



STEM Yeterlikleri Değerlendirme Çerçevesinin 4. sınıf Düzeyi için Uyarlanması

Serkan ARIKAN¹, Melek PESEN² ve Emine ERKTİN³

Özet

Alan yazında karşılaşılan çalışmalarda görülmektedir ki öğrencilerin STEM yeterlikleri değerlendirilirken STEM yeterliklerini ölçmek amacı ile oluşturulmuş, disiplinler arası bir ölçme aracı ile elde edilen puanlar yerine, büyük ölçekli testlerden veya başka amaçlar için geliştirilmiş sınavlardan elde edilen matematik ve/veya fen puanları kullanılmaktadır. Öğrencilerin STEM yeterlik puanlarını belirlemek amacı ile bir ölçme aracı geliştirebilmek için öncelikle STEM yeterlikleri değerlendirme çerçevesine ihtiyaç duyulmaktadır. Özellikle ilkökul seviyesinde, STEM yeterliklerini değerlendirebilmek amacıyla kullanılabilir bir değerlendirme çerçevesi bulunmamaktadır. Bu sebeple bu çalışmada, Arıkan ve diğerleri (2022) tarafından geliştirilmiş ve 8. sınıf düzeyinde geçerlik kanıtları sunulmuş olan STEM yeterlikleri değerlendirme çerçevesinin, 4. sınıf düzeyine uyarlanması amaçlanmıştır. Bu kapsamda, 4. sınıf düzeyine uyarlanan değerlendirme çerçevesi temel alınarak geliştirilen bir ölçme aracından elde edilen puanların güvenilirlik ve geçerlik kanıtları sunulmuştur. Bu çalışmada ortaya konan değerlendirme çerçevesi, STEM yeterliklerinin çok boyutlu ve bütünlük yapısı- fen, teknoloji, mühendislik ve matematik alanları- dikkate alınarak hazırlanmıştır. Ortaya konan faktör yapısı, 4. sınıf öğrencilerinden toplanan veriler kullanılarak doğrulayıcı faktör analizi ile değerlendirilmiştir. Geliştirilen ölçme aracında yer alan maddeler, madde tepki kuramı kullanılarak madde düzeyinde incelenmiştir. Bulgular, 8. sınıflar için geliştirilmiş olan STEM değerlendirme çerçevesinin 4. sınıflar için de kullanılabilirliğini göstermektedir.

Makale Bilgileri

Araştırma
Makalesi

Gönderim Tarihi
10/02/2023
Kabul Tarihi
12/10/2023
Yayın Tarihi
15/01/2024

Anahtar Kelimeler

STEM
değerlendirme
çerçevesi,
Rutin olmayan
problemler,
Doğrulayıcı
faktör analizi,
Madde analizi

¹ Boğaziçi Üniversitesi, ORCID: 0000-0001-9610-5496, serkan.arikan1@boun.edu.tr

² Boğaziçi Üniversitesi, ORCID: 0000-0003-1167-0707, melek.pesen@boun.edu.tr

³ Boğaziçi Üniversitesi, ORCID: 0000-0002-9428-7115, erktin@boun.edu.tr

Atıf:

Arıkan, S., Pesen, M. ve Erktin, E. (2023). STEM yeterlikleri değerlendirme çerçevesinin 4. sınıf düzeyi için uyarlanması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi [PAÜEFD]*, 60, 201-225. <https://doi.org/10.9779/pauefd.1249861>

Giriş

STEM Eğitiminin Önemi

Küresel ekonomide bir yer edinme endişesi, ülke yönetimlerini eğitim politikalarında STEM (Science, Technology, Engineering, Mathematics [Fen, Teknoloji, Mühendislik, Matematik]) eğitimini ön plana almak üzere harekete geçmeye yöneltmiştir (Kelly ve Knowles, 2016). STEM alanlarında başarılı olmak, işgücü piyasasında rekabet edebilmek ve ekonomik istikrara ve üstünlüğe sahip olmak ile ilişkilendirilmektedir (Biçer ve diğerleri, 2017; National Research Council [Ulusal Araştırma Konseyi], 2005; 2011; National Science and Technology Council [Ulusal Fen ve Teknoloji Konseyi], 2011). Bu çabalar STEM eğitimini popüler bir alana dönüştürmüş, birçok araştırmacı, politikacı, öğretmen ve yöneticinin odak noktası haline getirmiştir (Biçer ve diğerleri, 2017; Brown ve diğerleri, 2011; Capraro ve Çorlu, 2013). Ülkemizde de STEM alanlarında yetişmiş birey ihtiyacı, alan yazında, çeşitli eğitim raporlarında ve Millî Eğitim Bakanlığı'nın yol haritalarında vurgulanmıştır (Millî Eğitim Bakanlığı [MEB], 2016; Türk Sanayicileri ve İş İnsanları Derneği [TÜSİAD], 2014; Yıldırım ve Gelmez-Burakgazi, 2020).

STEM eğitimi, doğası gereği, fen, teknoloji, mühendislik ve matematiğin birbiriyle örtüştüğü ve bütünleştiği disiplinler arası bir öğretme ve öğrenme yaklaşımını yansıtmaktadır (Broggy ve diğerleri, 2017). Örneğin, matematik, bilimsel kavramların anlaşılması için çok önemlidir. Benzer şekilde, bilimsel düşünce yöntemleri ve bilimsel kavramlar matematik için çeşitli bağlamlar oluşturur. Mühendislik ve teknoloji matematiğe ve uygulamalarına dayanır. Araştırmalar, beynin matematik, fen ve ilgili disiplinlerin bir arada bulunduğu durumlar üzerine çalışması durumunda en iyi performansı gösterdiğini göstermektedir (Biçer ve diğerleri, 2017; Freudenrich ve Boyd, 2001). STEM eğitimi, öğrencilerin tüm STEM alanlarında etkili problem çözebilmeleri için eleştirel düşünme becerilerini geliştirmeyi amaçlamaktadır (White, 2014).

STEM Yeterliklerinin Değerlendirilmesi

Ölçme ve değerlendirme, öğretimde doğru kullanıldığında niteliği artırır, öğrenmeye katkı sağlar (Shulman, 2009). STEM eğitiminin niteliğini artırmak için de STEM yeterliklerinin sistematik olarak değerlendirilmesi ve izlenmesi gerekmektedir (Saxton ve diğerleri., 2014). STEM yeterliklerini ölçmek için kullanılacak biçimlendirici (formative) ve sonuca yönelik (summative) değerlendirme araçları, daha nitelikli STEM eğitimi için esastır. Biçimlendirici değerlendirme, öğrenme sürecinin her aşamasında sağlanan anında geri bildirim içerir. Sonuca yönelik değerlendirme ise öğrencilerin sahip olduğu becerileri ortaya koyarken, zaman içerisindeki gelişimlerini de izleme imkânı sunar. Alanyazında STEM eğitiminde biçimlendirici ve sonuca yönelik

değerlendirmenin etkinliğini artırmaya yönelik sınırlı sayıda çalışma vardır (Capraro ve Çorlu, 2013; Haudek ve diğerleri, 2011).

