

Aydınlatıcı Program Değerlendirme Modeli Program Değerlendirme Ölçeği (Öğrenci Formu): Geçerlik ve Güvenirlik Çalışması

Osman AKTAN^{1*} 
Beyza ÜÇÜNCÜ² 
Tuğbanur BİNGÖL^{3*} 
Özge GÖKBAYRAK⁴ 

¹Düzce Üniversitesi, Eğitim Fakültesi, Özel Eğitim Bölümü, Düzce, Türkiye
osmanaktan@duzce.edu.tr

²Düzce Ölçme Değerlendirme Merkezi, Düzce, Türkiye
beyza_3@outlook.com

³Fevzi Çakmak Ortaokulu, Van, Türkiye
bingoltugbanur@gmail.com

⁴Düzce Mesleki Eğitim Merkezi, Düzce, Türkiye
ozgeakgun462@gmail.com

Geliş tarihi: 18.07.2025
Kabul tarihi: 30.10.2025
Yayın tarihi:

Özet: Bu çalışmada, Öğrenme Güçlüğü Destek Eğitim Programı'nın (ÖGDEP) öğrenci görüşleri doğrultusunda değerlendirilmesi için, Aydınlatıcı Program Değerlendirme Modeli (APDM)'ne dayalı bir ölçek geliştirilmesi amaçlanmaktadır. Ölçeğin geliştirilme aşamasında, uzman görüşleri alınmış, kapsam ve yapı geçerliği pilot uygulama ile test edilmiştir. Bu amaç doğrultusunda araştırmacılar tarafından hazırlanan taslak ölçek maddeleri 2022-2023 eğitim-öğretim yılında üç çalışma grubu olarak ÖGDEP'den yararlanan toplam 502 öğrenciye uygulanmıştır. Ölçek geliştirme kapsamında uzman görüşleri doğrultusunda pilot uygulama gerçekleştirilmiş, ölçeğin kapsam ve yapı geçerliği test edilmiştir. Açıklayıcı Faktör Analizi sonucu meydana gelen ölçek yapısının geçerliği, Doğrulayıcı Faktör Analizi ile test edilmiştir. Analizler sonucunda, APDM Program Değerlendirme Ölçeği (Öğrenci Formu)'ni (APDMPDÖ-ÖF) oluşturan 16 madde ve tek faktörlü model doğrulanmıştır. Ölçeğin açıkladığı toplam varyans % 62.30 olarak tespit edilmiştir. Ölçek maddelerine ilişkin faktör yük değerleri 0.30 ile 0.88 arasında değişmektedir. Analizler sonucu, APDMPDÖ-ÖF'nin geçerli ve güvenilir sonuçlar verdiği ve APDM'ye dayalı olarak ÖGDEP'in değerlendirilmesine yönelik öğrenci görüşlerini ölçmek amacıyla kullanılabileceği söylenebilir.

Anahtar Kelime: Program Değerlendirme, Aydınlatıcı Program Değerlendirme Modeli, Öğrenci, Aydınlatıcı Değerlendirme Modeline Dayalı Program Değerlendirme Ölçeği

GİRİŞ

Günümüz kapsayıcı eğitim uygulamaları ile birlikte bireysel farklılıklar eğitim alanında önem kazanmaya başlamış, bireysel farklılıklar öğrenme süreçlerinde daha fazla gözetilen bir kavram olmuştur (Aktan, 2023). Özel eğitim, bireysel farklılıklara dayalı bir eğitim anlayışına dayanmakta olup; bireyin kişisel, sosyal ve öğrenme özellikleri dikkate alınarak, gereksinimlerine uygun öğretim süreçleri planlanmakta ve uygulanmaktadır (Zigmond & Kloo, 2017).

Özel eğitim; öğrenme, sosyal, duygusal, davranışsal ve fiziksel alanlarda yetersizlik yaşayan öğrencilere, özel olarak hazırlanmış programlar ve özel olarak yetiştirilmiş uzmanlar aracılığı ile araştırma temelli, sistematik öğretim ve destek hizmetlerinin sağlanmasıdır (Kauffman & Hallahan, 2011). Bu bağlamda özel eğitimin amacı, özel gereksinimli bireyleri ilgi ve ihtiyaçları doğrultusunda gelişimlerini en üst düzeyde sağlamak ve onları toplumda bağımsız bireyler olarak yer alabilecek şekilde yetiştirmektir (Bryant vd., 2019). Özel gereksinimli bireyler öncelikli olarak, genel eğitim sınıflarında akranları ile birlikte olmak üzere, yetersizlik tür ve derecesine göre farklı eğitim ortamlarında öğrenim görmektedirler (Armstrong vd., 2016). Gereksinim temelli sağlanan destek hizmetleri ile gerçekleştirilen kaynaştırma uygulamaları, özel gereksinimli öğrencilerin sosyal uyumları ile akademik gelişimlerinin sağlanması amacını taşımaktadır (Aktan, 2021; Hornby, 2011). Kaynaştırma uygulamalarının başarısı için öğrencilere sınıf içinde özel eğitim öğretmeni desteği, öğretim uyarlamaları, akran desteği; sınıf dışında danışmanlık, uzman desteği, destek eğitim odası desteği, ek öğretim destekleri gibi destek hizmetleri sağlanmaktadır (Aktan, 2020; Kutay, 2018; Paulsrud & Nilholm, 2023). Bu desteklerden biri de okul dışında desteklerden biri olup yetersizlik odaklı olarak geliştirilen destek eğitim programlarıdır.

Türkiye'de destek eğitim programları Milli Eğitim Bakanlığı (MEB) tarafından ilk defa 2008 yılında, alanda çalışan öğretim üyeleri, uzmanlar ve öğretmenlerin görüşleri doğrultusunda, özel gereksinimli öğrencilerin yetersizlik türleri ile gereksinimleri dikkate alınarak, yedi farklı yetersizlik alanında modüler programlar olarak hazırlanmıştır (MEB, 2008). Özel eğitimdeki yeni gelişmeler sonucu, MEB tarafından öğretim üyeleri, öğretmenler ile farklı alan uzmanlarının katılımı ile öğrenen merkezli, öğretmene rehber, güncel içerikler ile güncel yöntem ve tekniklere dayalı olarak, 2021 yılında yeni destek eğitim programları

Cite as (APA 7): Aktan, O., Üçüncü, B., Bingöl, T. ve Gökbayrak, Ö. (2026). Aydınlatıcı program değerlendirme modeli program değerlendirme ölçeği (öğrenci formu): Geçerlik ve güvenilirlik çalışması. *Trakya Eğitim Dergisi*, 16(3), 1437–1471. <https://doi.org/10.24315/tred.1745348>

geliştirilmiştir. Yenilenen destek eğitim programlarıyla özel gereksinimli bireylerin; öz bakım, günlük yaşam, dil ve iletişim, bilişsel, sosyal, duygusal ve bedensel gelişim alanlarındaki ilerlemeleri; temel akademik becerileri kazanmaları, milli ve manevi değerleri edinmeleri ile toplumla etkileşim kurmalarının ve bütünleşmelerinin sağlanması amaçlanmaktadır. Ayrıca, toplumda bağımsız bireyler olarak yaşamlarını sürdürebilmeleri desteklenmektedir. Destek eğitim programları; zihinsel yetersizlik, otizm, dil ve konuşma güçlüğü, işitme yetersizliği, görme yetersizliği, bedensel yetersizlik ve öğrenme güçlüğü olmak üzere yedi yetersizlik alanındaki bireylere yönelik olarak hazırlanmıştır (MEB, 2021).

ÖGDEP ile okuma, yazma ve matematik konularında problem yaşayan bireylerin; dil ve iletişim, algı, motor planlama, bellek, sosyal uyum, öz düzenleme, okuryazarlık, matematik becerilerinin desteklenmesi hedeflenmektedir. ÖGDEP, altı modülden ve toplam 1100 ders saatinden oluşmaktadır. Program; öğrenen merkezli program tasarımı esas alınarak, ilerlemecilik yaklaşımı ile pragmatist düşüncenin eğitime yansımaları doğrultusunda geliştirilmiştir. Modüler olarak hazırlan program ile öğrenen merkezli bir anlayışla ve bireyin performansına dayalı olarak, uygulamada önerilen yöntem ve teknikler aracılığıyla bireylerin desteklenmesi amaçlanmaktadır. Bireyin mevcut performansı ve eğitim gereksinimleri dikkate alınarak, ÖGDEP'e dayalı dayalı olarak her birey için Bireyselleştirilmiş Eğitim Programı (BEP) hazırlanmalı ve uygulanmalıdır (Aktan & Orakcı, 2026; MEB, 2021). ÖGDEP, özel eğitim ve rehabilitasyon merkezlerinde öğrenme güçlüğü tanısı bireylere uygulanmaktadır. İlk olarak 2021 yılında uygulanmaya başlanan programın etkililiğini ortaya koymak amacıyla, hem programı uygulayanlar hem de programdan yararlanan bireylerin görüşlerine dayalı olarak yapılan değerlendirmelerin, programa dayalı öğretimin niteliğine katkı sağlayabileceği düşünülmektedir.

Program değerlendirme, program geliştirmenin son aşamasıdır. Bu süreç; programın amaçlarına ne ölçüde ulaşıldığını belirlemek, programın işleyişine yönelik veriler toplamak ve programın etkililiği hakkında karar vermek amacıyla yürütülür (Chen, 2005; Oliva, 2009). Program değerlendirme, hazırlanan yeni bir eğitim programının uygulamadaki işleyişini görme ve sahadan elde edilen verilere dayanarak programı geliştirme için yürütülen dinamik bir süreçtir (Marsh & Willis, 2007). Program değerlendirme süreci, program ile değerlendirme modeli ve buna dayalı veri toplama araçlarının belirlendiği planlama aşaması, program değerlendirme modeline uygun veri toplama araçları ile verilerin toplanması, analizi ve yorumlanmasını içeren uygulama aşaması ile program değerlendirme verilerine dayalı programdaki eksikliklerin belirlendiği değerlendirme aşaması olmak üzere üç aşamada gerçekleştirilir. Program değerlendirmek amacıyla Tyler, Cronbach, Campbell ve Stanley, Stake, Stufflebeam, Scriven, Suchman, Alkin, Guba, Provan, Weiss gibi farklı kişiler tarafından çeşitli program değerlendirme modelleri geliştirilmiştir (Madaus vd., 2012). Katılımcı görüşleri odaklı program değerlendirme modellerinden biri de APDM'dir (Burden, 2017; Parlett & Hamilton, 1977).

APDM'de hazırlanan program; öğretmen, öğrenci, program geliştirme uzmanı ve alan uzmanı gibi doğrudan ilgili katılımcılar tarafından değerlendirilerek, programın aksayan yönlerinin belirlenmesi amaçlanmaktadır (Parlett & Hamilton, 1977). Bu modelde, programın süreci; programın uygulandığı kurumlar ve programdan yararlanan paydaşların programa yönelik değerlendirmeleri ele alınmakta, ayrıca programın öğrencilerin akademik ve sosyal gelişimlerine olan etkileri de ele alınmaktadır (Özudoğru & Adıgüzel, 2016). Bunun yanı sıra, programın uygulanmasında karşılaşılan problemler, programa öğretmen, öğrenci ve diğer paydaşların katılım düzeyleri, kullanılan araç-gereçlerin ve diğer etkenlerin programa etkisi ile programın öğrencilere olan yansımaları da değerlendirme kapsamına alınmaktadır (Uşun, 2016). Modelde, programın uygulanmasında yönelik resmi dökümanlar, raporlar, uygulamaya yönelik açıklamalar, donanımsal imkânları içeren öğretim sistemi ile öğretmen ve öğrencilerin etkileşimde bulundukları öğrenme ortamı olmak üzere iki temel kavram söz konusudur (Parlett & Hamilton, 1977). APDM, programdan beklenen çıktılar ile öğrenme ortamında gerçekleşen öğrenmeler arasında ilişkiye odaklanmaktadır.

Öğretim programlarında hedeflenen çıktılara ulaşabilmek için, programın etkili olarak uygulanması büyük önem taşır. Bu doğrultuda, programın uygulanması sürecinde karşılaşılan olumlu ya da olumsuz durumları belirlemek ve programın etkililiğini artırmak amacıyla, programdan doğrudan etkilenen katılımcıların deneyimleri ve duyguları dikkate alınmaktadır (Cahapay, 2020). Bu nedenle öğretmen, öğrenci, uzman ve aile gibi farklı paydaşların hem programın uygulanmasına yönelik görüşleri hem de sınıf içi öğretim süreçlerine ilişkin deneyimleri değerlendirme kapsamına alınmaktadır (Richards, 2001). Bu çerçevede, APDM; ideal bir eğitim programının nasıl olması gerektiğine dair nitelikli bilgiler sunmayı, katılımcı görüşleri doğrultusunda programın güçlü ve geliştirilmesi gereken yönlerini belirlemeyi ve programın öğrencilerin

akademik ve sosyal gelişimlerine katkılarını değerlendirmeyi amaçlar. Ayrıca model, mevcut programın başarıya da yetersizlik nedenlerine ilişkin ayrıntılı analizler yapılabilmesi için değerlendiricilere kapsamlı veri toplama imkânı sunar (Parlett & Hamilton, 1977). Bu amaçla, model kapsamında hem nitel (gözlem, görüşme, doküman incelemesi vb.) hem de nicel (test, anket, ölçek vb.) veri toplama yöntemlerinden yararlanılmaktadır (Ornstein & Hunkins, 2009).

Program değerlendirme çalışmalarında ölçeklerin kullanımı yaygın bir uygulamadır. Türkiye’de ilgili alayazını incelendiğinde, program değerlendirme araştırmalarına veri sağlamak amacıyla derslere yönelik öğretim programları için (Bal & Arslan, 2021; Karabay-Turan vd., 2019) ve genel olarak farklı amaçlarla hazırlanan programların değerlendirilmesine yönelik ölçme araçları geliştirildiği görülmektedir (Adıgüzel, 2015; Akıncı & Köse, 2019; Altınsoy & Küçüksüleymanoğlu, 2021; Aslan vd., 2016). Öğretim programlarının değerlendirilmesine ilişkin çalışmalar için geliştirilen ölçme araçlarında, değerlendirme modeli olarak özellikle CIPP modeli (Çopur vd., 2021; Kavan, 2023) ve APDM (Aktan vd., 2025; Adıgüzel & Özudoğru, 2014) temel alınmıştır. Yapılan araştırmalar incelendiğinde, alanyazında program değerlendirme amacıyla bazı ölçek geliştirme çalışmaları yapıldığı görülmektedir. APDM’nin esas alındığı iki ölçme aracı olduğu ve hedef kitlesinin ise öğretmenler olduğu görülmektedir. Bununla birlikte APDM’ye dayalı, öğrencilerin program hakkında görüşlerinin değerlendirilmesine yönelik bir ölçme aracının henüz geliştirilmediği tespit edilmiştir.

Araştırmanın Amacı

Program değerlendirme çalışmalarında gözlem, görüşme, doküman inceleme, ölçek ve anket gibi hem nicel hem de nitel ölçme araçlarından yararlanılmaktadır (Rao & Woolcock, 2003). Ancak gözlem, görüşme, doküman inceleme gibi nitel veri toplama yöntemleri, zaman alıcı olmaları ve yorumlayıcının öznel değerlendirmelerine açık yapıları nedeniyle bazı sınırlılıklar içermektedir (Chen, 2010). Bu nedenle, çok sayıda katılımcının programa ilişkin görüşlerinin objektif ve sistematik olarak toplanabilmesi için geçerli ve güvenilir ölçme araçlarına ihtiyaç duyulmaktadır. Öğrenme güçlüğü olan öğrenciler için hazırlanmış ÖGDEP bağlamında, programın öğrenci görüşleri doğrultusunda değerlendirilmesine olanak tanıyan geçerli ve güvenilir bir ölçeğin geliştirilmesi, değerlendirme sürecinin katılımcı odaklı ve objektif bir şekilde yürütülmesine önemli katkılar sağlayabilir.

APDM, bir programın etkililiğini yalnızca nicel çıktılar ile birlikte, programdan doğrudan etkilenen öğretmenler, öğrenciler, alan uzmanları ve program geliştiriciler gibi paydaşların algı, deneyim ve değerlendirmelerine dayalı olarak incelemeyi temel almaktadır (Parlett & Hamilton, 1977). Bu yaklaşım, programın uygulanma sürecinde ortaya çıkan dinamikleri, karşılaşılan güçlükleri ve uygulamanın gerçek bağlamını derinlemesine analiz etmeyi amaçlar. Program değerlendirmesi sonucunda elde edilen bulguların, uygulamada karşılaşılan sorunlara çözüm önerileri geliştirilmesine, programın güçlü ve geliştirilmesi gereken yönlerinin belirlenmesine ve dolayısıyla programın yeniden yapılandırılmasına katkı sağlaması beklenmektedir. Özellikle programın yararlanıcısı olan öğrencilerin görüşlerine dayalı elde edilen veriler, programın uygulama süreçlerinin iyileştirilmesi açısından önemli rehberlik işlevi görür. Bu bağlamda ÖGDEP’den yararlanan öğrencilerin programa ilişkin deneyim ve algılarının belirlenmesi, programın etkililiğinin anlaşılması ve öğrenci merkezli iyileştirme stratejilerinin geliştirilmesi açısından kritik öneme sahiptir. Bu bütüncül değerlendirme süreci, programın sürdürülebilirliği ve eğitim kalitesinin artırılması açısından sağlam bir temel oluşturmakta, farklı paydaşların bakış açılarını dikkate alan APDM program değerlendirme yaklaşımını desteklemektedir. Bu doğrultuda, bu araştırmanın temel amacı, ÖGDEP’den yararlanan öğrencilerin program hakkındaki görüşlerini sistematik biçimde ortaya koymak amacıyla APDM’ye dayalı bir ölçme aracı olan APDMPDÖ-ÖF’nin geliştirilmesidir.

