0.053)
I21	2.381 (0.114)	-.618 (0.042)	.598 (0.040)	1.572 (0.061)
I22	1.934 (0.098)	.268 (0.044)	.976 (0.051)	2.116 (0.098)
I24	3.201 (0.160)	-.041 (0.035)	.876 (0.039)	1.775 (0.064)
I25	3.450 (0.181)	.147 (0.034)	.915 (0.039)	1.608 (0.057)
I28	4.821 (0.273)	.132 (0.031)	.899 (0.036)	1.564 (0.051)

Note: Iteration = 163; LogLikelihood = - 13840.758; $p < .05$.

Although the ideal discrimination for each item in the scale items is accepted between .5 and 2, the range between .75 and 2.50 for the “a” parameter is also acceptable. I12, I21, and I22 achieved the ideal discrimination level among the 10 items under PBES actor 1. According to the estimations made according to GPCM (LogLikelihood, $p < .05$), item agreement was achieved. Item characteristic curves are shown in Figure 2.

Figure 2

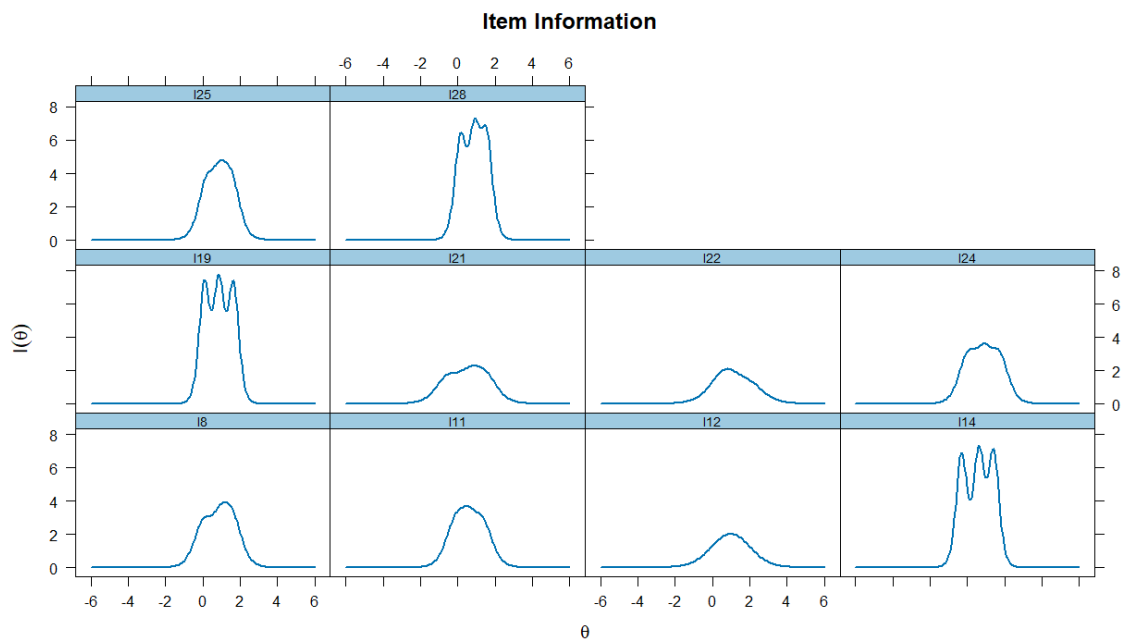
Item Characteristic Curves for Factor 1 Items of the PBES



Item characteristic curves show that almost all of the 10 items under Factor 1 work well with their options in the response set. Item information functions are shown in Figure 3. The item information function is a graphical representation that shows the range of the trait (the trait tried to be measured in the scale) in which the item best discriminates the individuals taking the measurement tool (Edelen & Reeve, 2007). The higher the peak of the curve in the item information function, the more information the item provides. When the item information functions of the items of Factor 1 of the PBES are analyzed, the items with the most information are 14, 19, 25, and 28. Items 12, 21, and 22 are less informative than the other items.

Figure 3

Item Information Functions of Factor 1 Items of the PBES



We tested local independence for Factor 2, Problem Behaviors Affecting Self-Learning, and found no item violating local independence. Table 7 shows the statistics calculated for the items in Factor 2.