STEM yeterlikleri ile ilgili yapılan değerlendirmelerle ilgili en önemli sorun disiplinler arası değerlendirme çerçevesinin eksikliğidir. Birçok çalışma, öğrencilerin matematik veya fen puanlarını büyük ölçekli sınavlardan almakta ve bu puanları STEM başarı göstergeleri olarak kullanmaktadır. Ancak bu değerlendirme araçları öğrencilerin STEM yeterliklerini ölçmek amacıyla oluşturulmamıştır. Bu değerlendirme araçlarında STEM'in disiplinler arası doğası eksiktir. Örneğin, Han vd. (2015), STEM eğitimi ile matematik başarısı arasındaki ilişkiyi araştırmak için Texas Bilgi ve Becerilerin Değerlendirilmesi (Texas Assessment of Knowledge and Skills) testini kullanmışlardır. Benzer şekilde, Ing (2014), STEM başarısını ölçmek için Amerikan Gençliğinin Boylamsal Çalışması (Longitudinal Study of American Youth) ölçeğinin matematik puanlarını kullanmıştır. Matematik ve/veya fen puanlarının STEM yeterlikleri için özel olarak geliştirilmeyen testlerden alınmış olması, iyi tanımlanmış çok boyutlu bir STEM değerlendirme çerçevesinin eksikliğini göstermektedir. STEM yeterliklerini değerlendirme çerçevesinin oluşturulmasının, bu yeterlikleri değerlendiren araçların geliştirilmesine ve doğrulanmasına rehberlik etmesi beklenmektedir. Hem belirli içerik alanlarına hem de STEM alanlarının disiplinler arası yapısına odaklanan bir STEM değerlendirme modelinin geliştirilmesine ihtiyaç vardır (Harwell ve diğerleri, 2015).

STEM yeterliklerini ölçmek üzere hazırlanan çalışmalardan birinde Saxton vd. (2014) ortak bir STEM ölçme sisteminin geliştirilmesi üzerinde çalışmışlardır. Güçlü bir teorik yapıyı vurgulayan bu çalışmada STEM boyutları arasındaki bağlantıların önemi belirtilmiştir. Ancak, bu çalışmada sunulan çerçeve veri toplanarak test edilmemiş ve değerlendirme aracı sunulmadan yalnızca başlangıç aşaması olarak değerlendirme sisteminin kavramsallaştırılması hedeflenmiştir. Başka bir çalışmada, Biçer vd. (2017), öğrencilerin STEM alanında yeterliklerini görmek için entegre bir STEM değerlendirme modeline ihtiyaç olduğunu belirtmişlerdir. Ancak bununla birlikte, öğrencilerin Texas Bilgi ve Becerilerin Değerlendirilmesi (Texas Assessment of Knowledge and Skills) testindeki matematik ve fen puanları arasındaki korelasyonu rapor etmişler ve bunun teknoloji ve mühendislik alanları ölçme araçlarına dahil edilene kadar yeterli bir STEM değerlendirme modeli olarak hizmet edebileceğini belirtmişlerdir. Harwell vd. (2015) ise mühendisliği bir araç olarak kullanarak matematik, fen ve teknoloji becerilerini ölçmeye çalışmışlardır. Yakın zamanda ise lise öğrencilerinin öz değerlendirme yaparak STEM yeterliklerini değerlendirdikleri çalışmalar bulunmaktadır (Nga ve diğerleri, 2022; Trung ve diğerleri, 2022). Bu çalışmaların hazırladıkları çerçeve 5 boyut (Bilgi toplama ve işleme, bilgi yönetimi ve kullanımı, çözümün uygulanması, teknik güvenlik, topluluk paylaşımı) ve 15 davranışsal

göstergeyi kapsamaktadır. STEM yeterliklerinin değerlendirilmesi amacı ile Arıkan vd. (2022) tarafından 8. sınıflar düzeyine uygun bir değerlendirme çerçevesi geliştirilmiş ve bu çerçeve temel alınarak hazırlanan bir ölçme aracının geçerlik ve güvenilirlik kanıtları sunulmuştur. Bu çerçevede fen, teknoloji, mühendislik ve matematik alanları ile ilgili alt boyutlar (fizik, kimya, biyoloji, modelleme, teknoloji ve mühendislik problemleri, kodlama, algoritmik düşünme, kavramlar, örüntüler ve argümantasyon) tanımlanmış ve aralarındaki ilişkiler ortaya konmuştur. Ancak alan yazında 4. sınıf düzeyine uygun bir STEM değerlendirme çerçevesi bulunmamaktadır. İlkokuldan ortaokula geçerken öğrencilerin sahip oldukları STEM yeterliklerinin geçerliği ortaya konmuş bir değerlendirme çerçevesi kapsamında ölçülmesi gerekmektedir.

STEM Yeterlikleri Değerlendirme Çerçevesi

Disiplinler arası yapının vurgulandığı STEM yaklaşımlarında her ne kadar bazı disiplinler öne çıksa da (English, 2015; Fitzallen, 2015) STEM ile ilgili pek çok çalışmada matematiğe yapılan vurgu dikkat çekicidir. Bu vurgu matematik becerisinin STEM yeterliklerini anlamada ve STEM ile ilgili kariyer seçimlerini tahmin etmede etkili olmasına ve matematik okuryazarlığının STEM problemlerini çözmedeki önemine dayanır (Enderson ve Ritz, 2016). Bütünleştirici bir STEM yaklaşımının güçlü bir matematik temeli gerektirdiği uzman kuruluşlarca da vurgulanmaktadır (National Council of Teachers of Mathematics (NCTM) [Ulusal Matematik Öğretmenleri Konseyi], 2018), Matematik diğer STEM disiplinleri için önemli bir dil işlevi gördüğünden (Bennett ve Ruchti, 2014; English, 2015; Schmidt ve Houang, 2007) ve bu disiplinlerde karşılaşılan problemleri çözmek için gerekli bir araç olduğundan (Enderson ve Ritz, 2016; English, 2015) STEM disiplinlerinin merkezinde yer alır. Matematik her STEM eğitim programında önemli bir rol oynar ve “Öğrencilerin okulda öğrendiği matematik, bütünleştirici STEM problemlerinin üstesinden gelmek için araç olarak kullanılacak içerikleri ve düşünme becerilerini içerir” (NCTM, 2018, s. 3, çev.). Tüm bu nedenlerden dolayı, STEM değerlendirme çerçevesinin de matematiği bir araç olarak içeren STEM problemlerini temsil edecek temellere sahip olması gerekmektedir. Arıkan ve diğerleri (2022) tarafından ortaya konan STEM değerlendirme çerçevesi de bütünleştirici bir yaklaşıma vurgu yapmakla birlikte, matematiği bir araç olarak ele almıştır.