YÖNTEM

Araştırma Modeli

Bu araştırma, ölçek geliştirme amaçlı betimsel bir araştırmadır. Ölçek, ÖGDEP’den yararlanan ve ilkökul, ortaokul ve lise kademelerinde öğrenim gören öğrenme güçlüğü tanısı almış öğrencilere yönelik olarak hazırlanmıştır. Araştırma kapsamında, APDM esas alınarak, öğrenme güçlüğü olan bireylerin ÖGDEP’ye ilişkin değerlendirmelerini ölçmeyi amaçlayan bir ölçek geliştirme sürecine yönelik işlem basamakları aşağıda sırasıyla sunulmuştur.

Ölçek Geliştirme Basamakları

Ölçek geliştirme sürecinde, aşağıda sıralanan işlem basamakları dikkate alınmıştır (Carpenter, 2018; DeVellis, 2017; Kyriazos & Stalikas, 2018; Price, 2017);

Ölçülecek yapı ve özelliklerin belirlenmesi: Ölçek geliştirme sürecinin ilk adımı olarak, ölçeğin amacı ve ölçülmek istenen özellikler belirlenmiştir. Bu bağlamda, ÖGDEP'nin APDM'ye göre değerlendirilmesi esas alınmıştır. APDM, programdan etkilenen öğretmen, öğrenci, uzman ve aile gibi farklı katılımcıların duygu ve deneyimlerinin yanı sıra, programın uygulanmasına yönelik görüşler ile sınıf içi öğretim uygulamalarını dikkate almaktadır (Richards, 2001). Model, programın hedef ve çıktıları yerine, program süreci ve işleyişine odaklanması yönüyle diğer program değerlendirme modellerinden ayrılmaktadır (Burden, 2017). Geliştirilecek ölçek ile öğrenme güçlüğü tanılı öğrencilerin, APDM çerçevesinde ÖGDEP'ye ilişkin duygu ve deneyimlerine dayalı değerlendirmelerinin ölçülmesi, böylece programın etkili uygulanmasına ilişkin geçerli ve güvenilir veri sağlamaya yönelik bir ölçme aracının geliştirilmesi amaçlanmıştır. Ölçek, ilkökul, ortaokul ve lisede öğrenim gören ve destek eğitim programlarından yararlanan öğrenme güçlüğü tanılı bireylere yönelik hazırlanmıştır. Farklı eğitim kademelerinde aynı programdan yararlanan öğrenciler olması nedeniyle, programın tüm eğitim kademelerinde öğrenim gören öğrenciler üzerinde etkisinin bütüncül olarak değerlendirilmesi amacıyla, her kademedeki veri toplanmıştır. Uygulama aşamasında ise öğrencilerin yaş ve gelişim düzeyleri dikkate alınarak, ölçeğin doğru ve sağlıklı uygulanabilmesi için öğrencilere uygulama öncesi gerekli açıklamalar yapılmıştır. Öğrencilere, ölçeğin nasıl doldurulacağı ve ölçek maddelerindeki derecelendirmelerin ne anlama geldiği konusunda öğretmenler ve araştırmacılar tarafından rehberlik sağlanmıştır.

Madde Yazımı

Taslak madde yazımı sürecinde kapsamlı bir literatür taramasından (Aktan vd., 2025; Çelikel, 2022; Gültekin & Demirci, 2020; Özudoğru, 2016; Özudoğru & Adıgüzel, 2016; Palabıyık, 2021; Parlett & Hamiltorn, 1977) ve alan uzmanlarının görüşlerinden yararlanılmıştır. APDM'nin amacı, uygulanan bir eğitim programının öğrenme çıktılarına ulaşip ulaşmadığını kontrol etmekle birlikte, aynı zamanda programın uygulanma sürecine ilişkin durumları ve programdan yararlanan kişilerin görüş ve deneyimlerini belirlemeyi amaçlamaktadır. (Parlett & Hamilton, 1977). Bu doğrultuda, ölçek maddelerinin hem programın öğrenme çıktıları hem de programın uygulanma sürecine ilişkin durumları da kapsayacak şekilde hazırlanması amaçlanmıştır. APDM'nin yaklaşımı ve program değerlendirme anlayışı doğrultusunda, ölçek maddeleri ilk olarak araştırmacılar tarafından bağımsız olarak yazılmıştır. Her bir araştırmacı tarafından oluşturulan maddeler, sonrasında ekip toplantılarında bir araya getirilmiş, içerik açısından benzer olan maddeler birleştirilmiş ve akademik tutarlılık sağlanarak yeniden yapılandırılmıştır. Madde geliştirme sürecinde; birinci araştırmacı 32, ikinci araştırmacı 27, üçüncü araştırmacı 29 ve dördüncü araştırmacı 25 taslak madde önermiştir. Tüm maddeler ekip üyeleri tarafından tek tek değerlendirilmiştir. Ortak olarak tüm araştırmacıların önerdiği 8 madde, kavramsal bütünlüğe uygunluğu açısından doğrudan kabul edilmiştir. Ek olarak, yapılan değerlendirme sonucunda, içerik açısından anlamlı, açık, ölçülebilir ve benzerlik göstermeyen 16 madde daha oy birliği ile seçilerek ölçeğe dahil edilmiştir. Bu doğrultuda 24 maddelik taslak ölçek formu hazırlanmıştır.

Uzman İncelemesi

Taslak ölçek formu, kapsam geçerliliğini sağlamak amacıyla alanında uzman kişiler tarafından değerlendirilmiştir. Bu süreçte; bir ölçme ve değerlendirme uzmanı, iki özel eğitim uzmanı, üç program geliştirme uzmanı ile bir de dil uzmanından görüş alınmıştır. Uzman görüşleri doğrultusunda yapılan değerlendirmeler sonucunda iki maddenin içerik açısından benzer ve aynı ifadeyi taşıdığı tespit edilerek birleştirilmiş, üç madde ise ölçeğin amacına uygun olmadığı gerekçesi ile formdan çıkarılmıştır. Değerlendirmeler sonucu, 21 maddelik nihai taslak ölçek formu elde edilmiştir. Ölçekte yer alan maddelere ilişkin derecelendirme ilk olarak, ilkökul öğrencileri için üçlü, ortaokul ve lise öğrencileri için beşli likert tipin derecelendirme formatı önerilmiştir. Ancak uzman görüşleri doğrultusunda, derecelendirmenin güvenilirliğini artırmak ve uygulama birliğini sağlamak amacıyla beşli likert derecelendirme kullanılmasına karar verilmiştir. Bu kararın alınmasında 4.sınıf ve üzerindeki öğrencilerin uygun dil ve tasarım koşullarında beşli likert derecelendirme seçenekleri arasında anlamlı ve güvenilir ayrımlar yapabildiklerine ilişkin araştırma bulguları da etkili olmuştur (Chambers & Johnston, 2002). Ayrıca, ölçek maddeleri dil ve tasarım açısından, katılımcı öğrencilerin yaş, bilişsel gelişim ve algı düzeylerine uygun olarak sade ve anlaşılır biçimde yapılandırılmıştır. Bu doğrultuda, tüm kademelerde öğrenim gören öğrenme güçlüğü tanılı öğrencilere yönelik ortak bir derecelendirme sistemi benimsenmiştir. Beşli likert tipi derecelendirme ölçeğinde, öğrencilerin madde

ifadesine katılma dereceleri, “hiç katılmıyorum (1)”, “katılmıyorum (2)”, “biraz katılıyorum (3)”, “katılıyorum (4)” ve “tamamen katılıyorum (5)” olarak belirtilmiştir.

Pilot Uygulama

Taslak ölçek formunda bulunan 21 maddenin kapsam ve anlaşılabilirliğini sınamak amacıyla ilkököl, ortaokul ve lise kademelerinde öğrenim gören öğrenme güçlüğü tanısı almış 18 öğrenci ile pilot uygulama gerçekleştirilmiştir. Pilot çalışma öncesinde, araştırmacılarından biri tarafından ölçeğin amacı öğrencilere açık ve sade bir dille açıklanmıştır. Ayrıca, ölçek maddeleri ile ölçeğin derecelendirme sistemi kısaca açıklanarak, öğrencilerin formu doğru bir şekilde doldurmaları için gerekli rehberlik sağlanmıştır. Pilot uygulama kapsamında, öğrencilerin madde anlaşılabilirliğine ve genel izlenimlerine ilişkin geri bildirim almak amacıyla ölçek formuna bir açık uçlu bir soruda eklenmiştir. Öğrenciler tarafından taslak ölçek formu 7-12 dakika arasında tamamlanmıştır. Açık uçlu soru formuna öğrencilerden her hangi maddeye ilişkin anlaşılma durumu dair bir bildirim yapılmamıştır. Pilot çalışma sonucunda elde edilen bulgular doğrultusunda, ölçek maddelerinin hedef kitle tarafından anlaşılır ve uygulanabilir olduğu değerlendirilmiştir.

Puanlama ve veri analiz yöntemi: Geliştirilen ölçek Likert tipi derecelendirme sistemine göre düzenlenmiştir. Her bir madde için puanlama; “Hiç katılmıyorum (1)”, “Katılmıyorum (2)”, “Kısmen katılıyorum (3)”, “Katılıyorum (4)” ve “Tamamen katılıyorum (5)” olarak belirlenmiştir. Ölçekten alınan puanlar, öğrencilerin ÖGDEP’ye yönelik görüşlerini yansıtmaktadır. Bu doğrultuda ölçekten yüksek puanlar alınması, öğrencilerin programı genel olarak olumlu değerlendirdiğini ve programın amacına uygun olarak uygulandığını göstermektedir. Ölçekten düşük puanlar alınması ise, öğrencilerin programın değerlendirilmesine yönelik görüşlerinin olumsuz olduğu ve uygulama sürecinde çeşitli sorunlar olduğuna işaret etmektedir. Geliştirilen ölçeğin yapı geçerliğini belirlemek için açımlayıcı ve doğrulayıcı faktör analizi, güvenilirliğini belirlemek için Cronbach Alfa, bileşik güvenilirlik ve test tekrar test yöntemleri kullanılmıştır.

Çalışma Grubu

Araştırmada veri toplamak amacıyla üç farklı çalışma grubundan veri toplanmıştır. Araştırmanın çalışma grupları ölçüt örnekleme yöntemi ile seçilmiştir. Ölçüt örneklemede araştırmacı, önceden belirlenen ölçütü karşılayan durumları çalışmasına dahil eder (Yıldırım & Şimşek, 2013). Araştırmanın çalışma grubu oluşturulurken, öğrencilerin Rehberlik ve Araştırma Merkezi tarafından eğitsel tanılama sonucunda öğrenme güçlüğü tanısı almış olmaları ve ÖGDEP’den yararlanıyor olmaları temel ölçütler olarak alınmıştır. Ölçek geliştirme çalışmalarında, geçerli analizlerin yapılabilmesi için çalışma grubunun büyüklüğü, ölçek madde sayısının beş ya da on katı olmalıdır (Hair vd., 2010; Tavşancıl, 2014). Araştırmacılar tarafından oluşturulan 21 maddelik taslak ölçek formunda yer alan maddelerin kapsam ve anlaşılabilirliğini sınamak amacıyla, ilk olarak ilkököl (4.sınıf), ortaokul ve lise kademelerinde öğrenim gören 18 öğrenme güçlüğü tanıli öğrenci ile pilot çalışma gerçekleştirilmiştir. Açımlayıcı Faktör Analizi (AFA) sürecinde, farklı eğitim kademelerinde (ilkokul, ortaokul, lise) öğrenim gören 224 öğrenme güçlüğü tanıli öğrenciden veri toplanmıştır. Doğrulayıcı Faktör Analizi (DFA) ise, AFA çalışma grubu ile aynı özelliklere sahip ancak farklı bireylerden oluşan 260 öğrencinin katılımı ile gerçekleştirilmiştir. Böylece araştırma yer alan her iki çalışma grubunda önerilen örneklem büyüklüğüne uygun olarak, ölçek madde sayısının yaklaşık on katı kadar katılımcı içerecek şekilde yapılandırılmıştır. Araştırmanın yer alan öğrencilerin demografik bilgileri Tablo 1’de verilmiştir.

Tablo 1

Çalışma Grubunu Oluşturan Öğrencilerin Demografik Özellikleri

Değişken		Pilot çalışma grubu n/%	AFA Grubu n/%	DFA Grubu n/%
Cinsiyet	Kadın	8 (44.40)	131 (58.48)	153 (58.85)
	Erkek	10 (55.60)	93 (41.52)	107 (41.15)
Kademe	İlkokul	5 (27.78)	32 (14.29)	48 (18.46)
	Ortaokul	7 (38.89)	117 (52.23)	130 (50.00)
	Lise	6 (33.33)	75 (33.48)	82 (31.54)
Program	1 yıl	6 (33.33)	86 (38.40)	116 (44.60)
yararlanma süresi	2 yıl	12 (66.67)	138 (61.60)	144 (55.40)
Total		18	224	260

Araştırma grubunu oluşturan öğrencilere ilişkin demografik veriler incelendiğinde, katılımcıların çoğunluğunu erkek öğrencilerin oluşturduğu, büyük bir kısmının ortaokul düzeyinde öğrenim gördüğü ve önemli bir bölümünün en az iki yıldır ÖGDEP'den yararlandıkları görülmektedir.

Verilerin Toplama ve Analizi

Araştırma verileri, 2022-2023 eğitim öğretim yılı bahar döneminde (Mart-Mayıs ayları arasında) Düzce ilinde bulunan çeşitli okullarda öğrenim gören öğrenme güçlüğü tanılı öğrencilerden toplanmıştır. Araştırmaya başlamadan önce, bilimsel araştırma yayın etik kurul izni alınmıştır. Etik kurul onayından sonra Düzce İl Milli Eğitim Müdürlüğü'nden araştırmanın sahada uygulanmasına yönelik uygulama izni alınmıştır. Katılımcıların 18 yaşından küçük bireyler olmaları nedeniyle ayrıca Veli Onam Formu alınmıştır. Araştırmada analizler öncesi AFA ve DFA veri seti, örneklem büyüklüğü, normallik, uç değerler, doğrusallık, çoklu bağlantı bakımından incelenmiştir. Uç değerleri belirlemek amacıyla ölçek maddelerine ilişkin minimum ve maximum değerler incelenmiş; her iki veri setinin de uygun olduğuna karar verilmiştir. Çalışma grubu verilerinin faktör analizine uygunluğu Kaiser-Meyer-Olkin (KMO) kat sayısı ve Bartlett Sphericity Testi uygulanmıştır (Field, 2018). Ölçeğin faktör belirleme, ilgili maddelerin toplanması ve faktörleşmesini sağlamak amacıyla temel bileşenler faktörleşme analizi yapılmıştır (Büyüköztürk, 2017). Faktörleşme yöntemi olarak madde yük değerlerinin varyanslarını en yüksek seviyeye çıkarmak amacıyla eksen döndürme tekniklerinden Varimax yönteminin kullanılması kararlaştırılmıştır (Acal vd., 2020; DeVellis, 2017; Tabachnick & Fidell, 2013). DFA aşamasında, ölçeğin model uyumunu test etmek amacıyla standart uyum indeksleri kullanılmış, bu doğrultuda alanyazındaki standart uyum indeks değerleri (Tablo 6) esas alınmıştır (Carvalho & Chima, 2014; Kline, 2014; Seçer, 2017; Shi vd., 2019; Tabachnick & Fidell, 2013; Vieira, 2011). Ölçeğin güvenirlik çalışmaları kapsamında ölçek maddeleri, bileşik güvenirlik, Cronbach's Alfa, test tekrar test yöntemi ile test edilmiştir. Ayrıca maddelerin ayırt edicilik özellikleri incelenmiştir.

Etik Konular

Bu araştırma, Düzce Üniversitesi Bilimsel Araştırma ve Yayın Etik Kurulu'nun 23.02.2023 tarihli ve 2023/58 sayılı kararıyla etik açıdan uygun bulunmuştur. Araştırma sürecinde bilimsel araştırma ve yayın etiği ilkelerine tam olarak uyulmuş; katılımcı ve velilerinden gönüllü onam formu alınmıştır. Ayrıca çalışmada kullanılan tüm fikir ve sanat eserleri için ilgili telif hakları düzenlemelerine riayet edilmiştir.

BULGULAR VE YORUM

Yapı Geçerliği

APDMPDÖ-ÖF'nin yapı geçerliği AFA ve DFA ile test edilmiştir. Bu doğrultuda, ölçeğin faktör yapısının ölçtüğü yapılar ile olan ilişkisi ve kuramsal temellere uygunluğu incelenmiştir.