Table 7

Item Fit-Indexes for PBES Factor 2 (Problem Behaviors Affecting Self-Learning)

Item	$S \chi^2$	GPCM df	RMSEA
I2	45.495	31	0.016
I7	40.743	26	0.018
I9	62.533	29	0.026
I13	41.415	28	0.016
I15	96.480	43	0.027
I17	48.369	27	0.021
I20	76.344	32	0.028
I23	52.812	28	0.022

RMSEA values for the eight items in Factor 2 of the PBES are 0.028 and below. The required level of fit for the items was obtained. The “a” (item discrimination) and “b” (item difficulty) parameters and standard errors were determined for the eight items in Factor 2. Table 8 presents the results.

Among the eight items under Factor 2 of the PBES, I2, I15, and I20 achieved the ideal discrimination level. According to the estimations made according to GPCM (LogLikelihood, $p < .05$), item agreement was achieved. Item characteristic curves are shown in Figure 4.

Table 8

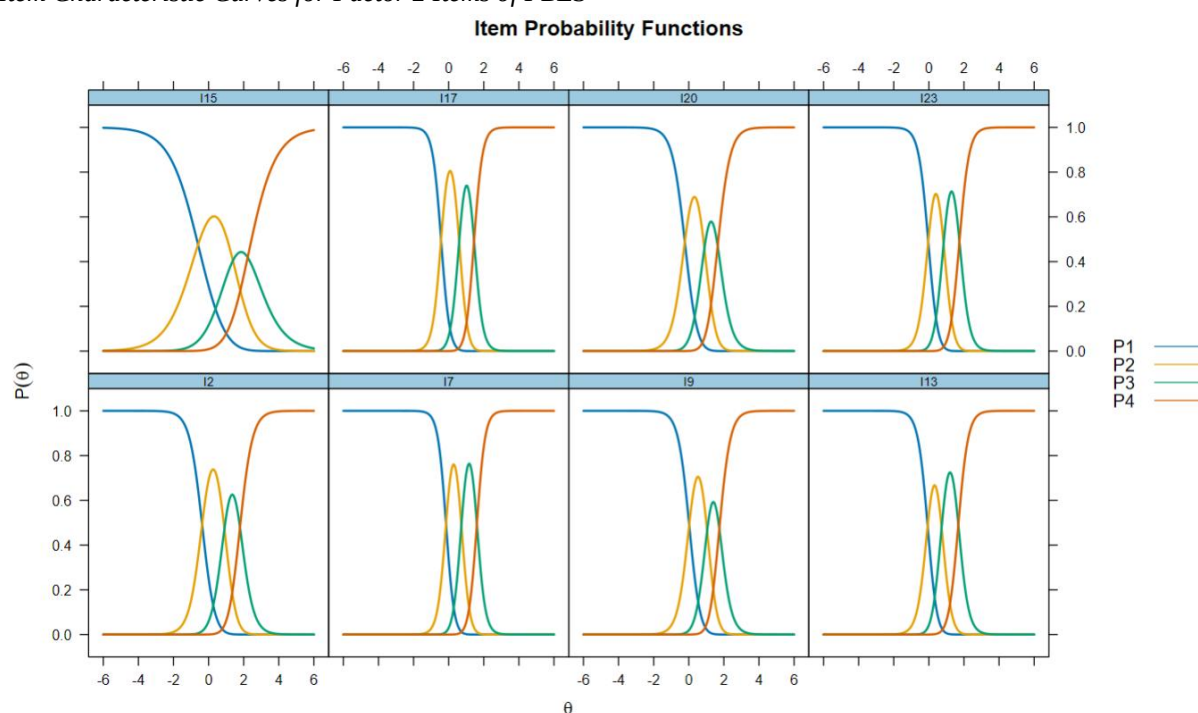
Item Parameters for PBES Factor 2 (Problem Behaviors Affecting Self-Learning)

Item	$a(SE)$	$b1(SE)$	$b2(SE)$	$b3(SE)$
I2	2.733 (0.135)	-0.370 (0.038)	0.902 (0.042)	1.789 (0.069)
I7	4.233 (0.229)	-0.151 (0.033)	0.725 (0.035)	1.609 (0.053)
I9	3.091 (0.162)	0.025 (0.035)	1.047 (0.043)	1.743 (0.066)
I13	3.504 (0.185)	-0.069 (0.034)	0.728 (0.037)	1.680 (0.059)
I15	1.169 (0.063)	-0.598 (0.059)	1.387 (0.079)	2.227 (0.133)
I17	4.109 (0.216)	-0.432 (0.035)	0.600 (0.034)	1.448 (0.048)
I20	2.699 (0.134)	-0.224 (0.037)	0.890 (0.042)	1.647 (0.063)
I23	3.572 (0.189)	-0.033 (0.034)	0.840 (0.038)	1.743 (0.061)

Note: Iteration = 124; LogLikelihood: - 11707.660; $p < .05$.