STEM değerlendirme çerçevesi oluşturulurken temel alınacak bilişsel alan tanımlanmalıdır. TIMSS (Trends in International Mathematics and Science Study, [Uluslararası Matematik ve Fen Eğilimleri Araştırması]) 2019 Değerlendirme Çerçevesinde (TIMSS 2019 Assessment Framework) tanımlanan bilişsel alanlardan akıl yürütme becerisi STEM'in doğası ile bağdaşır. TIMSS 2019'a göre akıl yürütme becerisi

“alışık olunmayan durumlar, karmaşık bağlamlar ve çok adımlı durumları kapsayacak şekilde rutin problemlerin çözümünün ötesine geçen” beceriler olarak tanımlanmaktadır (Mullis ve Martin, 2017, s. 22). TIMSS akıl yürütme becerisinde yer alan rutin olmayan problemleri tanımlarken aşına olmama durumunu öne çıkarmaktadır. Rutin olmayan problemlerin aksine, öğrencilerin herhangi bir ders kitabında veya derste karşılaşılabileceği ve bilinen bir algoritmayı kullanmalarını gerektiren problemler ise rutin problemler olarak adlandırılır. Bu tanımlara göre, rutin olmayan problemler, bir öğrencinin akıl yürütmesini gerektiren, öğrenciler için anlamlı ancak günlük hayatta sıklıkla karşılaşmadıkları durumları içeren problemlerdir (Milgram, 2007). Rutin olmayan problemlerin bir diğer ayırt edici yönü ise öngörülebilir ve tekrar tekrar öğretilen bir yaklaşımın olmamasıdır; birden fazla çözüm yolu olabilir (Elia ve diğerleri, 2009; Lee ve diğerleri, 2014; Pantziara ve diğerleri, 2009; Woodward ve diğerleri, 2012). Arıkan vd., (2022) tarafından ortaya konan STEM değerlendirme çerçevesi de oluşturulacak problemlerin aşına olunmayan durumları kapsayan rutin olmayan problemlerden kurgulanmasını önermiştir.

STEM yeterliklerini ölçmek üzere STEM alanlarının alt boyutları da tanımlanmalıdır. Arıkan vd. (2022) tarafından 8. sınıf düzeyine uygun olarak ortaya konan STEM değerlendirme çerçevesi matematik boyutunu, algoritmik düşünme, kavramlar, örüntüler ve argümantasyon ile; fen boyutunu fizik okuryazarlığı, kimya okuryazarlığı, biyoloji okuryazarlığı; ve teknoloji ve mühendislik boyutunu ise, modelleme, tekno-mühendislik problemleri ve kodlama alt boyutları ile ölçülebileceğini belirtmişlerdir.

Dördüncü sınıf düzeyinde STEM yeterliklerini değerlendirme amacıyla kullanılacak bir değerlendirme çerçevesine ihtiyaç duyulmaktadır. Bu kapsamda bu çalışmanın amacı Arıkan vd. (2022) tarafından önerilen STEM yeterlikleri değerlendirme çerçevesinin 4. sınıf düzeyine uyarlanması ve bu değerlendirme çerçevesi temel alınarak geliştirilen bir ölçme aracından elde edilen puanların güvenilirlik ve geçerlik kanıtlarının sunulmasıdır. Sekizinci sınıf düzeyine benzer olarak, değerlendirme çerçevesi dördüncü sınıf düzeyine uyarlanırken STEM'in çok boyutlu yapısını yansıtan, matematiği bir araç olarak kullanan ve rutin olmayan problemlere odaklanan bir yapı korunmuştur. Bu çalışmanın araştırma sorusu “4. sınıf düzeyine uyarlanan STEM yeterlikleri değerlendirme çerçevesi kapsamında tanımlanan boyutlar ve boyutlar arasındaki beklenen ilişkiler öğrenci yanıtları ile desteklenmekte midir?” şeklindedir.

Yöntem

Katılımcılar

Katılımcılar gönüllü olarak çalışmaya katılan 4. sınıf öğrencileridir. Çalışma kapsamında veriler pilot ve asıl çalışma olmak üzere iki farklı zamanda toplanmıştır. Pilot çalışmaya İstanbul'daki bir okuldan toplam 95 öğrenci katılmıştır. Asıl çalışmaya ise iki farklı şehirden ve altı okuldan toplam 205 öğrenci katılmıştır. Pilot uygulamaya katılanların %43'ü kız, %57'si erkek, asıl uygulamaya katılanların %53'ü kız, %47'si erkektir. Çalışma ile ilgili etik kurul izinleri alınmıştır.

Veri Toplama Aracı

Bu çalışmada Arıkan vd. (2022) tarafından ortaya konan STEM yeterlikleri değerlendirme çerçevesi ilk olarak dördüncü sınıf düzeyine uyarlanmış, ardından ortaya çıkan alt boyutlar temel alınarak bir ölçme aracı geliştirilmiştir. Dördüncü sınıf düzeyi gözetilerek oluşturulan maddeleri belirleyen STEM yeterlikleri çerçevesinin alt boyutları Tablo 1'de gösterilmiştir. Matematik boyutu Arıkan vd. (2022) önerdiği gibi algoritmik düşünme, kavramlar, örüntüler ve argümantasyon alt boyutlarından oluşmaktadır. Sekizinci sınıflar düzeyinde fen okuryazarlığı, fizik, kimya ve biyoloji ile ilgili okuryazarlık olarak üçe ayrılmışken, dördüncü sınıflar düzeyinde hepsi birleştirilerek fen okuryazarlığı olarak tek bir boyut olarak ele alınmıştır. Teknoloji ve mühendislik boyutunda da yine önerilen yapıdaki gibi modelleme, kodlama ve tekno-mühendislik problemleri boyutları temel alınmıştır. Arıkan vd. (2022) de vurgu yaptığı gibi STEM becerilerini ölçmek için öğrencilerin sıklıkla aşına oldukları problemler yerine akıl yürütme becerisi gerektiren rutin olmayan problemler temel alınarak maddeler hazırlanmıştır. Bu alt boyutlar aşağıda detaylı olarak açıklanmıştır. Bu boyutları temsil eden örnek sorular da çalışmanın sonunda verilmiştir. Bu araştırma, Muğla Sıtkı Koçman Üniversitesi İnsan Araştırmaları Etik Kurulunun 23/03/2018 tarihli 45 sayılı kararı ile alınan izinle yürütülmüştür.

Tablo 1

STEM Yeterlikleri Çevresi Alt Boyutları

Matematik	Fen	Teknoloji ve Mühendislik
Algoritmik Düşünme	Fen Okuryazarlığı	Modelleme
Kavramlar		Kodlama
Örüntüler		Tekno-Mühendislik Problemleri
Argümantasyon		

Algoritmik Düşünme

Algoritmik düşünme Brown (2015) tarafından “algoritmaları anlama, yürütme, değerlendirme ve oluşturma yeteneği” olarak tanımlanmıştır (s.1). Bir algoritma, açık, yürütülebilir ve gerçekleştirilecek adım adım sıralı talimatlar içerir (Brown, 2015). Algoritmik düşünme çoğunlukla hesaplamalı düşünme ve dolayısıyla bilgisayar bilimiyle ilişkilendirilmektedir (van Borkulo ve diğerleri, 2021).

Kavramlar

Kavramlar, genel olarak, öğrencilerin öğretim programlarında karşılarına çıkan matematiksel kavram ve ilkeleri bilmeleri ve uygulayabilmeleri ile ilgilidir. STEM değerlendirme çerçevesinde ise, kavramlar, öğrencilerin bilmeleri gereken kavramlardan yola çıkılarak fazla aşına olmadıkları yeni tanımlar çerçevesindeki okuryazarlıkları ve uygulayabilme becerileri ile ilgilidir.

Örüntüler

Örüntüleri tanıma, mevcut düzeni ve matematiksel yapıyı görebilmeyi ve bu yapıdan gerekli çıkarımları ve/veya kuralları elde edebilmeyi içerir (Blanton ve Kaput, 2011; Miller, 2019). Örüntü tanıma sorularında, öğrencilerden sayı veya şekil desenini tanımları ve desende takip eden terimleri bulmaları beklenmektedir (Arıkan ve diğerleri, 2022).