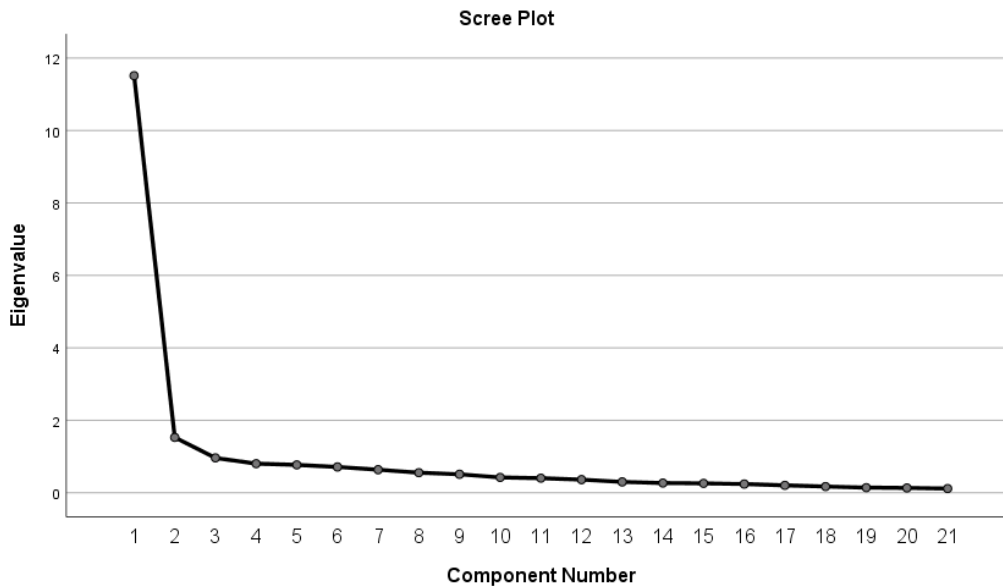
AFA

AFA, maddeler arasındaki ilişkilere dayalı faktör bulmayı sağlayan bir analiz yöntemidir (Büyüköztürk, 2017; Fabrigar & Wegener, 2011). AFA öncesi, KMO kat sayısı ve Bartlett Küresellik Testi sonuçları incelenmiştir. KMO kat sayısı, veri setindeki örneklem büyüklüğünün faktör analizi için yeterli olup olmadığını değerlendirmeye yönelik bir uygunluk testidir. KMO katsayısı, sıfır ile 1.00 arasında değerler alır, 1.00'e yakın değerler alması korelasyon örüntülerinin düzenli olduğunu dolayısıyla faktör analizinin ayrık ve güvenilir faktörler vereceği anlamına gelir. Bartlett Küresellik Testi, korelasyon matrisindeki değişkenler arasındaki korelasyonların sıfıra eşit olup olmadığı hipotezini test eder ve bu testin istatistiksel olarak anlamlı çıkması, verilerin faktör analizine uygunluğu için istenen bir durumdur. KMO kat sayısı .92, Bartlett Sphericity testi değerleri ise anlamlı bulunmuştur ($p < .01$). KMO katsayısı ve Bartlett küresellik testi sonuçları veri setinin faktör analizi için uygun olduğunu ortaya koymaktadır (Field, 2018; Howard, 2016; Morrison, 2009; Tabachnick & Fidell, 2013). Veri setine uygulanan ilk AFA uygulaması sonucunda ölçeğin iki faktörlü bir yapıya sahip olduğu belirlenmiştir. Bu yapıya ilişkin KMO kat sayısı, Bartlett Küresellik Testi değerleri yapının uygunluğunu ortaya koymaktadır. İlgili faktör yapısına ilişkin bulgular Tablo 2'de verilmiştir.

Tablo 2*ADMÖPDÖ Faktör Öz değerleri, Varyans ve Yığılımlı Varyans Değerleri*

Faktör	Özdeğer	Varyans	Yığılımlı Varyans
1	11.256	53.456	40.371
2	1.701	8.009	61.553

Tablo 1’de görüldüğü üzere, ilk AFA’da ortaya çıkan 2 faktörlü ölçek yapısına ilişkin özdeğerler birinci faktör için 11.256, ikinci faktör için ise 1.701 olarak hesaplanmıştır. Toplam varyansın % 53.45’i birinci faktör, % 8.01’u ise ikinci faktör tarafından açıklanmaktadır. Ölçek geliştirme çalışmalarında faktörlerin açıkladığı varyans oranlarının yüksek olması, ölçeğin ölçülmek istenen özelliği yansıttığı ve faktör yapısının güçlü olduğu anlamına gelmektedir (Tavşancıl, 2014). Alan yazında sosyal bilimler alanında varyans oranlarını % 30 ile % 60 arasında olması kabul görmektedir (Büyüköztürk, 2017; Tavşancıl, 2014). Bu bağlamda, birinci faktörün varyansın %53.46’sını tek başına açıklamasından ve ilk faktöre ait özdeğerin ikinci faktöre ait özdeğerin yaklaşık 11 katı olması, ölçeğin tek faktörlü olabileceğine işaret etmektedir. Ölçek boyutlarına ilişkin varyansın büyük çoğunluğunun tek faktörlü yapıya ilişkin olması nedeniyle, faktör yapısına ilişkin net karar vermek amacıyla Şekil 1’de ölçeğin yamaç grafiği (Scree Plot) oluşturulmuştur (Ledesma vd., 2007; Ledesma vd., 2015).

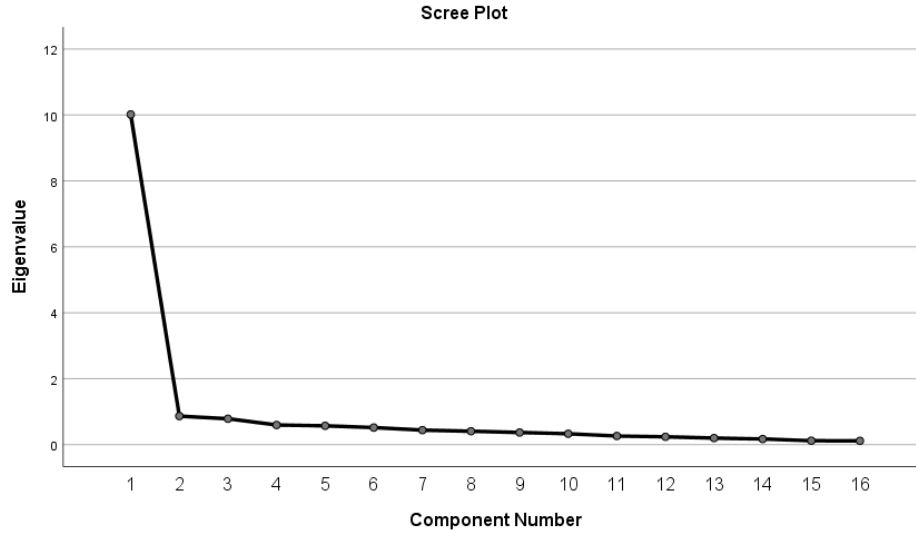
Şekil 1*Ölçek Maddelerine İlişkin Yamaç Birikinti Grafiği (Scree Plot)*

Ölçeğin faktör sayısını netleştirmek amacıyla yamaç grafiği ve paralel analizi yöntemlerinden yararlanılmıştır. İki faktörlü yapıya ilişkin yamaç grafiği (Scree Plot) incelendiğinde, öz değerlerin tek faktörde çok yüksek olması, diğer faktörün özdeğerinin çok düşük olması, ayrıca grafikte belirgin olan tek kırılmanın olması nedeniyle (Pallant vd., 2016; Silva vd., 2020), ölçeğin tek faktörlü bir yapıda değerlendirilmesini desteklemektedir. Ayrıca paralel analiz yöntemi ile veri setinden rast gele üretilmiş veri üzerinde yapılan analizlerde elde edilen özdeğerin, gerçek veri setindeki özdeğerden düşük olması da ölçeğin tek faktörlü yapısını desteklemektedir (Hayton vd., 2004). Ayrıca, ölçek maddelerinin, programın hem öğrenme çıktıları hem de programın uygulanma süreci ile ilişkili olduğu dikkate alındığında, kavramsal temellendirme açısından da tek faktörlü yapının uygun olduğu değerlendirilmiştir (Parlett & Hamilton, 1977). Ölçek tek faktörlü olarak değerlendirildikten sonra, tekrarlanan AFA sonucunda faktör yükü. 30’un altında olan 5 madde olduğu belirlenmiştir. Bu maddelerin ölçeğin yapısal geçerliliğini olumsuz etkileyebileceği düşünülerek, ilk aşamada, ölçek maddelerinin kademeli olarak azaltılması stratejisi ile çeşitli AFA denemeleri ile madde elenmesine gidilmiştir. Yapılan denemelerde, ilgili maddelerin faktör yükleri incelenmiş ve ilgili maddelerin faktör yüklerinin yine belirlenen sınır değerin (0.30) altında kaldığı belirlenmiştir. Söz konusu maddeler, ölçeğin

yapısal bütünlüğünü sağlamak amacıyla formdan çıkarılmıştır. (Büyüköztürk, 2017; Louangrath & Sutanapong, 2018; Raubenheimer, 2004; Şencan, 2005). Ölçekten 5 madde çıkarıldıktan sonra, elde edilen yeni faktör yapısı incelenmiş ve hem özdeğerler hem de yamaç grafiğinde ölçeğin tek faktörlü yapının korunduğu belirlenmiştir. Aşağıda Şekil 2’de ölçekte kalan 16 maddeye ilişkin yamaç grafiği verilmiştir.

Şekil 2

Kalan Ölçek Maddelerine İlişkin Yamaç Birikinti Grafiği (Scree Plot)



AFA sonucu madde yükleri. 30’un altında olan ve ölçek formundan çıkarılan 5 madde arasında yer alan “Destek eğitim programı sayesinde okuma becerilerim gelişti.”, “Destek eğitim programı sayesinde yazma becerilerim gelişti.” ve “Destek eğitim programı sayesinde matematik becerilerim gelişti.” Maddeleri yer almaktadır. Bu maddelerin düşük faktör yüklerine sahip olmasının, öğrencilerin farklı öğrenme güçlüğü türlerine sahip olmalarından kaynaklanabileceği düşünülmektedir. Özellikle, öğrencilerin farklı öğrenme güçlüğü türlerine sahip olmaları, bu maddelere verdikleri yanıtların tutarlılığını azaltmış olabilir. Ayrıca, öğrenme güçlüğü olan öğrenciler arasında karma öğrenme güçlüğü olarak ifade edilebilecek durumunun daha sınırlı gözlenmesi, bu maddelerin ölçeğin genel yapısına katkısını kısıtlamış olabilir. Bunun yanında ölçekten çıkarılan maddeler arasında yer alan "Programda öğrendiğim bilgileri okulda ve günlük yaşamda uygulayabiliyorum.", "Programda öğrendiğim bilgiler yeni bilgiler öğrenmemi kolaylaştırıyor." maddelerinin düşük faktör yüküne sahip olması ise, öğrencilerin bireysel farklılıkları ve bu konulardaki deneyim eksiklikleri ile ilişkilendirilebilir. Ölçeğin tek faktörlü yapısına ilişkin değerler Tablo 3’te verilmiştir.

Tablo 3

APDMPDÖ-ÖF Faktör Öz değerleri, Varyans ve Yığılımlı Varyans Değerleri

Faktör	Özdeğer	Varyans
1	9.969	62.303

Tek faktörlü ölçek için hesaplanan toplam varyans % 62.303 olarak bulunmuştur. Bu varyans oranı alanyazında kabul gören geçerli bir değerdir (Büyüköztürk, 2017; Tavşancıl, 2014). AFA sonucunda APDMPDÖ-ÖF’nin maddeleri tek faktör altında toplandığı için, herhangi bir döndürme işlemi yapılmamıştır. Tek faktörlü yapıdaki ölçekte toplanan 16 madde, toplam varyansın % 62.30’unu açıklamaktadır. Her bir faktöre ilişkin özdeğerin 1,00’dan yüksek olması, ayrıca faktör analizi ile ulaşılan varyans değerinin % 60’tan yüksek olması, ölçek yapısının ölçülmek istenen özelliği güçlü biçimde yansıttığını göstermektedir (Beavers vd., 2013; Büyüköztürk, 2017). Tek faktörlü olan ölçek yapılarında varyans oranının % 30 ve üzerinde olması yeterlidir (Tavşancıl, 2014). Ölçeğin tamamı için hesaplanan varyansın (% 62.30) alanyazını için oldukça

yüksek bir varyans değeri olduğu görülmektedir. Ölçekte yer alan 16 maddenin yapı özellikleri Tablo 4'te sunulmuştur.

Tablo 4

Maddelerin Faktör Yüklü Değerleri ve Madde Toplam Korelasyonları

Nihai Madde No	Madde No	Maddeler	Madde Toplam Korelasyonu	Faktör Yük Değeri
1	16	Sınıftaki öğrenme etkinliklerine katılmaktan hoşlanıyorum.	.781	.821
2	11	Matematik konularını öğrenmekten keyif alıyorum	.779	.782
3	15	Sınıftaki öğrenme etkinliklerine katılmakta sorun yaşıyorum.	.755	.799
4	10	Rehabilitasyon merkezinde yeni konular öğreneceğim için heyecanlanıyorum.	.748	.796
5	12	Okuma etkinlikleri yapmaktan keyif alıyorum.	.742	.753
6	9	Derslerde öğrendiğim konular ilgimi çekiyor.	.732	.803
7	18	Öğretmenim derste konuyu anlatırken, konu hakkında örnekler verir.	.669	.729
8	17	Derslerde kullanılan materyaller dikkatimi çekiyor	.656	.719
9	13	Yazma etkinliklerini öğrenmekten keyif alıyorum.	.601	.632
10	7	Destek eğitim programından yararlanmaktan dolayı mutluyum.	.574	.626
11	20	Öğretmenim yaptığım etkinliklerle ilgili, benimle fikirlerini paylaşır.	.528	.599
12	21	Sınıfta yapılan etkinliklerden sonra öğretmenim derste neler öğrendiğime ilişkin sorular sorar.	.477	.547
13	14	Dersleri öğretmenimin anlattığı süre içerisinde anlayabiliyorum.	.468	.476
14	8	Rehabilitasyon merkezinde anlatılan konuları, dersler için ayrılan sürelerde anlıyorum.	.426	.484
15	19	Öğretmenim yaptığım etkinliklerde beni takip eder, gözlemler (izler).	.368	.464
16	4	Destek eğitim programı sayesinde okuma becerilerim gelişti.	.302	.363

Tablo 4'te verildiği üzere ölçekte yer 16 maddenin faktör yük değerleri, 0.30 ile 0.78 arasında değişmekte olup, madde toplam test korelasyonları ise 0.36 ile 0.82 arasında değişmektedir. Bu bulgular, maddelerin yük değerleri bakımından uygunluğunu göstermektedir (Louangrath & Sutanapong, 2018; Punch, 2013; Raubenheimer, 2004).

DFA

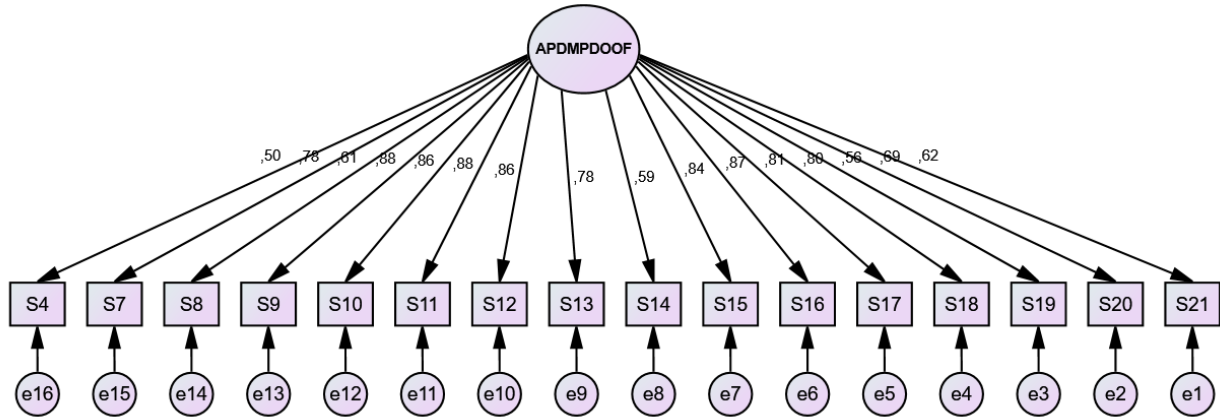
AFA ile ulaşılan tek faktör ve 16 maddeli ölçeğin uygunluğu, 260 kişilik öğrenci gruba uygulanarak, yapı DFA ile test edilmiştir. DFA sonucu ölçeğe ilişkin uyum indeks değerleri Tablo 5'te sunulmuştur.