Figure 4

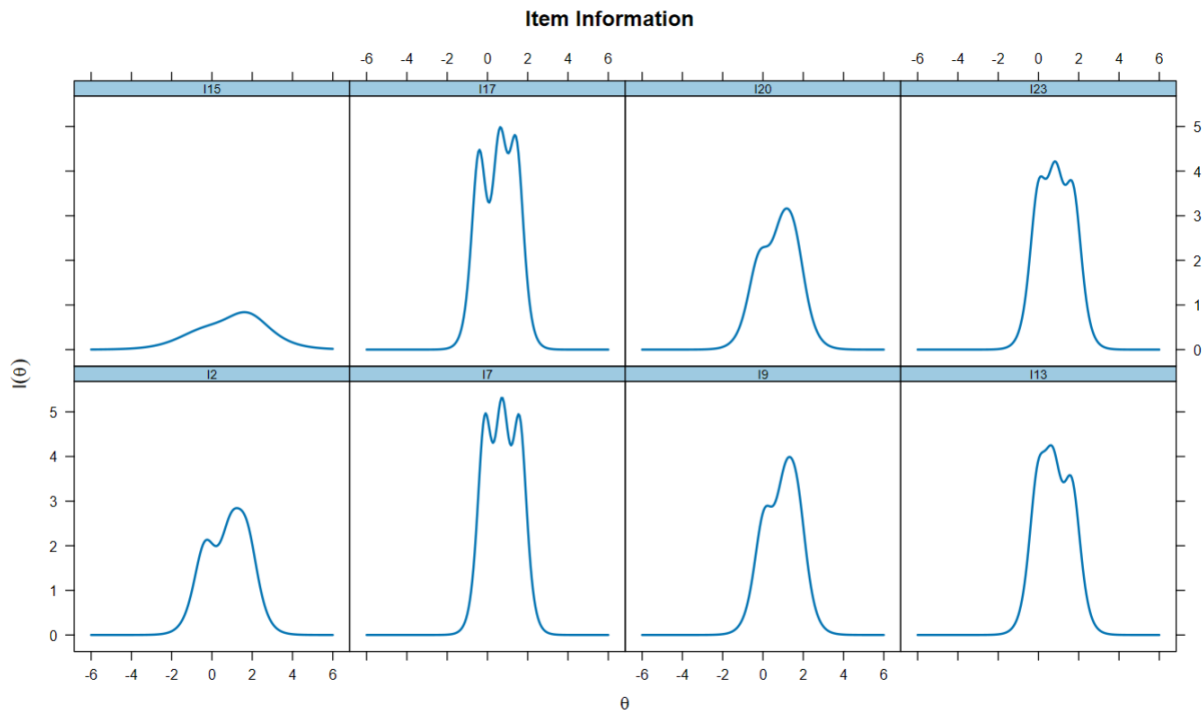
Item Characteristic Curves for Factor 2 Items of PBES



Item characteristic curves show that almost all of the eight items under Factor 2 work well with their options in the response set. Figure 5 depicts item information functions. When we look at the item information functions of the PBES Factor 2 items, the most informative items are 2, 7, 9, 13, 17, 20, and 23. Item 15 provides less information than the other items. The calculated marginal reliability coefficient is 893 for Factor 1 and 0.887 for Factor 2. As explained in the data analysis section, this value indicates a high level of reliability according to the literature.

Figure 5

Item Information Functions of PBES Factor 2 Items



Discussion

This study examined the validity and reliability of the Problem Behavior Evaluation Scale (PBES), which will assess students' problem behaviors based on teachers' observations and perceptions in school and classroom environments, according to classical test theory and item response theory. A total of 590 classroom teachers working in primary schools participated in the study planned for this purpose. In the first stage, we conducted a literature review on the purpose and subject of the scale. Based on the literature review and expert opinions, we created an item pool for the scale. We analyzed the validity evidence of the PBES through exploratory factor analysis according to classical test theory, and item characteristic curves, item information functions, and item discriminations according to item response theory. As a result, we determined that the PBES is a valid scale with 18 items and two factors. We determined the reliability of the PBES using Cronbach's Alpha and McDonald's Omega coefficients according to classical test theory, and the marginal reliability coefficient according to item response theory. The results showed that the PBES had a high level of reliability. With its 18-item and two-factor structure, the PBES is an uncomplicated, easy-to-respond, and valuable scale. As a result of the analysis, the first factor consisted of 10 items, and the second factor consisted of eight items.