Argümantasyon

Argümantasyon, matematiksel ifadelerin ve argümanların nitelikli bir şekilde oluşturulması ve başkaları tarafından oluşturulan argümanların doğruluğunun değerlendirilmesi ile ilintilidir (Rumsey ve Langrall, 2016). Matematiksel argümantasyon sürecinde yeni bilgi üretme ve gerekçelendirmeler sunarak başkalarını ikna etme eylemleri de gerçekleşir (Rumsey ve Langrall, 2016). Toulmin (2003), bir argümanın temel olarak iddia, veri ve gerekçeden oluştuğunu belirtmiştir. STEM değerlendirme çerçevesinde yer alan argümantasyon problemlerinde öğrencilere bir iddia sunulmuş ve öğrencilerden verilen verilere göre iddiayı değerlendirmeleri ve kararlarını matematiksel olarak gerekçelendirmeleri istenmektedir (Arıkan ve diğerleri, 2022).

Fen Okuryazarlığı

Fen okuryazarlığı modern hayattaki bilimsel metinlerin ve durumların daha iyi analiz edilip değerlendirilmesi ve karar verme süreçlerinde daha iyi akıl yürütme fırsatı edinmek adına önem taşımaktadır (Feinstein, 2011). STEM değerlendirme çerçevesinde, fen bilgisi konularında yer alan matematikle ilişkilendirilebilen fen okuryazarlığı becerisi temel alınmıştır. Hazırlanan maddelerde öğrencilere sunulan bilimsel kısa metinlerin irdelenmesi, ilişkilerin açığa çıkarılması ve verilen veriler doğrultusunda sonuca ulaşılması beklenmektedir (Arıkan ve diğerleri, 2022).

Modelleme

Matematiksel modeller birtakım işlemlerin, ilişkilerin ve kuralların kullanımıyla bir sistemi tanımlamak, açıklamak ve öngörmek adına kullanılmasını ifade eder (Doerr ve English, 2003). Modelleme, gerçek hayatta karşılaşılabilecek senaryolarının matematiksel ifadelere çevrilmesi olarak tanımlanır (Blum ve Ferri, 2009; Turner ve diğerleri, 2022). STEM değerlendirme çerçevesinde, modelleme alt boyutu için öğrencilerden verilen verileri kullanarak en iyi modeli belirlemeleri beklenmektedir (Arıkan ve diğerleri, 2022).

Tekno Mühendislik Problemleri

Tekno mühendislik problemleri optimizasyon problemleri bağlamında öğrencilere sunulmuştur. Bu bağlamda STEM değerlendirme çerçevesinde öğrencilerin verilen bilgilere dayanarak çeşitli olasılıkları göz önünde bulundurmaları ve sunulan kısıtlamalar dahilinde en uygun yanıtı oluşturmaları beklenmektedir (Arıkan ve diğerleri, 2022).

Kodlama

Kodlama, programlama becerileri ile birlikte sayısal düşünme gerektiren, öğrencileri mühendislik kariyerlerine hazırlayan ve cebirsel düşünme becerilerini geliştiren önemli bir alan olarak görülmektedir (Miller, 2019; Wong ve diğerleri, 2015). 21. Yy. becerisi olarak sayısal düşünme, örüntüyü fark etme, parçalara ayırma, soyutlama ve algoritmalar oluşturma becerilerini kapsamaktadır (Miller, 2019). STEM değerlendirme çerçevesinde kodlama boyutu verilen kodu yürütmeyi gerektiren sorularla öğrencilere sunulmuştur. Öğrencilerden mevcut olan girdiyi kullanmaları, ifadelerin doğruluğunu değerlendirerek kodu çalıştırmaları ve ilkökul matematiği bağlamında istenen çıktıyı bulmaları beklenmektedir (Arıkan ve diğerleri, 2022).

Burada tanımlanan STEM değerlendirme çerçevesi kullanılarak, uygulanabilirlik de göz önünde bulundurulmuş ve her boyuttan ikişer madde hazırlanmıştır ve Tablo 2'de gösterilmiştir.

Tablo 2

Pilot ve Asıl Uygulama Belirtke Tablosu

Alt Boyutlar	Maddeler
Algoritmik Düşünme	S1, S2
Kavramlar ve Prensipler	S3, S4
Örüntüler	S5, S6
Argümantasyon	S7, S8
Fen Okuryazarlığı	S9, S10

Modelleme	S11, S12
Kodlama	S13, S15
Tekno Mühendislik	S14, S16
<hr/>	
Toplam Madde Sayısı	16
<hr/>	

Veri Analizi

Bu bölümde iç tutarlılık, madde analizi ve doğrulayıcı faktör analizi için izlenen veri analizi yöntemleri açıklanmıştır. Güvenirlik, maddelerin iç tutarlılığını dikkate alan Cronbach alfa değeri kullanılarak raporlanmıştır. Alfa değeri 0.70'ten küçük ise iç tutarlılıkta sorun olduğu düşünülürken; 0.70 ile 0.80 arası kabul edilebilir; 0.80 ile 0.90 arası iyi; ve 0.90'ın üstü ise mükemmel olarak değerlendirilir (Cicchetti, 1994). Alfa değerini hesaplamak için SPSS 22.0 kullanılmıştır.

Pilot uygulama madde analizleri, örneklem sayısı da göz önünde bulundurularak, Klasik Test Kuramına (KTK) dayalı bir şekilde gerçekleştirilmiştir (Hambleton ve Jones, 1993). Problemlili maddeleri, eğer varsa, tespit etmek ve revize etmek için madde güçlük ve madde ayırt edicilik değerleri hesaplanmıştır. Madde güçlüğü doğru yanıtların yüzdesini gösterir ve 0 ile 1 arasında değişir. Düşük değer bir maddenin zor olduğunu, yüksek değer ise bir maddenin örneklem için kolay olduğunu gösterir (Crocker ve Algina, 2008). Madde ayırt ediciliği ise, bir maddenin yüksek ve düşük başarılı öğrencileri birbirinden ayırt etme derecesini gösterir. Madde ayırt ediciliği -1 ile 1 arasında değişir ve 0.30'den yüksek değerler istenir (Crocker ve Algina, 2008). Bu çalışmada madde ayırt edicilik indeksi olarak düzeltilmiş madde-toplam korelasyonu (Rir) kullanılmıştır. KTK analizleri ITEMAN 4.4 programı ile gerçekleştirilmiştir.