Tablo 5

DFA Uyum İndeksleri

X^2/sd	SRMR	RMSEA	CFI	GFI	AGFI	IFI	NFI	NNFI
2.72	.08	.069	.95	.95	.90	.92	.90	.92

Aşağıda Şekil 3'te APDMPDÖ-ÖF'ye ilişkin DFA diyagramı yer almaktadır.

Şekil 3.*APDMPDÖ-ÖF DFA Diyagramı (standartlaştırılmış değerler)*

DFA için uyum indekslerinin değerlendirilmesinde, Tablo 6’da sunulan alanyazın referans değerleri temel alınmıştır.

Tablo 6*Uyum İndeksi Değerlendirme Ölçütleri*

Uyum İndeksleri	Mükemmel uyum	Kabul Edilebilir Uyum
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 3$
RMSEA	$0 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$
RMR	$0 \leq RMR \leq .05$	$.05 < RMR \leq .10$
SRMR	$0 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$
NFI	$.95 \leq NFI \leq 1.00$	$.90 \leq NFI \leq .95$
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI \leq .95$
AGFI	$.90 \leq AGFI \leq 1.00$	$.85 \leq AGFI \leq .90$
CFI	$.97 \leq CFI \leq 1.00$	$.95 \leq CFI \leq .97$

Kaynaklar: (Carvalho & Chima, 2014; Chan vd., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi vd., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011).

Ölçeğin uyum indeks değerleri şu şekilde özetlenebilir: $\chi^2/sd=2.72$, SRMR= .008, RMSEA= .069, CFI = .95, GFI= .95, AGFI = .90, IFI = .92, NFI= .90, NNFI= .92. Bu değerler alanyazında genel kabul kabul gören uyum indekslerine uygun olup, özellikle SRMR değerinin mükemmel uyumu gösterdiği görülmektedir (Carvalho & Chima, 2014; Chan vd., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi vd., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011). Bu bulgular doğrultusunda, APDMPDÖ-ÖF’nin tek boyutlu yapısının doğrulandığı görülmektedir.

Geçerlik ve Güvenirlik

Analiz sonuçları, yüksek güvenirlik değerlerine sahip olduğu belirlenmiştir. Bu doğrultuda APDMPDÖ-ÖF’nin güvenirlik analizleri Tablo 7’de sunulmuştur.

Tablo 7*APDMPDÖ-ÖF Güvenirlik Katsayıları*

Ölçek	Cronbach's Alpha Güvenirlik Katsayısı	Bileşik Güvenirlik (composite reliability, cR)	Test Tekrar Test
APDMPDÖ-ÖF	.94	.93	.90

APDMPDÖ-ÖF'nin on altı maddelik tek faktörlü yapısı için Cronbach's Alpha güvenirlik düzeyi .94; belirlenmiştir. Bileşik güvenirlik, ölçeğin iç tutarlılığını ortaya koyan bir güvenirlik katsayısı olup, DFA sonucu oluşan faktör yükleri ile hata varyansları kullanılarak hesaplanmaktadır (Peterson & Kim, 2013; Yang & Green, 2011). APDMPDÖ-ÖF'nin tek faktörlü yapısı için bileşik güvenirlik katsayısı ise .93 olarak belirlenmiştir. Ölçeklerin kararlılık katsayıları ise test tekrar yöntemi kullanılarak belirlenmektedir. Test tekrar test yöntemi, AFA ve DFA sonucu elde edilen ölçme aracının, aynı çalışma grubuna, aynı koşullarda fakat iki farklı zamanda uygulanmasıdır. İki uygulama sonrası madde korelasyon katsayıları kullanılarak ölçeğin test tekrar test güvenirlik katsayısı hesaplanır (Leppink & Pérez-Fuster, 2017; Seçer, 2017; Weir, 2005). Tek faktörlü ve 16 maddeli ölçek, 72 öğrenciye yirmi gün ara ile uygulanmıştır. İki uygulama arasında korelasyon değerleri hesaplanmış ve ölçek test-tekrar test güvenirlik değeri .91 olarak belirlenmiştir. Tüm bu bulgular güvenirlik sonuçları dikkate alındığında, ölçeğin güvenirlik değerlerinin oldukça yüksek olduğu belirlenmiştir (Büyüköztürk, 2017; DeVellis, 2014; Özdamar, 2016; Tavşancıl, 2014).

Ölçek Maddelerinin Ayırtediciliği

APDMPDÖ-ÖF'de yer alan maddelerin ayırt edicilik düzeylerini belirlemek amacıyla, %27'lik alt-üst gruplar arasındaki farkların anlamlı olup olmadığı incelenmiştir. Alt-üst gruplar arasındaki farklara ilişkin analiz sonuçları Tablo 8'de sunulmuştur.

Tablo 8*APDMPDÖ-ÖF Madde-Toplam Test Puanı Korelasyonları ile %27'lik Alt ve Üst Grup Karşılaştırmaları*

Madde No	%27 Grup Alt n= 130		%27 Grup Üst n= 130		t	Düzeltilmiş Madde Toplam Test Puanı Korelasyonu
	$\bar{x}$	SS	$\bar{x}$	SS		
1	4.48	.71	4.38	.70	.830*	.39
2	4.64	.70	4.64	.74	.809*	.81
3	4.20	.99	4.29	.99	-.430*	.54
4	4.60	.71	4.71	.64	-1.000*	.84
5	4.65	.75	4.72	.67	-.587*	.86
6	4.68	.80	4.68	.75	.929*	.88
7	4.65	.75	4.61	.78	.328*	.86
8	4.48	.91	4.54	.86	-.381*	.79
9	4.27	.99	4.51	.88	-1.451*	.69
10	4.60	.71	4.67	.67	-.610*	.82
11	4.65	.75	4.70	.66	-.355*	.88
12	4.57	.77	4.62	.81	-.425*	.87
13	4.64	.66	4.62	.72	.122*	.82
14	4.07	.51	3.98	.52	.971*	.69
15	4.54	.65	4.71	.61	-1.597*	.71
16	4.50	.58	4.55	.65	-.547*	.58

Ölçeği oluşturan tek faktörlü yapıya ait her bir maddenin ayırt edicilik düzeyini test etmek amacıyla, DFA çalışma grubundan elde edilen toplam puanlar küçükten büyüğe sıralanmış, %27'lik alt ve %27'lik üst grup arasında bağımsız örneklem için t-testi uygulanmıştır. Yapılan analizler sonucunda, her bir maddeye ilişkin alt-üst grup ortalamaları arasındaki farkın istatistiksel olarak anlamlı ($p < .01$) olduğu belirlenmiştir. Bu

bulgu, ölçeğin tüm maddelerinin yüksek ayırt edicilik düzeyine sahip olduğunu ortaya koymaktadır (Büyüköztürk, 2017; Kubiszyen & Borich, 2013).

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu araştırmanın amacı, ÖGDEP'in, APDM çerçevesinde, programdan doğrudan yararlanan öğrencilerin algılarına dayalı olarak değerlendirilmesine imkân tanıyan geçerli ve güvenilir bir ölçme aracı geliştirmektir. APDM, yalnızca programın çıktılarıyla sınırlı kalmayan; aynı zamanda amaçları, içeriği, uygulanma süreci ve program etkileşiminin, katılımcılar tarafından nasıl deneyimlendiğine ve yorumlandığına odaklanan niteliksel bir değerlendirme yaklaşımıdır (Parlett & Hamilton, 1977). Bu model, paydaş görüşlerinin merkeze alınmasını esas almakta ve öğrenci algılarının program değerlendirme sürecine dahil edilmesini, programın etkililiğinin daha bütüncül ve bağlamsal olarak anlaşılmasına katkı sağlayacak bir unsur olarak görmektedir. Bu kuramsal yapı temel alınarak, araştırmacılar tarafından modelin bileşenlerine dayalı biçimde 24 maddelik taslak madde havuzu oluşturulmuştur. Geliştirilen taslak, içerik geçerliğini sağlamak amacıyla alan uzmanlarının değerlendirilmesine sunulmuştur. Uzman değerlendirmeleri doğrultusunda, içerik açısından yetersiz bulunan üç madde formdan çıkarılmış ve böylece taslak ölçek formu 21 maddeye indirgenmiştir. Bu süreç, ölçeğin hem içerik geçerliğini artırmak hem de kuramsal yapıyla bütünlüğünü sağlamak açısından önemli görülmektedir (DeVellis, 2017).

APDMPDÖ-ÖF'nin yapı geçerliği AFA ve DFA ile incelenmiştir. AFA sonucu, tek faktörlü yapıda toplanan 16 madde toplam varyansın % 62.30'unu açıkladığı belirlenmiştir. Ölçek geliştirme çalışmalarında, her bir faktöre ilişkin özdeğerin 1,00'in üzerinde olması ve faktör analizi ile ulaşılan toplam varyans değerinin %60'tan yüksek olması, ölçeğin yapı geçerliğinin güçlü olduğuna işaret etmektedir (Beavers vd., 2013; Büyüköztürk, 2017). Ayrıca, tek faktörlü ölçeklerde açıklanan varyans oranının % 30 ve üzerinde olması yeterli görülmektedir (Tavşancıl, 2014). Bu bağlamda ölçeğin tamamı için hesaplanan varyans oranının (% 62.30) alanyazını için oldukça yüksek bir varyans oranı olduğu ve dolayısıyla ölçeğin yapı geçerliğinin yüksek olduğu görülmektedir. Elde edilen tek faktörlü ölçek yapısı DFA ile test edilmiştir. DFA'da ulaşılan sonuçlara göre ($X^2/sd=2.72$, SRMR= .008, RMSEA= .069, CFI = .95, GFI= .95, AGFI = .90, IFI = .92, NFI= .90, NNFI= .92). ADMÖPDÖ-ÖF'nin uyum indeks değerlerinin alanyazını için genel kabul edilebilir değerler olduğu, SRMR'de mükemmel uyum olduğu belirlenmiştir (Carvalho & Chima, 2014; Chan vd., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi vd., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011). Bu doğrultuda APDMPDÖ-ÖF'nin tek faktörlü yapısının doğrulandığı sağlandığı belirlenmiştir.

APDMPDÖ-ÖF'nin on altı maddelik tek faktörlü yapısı için Cronbach's Alpha güvenilirlik kat sayısı .94; bileşik güvenilirlik katsayısı ise .93 olarak belirlenmiştir. Ölçeklerin kararlılık düzeyleri, test tekrar yöntemi ile belirlenmektedir. Test tekrar test yöntemi, AFA ve DFA sonucu elde edilen ölçme aracının, aynı çalışma grubuna, aynı koşullarda fakat iki farklı zamanda uygulanmasıdır. Uygulamalardan elde edilen sonuçlar madde korelasyon katsayıları aracılığıyla karşılaştırılarak ölçeğin test tekrar test güvenilirlik katsayısı hesaplanır (Leppink & Pérez-Fuster, 2017; Seçer, 2017; Weir, 2005). Bu kapsamda, tek faktörlü ve 16 maddeli ölçek, 72 öğrenciye yirmi gün ara ile uygulanmıştır. Uygulamalar arasındaki korelasyon katsayısı .91 olarak hesaplanmıştır. Elde edilen bulgular doğrultusunda, ölçeğin güvenilirlik düzeyinin yüksek olduğu sonucuna ulaşılmıştır (Büyüköztürk, 2017; DeVellis, 2014; Özdamar, 2016; Tavşancıl, 2014).

APDMPDÖ-ÖF'de yer alan maddelerin ayırt edicilik düzeylerini belirlemek amacıyla, %27'lik alt ve üst gruplar temel alınarak maddelerin ayırt edicilik düzeyleri incelenmiştir. Yapılan analiz sonucunda, ölçek maddelerinin her biri için alt grup ve üst grup katılımcılar arasında istatistiksel olarak anlamlı düzeyde fark olduğu ($p < .01$) saptanmıştır. Bu sonuç, her bir maddenin, düşük ve yüksek düzeyde programa ilişkin değerlendirmelere sahip bireyleri birbirinden ayırt edebildiğini göstermektedir. Başka bir deyişle, ölçek maddeleri, öğrencilerin programı değerlendirme düzeylerine duyarlı biçimde tepki vermelerine olanak tanımakta ve bu yönüyle ölçme aracının yapı geçerliği açısından önemli bir gösterge sunmaktadır. Madde düzeyinde elde edilen anlamlı farklılıklar, ölçeğin hedeflediği yapıyı ölçme konusundaki duyarlılığını ve ölçüm gücünü desteklemektedir (Büyüköztürk, 2017; Kubiszyen & Borich, 2013).

Araştırma bulguları doğrultusunda, APDMPDÖ-ÖF'nin APDM temel alınarak geliştirilmiş; öğrencilerin programı değerlendirmeye yönelik algılarını ölçmeyi amaçlayan, geçerlik ve güvenilirlik açısından bilimsel yeterliliğe sahip bir ölçme aracı olduğu kanıtlanmıştır. Ölçeğin geçerlik analizlerinden elde edilen sonuçlar, yapı geçerliğinin sağlandığını; güvenilirlik analizleri ise ölçeğin tutarlı ve güvenilir ölçümler

sunduğunu ortaya koymuştur. Bu doğrultuda, APDMPDÖ-ÖF'nin öğrencilerin programlara yönelik değerlendirmelerini çok boyutlu bir yaklaşımla ele almak isteyen araştırmacı ve uygulayıcılar için uygun bir değerlendirme aracı olduğu söylenebilir. Araştırmanın bulgularına dayanarak, APDMPDÖ-ÖF'nin farklı katılımcı gruplarında uygulanarak geçerlik ve güvenilirlik çalışmalarının genişletilmesi önerilmektedir. Ayrıca, gelecekte yapılacak çalışmalarda, ölçek kullanılarak programdan yararlanan öğrencilerin demografik özellikleri bağlamında program değerlendirmeye ilişkin öğrenci görüşlerinin incelenmesine odaklanması önerilmektedir. Bu sayede, program değerlendirme süreçlerine daha bütüncül ve derinlikli bir katkı sağlanabileceği düşünülmektedir.

KAYNAKÇA

- Acal, C., Aguilera, A. M., & Escabias, M. (2020). New modeling approaches based on varimax rotation of functional principal components. *Mathematics*, 8(11),1-15. <https://doi.org/10.3390/math8112085>
- Adıgüzel, O. C. ve Özüdoğru, F. (2014). İlkokul 2. sınıf İngilizce öğretim programına yönelik Aydınlatıcı Değerlendirme Modeline Dayalı Program Değerlendirme Ölçeği çalışması. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 4(2), 124-136.
- Adıgüzel, O. C. (2015). Development of a Scale to Evaluate the Pedagogical Formation Program Implemented with Turkish Prospective Teachers. *Education Research and Review*, 10(2), 101-110. <http://dx.doi.org/10.5897/ERR2014.2029>
- Akıncı, M., & Köse, E. (2019). Validity and Reliability Study of Teacher Training Curriculum Evaluation Scale (TTCES). *Uluslararası Bilimsel Araştırmalar Dergisi (IBAD)*, 4(2), 236-252. <http://dx.doi.org/10.21733/ibad.534751>
- Aktan, O. (2020). Determination of educational needs of teachers regarding the education of inclusive students with learning disability. *International Journal of Contemporary Educational Research*, 7(1), 149-164. <https://doi.org/10.33200/ijcer.638362>
- Aktan, O. (2021). Özel gereksinimli bireylere yönelik Öğretmen Sosyal Kabul Ölçeği'nin geliştirilmesi, geçerlik ve güvenilirlik çalışması. *Trakya Eğitim Dergisi*, 11(1), 315-332. <https://doi.org/10.24315/tred.71298>
- Aktan, O. (2023). Kapsayıcı Eğitim. S.F.Büyükalan (Ed.), *Öğrenme öğretme kuram ve yaklaşımları içinde* (4.Baskı, s. 221-265). Pegem Akademi.
- Aktan, O., Üçüncü, B., Bingöl, T., & Gökbayrak, Ö. (2025). Aydınlatıcı Değerlendirme Modeli Değerlendirme Ölçeği'nin geliştirilmesi: Geçerlik ve güvenilirlik analizi. *Anadolu Eğitim Liderliği ve Öğretim Dergisi*, 13(1), 49-64.
- Aktan, O., & Orakcı, Ş. (2026). Teachers' Curriculum Literacy and Individualised Education Program Competencies in Turkey: A Mixed Method Study. *International Journal of Disability, Development and Education*, 73(5), 949-975. <https://doi.org/10.1080/1034912X.2025.2605504>
- Altınsoy, T. & Küçüksüleymanoğlu, R. (2021). 5. Sınıf Müzik Öğretim Programı Değerlendirme Ölçeğinin Geliştirilmesi. *Necmettin Erbakan Üniversitesi Ereğli Eğitim Fakültesi Dergisi*, 3 (1), 82-96. Retrieved from <https://dergipark.org.tr/tr/pub/neueefd/issue/62544/714944>
- Armstrong, F., Armstrong, D., & Barton, L. (2016). *Inclusive education: Policy, contexts and comparative perspectives*. Routledge.
- Aslan, M., Soyalt, H., Karahan, O. & Altuntaş, M. (2016). Okul Öncesi Eğitim Programı Değerlendirme Ölçeği'nin geliştirilmesi. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 657-683.