By the characteristics and scope of the items, the first factor was named "problem behaviors affecting classroom climate," and the second factor was named "problem behaviors affecting self-learning." Previous studies included different items in the developed or adapted scales with different factor numbers and names. For example, the TOCA-R developed by Werthamer-Larsson et al. (1991), which aims to assess students' classroom adaptive behaviors based on teachers' observations, consists of three factors (concentration problems, disruptive/aggressive behavior, and shy behavior) with 43 items. The TOCA-C, developed by Leaf et al. (2002) and examined for validity and factor structure by Koth et al. (2009), consists of three factors (concentration problems, disruptive behavior, and prosocial behavior) with 25 items. The adaptation of TOCA-C into Turkish with the data obtained from preschools consists of 21 items with the same factors as TOCA-C (Aydos et al., 2014). When examining the literature, we see that the scales developed or adapted generally directly measure students' classroom adaptation behaviors (appropriate and inappropriate). Therefore, a third factor is positive social behaviors or a similar factor. However, since the PBES aims to assess problem behaviors, positive social behaviors are not included among the factors directly. The PBES developed in this study differs from previously developed scales. On the other hand, although the Revised Problem Behavior Checklist, which was developed by Quay & Peterson (1996) and whose

psychometric properties were examined in Turkish by Kaner & Çiçekçi (2000), aims to measure problem behaviors directly, the number of items is enough to create difficulties in practice. When examining the scales, we see that the original scale has six factors with 89 items (Conduct Disorder, Socialized Aggression, Attention Problems - Playfulness, Anxiety - Withdrawal, Psychotic Behavior, Motor Tension), while the Turkish adaptation has three factors with 56 items (Conduct Disorder - Hyperactivity and Impulsivity, Depression - Attention Deficit and Socialized Aggression). In this respect, the PBES is a relatively easy scale with its 18-item two-factor structure. In addition, the PBES is a handy scale for practice environments that teachers can use to identify problem behaviors in the classroom without the need for an additional evaluator. While the literature emphasizes the positive aspects of direct observation as a behavior assessment method and its difficulties in requiring additional labor and time, behavior assessment scales that teachers can easily use are associated with efficiency (Chafouleas et al., 2009). From these perspectives, PBES is a practical and helpful scale that can be used to identify students' problem behaviors in practice settings.

One of the two factors of the PBES developed in the current study was named "Problem Behaviors Affecting Classroom Climate." The term classroom climate in the name of this factor can be used with different names such as classroom atmosphere, classroom environment, and learning environment in the literature (Adelman & Taylor, 2005). Classroom climate is defined as the psychological, social, and physical effects created by teacher-student and student-student relationships in the classroom, classroom rules, and the physical structure of the classroom (Açıkgöz, 1998). In the international and national literature, there are various scales and scale adaptation studies prepared for different school levels aiming to directly assess classroom climate (e.g., Çengel & Türkoğlu, 2015; Dwyer et al., 2004; Özbay & Şahin, 2000; Sağkal et al., 2015). When the literature is examined, it is generally seen that the dimensions, including the teacher's classroom management and teaching style and teacher-student interaction, constitute the classroom climate (Evans et al., 2009). In studies, classroom climate is associated with variables such as class participation and motivation (Ryan & Patrick, 2001), course success (Pierce, 1994), delinquency (Sprott, 2004), attitudes toward the course, and self-esteem (Chionh & Fraser, 2009), and mental health (Somersalo et al., 2002). The items that make up the "Problem Behaviors Affecting Classroom Climate" factor of the PBES developed in this study also support the content specified in the literature.

The second factor of the PBES relates to students' problem behaviors that negatively affect their learning. Examining the items in this factor, we see that they relate to problem behaviors concerning students' participation in class and their task fulfillment behaviors necessary for success. The literature emphasizes that what students learn at school is directly related to the level of their participation in course activities (Brophy, 1979; Greenwood et al., 2002). At the same time, studies reveal that high levels of participation are associated with success, while low levels of participation are associated with problem behaviors (Baker et al., 2008; Noltemeyer et al., 2015). Low levels of engagement in course activities pose various challenges for the students, other students in the class, and the teacher (Alter et al., 2013). In addition, low engagement levels are considered an important indicator of students' long-term success (Skinner et al., 1998). These behaviors, which lead to negative consequences for teachers and students, can be measured by the second factor of the PBES. As a result, with the two factors of the PBES, problem behaviors that can directly affect classroom climate, teacher effectiveness, and students' short and long-term achievement can be assessed accurately and practically.