Asıl uygulamanın madde analizleri ise Madde Tepki Kuramı (MTK) ile gerçekleştirilmiştir. Açık uçlu soruların madde parametreleri genelleştirilmiş kısmi puan (Generalized Partial Credit) metodu ile kestirilirken, çoktan seçmeli sorular için madde güçlüğü, ayırt ediciliğin ve şans parametresinin hesaplandığı üç parametrelili lojistik model (b, a ve c değerleri) kullanılmıştır (Guyer ve Thompson, 2014). MTK'da madde güçlüğü, b, şans parametresinin sıfır olduğu durumlarda, maddeyi %50 olasılıkla doğru yanıtlamak için gerekli olan beceri düzeyi olarak tanımlanır. Madde güçlüğü genel olarak -3 ile +3 arasında değer alırken, yüksek değerler sorunun zor olduğunu, düşük değerler ise sorunun kolay olduğunu belirtir (Guyer ve Thompson, 2014). Ayırt edicilik, a değeri, madde karakteristik eğrisinde b değerine denk gelen noktadaki eğimi hesaplanarak kestirilir. Ayırt edicilik değerinin yüksek olması beklenir. Şans parametresi, c, şans faktörüyle sorunun

doğru yanıtlanma olasılığını yansıtır. Beceri düzeyi çok düşük olan öğrencilerin soruyu doğru yanıtlanma olasılığıdır. Şans parametresi için beklenen değer, seçenek sayısının 1'e bölünmesiyle ortaya çıkan sayıdır. Şans parametresi, beklenen değerden yüksek olduğunda sorudaki seçeneklerin bazılarının kolay elendiği ve soruyu yanıtlamada şans faktörünün öne çıktığı söylenir (Guyer ve Thompson, 2014). Yanlış, kısmi doğru ve tam doğru (0-1-2) şeklinde puanlanan açık uçlu sorularda a ve iki farklı b parametresi (b1 ve b2) kestirilir. b1, 0 puandan 1 puana geçebilmek için gerekli olan beceriyi, b2 ise 1 puandan 2 puana geçebilmek için gerekli olan beceriyi göstermektedir. Madde Tepki Kuramı analizleri IRTPRO 4.20 programı ile yapılmıştır (Cai ve diğerleri, 2011).

STEM değerlendirme çerçevesinde önerilen modelin hazırlanan ölçme aracındaki öğrenci cevaplarıyla uyumunu belirlemek adına doğrulayıcı faktör analizi kullanılmıştır (Byrne, 1998; Kline, 1998; Schumacker ve Lomax, 2004). Doğrulayıcı faktör analizi (DFA) bilişsel ve psikolojik yapıların faktör yapılarını değerlendirerek yapı geçerliğine dair kanıtlar sunmada kullanılır (DiStefano ve Hess, 2005). Öğrenci yanıtlarından elde edilen kovaryans matrisi ile ele alınan modelin oluşturduğu kovaryans matrisinin benzer olması yani farkın minimum olması beklenmektedir (Ullman, 2001). Model ile verinin benzer olup olmadığına uyum iyiliği indislerine bakılarak karar verilmektedir. Bu uyum iyiliği indislerinden Root Mean Square Error of Approximation değerinin (RMSEA) 0.08 ve altı olması (Byrne, 1998); Comparative Fit Index (CFI) ve Tucker-Lewis Index (TLI) değerinin 0.95 ve üzeri olması (Ullman, 2001), oluşturulan modelin iyi olduğunu göstermektedir. Bu çalışmada MPLUS 7.4 programı kullanılarak DFA gerçekleştirilmiştir.

Bulgular

İç Tutarlılık

Pilot ve asıl uygulamada elde edilen verilerin güvenilirliği için Cronbach Alpha iç tutarlılık değerleri Tablo 3'te verilmektedir. Hem pilot hem de asıl uygulamada elde edilen güvenilirlik değerleri 0.70 civarı olduğu için verilerin güvenilirliği kabul edilebilirdir. Asıl uygulama değeri pilot uygulama değerinden daha yüksek çıkmıştır.

Tablo 3

Test Puanlarının İç Tutarlılık Değerleri

Uygulama	Kişi Sayısı	Soru Sayısı	Cronbach Alpha
Pilot	95	16	.69
Asıl	205	16	.75

Madde Analizi

Pilot Uygulama

Pilot uygulama madde analizi sonuçları Tablo 4'te verilmektedir. Pilot uygulamada 8 alt boyuttan 16 soru kullanılmıştır. Madde güçlük değerleri 0.26 ile 0.88 arasında değişmektedir. Ortalama madde güçlüğü 0.47 olarak hesaplanmıştır. Bu değerler testin zorluğunun gruba uygun olduğunu göstermektedir. Ayırt edicilik değerleri ise 0.07 ile 0.51 arasında değişmektedir. Ortalama ayırt edicilik değeri 0.26 olarak hesaplanmıştır. Negatif ayırt edicilik değeri olmaması olumlu olmakla birlikte altı sorunun ayırt edicilik değeri 0.20'den küçük, iki sorunun ayırt edicilik değeri ise 0.20 ile 0.30 arasındadır. Pilot uygulamanın amacı gereği kapsam geçerliği korunarak, pilot uygulamada elde edilen veriler incelenerek ve sorular arası uyuma bakılarak gerekli düzeltmeler yapılmış ve asıl uygulama formu elde edilmiştir.

Tablo 4

Pilot Uygulama Madde Analizi Sonuçları

Soru No	Soru Alt Boyutu	Soru Formatı	Madde Güçlüğü (p)	Madde Ayırt ediciliği (Rir)
1	Algoritmik Düşünme	AU	.83	.36
2	Algoritmik Düşünme	AU	.86	.19*
3	Kavramlar	ÇS	.40	.36
4	Kavramlar	ÇS	.30	.07*
5	Örüntüler	ÇS	.38	.11*
6	Örüntüler	ÇS	.88	.08*
7	Argümantasyon	AU	.44	.32
8	Argümantasyon	ÇS	.41	.32
9	Fen Okuryazarlığı	ÇS	.65	.24*
10	Fen Okuryazarlığı	ÇS	.76	.51
11	Modelleme	ÇS	.54	.26*
12	Modelleme	ÇS	.37	.10
13	Kodlama	ÇS	.26	.30
14	Tekno Mühendislik	ÇS	.48	.30
15	Kodlama	ÇS	.31	.18*
16	Tekno Mühendislik	ÇS	.34	.40

Not. ÇS: Çoktan Seçmeli; AU: Açık uçlu; * Rir < 0,30

Asıl Uygulama

Asıl uygulama madde analizi sonuçları Tablo 5'te verilmektedir. Asıl uygulamada 8 alt boyuttan 16 soru kullanılmıştır. Bu soruların klasik test kuramına göre hesaplanan madde güçlüğü değerleri (p), 0.27 ile 0.80 arasında değişmektedir. Madde Tepki Kuramı madde güçlüğü değerleri (b) ise -2.17 ile 2.34 arasında değişmektedir. b değerlerinin

böyle geniş bir aralıkta değişmesi sorulardaki güçlük dağılımının gruba uygun olduğu şeklinde yorumlanabilir. Ayırt edicilik değerleri (a) ise 0.52 ile 3.67 arasında değişmektedir. Ortalama ayırt edicilik değeri 1.56 olarak hesaplanmıştır. Genel olarak ayırt edicilik değerleri iyidir. Şans parametresi değerleri (c) 0.14 ile 0.24 arasında değişmektedir. 0.25'ten büyük değer olmaması soruların yanıtlarının şans eseri bulunma olasılıklarının beklenenden düşük olduğunu göstermektedir. Genel olarak madde parametrelerinin iyi olduğu görülmektedir.