- Bal, A. P.ve Arslan, U. (2021). Ortaokul Matematik Öğretim Programını Değerlendirme Ölçeğinin geliştirilmesi: Öğrenci boyutu. *Akdeniz Eğitim Araştırmaları Dergisi*, 15(37), 26-40. <https://doi.org/10.29329/mjer.2021.380.2>
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1), 1-13. <https://doi.org/10.7275/qv2q-rk76>
- Bryant, D. P., Bryant, B. R., & Smith, D. D. (2019). *Teaching students with special needs in inclusive classrooms* (1th ed.). SAGE Publications Ltd.
- Burden, B. (2017). Illuminative evaluation. In *Frameworks for practice in educational psychology: A textbook trainees and practitioners*, edited by B. Kelly, L. M. Woolfson, and J. Boyle, 291– 308. 2nd ed. London: Jessica Kingsley
- Büyüköztürk, Ş. (2017). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (22. baskı). Ankara: Pegem.
- Cahapay, M. B. (2020). The responsiveness of bachelor of elementary education curriculum: An illuminative evaluation. *International Journal of Evaluation and Research in Education*, 9(3), 743-750. <https://doi.org/10.11591/ijere.v9i3.20649>
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication methods and measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Carvalho, J., & Chima, F. O. (2014). Applications of structural equation modeling in social sciences research. *American International Journal of Contemporary Research*, 4 (1), 6-11.
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of pediatric psychology*, 27(1), 27-36.
- Chan, F., Lee, G. K., Lee, E. J., Kubota, C. & Allen, C. A. (2007). Structural equation modeling in rehabilitation counseling research. *Rehabilitation Counseling Bulletin*, 51 (1), 53- 66. <https://doi.org/10.1177/00343552070510010701>
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.
- Chen, H. T. (2010). The bottom-up approach to integrative validity: a new perspective for program evaluation. *Evaluation and program planning*, 33(3), 205-214. <https://doi.org/10.1016/j.evalproplan.2009.10.002>
- Çelikel, F. (2022). *Ortaokul matematik dersi öğretim programının matematiksel yetkinlik bağlamında aydınlatıcı program değerlendirme modeli ile değerlendirilmesi* [Yayınlanmamış doktora tezi]. Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Çopur, E., Türkmenoğlu, M., Artut Dinç, P., & Bal, A. P. (2021). Stufflebeam'ın CIPP modeline göre ilköğretim matematik dersi öğretim programını değerlendirme ölçeği. *International Journal of Educational Spectrum*, 3(2), 150-175. <https://doi.org/10.47806/ijesacademic.903747>
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: Sage.
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics*. The USA: Sage
- Gültekin, M., & Demirci, C. (2020). Evaluation of the English language preparatory school curriculum in the lights of illuminative evaluation model. *European Journal of English Language Teaching*, 5(3), 103-131.

- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). SEM: An introduction. *Multivariate data analysis: A global perspective*, 5(6), 629-686.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2), 191-205. <https://doi.org/10.1177/1094428104263675>
- Hornby, G. (2011). Inclusive education for children with special educational needs: A critique. *International Journal of Disability, Development and Education*, 58(3), 321-329. <https://doi.org/10.1080/1034912X.2011.598678>.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?. *International journal of human-computer interaction*, 32(1), 51-62.
- Jöreskog, K. & Sörbom, D. (1998). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*, 4th ed. Scientific Software International, Hillsdale.
- Karabay-Turan, A., Kuşdemir-Kayıran, B. ve Bal, A. P. (2019). Türkçe Dersi Öğretim Programını Değerlendirme Ölçeği: Geçerlik ve güvenirlik çalışması. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 19(4), 1654-1666. <https://dx.doi.org/10.17240/aibuefd.2019..-543058>
- Kauffman, J. M., & Hallahan, D. P. (Eds.). (2011). *Handbook of special education*. Routledge.
- Kavan, N. (2023). Türkçe Dersi Öğretim Programı CIPP Modeli Değerlendirme Ölçeği: Geçerlik ve Güvenirlik Çalışması. *Türkiye Eğitim Dergisi*, 8(1), 159-185. <https://doi.org/10.54979/turkegitimdergisi.1306033>
- Kline, P. (2014). *An easy guide to factor analysis*. Routledge.
- Kubiszyen, T. & Borich, G. (2013). *Educational testing and measurement: Classroom application and practice*. (10th ed.) Indianapolis: John Wiley ve Sons, INC
- Kutay, V. (2018). Inclusive education in Turkey. *Studies in Educational Research and Development*, 2(2), 144-162.
- Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. *Psychology*, 9(11), 2531-2560.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy to use computer program for carrying out a parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1-11. <https://doi.org/10.7275/wjnc-nm63>
- Ledesma, R. D., Mora, P. V., & Macbeth, G. (2015). The scree test and the number of factors: A dynamic graphics approach. *The Spanish Journal of Psychology*, 18(E11), 1-10. <https://doi.org/10.1017/sjp.2015.13>
- Leppink, J., & Pérez-Fuster, P. (2017). We need more replication research—A case for test-retest reliability. *Perspectives on Medical Education*, 6(3), 158-164. <https://doi.org/10.1007/s40037-017-0347-z>
- Louangrath, P. I., & Sutanapong, C. (2018). Validity and reliability of survey scales. *International Journal of Research & Methodology in Social Science*, 4(3), 99-114.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (Eds.). (2012). *Evaluation models: Viewpoints on educational and human services evaluation* (Vol. 6). Springer Science & Business Media.
- Marsh, C. J. & Willis, G. (2007). *Curriculum: Alternative approaches, ongoing issues*. New Jersey: Pearson Prentice Hall.

- MEB. (2008). *Özel öğrenme güçlüğü destek eğitim programı* (s.4-8).Ankara
- MEB. (2021). *Öğrenme Güçlüğü Olan Bireyler İçin Destek Eğitim Programı*. https://ookgm.meb.gov.tr/meb_iys_dosyalar/2021_07/09112608_21130110_Ogrenme_Guclugu.pdf adresinden erişilmiştir.
- Morrison, J. T. (2009). Evaluating factor analysis decisions for scale design in communication research. *Communication Methods and Measures*, 3(4), 195–215.
- Oliva, P.F. (2009). *Developing the curriculum*. New York: Pearson Allyn and Bacon
- Ornstein, A. C., & Hunkins, F. P. (2009). *Curriculum: Foundations, principles, and issues* (5th ed.). Columbus, OH: Allyn & Bacon
- Özdamar, K. (2016). *Ölçek ve test geliştirme yapısal eşitlik modellemesi*. Eskişehir: Nisan.
- Özüdoğru, F. (2016). *İlkokul 2. sınıf İngilizce öğretim programının diller için Avrupa ortak başvuru metni doğrultusunda aydınlatıcı değerlendirme modeli ile değerlendirilmesi* [Yayınlanmamış doktora tezi]. Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Özüdoğru, F., & Adıgüzel, O. C. (2016). Aydınlatıcı Program Değerlendirme Modeli. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 16 (Özel Sayı), 25-34.
- Palabıyık, T. (2021). *Aydınlatıcı program değerlendirme modeline dayalı olarak ikinci sınıf İngilizce öğretim programının öğretmen görüşlerine göre değerlendirilmesi* [Yayınlanmamış yüksek lisans tezi]. Mardin Artuklu Üniversitesi, Lisansüstü Eğitim Enstitüsü
- Pallant, J. F., Haines, H. M., Green, P., Toohill, J., Gamble, J., Creed, D. K., & Fenwick, J. (2016). Assessment of the dimensionality of the Wijma delivery expectancy/experience questionnaire using factor analysis and Rasch analysis. *BMC pregnancy and childbirth*, 16, 1-11. <https://doi.org/10.1186/s12884-016-1157-8>
- Parlett, M. & Hamilton, D. (1977). Evaluation as Illumination: A New Approach to the Study of Innovative Programs. Parlett, M. & Dearden, G. (Eds.), *Introduction to Illuminative Evaluation: Studies in Higher Education* içinde (ss. 9-29). California: Pacific Soundings Press.
- Paulsrud, D., & Nilholm, C. (2023). Teaching for inclusion—a review of research on the cooperation between regular teachers and special educators in the work with students in need of special support. *International Journal of Inclusive Education*, 27(4), 541-555. <https://doi.org/10.1080/13603116.2020.1846799>
- Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194–198. <https://doi.org/10.1037/a0030767>
- Price, L. R. (2017). *Psychometric Methods: Theory into Practice*. New York: The Guilford Press
- Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. Sage.
- Rao, V., & Woolcock, M. (2003). Integrating qualitative and quantitative approaches in program evaluation. In Bourguignon F., Pereira Da Silva L. (Eds.), *Evaluating the poverty and distributional impact of economic policies: A general introduction*. World Bank and Oxford University Press.
- Raubenheimer, J. E. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, 30 (4), 59-64.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge university press.
- Seçer, İ. (2017). *SPSS ve LISREL ile pratik veri analizi*. Anı Yayıncılık.

- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and psychological measurement*, 79(2), 310-334. <https://doi.org/10.1177/0013164418783530>
- Silva, R., Oliveira, D., Santos, D. P., Santos, L.F.D., Wilson, R. E. & Bedo, M. (2020). Criteria for choosing the number of dimensions in a principal component analysis: an empirical assessment. In *Anais do XXXV Simposio Brasileiro de Bancos de Dados*, 145–150. Porto Alegre, Brazil: Computacao Sociedade Brasileira de Computacao. <https://sol.sbc.org.br/index.php/sbbd/article/view/13632>
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*. Ankara: Seçkin.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. The USA: Pearson Education.
- Tavşancıl, E. (2014). *Tutumların ölçülmesi ve SPSS ile veri analizi*, Ankara: Nobel.
- Ullman, J. B. (2006). Structural equation modeling: reviewing the basics and moving forward. *Journal of personality assessment*, 87 (1), 35- 50. https://doi.org/10.1207/s15327752jpa8701_03.
- Uşun, S. (2016). Eğitimde program değerlendirme süreçler-yaklaşımlar ve modeller (2. baskı). Ankara: Anı.
- Vieira A. L. (2011). *Interactive LISREL in practice, getting started with a SIMPLIS Approach*. London: Springer. <https://doi.org/10.1007/978-3-642-18044-6>.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research*, 19 (1), 231-240. <https://doi.org/10.1519/15184.1>.
- Yang, Y., & Green, S.B. (2011). Coefficient Alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29 (4) 377-392. <https://doi.org/10.1177/0734282911406668>.
- Yıldırım, A. & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. (9. Baskı). Ankara: SeçkinYayınılık.
- Zigmond, N. P., & Kloo, A. (2017). *General and special education are (and should be) different*. In *Handbook of special education* (pp. 249-261). Routledge.

Illuminative Evaluation Model Programme Evaluation Scale (Student Form): A Validity and Reliability Study

Osman AKTAN^{1*} 

Beyza ÜÇÜNCÜ² 

Tuğbanur BİNGÖL^{3*} 

Özge GÖKBAYRAK⁴ 

¹Düzce University, Faculty of Education,
Department of Special Education, Düzce,
Türkiye
osmanaktan@duzce.edu.tr

²Düzce Assessment and Evaluation
Center, Düzce, Türkiye
beyza_3@outlook.com

³Fevzi Çakmak Middle School, Van,
Türkiye
bingoltugbanur@gmail.com

⁴Düzce Vocational Education Center,
Düzce, Türkiye
ozgeakgun462@gmail.com

*Corresponding Author

Received: 18.07.2025
Accepted: 30.10.2025
Available Online: xx.xx.202x

Abstract: This study aims to develop a scale based on the Illuminative Evaluation Model (IEM) to evaluate the Specific Learning Disabilities Support Education Programme (SLDSEP) from the perspective of students. During the scale development process, expert opinions were obtained, and content and construct validity were tested through a pilot administration. For this purpose, draft scale items drafted by the researchers were administered to a total of 502 students benefiting from the SLDSEP across three distinct study groups during the 2022–2023 academic year. Following the pilot administration conducted in line with expert reviews, the content and construct validity of the scale were evaluated. The validity of the scale structure emerging from Exploratory Factor Analysis (EFA) was subsequently tested using Confirmatory Factor Analysis (CFA). Analysis results confirmed a 16-item, single-factor model for the Illuminative Evaluation Model Programme Evaluation Scale Student Form (IEMPES-SF). The total variance explained by the scale was determined to be 62.30%. Factor loadings for the scale items ranged between 0.30 and 0.88. Based on these findings, the IEMPES-SF yields valid and reliable estimates and can be effectively utilized to measure student perceptions regarding the evaluation of the SLDSEP grounded in the IEM framework.

Keywords: Programme Evaluation, Illuminative Evaluation Model, Student, Illuminative Evaluation Model Programme Evaluation Scale

INTRODUCTION

In contemporary inclusive education practices, individual differences have gained paramount importance within the educational domain, becoming a concept that is increasingly accommodated throughout learning processes (Aktan, 2023). Special education is fundamentally anchored in an instructional paradigm responsive to individual differences; accordingly, instructional processes are planned and implemented by taking into account the personal, social, and learning characteristics of individuals to satisfy their unique needs (Zigmond & Kloo, 2017).

Special education entails the provision of evidence-based, systematic instruction and support services to students experiencing academic, social, emotional, behavioral, or physical disabilities, facilitated by specially designed programmes and highly trained professionals (Kauffman & Hallahan, 2011). In this context, the primary objective of special education is to maximize the development of individuals with special needs in accordance with their interests and requirements, thereby preparing them to function as independent members of society (Bryant et al., 2019). Depending on the type and severity of their disabilities, individuals with special needs receive instruction in various educational settings, primarily alongside their peers within general education classrooms (Armstrong et al., 2016). Inclusive practices, executed through needs-based support services, aim to foster both the social integration and academic development of students with special needs (Aktan, 2021; Hornby, 2011). To ensure the success of inclusive education, diverse support services are provided, including co-teacher support, instructional adaptations, and peer tutoring within the classroom, as well as counseling, specialist consultation, resource room support, and supplementary instruction outside the classroom (Aktan, 2020; Kutay, 2018; Paulsrud & Nilholm, 2023). Among these out-of-school interventions are specialized support education programmes developed to target specific disability categories.

In Turkey, support education programmes were first introduced by the Ministry of National Education (MoNE) in 2008. Developed in consultation with faculty members, specialists, and practitioners, these modular programmes spanned seven distinct disability categories designed around the specific impairments and clinical needs of students with special needs (MoNE, 2008). Driven by contemporary advancements in special

education, MoNE collaborated with academic researchers, educators, and multidisciplinary experts to design updated support education programmes in 2021. This revised framework adopts a learner-centered paradigm, offers pedagogical guidance for teachers, incorporates current curriculum contents, and utilizes evidence-based methods and techniques. The modernized support education programmes aim to foster progress in self-care, daily living, language and communication, cognitive, socio-emotional, and motor development domains among individuals with special needs. Furthermore, the framework intends to help these individuals acquire foundational academic skills, assimilate national and moral values, and achieve social interaction and integration, thereby enabling them to maintain their lives as independent individuals within society. These support education programmes are tailored for individuals across seven disability domains: intellectual disability, autism spectrum disorder, speech and language impairment, hearing impairment, visual impairment, physical disability, and specific learning disabilities (MoNE, 2021).

The SLDSEP targets individuals experiencing difficulties in reading, writing, and mathematics, with the objective of fostering their language and communication, perception, motor planning, memory, social adaptation, self-regulation, literacy, and mathematical skills. Consisting of six modules totaling 1,100 instructional hours, the programme was designed using a learner-centered curriculum design model, reflecting progressive pedagogy and pragmatic educational philosophy. This modular structure aims to support individuals through recommended instructional methods and techniques within a learner-centered paradigm tailored to individual performance. Grounded in the SLDSEP framework, an Individualised Education Programme (IEP) must be designed and implemented for each student, taking into account their current performance levels and educational needs (Aktan & Orakcı, 2026; MoNE, 2021). The SLDSEP is delivered to individuals diagnosed with learning disabilities within special education and rehabilitation centres. To demonstrate the effectiveness of this programme, which was first implemented in 2021, evaluations based on the perspectives of both practitioners and beneficiaries are critical, as they may enhance the overall quality of programme-based instruction.

Programme evaluation constitutes the final stage of curriculum development. This process is conducted to determine the extent to which programme objectives have been achieved, to gather empirical data regarding operational dynamics, and to make evidence-based decisions concerning programme effectiveness (Chen, 2005; Oliva, 2009). As a dynamic enterprise, programme evaluation provides insights into the field-based execution of a newly developed educational curriculum, facilitating iterative improvements based on empirical data (Marsh & Willis, 2007). The evaluation process typically unfolds across three sequential phases: the planning phase, wherein the curriculum, the evaluation design, and corresponding data collection instruments are selected; the implementation phase, which involves collecting, analyzing, and interpreting data via tools aligned with the selected framework; and the assessment phase, which identifies curricular deficiencies based on the accumulated evaluation data. Various evaluation models have been developed by theorists such as Tyler, Cronbach, Campbell and Stanley, Stake, Stufflebeam, Scriven, Suchman, Alkin, Guba, Provas, and Weiss to guide these systematic inquiries (Madaus et al., 2012). One stakeholder-oriented approach that prioritizes participant perspectives is the IEM (Burden, 2017; Parlett & Hamilton, 1977).