This scale can be used in primary schools in Türkiye to evaluate students' problem behaviors based on teachers' observations and perceptions. The scale is thought to support scientific studies on problem behaviors. In addition, it is hoped that the PBES will contribute to the rapid identification of students' problem behaviors in primary schools and to making necessary arrangements.

As a result, we conducted a scale development study in this research. The study employed classroom teachers working in primary schools. The study can be extended to different school types and levels, teachers with different teaching specialties, and special education teachers. We finalized the scale based solely on teachers' responses. Observations and qualitative interviews can be conducted in schools, and the scope of the study can be expanded by examining teacher minutes, student outcomes, etc. Although the number of participants reached during the data collection process of the scale seems sufficient for the analysis, the validity and reliability of the scale can be tested again with different participants in future studies.

References

- Achenbach, T. M. (1978). The child behavior profile: I. boys aged 6-11. *Journal of Consulting and Clinical Psychology*, 46(3), 478-488. <https://doi.org/10.1037/0022-006X.46.3.478>
- Achenbach, T. M., & Edelbrock, C. S. (1979). The Child Behavior Profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology*, 47(2), 223-233. <https://doi.org/10.1037/0022-006X.47.2.223>
- Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, 21(8), 265-271. <https://doi.org/10.1542/pir.21-8-265>
- Açıkgöz, K. Ü. (1998). *Etkili öğrenme ve öğretme [Effective learning and teaching]* (2nd ed.). Kanyılmaz Matbaası.
- Adelman, H. S., & Taylor, L. (2005). Classroom climate. In S. W. Lee (Ed.), *Encyclopedia of school psychology* (pp. 88-90). Sage Publications.
- Akkök, F., Aşkar, P., & Sucuoğlu, B. (1986). *Öğretmen gözlem formu el kitabı [Teacher observation form handbook]*. Türk Eğitim Derneği Yayınları.
- Aloe, A. M., Shisler, S. M., Norris, B. D., Nickerson, A. B., & Rinker, T. W. (2014). A multivariate meta-analysis of student misbehavior and teacher burnout. *Educational Research Review*, 12, 30-44. <https://doi.org/10.1016/j.edurev.2014.05.003>
- Alter, P., Walker, J. N., & Landers, E. (2013). Teachers' perceptions of students' challenging behavior and the impact of teacher demographics. *Education and Treatment of Children*, 36(4), 51-69. <https://www.jstor.org/stable/42900226>
- Aman, G. M., Singh, N. N., Stewart, A. W., & Field, C. J. (1985a). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, 89(5), 485-491.
- Aman, G. M., Singh, N. N., Stewart, A. W., & Field, C. J. (1985b). Psychometric characteristics of the aberrant behavior checklist. *American Journal of Mental Deficiency*, 89(5), 492-502.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173. <https://doi.org/10.1007/BF02294170>
- Aydos, E. H., Uysal, H., & Akman, B. (2014). Sınıfa Uyum Öğretmen Gözlemi-Kontrol Listesi'nin Türkçeye uyarlama çalışması [Teacher Observation Of Classroom Adaptation-Checklist Turkish adaptation study]. *Kastamonu Eğitim Dergisi*, 23(2), 809-824. <https://dergipark.org.tr/en/pub/kefdergi/issue/22599/241434>
- Bakanlar Kurulu Kararı [Council of Ministers Decision] (2002). İstatistiki bölge birimleri sınıflandırması [Classification of istatistical territorial units]. 28/08/2002 Tarihli ve 2002/4720 Sayılı Kararname <https://www.resmigazete.gov.tr/eskiler/2002/09/20020922.htm>
- Baker, J. A., Clark, T. P., Maier, K. S., & Viger, S. (2008). The differential influence of instructional context on the academic engagement of students with behavior problems. *Teaching and Teacher Education*, 24, 1876-1883. <https://doi.org/10.1016/j.tate.2008.02.019>
- Bentler P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bradshaw, C. P., & Kush, J. M. (2020). Teacher observation of classroom adaptation-checklist: Measuring children's social, emotional, and behavioral functioning. *Children & Schools*, 42(1), 29-40. <https://doi.org/10.1093/cs/cdz022>
- Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2016). *Direct behavior rating (DBR): Linking assessment, communication, and intervention*. Guilford Press.

- Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71(6), 733-750. <https://doi.org/10.1037/0022-0663.71.6.733>
- Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. A. (2007). Generalizability and dependability of direct behavior ratings to assess social behavior of preschoolers. *School Psychology Review*, 36(1), 63-79. <https://doi.org/10.1080/02796015.2007.12087952>
- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct behavior rating (DBR) an emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, 34(4), 195-200. <https://doi.org/10.1177/1534508409340391>
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, 48(3), 219-246. <https://doi.org/10.1016/j.jsp.2010.02.001>
- Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. <https://doi.org/10.18637/jss.v048.i06>
- Charles, C. M. (2011). *Building classroom discipline*. Pearson.
- Chionh, Y. H., & Fraser, B. J. (2009). Classroom environment, achievement, attitudes and self-esteem in geography and mathematics in Singapore. *International Research in Geographical and Environmental Education*, 18(1), 29-44. <https://doi.org/10.1080/10382040802591530>
- Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology*, 28(6), 693-710. <https://doi.org/10.1080/01443410802206700>
- Cohen, R. J., & Swerdlik, M. E. (2013). Test development (A. Kan, Trans.). In E. Tavşancıl (Ed.), *Psychological testing and assessment: Introduction to tests and measurement* (pp. 233-277). Nobel Publishing. (Original work published 2010)
- Çengel, M., & Türkoğlu, A. (2015). Öğrenciler tarafından algılanan sınıf iklimi ölçeği: Geçerlik ve güvenirlik çalışması [Classroom climate perceived by students scale: A validity and reliability study]. *Eğitimde Kuram ve Uygulama*, 11(4), 1240-1257. <https://dergipark.org.tr/en/pub/eku/issue/5467/74233>
- Daniels, B., Briesch, A. M., Volpe, R. J., & Owens, J. S. (2021). Content validation of direct behavior rating multi-item scales for assessing problem behaviors. *Journal of Emotional and Behavioral Disorders*, 29(2), 71-82. <https://doi.org/10.1177/1063426619882345>
- DeVellis, R. F. (2014). *Ölçek geliştirme, kuram ve uygulamalar [Scale development, theory and applications]*. (T. Totan, Trans.; 3. ed.). Nobel Yayıncılık. (Original work published 2011)
- Dwyer, K. K., Bingham, S. G., Carlson, R. E., Prisbell, M., Cruz, A. M., & Fus, D. A. (2004). Communication and connectedness in the classroom: Development of the connected classroom climate inventory. *Communication Research Reports*, 21(3), 264-272. <https://doi.org/10.1080/08824090409359988>
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16, 5-18. <https://doi.org/10.1007/s11136-007-9198-0>
- Erol, N., Arslan, B. L., & Akçakin, M. (1995). The adaptation and standardization of the Child Behavior Checklist among 6-18 year old Turkish children. In J. Sergeant (Ed.) *Eunethydis: European approaches to hyperkinetic disorder*. Fotoratar, 97-113.
- Erol, N., & Şimşek, Z. (1998). Türkiye ruh sağlığı profili: Çocuk ve gençlerde ruh sağlığı (Erol, N., Kılıç, C., Ulusoy, M., Keçeci, M., Şimşek, Z.) Türkiye ruh sağlığı profili [Türkiye mental health profile]. T.C. Sağlık Bakanlığı Sağlık Projesi Genel Müdürlüğü.
- Evans, I. M., Harvey, S. T., Buckley, L., & Yan, E. (2009). Differentiating classroom climate concepts: Academic, management, and emotional environments. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 4(2), 131-146. <https://doi.org/10.1080/1177083X.2009.9522449>

- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.
- Flannery, W. P., Reise, S. P., & Widaman, K. F. (1995). An item response theory analysis of the general and academic scales of the self-description questionnaire II. *Research in Personality*, 29(2), 168-188. <https://doi.org/10.1006/jrpe.1995.1010>
- Ford, T., Hayes, R., Byford, S., Edwards, V., Fletcher, M., Logan, S., Norwich, B., Pritchard, W., Allen, K., Allwood, M., Ganguli, P., Grimes, K., Hansford, L., Longdon, B., Norman, S., Price, A., Russell, A. E., & Ukoumunne, O. C. (2019). Training teachers in classroom management to improve mental health in primary school children: The STARS cluster RCT. *Public Health Research*, 7(6), 1-150. <https://doi.org/10.3310/phr07060>
- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, 31(3), 328-349. <https://doi.org/10.1080/02796015.2002.12086159>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. Pearson Education Limited.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publication.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological test: A progress report. *European Journal of Psychological Assessment*, 10(3), 229-244. <https://eric.ed.gov/?id=ED399291>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60. <https://academic-publishing.org/index.php/ejbrm/article/view/1224>
- Hu, L. T., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, 34, 111-117. <https://doi.org/10.1177/001316447403400115>
- Kanay, A., & Girli, A. (2008). Analyzing the relations between adaptive behaviors, self concept and academic achievements of elementary school students aged between 9-13 years old with attention deficit hyperactivity disorder. *Journal of Buca Faculty of Education*, 23, 184-191. <https://dergipark.org.tr/en/pub/deubefd/issue/25428/268284>
- Kaner, S., & Çiçekçi, A. U. (2000). Gözden geçirilmiş davranış problemleri kontrol listesinin Türkçeye uyarlanması [Adaptation of the revised behavior problems checklist to Turkish]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 2(4), 23-34. https://doi.org/10.1501/Ozlegt_0000000048
- Kaner, S., Büyüköztürk, Ş., İşeri, E., Ak, A., & Özyayın, L. (2011). Conners Öğretmen Dereceleme Ölçeği-Yenilenmiş/Uzun: Türk çocukları için psikometrik özelliklerinin değerlendirilmesi [Conners Teacher Rating Scale-Revised/Long Form: Evaluation of Psychometric Properties for Turkish Children]. *Klinik Psikiyatri Dergisi*, 14(3), 150-163.
- Kellam, S. G., Branch, J. D., Agrawal, K. C., & Ensminger, M. E. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention and evaluation*. University of Chicago Press.
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, 10, 165-185. <https://doi.org/10.1017/S0954579498001564>
- Kline, T. J. B. (2005). *Psychological testing, a practical approach to design and evaluation*. Sage.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151-162. <https://doi.org/10.32614/RJ-2014-031>

- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation-checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*, 42(1), 15-30. <https://doi.org/10.1177/0748175609333560>
- Kremer, K. P., Flower, A., Huang, J., & Vaughn, M. G. (2016). Behavior problems and children's academic achievement: A test of growth-curve models with gender and racial differences. *Children and Youth Services Review*, 67, 95-104. <https://doi.org/10.1016/j.childyouth.2016.06.003>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage Publications.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Leaf, P. J., Schultz, D., Keys, S., & Ialongo, N. (2002). *The Teacher Observation of Classroom Adaptation-Checklist (TOCA-C)*. Johns Hopkins Center for the Prevention of Youth Violence.
- Lewis, T. J., Newcomer, L. L., Trussell, R., & Richter, M. (2006). Schoolwide positive behavior support: Building systems to develop and maintain appropriate social behavior. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73-95). Lawrence Erlbaum Associates.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410. <https://doi.org/10.1037/0033-2909.103.3.391>
- Merrell, K. W. (1993). Using behavior rating scales to assess social skills and antisocial behavior in school settings: Development of the school social behavior scales. *School Psychology Review*, 22(1), 115-133. <https://doi.org/10.1080/02796015.1993.12085641>
- Miles, M. B., Huberman, A. M. & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Milli Eğitim Bakanlığı [Ministry of National Education]. (2023). National Education Statistics Formal Education 2022/2023 https://sgb.meb.gov.tr/www/icerik_goruntule.php?KNO=508
- Nash, P., Schlosser, A., & Scarr, T. (2016). Teachers' perceptions of disruptive behaviour in schools: A psychological perspective. *Emotional & Behavioural Difficulties*, 21(2), 167-180. <https://doi.org/10.1080/13632752.2015.1054670>
- Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2), 224-240. <https://doi.org/10.17105/spr-14-0008.1>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.
- Ødegård, M., & Solberg, S. (2024). Identifying teachers' reactive strategies towards disruptive behavior in classrooms. *Teaching and Teacher Education*, 145, 1-9. <https://doi.org/10.1016/j.tate.2024.104627>
- Özbay, Y., & Şahin, M. (2000). Empatik Sınıf Atmosferi Ölçeği (ESATÖ): Geçerlik ve güvenirlik çalışması [Validity and reliability of Intraclassroom Empathic Attitude Scale (IEAS)]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 19(19), 104-113. https://dergipark.org.tr/tr/pub/hunefd/issue/7819/102747#article_cite
- Pas, E., Cash, A., O'Brennan, L., Debnam, K., & Bradshaw, C. (2015). Profiles of classroom behavior in high schools: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology*. 53(2), 137-148. <https://doi.org/10.1016/j.jsp.2014.12.005>
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis*. Sage Publications.
- Pierce, C. (1994). Importance of classroom climate for at-risk learners. *The Journal of Educational Research*, 88(1), 37-42. <https://www.jstor.org/stable/27541952>
- Quay, H. C., & Peterson, D. R. (1996). *Revised behavior problem checklisit*. Professional manual.