Tablo 5*Asıl Uygulama Madde Analizi Sonuçları*

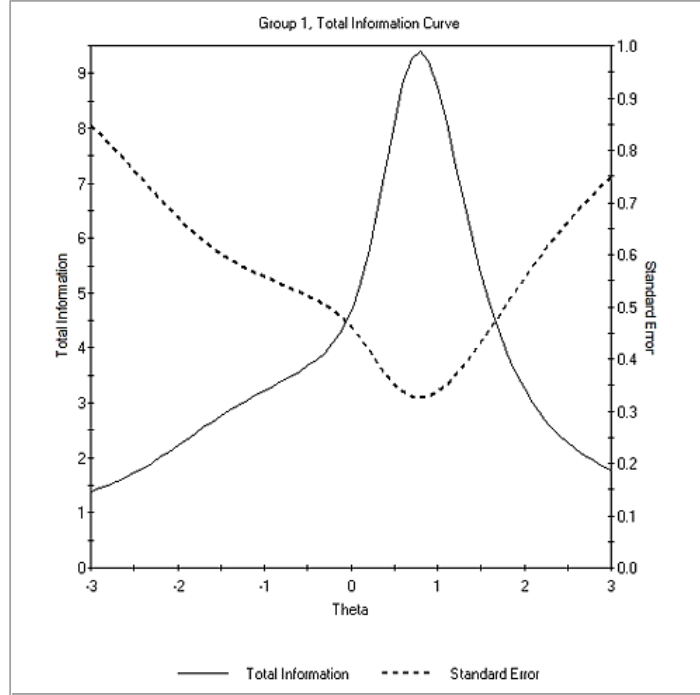
Soru No	Soru Alt Boyut	Soru Forması	Madde Güçlüğü (P)	Madde Güçlüğü (b)	Madde Ayırt ediciliği (a)	Şans Parametresi (c)
1	Algoritmik Düşünme	AU	.69	-.28	1.37	UD
2	Algoritmik Düşünme	AU	.76	-.62	1.77	UD
3	Kavramlar	ÇS	.33	1.40	1.27	.16
4	Kavramlar	ÇS	.27	2.34	.82	.14
5	Örüntüler	ÇS	.47	.71	.85	.16
6	Örüntüler	ÇS	.80	-2.17	.52	.20
7	Argümantasyon	AU	.31	.94	1.82	UD
8	Argümantasyon	ÇS	.43	.71	1.25	.15
9	Fen Okuryazarlığı	ÇS	.57	.45	.71	.24
10	Fen Okuryazarlığı	ÇS	.70	-.56	1.34	.17
11	Modelleme	ÇS	.37	1.81	.65	.16
12	Modelleme	ÇS	.41	1.09	1.13	.19
13	Kodlama	ÇS	.43	.86	2.26	.23
14	Tekno Mühendislik	ÇS	.40	.80	3.14	.20
15	Kodlama	ÇS	.43	.66	3.67	.20
16	Tekno Mühendislik	ÇS	.41	.72	2.39	.18

ÇS: Çoktan Seçmeli; AU: Açık uçlu; UD: Uygun Değil

Test Bilgi Fonksiyonu

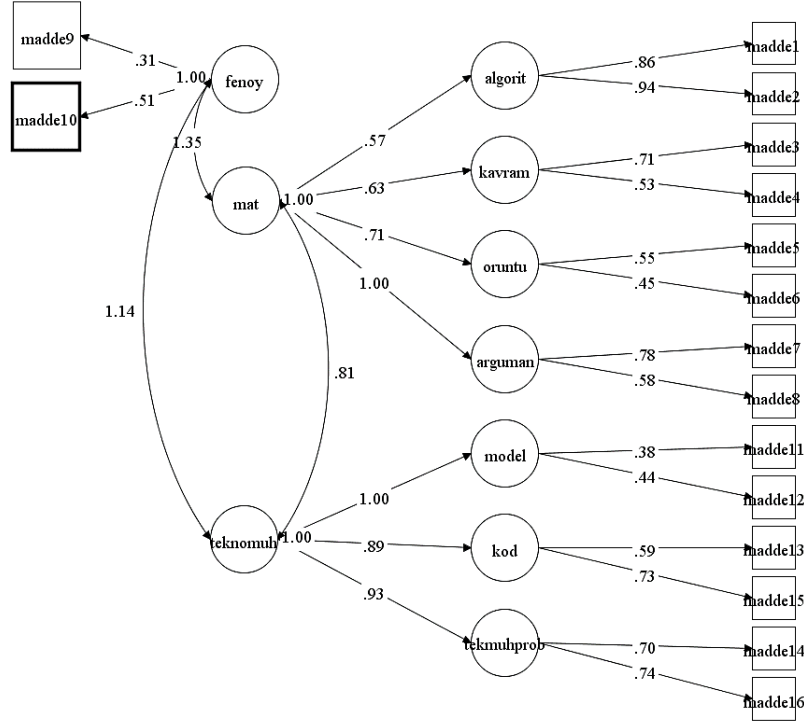
Asıl uygulama sonuçlarından Madde Tepki Kuramı ile elde edilen test bilgi fonksiyonu Şekil 1'de verilmektedir. Test bilgi fonksiyonu, test genelinde soruların ayırt ediciliğinin en fazla olduğu yetenek düzeyini ve testin en iyi hitap ettiği öğrenci grubunu göstermektedir. Geliştirilen test -0.50 ile 2.00 arasında yeteneğe sahip olan öğrenci grubu için daha az standart hataya sahiptir.

Şekil 1
Test Bilgi Eğrisi



Doğrulayıcı Faktör Analizi

Doğrulayıcı faktör analizi ile asıl uygulamada ölçülmesi hedeflenen yapının öğrenci yanıtları ile desteklenme düzeyi ile ilgili bulgular sunulmaktadır. İlk olarak Şekil 2'de verilen testin geliştirildiği sekiz faktörlü yapının öğrenci yanıtları ile uyumuna bakılmıştır. Sekiz faktörlü yapı için elde edilen değerler kabul edilebilir düzeyde bir uyum olduğunu göstermektedir (CFI = 0.924, TLI = 0.905, RMSEA = 0.053). Bu bulgu, 4. sınıflar için STEM yeterlikleri değerlendirme çerçevesi ile önerilen alt boyutlar temel alınarak geliştirilen bir testteki yapıların öğrenci yanıtları tarafından da desteklendiğini göstermektedir. Testin hedeflenen alt boyutları ayrı ayrı boyutlar olarak ortaya çıkmaktadır. Faktör yükleri incelendiğinde ise en düşük değer 0.306 olduğu görülmüştür. Alternatif bir model olarak 16 maddenin tek boyutlu olup olmadığı incelenmiş ve elde edilen uyum değerleri, tek boyutlu bir yapının desteklenmediğini göstermektedir (CFI = 0.820, TLI = 0.793, RMSEA = 0.078). Sonuç olarak, önerilen STEM değerlendirme çerçevesinin öğrenci yanıtları ile desteklenmektedir (bkz. Tablo 6).