In the IEM framework, the curriculum is evaluated directly by primary stakeholders such as teachers, students, curriculum developers, and domain experts, with the primary objective of identifying operational bottlenecks and systemic deficiencies (Parlett & Hamilton, 1977). This model scrutinizes curricular processes, institutional contexts, and stakeholder evaluations, while concurrently examining the broader impact of the curriculum on students' academic and socio-emotional development (Özüdoğru & Adıgüzel, 2016). Furthermore, the evaluation encompasses the pedagogical challenges encountered during implementation, the engagement levels of teachers, students, and other stakeholders, the educational impact of instructional materials, and the qualitative reflections of the curriculum on student experiences (Uşun, 2016). The model operates around two core concepts: the instructional system, which comprises official documents, administrative reports, implementation guidelines, and material infrastructure; and the learning milieu, which represents the social-psychological and material environment where teachers and students interact (Parlett & Hamilton, 1977). Ultimately, the IEM focuses on the complex interplay between the intended outcomes of a curriculum and the actual learning that manifests within the learning milieu.

To achieve the targeted outcomes in educational curricula, effective implementation is of paramount importance. Accordingly, to identify positive or negative occurrences during implementation and to enhance

curricular efficacy, the experiences and emotions of the participants directly affected by the curriculum are meticulously considered (Cahapay, 2020). For this reason, both the perspectives of various stakeholders, such as teachers, students, specialists, and families, regarding curriculum implementation and their experiences within classroom instructional processes are integrated into the evaluation scope (Richards, 2001). Within this framework, the IEM aims to provide high-quality insights into the ideal characteristics of an educational curriculum, to identify its strengths and areas requiring improvement based on participant feedback, and to evaluate its contributions to the academic and social development of students. Furthermore, the model offers evaluators comprehensive data collection opportunities to conduct detailed analyses of the reasons underlying curriculum success or deficiency (Parlett & Hamilton, 2017). To this end, both qualitative methods, such as observation, interview, and document analysis, and quantitative tools, including tests, questionnaires, and scales, are utilized within the scope of the model (Ornstein & Hunkins, 2009).

The utilization of scales in programme evaluation research is a widespread practice. An examination of the relevant literature in Turkey indicates that measurement instruments have been developed to provide empirical data for subject-specific curricula (Bal & Arslan, 2021; Karabay-Turan et al., 2019) and for the evaluation of curricula designed for various general purposes (Adıgüzel, 2015; Akıncı & Köse, 2019; Altınsoy & Küçüksüleymanoğlu, 2021; Aslan et al., 2016). Among the measurement instruments developed for curriculum evaluation studies, the Context, Input, Process, and Product model (Çopur et al., 2021; Kavan, 2023) and the IEM (Aktan et al., 2025; Adıgüzel & Özüdoğru, 2014) have been prominently adopted as theoretical evaluation frameworks. A review of these studies demonstrates that existing scale development initiatives for program evaluation purposes remain limited. It is observed that there are two measurement instruments based on the IEM, both of which target teachers as their audience. Consequently, a measurement instrument designed to evaluate student perceptions of curricula based on the IEM has yet to be developed.

Purpose of the Study

Programme evaluation studies leverage both quantitative and qualitative measurement instruments, including observations, interviews, document analyses, scales, and questionnaires (Rao & Woolcock, 2003). Nevertheless, qualitative data collection methods such as observation, interview, and document analysis possess certain limitations due to being time-consuming and susceptible to the subjective interpretations of the evaluator (Chen, 2010). Consequently, valid and reliable measurement instruments are required to systematically and objectively gather the perspectives of a large volume of participants regarding educational curricula. Within the context of the SLDSEP designed for students with specific learning disabilities, developing a valid and reliable scale that facilitates curriculum evaluation based on student opinions may significantly contribute to executing the evaluation process in a participant-oriented and objective manner.

The IEM is fundamentally based on examining the effectiveness of a curriculum through quantitative outcomes alongside the perceptions, experiences, and evaluations of primary stakeholders directly impacted by the intervention, such as teachers, students, domain specialists, and curriculum developers (Parlett & Hamilton, 1977). This approach aims to analyze deeply the dynamics emerging during implementation, the challenges encountered, and the actual context of the practice. The findings obtained from programme evaluation are anticipated to contribute to generating solutions for operational problems, identifying the strengths and areas requiring improvement, and thus restructuring the curriculum. Specifically, data obtained from students as beneficiaries of the curriculum serve an important formative guidance function for refining implementation processes. In this regard, determining the experiences and perceptions of students utilizing the SLDSEP is critical for understanding programme effectiveness and developing learner-centered improvement strategies. This holistic evaluation process builds a robust foundation for programme sustainability and educational quality enhancement, supporting the IEM approach that incorporates diverse stakeholder perspectives. Accordingly, the primary purpose of this study is to develop the IEMPES-SF, a measurement instrument grounded in the IEM framework, to systematically capture the perspectives of students utilizing the SLDSEP.

METHOD

Research Design

This study is a descriptive investigation designed for scale development. The scale was developed for students diagnosed with specific learning disabilities who attend primary, middle, or high school and utilize the SLDSEP. Within the scope of this study, the systematic procedural steps executed to develop a measurement instrument based on the IEM, aimed at assessing the evaluations of individuals with learning disabilities regarding the SLDSEP, are delineated below.

Scale Development Process

During the scale development process, the following sequential operational steps were taken into consideration (Carpenter, 2018; DeVellis, 2017; Kyriazos & Stalikas, 2018; Price, 2017):

Conceptualization of the Target Construct and Attributes

As the foundational step of the scale development process, the primary purpose of the scale and the psychological attributes intended for measurement were operationalized. In this context, the evaluation of the SLDSEP was grounded on the core tenets of the IEM. The IEM accommodates the perceptions, experiences, and emotions of diverse participants impacted by a curriculum, such as teachers, students, specialists, and families, while concurrently examining implementation processes and classroom pedagogical practices (Richards, 2001). The model diverges from alternative curriculum evaluation frameworks in that it prioritizes internal curricular processes and operational dynamics over predetermined objectives and final outputs (Burden, 2017). The scale intended for development aims to measure the evaluations of students diagnosed with learning disabilities based on their subjective experiences and perceptions within the IEM framework, thereby producing a psychometrically sound measurement instrument to supply valid and reliable data regarding effective curriculum execution. The target population comprised individuals diagnosed with specific learning disabilities across primary, middle, and high school tiers who receive services via support education programmes. Because students across these distinct educational levels utilize the exact same curriculum, empirical data were gathered from each tier to comprehensively assess the impact of the programme from a holistic perspective. During the administration phase, developmental variations and age levels of the participants were carefully considered; consequently, necessary instructions were delivered to the students prior to administration to ensure psychometric integrity and accurate data screening. Teachers and the researchers provided systematic guidance to the students concerning how to complete the instrument and the operational meaning of the item rating scales.

Item Generation

During the draft item generation phase, an extensive literature review (Aktan et al., 2025; Çelikel, 2022; Gültekin & Demirci, 2020; Özüdoğru, 2016; Özüdoğru & Adıgüzel, 2016; Palabıyık, 2021; Parlett & Hamilton, 1977) and opinions of domain experts were utilized. The core objective of the IEM is to verify whether an implemented educational curriculum achieves its targeted learning outcomes while concurrently identifying operational dynamics throughout implementation alongside the views and experiences of the program beneficiaries (Parlett & Hamilton, 1977). In this direction, it was intended that the scale items cover both the learning outcomes of the curriculum and the conditions pertaining to the implementation process. Aligned with the theoretical approach and programme evaluation philosophy of the IEM, the pool of items was initially drafted independently by the researchers. Items generated by each researcher were subsequently aggregated during team consensus meetings, where conceptually overlapping items were collapsed and restructured to ensure academic consistency. Throughout the item generation process, the first researcher proposed 32 items, the second proposed 27, the third proposed 29, and the fourth proposed 25 draft items. All statements were reviewed individually by the team members. Eight items recommended in common by all researchers were directly retained due to their conceptual alignment. Additionally, based on the evaluation, 16 more items that were found to be conceptually meaningful, clear, measurable, and non-redundant were unanimously selected for inclusion, yielding a 24-item initial draft scale form.

Expert Review

The draft scale form was evaluated by a panel of experts to ensure content validity. During this process, feedback was obtained from one measurement and evaluation expert, two special education specialists, three curriculum development experts, and one language specialist. Based on the qualitative evaluations of these experts, two items were identified as conceptually redundant and were merged into a single statement, while three items were removed from the form because they did not align with the overarching purpose of the scale. Following these revisions, a 21-item final draft scale form was obtained. Regarding the item scoring system, a three-point Likert-type response format was initially proposed for primary school students, and a five-point format was suggested for middle and high school students. However, based on the expert consensus, a uniform five-point Likert scale was adopted to optimize measurement reliability and ensure standardized administration across groups. This decision was supported by empirical evidence indicating that students in the fourth grade and above can make meaningful and reliable distinctions among five-point Likert options when appropriate linguistic and typographic adaptations are provided (Chambers & Johnston, 2002). Furthermore, the scale items were structured in a parsimonious and comprehensible manner regarding language and layout to match the age, cognitive development, and comprehension levels of the participating students. Accordingly, a unified response system was used for all students diagnosed with learning disabilities across educational tiers. On this five-point Likert-type scale, students' level of agreement with each statement was scored as strongly disagree (1), disagree (2), partially agree (3), agree (4), and strongly agree (5).

Pilot Administration

To evaluate the comprehensiveness and comprehensibility of the 21 items within the draft instrument, a pilot administration was conducted with 18 students diagnosed with specific learning disabilities across primary, middle, and high school tiers. Prior to the pilot session, one of the researchers explained the primary objective of the scale to the participants using parsimonious and accessible language. Furthermore, the individual scale statements and the corresponding Likert response system were briefly introduced, ensuring that necessary guidance was provided to facilitate accurate completion. Within the scope of this pilot phase, a single open-ended question was appended to the instrument to systematically elicit participant feedback regarding item readability and their overarching impressions. The students completed the draft scale within an interval spanning 7 to 12 minutes. No indication of ambiguity or misinterpretation regarding any specific item expression was reported in the open-ended feedback sections. Based on the evidence gathered from this pilot administration, the scale items were determined to be comprehensible and psychometrically feasible for the target population.

Scoring and Data Analysis

The developed measurement instrument was structured using a Likert-type response format. For each constituent item, the scoring continuum was defined as strongly disagree (1), disagree (2), partially agree (3), agree (4), and strongly agree (5). The composite scores obtained from the scale reflect student perceptions regarding the evaluation of the SLDSEP. Higher scores derived from the instrument indicate that students possess generally favorable perceptions of the curriculum and that the programme is implemented in alignment with its intended objectives. Conversely, lower scores indicate unfavorable student evaluations of the curriculum, suggesting operational challenges within the implementation process. To establish the construct validity of the developed scale, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were utilized; measurement reliability was evaluated using Cronbach's alpha, composite reliability, and test-retest reliability methods.

Study Groups

To gather empirical data for psychometric evaluation, evidence was accumulated from three distinct study groups. The participants were sampled using criterion sampling, a purposive sampling method wherein the researcher incorporates cases that satisfy a predetermined set of criteria (Yıldırım & Şimşek, 2013). In structuring the study groups, the primary inclusion criteria required that the students had received an official educational diagnosis of specific learning disabilities by a Guidance and Research Centre and were active beneficiaries of the SLDSEP. In scale development research, establishing valid mathematical models requires that the sample size equals approximately five to ten times the number of scale items (Hair et al., 2010;

Tavşancıl, 2014). To examine the initial item clarity and functionality of the 21 draft items drafted by the researchers, a pilot study was first executed with 18 students diagnosed with specific learning disabilities enrolled in primary (fourth grade), middle, and high school levels. For the EFA phase, empirical data were gathered from 224 students with specific learning disabilities across identical educational tiers (primary, middle, and high school). The subsequent CFA phase was conducted with an independent cross-validation sample of 260 students possessing equivalent demographic characteristics to the EFA group. Thus, both primary study groups were structured to contain approximately ten times the number of scale items, strictly satisfying the recommended sample size thresholds. The demographic characteristics of the participating students are presented in Table 1.

Table 1*Demographic Characteristics of the Students Comprising the Study Groups*

Variable		Pilot Study Group n/%	EFA Group n/%	CFA Group n/%
Gender	Female	8 (44.40)	131 (58.48)	153 (58.85)
	Male	10 (55.60)	93 (41.52)	107 (41.15)
Grade Level	Primary School	5 (27.78)	32 (14.29)	48 (18.46)
	Middle School	7 (38.89)	117 (52.23)	130 (50.00)
	High School	6 (33.33)	75 (33.48)	82 (31.54)
Duration of SLDSEP	1 Year	6 (33.33)	86 (38.40)	116 (44.60)
Enrollment	2 Years	12 (66.67)	138 (61.60)	144 (55.40)
Total		18	224	260

An examination of the demographic characteristics of the participants indicates that the majority of the students in the study groups were female, a substantial proportion were enrolled at the middle school level, and a significant segment had been active beneficiaries of the SLDSEP for at least two years.

Data Collection and Analysis

Empirical data for this study were gathered during the spring semester of the 2022–2023 academic year, spanning from March to May, from students diagnosed with specific learning disabilities enrolled in various schools within the Düzce province. Prior to the initiation of data collection, institutional ethics committee approval for scientific research and publication was obtained. Following ethical clearance, administrative research permission was secured from the Düzce Provincial Directorate of National Education for field implementation. Because the participants were minors under the age of 18, parental informed consent forms were systematically obtained. Prior to conducting the primary statistical analyses, both the EFA and CFA datasets were screened for sample size adequacy, normality, multivariate outliers, linearity, and multicollinearity. To detect potential outliers, the minimum and maximum values of the scale items were examined, confirming that both datasets were psychometrically sound and clear of anomalous data entries. The suitability of the dataset for factor analysis was evaluated using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity (Field, 2018). To determine the underlying factor structure and facilitate item clustering, principal components analysis was executed (Büyüköztürk, 2017). To optimize the variance of the item factor loadings and achieve simple structure, the Varimax orthogonal rotation technique was implemented (Acal et al., 2020; DeVellis, 2017; Tabachnick & Fidell, 2013). During the CFA phase, standard goodness-of-fit indices were utilized to test the structural model fit of the scale, evaluated against the conventional threshold criteria in the psychometric literature presented in Table 6 (Carvalho & Chima, 2014; Kline, 2014; Seçer, 2017; Shi et al., 2019; Tabachnick & Fidell, 2013; Vieira, 2011). Within the scope of the reliability analysis, the scale items were evaluated using composite reliability, Cronbach's alpha, and test-retest reliability methods. Additionally, item discrimination indices were examined to scrutinize item performance.

Ethical Considerations

This investigation was approved on ethical grounds by the Düzce University Scientific Research and Publication Ethics Committee with the decision dated February 23, 2023, numbered 2023/58. All principles of scientific research and publication ethics were rigorously maintained throughout the research process, and voluntary informed consent was obtained from both the participants and their parents or legal guardians. Furthermore, intellectual property and copyright regulations were strictly followed for all existing scholarly works utilized in this study.

FINDINGS

Construct Validity

The construct validity of the IEMPES-SF was evaluated using EFA and CFA. Within this scope, the underlying factor structure of the scale, its relationships with the observed dimensions, and its alignment with theoretical foundations were scrutinized.

Exploratory Factor Analysis

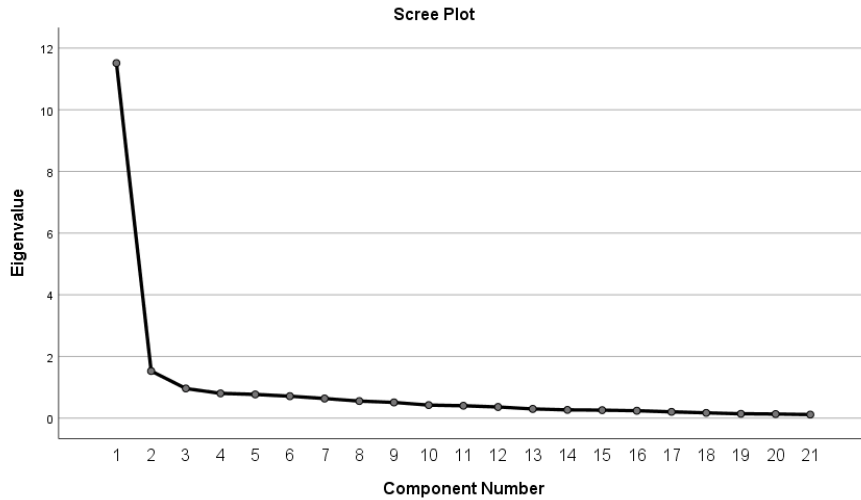
EFA is a statistical technique employed to identify latent dimensions based on the patterns of correlations among measured items (Büyüköztürk, 2017; Fabrigar & Wegener, 2011). Prior to executing the EFA, the KMO measure of sampling adequacy and Bartlett's test of sphericity were examined. The KMO index assesses whether the sample size within the dataset is adequate for factor analysis. KMO values range from 0 to 1.00, where values approaching 1.00 indicate structured correlation patterns, signifying that the factor analysis will yield distinct and stable factors. Bartlett's test of sphericity evaluates the null hypothesis that the correlation matrix is an identity matrix, implying that the variables are orthogonal; a statistically significant test result is required to demonstrate the suitability of the data for factor analysis. The KMO coefficient was determined to be .92, and the Bartlett's test of sphericity was statistically significant ($p < .01$). These KMO and Bartlett's test results demonstrate that the dataset meets the necessary assumptions for factor analysis (Field, 2018; Howard, 2016; Morrison, 2009; Tabachnick & Fidell, 2013). The initial EFA conducted on the dataset revealed a two-factor structural model. The KMO coefficient and Bartlett's test of sphericity values associated with this structure support the adequacy of the model. The empirical findings regarding this factor structure are presented in Table 2.