- Robinson, S. L., & Ricord-Griesemer, S. M. (2006). Helping individual students with problem behavior. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 73-95). Lawrence Erlbaum Associates.
- Ross, R., Starrett, A., & Irvin, M. J. (2023). Examining the relationships between kindergarteners' demographic characteristics and behavior problems across geographic locale. *Early Childhood Education Journal*, 51(8), 1427-1439. <https://doi.org/10.1007/s10643-022-01391-6>
- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2), 437-460. <https://doi.org/10.3102/00028312038002437>
- Sağkal, A. S., Kabasakal, Z. T., & Törnüklü, A. (2015). Sınıf iklimi envanteri'nin (SİE) Türkçe'ye uyarlanması [Turkish adaptation of the connected classroom climate inventory (CCCI)]. *İlköğretim Online*, 14(4), 1179-1192.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351-380. <https://www.jstor.org/stable/42899983>
- Skinner, E. A., Zimmer-Gembeck, M. J., & Connell, J. P. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, 63(2-3), v-220. <https://doi.org/10.2307/1166220>
- Somersalo, H., Solantaus, T., & Almqvist, F. (2002). Classroom climate and the mental health of primary school children. *Nordic Journal of Psychiatry*, 56(4), 285-290. <https://doi.org/10.1080/08039480260242787>
- Sprott, J. B. (2004). The development of early delinquency: Can classroom and school climates make a difference? *Canadian Journal of Criminology and Criminal Justice*, 46(5), 553-572. <https://doi.org/10.3138/cjccj.46.5.553>
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325. <https://link.springer.com/article/10.1007/BF02295289>
- Sucuoğlu, N. (2003). Sorun davranışlar kontrol listesinin Türkçe formunun psikometrik özelliklerinin incelenmesi [Examining the psychometric properties of the Turkish form of the problem behavior checklist]. *Türk Psikoloji Dergisi*, 18(52), 77-91. <https://psikiyatridizini.net/viewarticle.aspx?articleid=21399&tammetinvar=yes>
- Sucuoğlu, B., & Kargın, T. (2010). *İlköğretimde kaynaştırma uygulamaları: Yaklaşımlar, yöntemler, teknikler* [Inclusive practices in primary education: Approaches, methods, techniques]. Kök Yayıncılık.
- Şeker, H., & Gençdoğan, B. (2014). *Psikolojide ve eğitimde ölçme aracı geliştirme* [Developing measurement tools in psychology and education]. Nobel Yayıncılık.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson Education.
- Toraman, Ç., & Özdemir H. F. (2023). Evren ve örneklem. In A. İlğan & A. Ekinci (Eds.), *Eğitimde araştırma yöntemleri* [Research methods in education] (pp.135-167). Nobel Akademik Yayıncılık.
- Vieira, A. L. (2011). Interactive LISREL in practice, getting started with a SIMPLIS Approach. Springer. <https://doi.org/10.1007/978-3-642-18044-6>
- Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review*, 41(3), 246-261. <https://doi.org/10.1080/02796015.2012.12087506>
- Volpe, R. J., & Briesch, A. M. (2015). Multi-item direct behavior ratings: Dependability of two levels of assessment specificity. *School Psychology Quarterly*, 30(3), 431-442. <http://dx.doi.org/10.1037/spq0000115>

- Volpe, R. J., & Briesch, A. M. (2016). Dependability of two scaling approaches to direct behavior rating multi-item scales assessing disruptive classroom behavior. *School Psychology Review*, 45(1), 39-52. <https://doi.org/10.17105/SPR45-1.39-52>
- Warner, R. M. (2013). *Applied statistics, from bivariate through multivariate techniques*. Sage Publications.
- Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). *Teacher observation of classroom adaptation-revised (TOCA-R) [Database record]*. APA PsycTests. <https://doi.org/10.1037/t31163-000>
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187-213. <https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>
- Yukay-Yüksel, M. (2009). Okul sosyal davranış ölçeklerinin (OSDÖ) Türkçeye uyarlanması [Adaptation of school social behavior scales (SSBS) to Turkish]. *Kuram ve Uygulamada Eğitim Bilimleri*, 9(3), 1605-1650. <https://tr-scales.arabpsychology.com/wp-content/uploads/pdf/okul-sosyal-davranis-olcekleri-b-formu-olumsuz-sosyal-davranislar-toad.pdf>
- Zhao, Y. (2008). *Approaches for addressing the fit of item response theory models to educational test data*. [Doctoral dissertation, University of Massachusetts Amherst]. <https://www.proquest.com/docview/220010848?fromopenview=true&pqorigsite=gscholar&sourcetype=Dissertations%20&%20Theses>