Şekil 2*STEM Değerlendirme Çerçevesi Modeli***Tablo 6***Doğrulayıcı Faktör Analizi Uyum İndeksleri*

Uygulama	χ^2	df	χ^2/df	TLI	CFI	RMSEA (90% CI)
Asıl uygulama (8 Faktör)	150.966*	96	1.572	.905	.924	.053 (.036; .069)
Alternatif model (1 Faktör)	234.202*	104	2.252	.793	.820	.078 (.065; .092)

* $p < 0.05$ **Tartışma**

STEM eğitiminin ve STEM yeterliklerinin önemi dünya genelinde giderek daha fazla öne çıkmaktadır. STEM alanlarında akademik olarak başarılı olan öğrencilerin modern toplumlara öncülük etmeleri beklenmektedir. STEM alanlarında başarılı olmak, ülkelerin işgücü piyasasında küresel olarak rekabet edebilmeleriyle, ekonomik istikrar ve üstünlüğe sahip olmalarıyla ilişkilendirilmektedir (Bicer ve diğerleri, 2017; National Science and Technology Council, 2011). Bu sebeple,

lkeler eđitim hayatlarına STEM kariyerlerinde devam edebilecek đrenci sayısını arttırmayı hedeflemektedir (Carter, 2007). Bu da STEM eđitiminin olanaklarından daha fazla yararlanmak iin gerek okul mfredatlarında gerekse mfredat dıŐı alıŐmalarda STEM eđitiminin kalitesini artırmaya ynelik abalara (Moreno ve diđerleri, 2016) ve STEM baŐarısı ile iliŐkili olan faktrler hakkında daha fazla sayıda araŐtırma yapılmasına yol amaktadır (Bier ve diđerleri, 2014). STEM eđitiminin niteliđini deđerlendirmek, STEM baŐarısını incelemek, đrencilerin STEM yeterliklerini belirleyerek geliŐimlerini inceleyebilmek iin geerli ve gvenilir puanlar elde edilebilen lme aralarına ihtiya vardır. Ancak, STEM araŐtırmalarıyla ilgili temel sorunlardan biri, STEM yeterliklerini belirleyecek ve deđerlendirecek iyi tanımlanmıŐ bir deđerlendirme erevesine bađlı olarak lme aralarının geliŐtirilmemiŐ olmasıdır. Mevcut araŐtırmalarda, đrencilerin STEM baŐarısını araŐtırmak iin ulusal veya uluslararası sınavlardan elde ettikleri matematik veya fen puanlarını kullanılmaktadır (Bicer ve diđerleri, 2014; Bicer ve diđerleri, 2017; Han ve diđerleri, 2015; Ing, 2014). YerleŐtirme sınavları, uluslararası karŐılaŐtırma sınavları veya ders notları gibi diđer amalar iin geliŐtirilen baŐarı testlerinden alınan puanların kullanılması, STEM ile ilgili araŐtırmaların niteliđini sınırlamaktadır. Yakın zamanda Arıkan vd. (2022) tarafından sekizinci sınıflar dzeyinde geliŐtirilmiŐ bir STEM yeterlikleri deđerlendirme erevesi bulunmaktadır. İlkokul seviyesi iin STEM yeterlikleri deđerlendirme erevesini alan yazına kazandırmak amacı ile bu alıŐmada bahsi geen deđerlendirme erevesi drdnc sınıf dzeyine uyarlanmıŐtır.

Bu alıŐmada STEM deđerlendirme erevesi ampirik olarak test edilmiŐ ve đrenci verileri nerilen yapıyı 3 ana boyut ve 10 alt boyut ile desteklemiŐtir. Bu ereveyi uyarlarlarken, boyutların kendi iindeki ve birbirleri arasındaki iliŐkiler nce teorik olarak deđerlendirilmiŐ, ardından ampirik olarak test edilmiŐtir. erevenin yapısını test etmek iin rutin olmayan problem maddeleri geliŐtirilmiŐtir. Farklı STEM boyutları đrenciler iin eŐitli problem senaryoları sađlamıŐtır. Bu alıŐma ile akıl yrtme becerisine odaklanan btnleŐik ve ok boyutlu STEM deđerlendirme modeli dođrulanmıŐtır. Modelde, matematik ana boyutunun altında algoritmik dŐnme, kavramlar, rntleri tanıma ve argmantasyon alt boyutları yer alır. Fen ana boyutu, fen okuryazarlıđı alt boyutuna sahiptir. Teknoloji ve mhendislik ana boyutu, modelleme, tekno-mhendislik problemleri ve kodlama alt boyutlarından oluŐmaktadır. Elde edilen đrenci verileri uyarlanan modeli desteklediđinden, bu model baŐka alıŐmalarda drdnc sınıf đrencilerinin STEM yeterliklerini lmek iin kullanılabilir. Bu nedenle, erevenin uygulanabilirliđi, erevenin yerel bađlamıyla ve geliŐtirilen test ile sınırlı deđildir. ereve, uluslararası byk lekli deđerlendirmelerde tanımlanan beceriler gzden geirilerek geliŐtirildiđinden evrensel olarak da kullanılabileređi dŐnlmektedir.

Sonuç

Bu çalışmanın önemi, STEM alt boyutları arasındaki ilişkileri belirlemede, her bir alt boyutun tanımlamada ve bunlara dair maddelerin örneklenerek bir değerlendirme çerçevesi oluşturmasında yatmaktadır. Geliştirilen test, herhangi bir araştırmacı veya öğretmen tarafından STEM alanında öğrenci yeteneklerini değerlendirmek için kullanılabilir. Ayrıca, öne sürülen çerçeve kullanılarak öğretmenler veya araştırmacılar kendi özel testlerini geliştirebilirler. Bu çerçeve 4. ve 8. sınıf konularına bu sınıf seviyesindeki öğrencilerin verilerine dayalı olarak geliştirilip doğrulandığından testin içeriği bu sınıf seviyesindeki konuları kapsayacak şekilde genel tutulmuştur. Daha düşük veya daha yüksek kademedeki öğrenci grupları için aynı alt boyutlara dayalı testler geliştirmek mümkündür. Bu çerçeve kullanılarak lise öğrencileri için geliştirilebilecek testlerin benzer becerileri ilgili konu boyutları dikkate alarak kapsayan maddeler içermesi beklenir. Dördüncü sınıftan daha küçük ilköğretim öğrencilerine yönelik bir test geliştirmek için ise benzer bir yaklaşım kullanılabilir.

Burada ortaya konan STEM değerlendirme çerçevesi kullanılarak, STEM alt boyutları arasındaki bağıntılarını belirleme konusunda daha fazla sayıda çalışma yapılması beklenmektedir. Bu çalışmada ortaya konan değerlendirme çerçevesi sayesinde STEM yeterliklerini gelişimsel olarak izlemeyi amaçlayan araştırmalara olanak sağlanabilir. Ayrıca, burada tanımlanan boyutlara odaklanan yeni çalışmalar da bu boyutların ölçülmesi ve bu boyutlarla diğer faktörler arasındaki ilişkilerin ortaya konması açısından önemlidir.

Sınırlılıklar

Bu çalışmadaki en önemli sınırlılık her alt boyut için iki soru kullanılmasıdır. 4. sınıf öğrencilerine çok uzun bir test vermemek adına alt boyutlar için ikişer soru tercih edilmiştir. Başka çalışmalarda alt boyutlarda daha çok sayıda maddenin kullanıldığı testler ile bu yapı yeniden test edilebilir. Diğer bir sınırlılık ise pilot ve son çalışmanın örneklemeleri uygun örneklemeyle dayalı olarak seçilmiştir. Veriler araştırmaya katılmayı kabul eden okullardan edinilmiştir. Veriler istatistiksel analiz yapmak için yeterli olsa da, bulguların farklı kültürlerden daha büyük rastgele örneklemeye dayalı olarak tekrarlanması, bulguların genelleştirilebilmesi adına önemli kanıtlar sağlayabilir.