Table 2

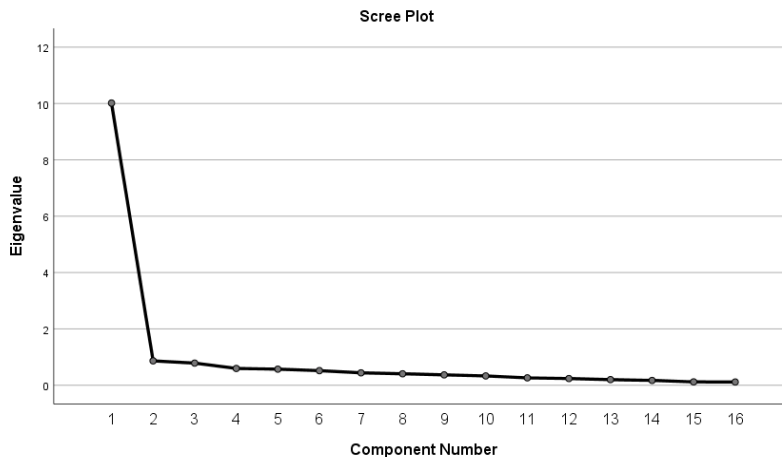
IEMPES Eigenvalues, Variance Explained, and Cumulative Variance Values

Factor	Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
1	11.256	53.456	40.371
2	1.701	8.009	61.553

As illustrated in Table 2, the eigenvalues for the two-factor scale structure emerging from the initial EFA were calculated as 11.256 for the first factor and 1.701 for the second factor. The first factor explained 53.45% of the total variance, whereas the second factor accounted for 8.01%. In scale development research, a high proportion of explained variance indicates that the measurement instrument accurately reflects the target construct and possesses a robust structural framework (Tavşancıl, 2014). Within the social sciences literature, an explained variance ratio ranging between 30% and 60% is widely accepted as satisfactory (Büyüköztürk, 2017; Tavşancıl, 2014). In this context, because the first factor independently accounted for 53.46% of the variance, and its eigenvalue was approximately 11 times greater than that of the second factor, the statistical evidence strongly suggests a unidimensional structure. Since the vast majority of the variance across the dimensions was subsumed under this single latent construct, a scree plot was generated in Figure 1 to make a definitive decision regarding the optimal factor solution (Ledesma et al., 2007; Ledesma et al., 2015).

Figure 1*Scree Plot of the Scale Items*

To refine the optimal number of factors for the scale, both scree plot and parallel analysis methods were utilized. When the scree plot for the initial structure was examined, the exceptionally high eigenvalue for the first factor, the negligible eigenvalue for the remaining factor, and the presence of only one distinct break in the curve strongly supported evaluating the scale as a single-factor structure (Pallant et al., 2016; Silva et al., 2020). Furthermore, parallel analysis demonstrated that the random eigenvalues generated from simulated datasets were consistently lower than the empirical eigenvalues obtained from the actual dataset, further corroborating the unidimensional model (Hayton et al., 2004). Considering that the scale items are conceptually interrelated with both the learning outcomes and the systemic processes of the curriculum, a single-factor structure was also deemed theoretically appropriate based on the core tenets of the IEM (Parlett & Hamilton, 1977). After defining the scale as unidimensional, a subsequent EFA revealed that five items exhibited factor loadings below the conventional threshold of 0.30. Recognizing that these items could compromise the structural validity of the instrument, a stepwise elimination strategy was implemented through iterative EFA trials. Throughout these iterations, the factor loadings of the problematic items consistently remained below the specified cut-off criteria (0.30); therefore, these statements were systematically excluded from the scale to ensure structural integrity (Büyüköztürk, 2017; Louangrath & Sutanapong, 2018; Raubenheimer, 2004; Şencan, 2005). Following the removal of these five items, the newly emerged factor structure was re-examined, confirming that the single-factor model was robustly maintained across both the calculated eigenvalues and visual scree indicators. The scree plot for the remaining 16 items of the scale is illustrated in Figure 2.

Figure 2*Scree Plot of the Remaining Scale Items*

Among the five items removed from the scale because their EFA factor loadings fell below 0.30 were the statements "Thanks to the support education programme, my reading skills improved," "Thanks to the support education programme, my writing skills improved," and "Thanks to the support education programme, my mathematics skills improved." The low factor loadings observed for these specific items are considered to stem from the clinical heterogeneity of the participants' specific learning disability profiles. In particular, because the students presented with distinct subtypes of learning disabilities, the internal consistency of their responses to these skill-specific statements may have been attenuated. Furthermore, the limited prevalence of co-occurring or combined learning disability profiles among the sampled students may have restricted the psychometric contribution of these items to the overarching structure of the scale. Additionally, the low factor loadings obtained for the other excluded items, specifically "I can apply the knowledge I learned in the programme at school and in daily life" and "The knowledge I learned in the programme makes it easier for me to learn new information," can be attributed to the individual differences of the students and their lack of cumulative experience in these functional generalization domains. The psychometric parameters concerning the finalized single-factor model of the scale are presented in Table 3.

Table 3

IEMPES-SF Eigenvalues, Variance Explained, and Cumulative Variance Values

Factor	Eigenvalue	Variance Explained (%)
1	9.969	62.303

The total variance explained calculated for the single-factor scale model was found to be 62.303%. This proportion of variance is a highly satisfactory value widely accepted within the psychometric literature (Büyüköztürk, 2017; Tavşancıl, 2014). Because the items of the IEMPES-SF clustered under a single latent dimension as a result of the EFA, no rotation procedure was executed. The 16 items retained in this single-factor structure accounted for 62.30% of the total variance. An eigenvalue greater than 1.00 for the factor, coupled with an explained variance ratio exceeding 60%, demonstrates that the scale structure robustly reflects the targeted latent construct (Beavers et al., 2013; Büyüköztürk, 2017). For unidimensional scale structures, an explained variance ratio of 30% or higher is considered sufficient (Tavşancıl, 2014). Consequently, the cumulative variance calculated for the entire scale (62.30%) represents an exceptionally high value within the social sciences literature. The psychometric properties and structural characteristics of the 16 items retained in the scale are presented in Table 4.

Table 4

Item Factor Loadings and Item-Total Correlations

Final Item No	Draft Item No	Items	Item-Total Correlation	Factor Loading
1	16	I enjoy participating in classroom learning activities.	.781	.821
2	11	I enjoy learning mathematics topics.	.779	.782
3	15	I experience difficulties participating in classroom learning activities.	.755	.799
4	10	I feel excited that I will learn new topics at the rehabilitation centre.	.748	.796
5	12	I enjoy doing reading activities.	.742	.753
6	9	The topics I learn in lessons interest me.	.732	.803
7	18	While explaining the topic in class, my teacher provides examples regarding the topic.	.669	.729
8	17	The materials used in lessons attract my attention.	.656	.719
9	13	I enjoy learning writing activities.	.601	.632
10	7	I am happy to benefit from the support education programme.	.574	.626
11	20	My teacher shares their ideas with me regarding the activities I perform.	.528	.599

Table 4 (Continued)

Final Item No	Draft Item No	Items	Item-Total Correlation	Factor Loading
12	21	After the activities performed in class, my teacher asks questions regarding what I have learned in the lesson.	.477	.547
13	14	I can comprehend the lessons within the duration explained by my teacher.	.468	.476
14	8	I comprehend the topics explained at the rehabilitation centre within the time allocated for lessons.	.426	.484
15	19	My teacher monitors and observes (tracks) me during the activities I perform.	.368	.464
16	4	Thanks to the support education programme, my reading skills improved.	.302	.363

As presented in Table 4, the factor loadings of the 16 items retained in the scale ranged from .36 to .82, while the corrected item-total correlations ranged from .30 to .78. These empirical findings demonstrate that the items possess psychometrically appropriate factor loadings and strong structural integrity (Louangrath & Sutanapong, 2018; Punch, 2013; Raubenheimer, 2004).

Confirmatory Factor Analysis

The structural validity and structural fit of the unidimensional, 16-item scale model established through EFA were cross-validated by conducting a CFA with an independent sample of 260 students. The goodness-of-fit indices obtained from the CFA are presented in Table 5.

Table 5

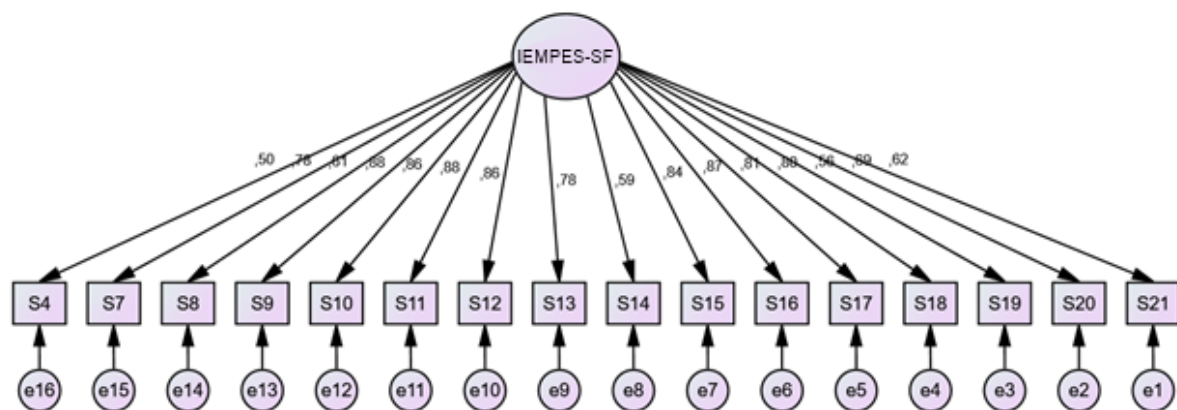
CFA Goodness-of-Fit Indices

χ^2/df	SRMR	RMSEA	CFI	GFI	AGFI	IFI	NFI	NNFI
2.72	.08	.069	.95	.95	.90	.92	.90	.92

Figure 3 presents the CFA diagram illustrating the factor structure and standardized pathways for the IEMPES-SF.

Figure 3

IEMPES-SF CFA Diagram (Standardized Estimates)



To evaluate the goodness-of-fit indices for the CFA, the literature reference values presented in Table 6 were utilized.

Table 6*Goodness-of-Fit Index Evaluation Criteria*

Fit Indices	Excellent Fit	Acceptable Fit
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 3$
RMSEA	$0 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$
RMR	$0 \leq RMR \leq .05$	$.05 < RMR \leq .10$
SRMR	$0 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$
NFI	$.95 \leq NFI \leq 1.00$	$.90 \leq NFI \leq .95$
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI \leq .95$
AGFI	$.90 \leq AGFI \leq 1.00$	$.85 \leq AGFI \leq .90$
CFI	$.97 \leq CFI \leq 1.00$	$.95 \leq CFI \leq .97$

References: (Carvalho & Chima, 2014; Chan vd., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi vd., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011).

The goodness-of-fit indices for the scale can be summarized as follows: : $\chi^2/sd=2.72$, SRMR= .008, RMSEA= .069, CFI = .95, GFI= .95, AGFI = .90, IFI = .92, NFI= .90, NNFI= .92. These values align closely with the widely accepted threshold criteria in the psychometric literature, with the SRMR value in particular demonstrating an excellent structural model fit (Carvalho & Chima, 2014; Chan et al., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi et al., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011). In light of these empirical findings, the unidimensional construct validity of the IEMPES-SF stands robustly confirmed.

Reliability and Validity

The statistical analysis indicates that the measurement instrument possesses high reliability coefficients. Within this scope, the reliability parameters computed for the IEMPES-SF are presented in Table 7.

Table 7*IEMPES-SF Reliability Coefficients*

Scale	Cronbach's Alpha Reliability Coefficient	Composite Reliability (CR)	Test-Retest Reliability
IEMPES-SF	.94	.93	.90

The Cronbach's alpha reliability coefficient for the 16-item, single-factor structure of the IEMPES-SF was determined to be .94. Composite reliability (CR) is an internal consistency estimate calculated using the standardized factor loadings and error variances derived from the CFA (Peterson & Kim, 2013; Yang & Green, 2011). The composite reliability coefficient for the unidimensional structure of the IEMPES-SF was calculated as .93. Furthermore, the stability coefficient of the instrument was evaluated using the test-retest reliability method. The test-retest method involves administering the finalized measurement scale obtained from the EFA and CFA phases to the same participant group under identical conditions at two distinct points in time, after which the correlation coefficient between the two administrations is computed (Leppink & Pérez-Fuster, 2017; Seçer, 2017; Weir, 2005). The 16-item, single-factor scale was re-administered to a subset of 72 students with a 20-day interval. The correlation between the two measurement points was calculated, yielding a test-retest reliability value of .91. Taken together, these psychometric indices demonstrate that the measurement scale possesses an exceptionally high level of reliability (Büyüköztürk, 2017; DeVellis, 2014; Özdamar, 2016; Tavşancıl, 2014).

Item Discrimination

To determine the discrimination levels of the items retained in the IEMPES-SF, the significance of the differences between the mean scores of the upper and lower 27% groups was examined. The results of the independent samples t-test regarding the differences between the upper and lower groups are presented in Table 8.

Table 8

IEMPES-SF Item-Total Test Score Correlations and Upper-Lower 27% Group Comparisons

Item No	Lower 27% (n=130)		Upper 27% (n=130)		t	Corrected Item-Total Correlation
	$\bar{x}$	SD	$\bar{x}$	SD		
1	4.48	.71	4.38	.70	.830*	.39
2	4.64	.70	4.64	.74	.809*	.81
3	4.20	.99	4.29	.99	-.430*	.54
4	4.60	.71	4.71	.64	-1.000*	.84
5	4.65	.75	4.72	.67	-.587*	.86
6	4.68	.80	4.68	.75	.929*	.88
7	4.65	.75	4.61	.78	.328*	.86
8	4.48	.91	4.54	.86	-.381*	.79
9	4.27	.99	4.51	.88	-1.451*	.69
10	4.60	.71	4.67	.67	-.610*	.82
11	4.65	.75	4.70	.66	-.355*	.88
12	4.57	.77	4.62	.81	-.425*	.87
13	4.64	.66	4.62	.72	.122*	.82
14	4.07	.51	3.98	.52	.971*	.69
15	4.54	.65	4.71	.61	-1.597*	.71
16	4.50	.58	4.55	.65	-.547*	.58

* $p < .05$

To evaluate the discrimination level of each item constituting the single-factor scale structure, the cumulative scores obtained from the CFA study group were ranked in ascending order, and an independent samples t-test was performed between the lower 27% and upper 27% groups. The results of the analysis revealed that the differences between the mean scores of the lower and upper groups were statistically significant for every single item ($p < .01$). This finding empirically demonstrates that all items in the scale possess a high level of item discrimination (Büyüköztürk, 2017; Kubiszyn & Borich, 2013).

CONCLUSION, DISCUSSION, AND RECOMMENDATIONS

The primary objective of this study was to develop a valid and reliable measurement instrument based on the Illuminative Evaluation Model (IEM), designed to evaluate the Support Education Programme for Students with Specific Learning Disabilities (SEPSLD) through the direct perceptions of the participating students. Rather than being restricted solely to program outcomes, the IEM operates as a qualitative and holistic evaluation framework that focuses on how the program's objectives, content, implementation processes, and instructional interactions are experienced and interpreted by its active participants (Parlett & Hamilton, 1977). This model places stakeholder perspectives at the center of inquiry and views the integration of student perceptions into the curriculum evaluation process as an essential element that facilitates a more comprehensive, nuanced, and contextual understanding of program effectiveness.

Grounded in this theoretical framework, an initial 24-item draft pool was generated by the researchers to represent the core components of the evaluation model. To establish content validity, this preliminary draft was submitted for expert review. Based on the qualitative feedback and consensus of the field experts, three items deemed conceptually or operationally inadequate were excluded from the instrument, thereby reducing the draft scale to 21 items. This developmental phase is considered crucial for enhancing the content validity

of the instrument while ensuring its structural alignment with the underlying theoretical construct (DeVellis, 2017).

The structural validity of the IEMPES-SF was rigorously evaluated using EFA and CFA. The EFA results revealed a unidimensional structure comprising 16 items, which explained 62.30% of the total variance. In scale development studies, having an eigenvalue greater than 1.00 for each factor and a total explained variance exceeding 60% is considered a strong indicator of robust structural validity (Beavers et al., 2013; Büyüköztürk, 2017). Moreover, for unidimensional scales, an explained variance ratio of 30% or higher is deemed statistically sufficient (Tavşancıl, 2014). Within this psychometric framework, the cumulative variance calculated for the entire scale (62.30%) represents an exceptionally high value in the literature, thereby demonstrating the high structural validity of the instrument. The single-factor model obtained from the EFA was subsequently cross-validated using CFA. According to the goodness-of-fit indices yielded by the CFA $\chi^2/df=2.72$, SRMR= .008, RMSEA= .069, CFI = .95, GFI= .95, AGFI = .90, IFI = .92, NFI= .90, NNFI= .92), the fit parameters of the IEMPES-SF fall well within the generally accepted thresholds in the literature, with the SRMR value indicating an excellent model fit (Carvalho & Chima, 2014; Chan et al., 2007; Jöreskog & Sörbom, 1993; Kline, 2014; Seçer, 2017; Shi et al., 2019; Tabachnick & Fidell, 2013; Ullman, 2006; Vieira, 2011). Consequently, these empirical findings firmly confirm and support the unidimensional structure of the IEMPES-SF.

The Cronbach's alpha reliability coefficient for the 16-item, single-factor structure of the IEMPES-SF was determined to be .94, while its composite reliability coefficient was calculated as .93. The stability levels of measurement scales are typically evaluated using the test-retest method. This approach involves administering the instrument derived from the EFA and CFA phases to the same study group under identical conditions at two distinct points in time. The test-retest reliability coefficient of the scale is then calculated by comparing the results obtained from both administrations using correlation analyses (Leppink & Pérez-Fuster, 2017; Seçer, 2017; Weir, 2005). Within this scope, the 16-item, single-factor scale was administered to 72 students with a 20-day interval. The correlation coefficient between the two administrations was calculated as .91. Based on these empirical findings, it was concluded that the scale possesses a high level of reliability and measurement stability over time (Büyüköztürk, 2017; DeVellis, 2014; Özdamar, 2016; Tavşancıl, 2014).

To evaluate the discrimination levels of the items retained in the IEMPES-SF, an analysis was conducted based on the upper and lower 27% groups. The results of this analysis revealed a statistically significant difference ($p < .01$) between the participants in the lower and upper groups for each individual item on the scale. This finding demonstrates that every item is capable of distinguishing between individuals with low and high levels of program evaluation perceptions. In other words, the scale items respond sensitively to the variation in students' levels of curriculum evaluation, providing an important indicator for the structural validity of the measurement instrument. The significant differences obtained at the item level firmly support the instrument's sensitivity and measurement power in capturing the targeted latent construct (Büyüköztürk, 2017; Kubiszyn & Borich, 2013).

In light of the empirical findings, it has been demonstrated that the IEMPES-SF, which was developed on the basis of the Illuminative Evaluation Model, is a scientifically competent measurement instrument with robust validity and reliability parameters specifically designed to capture student perceptions regarding curriculum evaluation. The results derived from the validity analyses confirm that structural validity has been firmly established, while the reliability analyses indicate that the scale yields highly consistent and dependable measurements. Consequently, the IEMPES-SF can be considered an exceptionally suitable assessment tool for researchers and practitioners who wish to address students' program evaluations through a comprehensive lens. Based on the findings of this study, it is recommended that the validity and reliability of the IEMPES-SF be further expanded by administering it to diverse participant groups and settings. Additionally, future studies utilizing this scale could focus on examining student perspectives on curriculum evaluation in relation to the demographic characteristics of the students benefiting from the program. Such endeavors are expected to provide a more holistic, nuanced, and in-depth contribution to curriculum evaluation processes within the field.

REFERENCES

- Acal, C., Aguilera, A. M., & Escabias, M. (2020). New modeling approaches based on varimax rotation of functional principal components. *Mathematics*, 8(11),1-15. <https://doi.org/10.3390/math8112085>

- Adıgüzel, O. C. ve Özüdoğru, F. (2014). A Study on Program Evaluation Scale Oriented to Illuminative Evaluation Model for the Primary School 2nd Grade English Curriculum. *Trakya Journal of Education*, 4(2), 124-136.
- Adıgüzel, O. C. (2015). Development of a Scale to Evaluate the Pedagogical Formation Program Implemented with Turkish Prospective Teachers. *Education Research and Review*, 10(2), 101-110. <http://dx.doi.org/10.5897/ERR2014.2029>
- Akıncı, M., & Köse, E. (2019). Validity and Reliability Study of Teacher Training Curriculum Evaluation Scale (TTCES). *IBAD Journal of Social Sciences*, 4(2), 236-252. <http://dx.doi.org/10.21733/ibad.534751>
- Aktan, O. (2020). Determination of educational needs of teachers regarding the education of inclusive students with learning disability. *International Journal of Contemporary Educational Research*, 7(1), 149-164. <https://doi.org/10.33200/ijcer.638362>
- Aktan, O. (2021). Development of the Teacher Social Acceptance Scale for Individuals with Special Needs, Validity and Reliability Study. *Trakya Journal of Education*, 11(1), 315-332. <https://doi.org/10.24315/tred.71298>
- Aktan, O. (2023). Inclusive education. In S. F. Büyükalın (Ed.), *Learning teaching theories and approaches* (4th ed., pp. 221-265). Pegem Akademi.
- Aktan, O., Üçüncü, B., Bingöl, T., & Gökbayrak, Ö. (2025). Aydınlatıcı Değerlendirme Modeli Değerlendirme Ölçeği'nin geliştirilmesi: Geçerlik ve güvenirlik analizi. *Anadolu Eğitim Liderliği ve Öğretim Dergisi*, 13(1), 49-64.
- Aktan, O., & Orakcı, Ş. (2026). Teachers' Curriculum Literacy and Individualised Education Program Competencies in Turkey: A Mixed Method Study. *International Journal of Disability, Development and Education*, 73(5), 949-975. <https://doi.org/10.1080/1034912X.2025.2605504>
- Altınsoy, T. & Küçüksüleymanoğlu, R. (2021). Development of the 5th Grade Music Course Curricula Evaluation Scale. *Journal of Necmettin Erbakan University Ereğli Faculty of Education*, 3 (1), 82-96. Retrieved from <https://dergipark.org.tr/tr/pub/neueefd/issue/62544/714944>
- Armstrong, F., Armstrong, D., & Barton, L. (2016). *Inclusive education: Policy, contexts and comparative perspectives*. Routledge.
- Aslan, M., Soyalt, H., Karahan, O. & Altuntaş, M. (2016). The Development of Preschool Curriculum Evaluation Scale. *Van Yuzuncu Yıl University Journal of Education Faculty*, 13(1), 657-683.
- Bal, A. P. & Arslan, U. (2021). Developing The Secondary School Mathematics Curriculum Evaluation Scale: Student Dimension. *Mediterranean Journal of Educational Research*, 15(37), 26-40. <https://doi.org/10.29329/mjer.2021.380.2>
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1), 1-13. <https://doi.org/10.7275/qv2q-rk76>
- Bryant, D. P., Bryant, B. R., & Smith, D. D. (2019). *Teaching students with special needs in inclusive classrooms* (1th ed.). SAGE Publications Ltd.
- Burden, B. (2017). Illuminative evaluation. In *Frameworks for practice in educational psychology: A textbook trainees and practitioners*, edited by B. Kelly, L. M. Woolfson, and J. Boyle, 291– 308. 2nd ed. London: Jessica Kingsley
- Büyüköztürk, Ş. (2017). *Data Analysis Handbook for the Social Sciences: Statistics, Research Design, SPSS Applications, and Interpretation* (22nd ed.). Pegem Academy

- Cahapay, M. B. (2020). The responsiveness of bachelor of elementary education curriculum: An illuminative evaluation. *International Journal of Evaluation and Research in Education*, 9(3), 743-750. <https://doi.org/10.11591/ijere.v9i3.20649>
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication methods and measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Carvalho, J., & Chima, F. O. (2014). Applications of structural equation modeling in social sciences research. *American International Journal of Contemporary Research*, 4 (1), 6-11.
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of pediatric psychology*, 27(1), 27-36.
- Chan, F., Lee, G. K., Lee, E. J., Kubota, C. & Allen, C. A. (2007). Structural equation modeling in rehabilitation counseling research. *Rehabilitation Counseling Bulletin*, 51 (1), 53- 66. <https://doi.org/10.1177/00343552070510010701>
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.
- Chen, H. T. (2010). The bottom-up approach to integrative validity: a new perspective for program evaluation. *Evaluation and program planning*, 33(3), 205-214. <https://doi.org/10.1016/j.evalprogplan.2009.10.002>
- Çelikel, F. (2022). *Evaluation of the middle school mathematics curriculum in terms of mathematical competence using the Illuminative Evaluation Model* [Unpublished doctoral dissertation]. Mersin University
- Çopur, E., Türkmenoğlu, M., Artut Dinç, P., & Bal, A. P. (2021). Scale for Evaluating Primary School Mathematics Curriculum According to Stufflebeam's CIPP Model. *International Journal of Educational Spectrum*, 3(2), 150–175. <https://doi.org/10.47806/ijesacademic.903747>
- DeVellis, R. F. (2017). *Scale Development: Theory and Applications* (4th ed.). Thousand Oaks, CA: Sage.
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics*. The USA: Sage
- Gültekin, M., & Demirci, C. (2020). Evaluation of the English language preparatory school curriculum in the lights of illuminative evaluation model. *European Journal of English Language Teaching*, 5(3), 103-131.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). SEM: An introduction. *Multivariate data analysis: A global perspective*, 5(6), 629-686.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2), 191-205. <https://doi.org/10.1177/1094428104263675>
- Hornby, G. (2011). Inclusive education for children with special educational needs: A critique. *International Journal of Disability, Development and Education*, 58(3), 321-329. <https://doi.org/10.1080/1034912X.2011.598678>.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?. *International journal of human-computer interaction*, 32(1), 51-62.
- Jöreskog, K. & Sörbom, D. (1998). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*, 4th ed. Scientific Software International, Hillsdale.

- Karabay-Turan, A., Kuşdemir-Kayıran, B., & Bal, A. P. (2019). Turkish Language Curriculum Evaluation Scale: A validity and reliability study. *Bolu Abant İzzet Baysal University Journal of Faculty of Education*, 19(4), 1654–1666. <https://doi.org/10.17240/aibuefd.2019...-543058>
- Kauffman, J. M., & Hallahan, D. P. (Eds.). (2011). *Handbook of special education*. Routledge.
- Kavan, N. (2023). Türkçe Dersi Öğretim Programı CIPP Modeli Değerlendirme Ölçeği: Geçerlik ve Güvenirlilik Çalışması. *Türkiye Eğitim Dergisi*, 8(1), 159-185. <https://doi.org/10.54979/turkegitimdergisi.1306033>
- Kline, P. (2014). *An easy guide to factor analysis*. Routledge.
- Kubiszyen, T. & Borich, G. (2013). *Educational testing and measurement: Classroom application and practice*. (10th ed.) Indianapolis: John Wiley ve Sons, INC
- Kutay, V. (2018). Inclusive education in Turkey. *Studies in Educational Research and Development*, 2(2), 144-162.
- Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. *Psychology*, 9(11), 2531-2560.
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy to use computer program for carrying out a parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1–11. <https://doi.org/10.7275/wjnc-nm63>
- Ledesma, R. D., Mora, P. V., & Macbeth, G. (2015). The scree test and the number of factors: A dynamic graphics approach. *The Spanish Journal of Psychology*, 18(E11), 1–10. <https://doi.org/10.1017/sjp.2015.13>
- Leppink, J., & Pérez-Fuster, P. (2017). We need more replication research—A case for test-retest reliability. *Perspectives on Medical Education*, 6(3), 158–164. <https://doi.org/10.1007/s40037-017-0347-z>
- Louangrath, P. I., & Sutanapong, C. (2018). Validity and reliability of survey scales. *International Journal of Research & Methodology in Social Science*, 4(3), 99-114.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (Eds.). (2012). *Evaluation models: Viewpoints on educational and human services evaluation* (Vol. 6). Springer Science & Business Media.
- Marsh, C. J. & Willis, G. (2007). *Curriculum: Alternative approaches, ongoing issues*. New Jersey: Pearson Prentice Hall.
- MoNE. (2008). *Support Education Program for Individuals with Specific Learning Disabilities* (pp. 4–8). Ankara, Türkiye.
- MoNE. (2021). *Support Education Program for Individuals with Learning Disabilities*. https://ookgm.meb.gov.tr/meb_iys_dosyalar/2021_07/09112608_21130110_Ogrenme_Guclugu.pdf
- Morrison, J. T. (2009). Evaluating factor analysis decisions for scale design in communication research. *Communication Methods and Measures*, 3(4), 195–215.
- Oliva, P.F. (2009). *Developing the curriculum*. New York: Pearson Allyn and Bacon
- Ornstein, A. C., & Hunkins, F. P. (2009). *Curriculum: Foundations, principles, and issues* (5th ed.). Columbus, OH: Allyn & Bacon
- Özdamar, K. (2016). *Scale and Test Development: Structural Equation Modeling*. Nisan Publishing.

- Özüdoğru, F. (2016). *Evaluation of the primary school second-grade English language curriculum based on the Common European Framework of Reference for Languages using the Illuminative Evaluation Model* [Unpublished doctoral dissertation]. Anadolu University
- Özüdoğru, F., & Adıgüzel, O. C. (2016). Illuminative Program Evaluation Model. *Anadolu University Journal of Social Sciences*, 16(Special Issue), 25–34. <https://doi.org/10.18037/ausbd.417424>
- Palabıyık, T. (2021). *Evaluation of the second-grade English curriculum based on the Illuminative Evaluation Model according to teachers' views* [Unpublished master's thesis]. Mardin Artuklu University.
- Pallant, J. F., Haines, H. M., Green, P., Toohill, J., Gamble, J., Creedy, D. K., & Fenwick, J. (2016). Assessment of the dimensionality of the Wijma delivery expectancy/experience questionnaire using factor analysis and Rasch analysis. *BMC pregnancy and childbirth*, 16, 1-11. <https://doi.org/10.1186/s12884-016-1157-8>
- Parlett, M. & Hamilton, D. (1977). Evaluation as Illumination: A New Approach to the Study of Innovatory Programs. Parlett, M. & Dearden, G. (Eds.), *Introduction to Illuminative Evaluation: Studies in Higher Education* içinde (ss. 9-29). California: Pacific Soundings Press.
- Paulsrud, D., & Nilholm, C. (2023). Teaching for inclusion—a review of research on the cooperation between regular teachers and special educators in the work with students in need of special support. *International Journal of Inclusive Education*, 27(4), 541-555. <https://doi.org/10.1080/13603116.2020.1846799>
- Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194–198. <https://doi.org/10.1037/a0030767>
- Price, L. R. (2017). *Psychometric Methods: Theory into Practice*. New York: The Guilford Press
- Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. Sage.
- Rao, V., & Woolcock, M. (2003). Integrating qualitative and quantitative approaches in program evaluation. In Bourguignon F., Pereira Da Silva L. (Eds.), *Evaluating the poverty and distributional impact of economic policies: A general introduction*. World Bank and Oxford University Press.
- Raubenheimer, J. E. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, 30 (4), 59-64.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge university press.
- Seçer, İ. (2017). *Practical data analysis with SPSS and LISREL: Analysis and reporting*. Anı Publishing.
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and psychological measurement*, 79(2), 310-334. <https://doi.org/10.1177/0013164418783530>
- Silva, R., Oliveira, D., Santos, D. P., Santos, L.F.D., Wilson, R. E. & Bedo, M. (2020). Criteria for choosing the number of dimensions in a principal component analysis: an empirical assessment. In *Anais do XXXV Simposio Brasileiro de Bancos de Dados*, 145–150. Porto Alegre, Brazil: Computacao Sociedade Brasileira de Computacao. <https://sol.sbc.org.br/index.php/sbbd/article/view/13632>
- Şencan, H. (2005). *Reliability and validity in social and behavioral measurements*. Seçkin Publishing.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. The USA: Pearson Education.
- Tavşancıl, E. (2014). *Measuring attitudes and data analysis with SPSS*. Nobel.
- Ullman, J. B. (2006). Structural equation modeling: reviewing the basics and moving forward. *Journal of personality assessment*, 87 (1), 35- 50. https://doi.org/10.1207/s15327752jpa8701_03

- Uşun, S. (2016). Educational program evaluation: Processes, approaches, and models (2nd ed.). Anı.
- Vieira A. L. (2011). *Interactive LISREL in practice, getting started with a SIMPLIS Approach*. London: Springer. <https://doi.org/10.1007/978-3-642-18044-6>.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research*, 19 (1), 231-240. <https://doi.org/10.1519/15184.1>.
- Yang, Y., & Green, S.B. (2011). Coefficient Alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29 (4) 377-392. <https://doi.org/10.1177/0734282911406668>.
- Yıldırım, A. & Şimşek, H. (2013). *Qualitative research methods in the social sciences (9th ed.)*. Ankara: SeçkinYayıncılık.
- Zigmond, N. P., & Kloo, A. (2017). *General and special education are (and should be) different*. In Handbook of special education (pp. 249-261). Routledge.