Etik Kurul İzin Bilgisi: *Bu araştırma, Muğla Sıtkı Koçman Üniversitesi İnsan Araştırmaları Etik Kurulunun 23/03/2018 tarihli 45 sayılı kararı ile alınan izinle yürütülmüştür.*

Yazar Çıkar Çatışması Bilgisi: *Yazarlar çıkar çatışması olmadığını beyan etmektedir.*

Yazar Katkısı: *Yazarlar çalışmaya eşit oranda katkı sağlamıştır.*

Destek: Bu çalışma TÜBİTAK 2218 projesi kapsamında elde edilen verilerden hazırlanmıştır.

Kaynakça

- Arıkan, S., Erkin, E., & Pesen, M. (2022). Development and validation of a STEM competencies assessment framework. *International Journal of Science and Mathematics Education*, 20, 1-24. <https://doi.org/10.1007/s10763-020-10132-3>
- Bennett, C. A., & Ruchti, W. (2014). Bridging STEM with mathematical practices. *Journal of STEM Teacher Education*, 49(1), 17–28. <https://doi.org/10.30707/JSTE49.1Bennett>
- Bicer, A., Capraro, R. M., & Capraro, M. M. (2017). Integrated STEM assessment model. *Eurasia Journal of Mathematics Science and Technology Education*, 13(7), 3959–3968. <https://doi.org/10.12973/eurasia.2017.00766a>.
- Bicer, A., Navruz, B., Capraro, R. M., & Capraro, M. M. (2014). STEM schools vs. non-STEM schools: Comparing students mathematics state based test performance. *International Journal of Global Education*, 3(3), 8–18.
- Blanton, M., & Kaput, J. (2011). Functional thinking as a route into algebra in the elementary grades. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 5–23). Berlin: Springer.
- Blum, W., & Ferri, R. B. (2009). Mathematical modelling: Can it be taught and learnt? *Journal of Mathematical Modelling and Application*, 1(1), 45–58.
- Broggy, J., O'Reilly, J., & Erduran, S. (2017). Interdisciplinarity and science education. In B. Akpan & K. Taber (Eds.), *Science Education. New directions in mathematics and science education* (pp. 81–90). Sense Publishers.
- Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. *Technology and Engineering Teacher*, 70(6), 5–9.
- Brown, W. (2015). *Introduction to algorithmic thinking*. Retrieved from <https://raptor.martincarlisle.com/Introduction%20to%20Algorithmic%20Thinking.doc>.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates, Inc.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Capraro, R.M., & Corlu, M.S. (2013). Changing views on assessment for STEM project-based learning. In R. M. Capraro, M. M. Capraro, & J. Morgan (Eds.), *STEM project-based learning: An integrated Science, Technology, Engineering, and Mathematics (STEM) approach* (2nd ed., pp. 109–118). Rotterdam, The Netherlands: Sense Publishers.
- Carter, L. (2007). Globalization and science education: The implications of science in the new economy. *Journal of Research in Science Teaching*,

- 45(5), 617–633. <https://doi.org/10.1002/tea.20189>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>.
- Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. Mason, OH: Cengage Learning
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23, 225–241.
- Doerr, H., & English, L. D. (2003). A modeling perspective on students' mathematical reasoning about data. *Journal for Research in Mathematics Education*, 34(2), 110–136.
- Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education*, 41(5), 605–618. <https://doi.org/10.1007/s11858-009-0184-6>.
- Enderson, M.C., & Ritz, J. (2016). STEM in general education: Does mathematics competence influence course selection. *Journal of Technology Studies*, 42(1), 30–41. <https://doi.org/10.21061/jots.v42i1.a.3>.
- English, L.D. (2015). STEM: Challenges and opportunities for mathematics education. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 4–18). Hobart, Australia: PME.
- Feinstein, N. (2011). Salvaging science literacy. *Science Education*, 95(1), 168–185.
- Fitzallen, N. (2015). STEM education: What does mathematics have to offer? In M. Marshman (Ed.), *Mathematics education in the margins. Proceedings of the 38th annual conference of the Mathematics Education Research Group of Australasia* (pp. 237–244). Sydney, Australia: Mathematics Education Research Group of Australasia (MERGA).
- Freudenrich, C., & Boyd, R. (2001). *How your brain works*. Howstuffworks. <http://www.howstuffworks.com/brain1.htm> adresinden 13 Aralık 2019 tarihinde alınmıştır.
- Guyer, R., & Thompson, N. A. (2014). *User's Manual for Xcalibre item response theory calibration software, version 4.2.2 and later*. Woodbury MN: Assessment Systems Corporation.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>.
- Han, S., Capraro, R., & Capraro, M. M. (2015). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of

- student factors on achievement. *International Journal of Science and Mathematics Education*, 13(5), 1089–1113. <https://doi.org/10.1007/s10763-014-9526-0>.
- Harwell, M., Moreno, M., Phillips, A., Guzey, S. S., Moore, T. J., & Roehrig, G. H. (2015). A study of STEM assessments in engineering, science, and mathematics for elementary and middle school students. *School Science and Mathematics*, 115(2), 66–74. <https://doi.org/10.1111/ssm.12105>.
- Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., & Urban- Lurain, M. (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: Forging a national network. *CBE—Life Sciences Education*, 10(2), 149–155.
- Ing, M. (2014). Can parents influence children's mathematics achievement and persistence in STEM careers? *Journal of Career Development*, 41(2), 87–103. <https://doi.org/10.1177/0894845313481672>.
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, 3(11), 1–11. <https://doi.org/10.1186/s40594-016-0046-z>.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. The Guilford Press.
- Lee, N. H., Yeo, D. J. S., & Hong, S. E. (2014). A metacognitive-based instruction for primary four students to approach non-routine mathematical word problems. *ZDM*, 46(3), 465–480. <https://doi.org/10.1007/s11858-014-0599-6>.
- MEB. (2016). *STEM Eğitimi Raporu*. Ankara: Millî Eğitim Bakanlığı- Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü (YEĞİTEK).
- Milgram, R. J. (2007). What is mathematical proficiency? In AH Schoenfeld (Ed.), *Assessing mathematical proficiency* (pp. 31–58). Cambridge University Press. <https://doi.org/10.1017/CBO9780511755378.007>.
- Miller, J. (2019). STEM education in the primary years to support mathematical thinking: Using coding to identify mathematical structures and patterns. *ZDM*, 51(6), 915–927.
- Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology*, 25(6), 889–897. <https://doi.org/10.1007/s10956-016-9643-3>.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. <http://timssandpirls.bc.edu/timss2019/frameworks/> adresinden 13 Aralık 2019 tarihine alınmıştır.
- National Council of Teachers of Mathematics. (2018). *Building STEM education on a sound Mathematical Foundation. Position statement*. Retrieved from <https://www.nctm.org/Standards-and-Positions/NCTM-Position-Statements/>.
- National Research Council. (2005). *How students learn: Mathematics in the*

- classroom*. The National Academies Press.
- National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. The National Academies Press.
- National Science and Technology Council. (2011). *The federal science, technology engineering, and mathematics (STEM) education portfolio*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/costem_federal_stem_education_portfolio_report.pdf
- Nga, N. T., Quỳnh, T. T. X., Uyên, N. P., & Trung, T.T. (2022). Một số nghiên cứu về năng lực STEM trên thế giới và đề xuất khung năng lực STEM cho học sinh phổ thông tại Việt Nam (An overview study on STEM competencies in the world and propose a STEM competency framework for high school students in Vietnam), *Tạp chí Giáo dục*, 22, 48-53.
- Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics*, 72(1), 39–60. <https://doi.org/10.1007/s10649-009-9181-5>
- Rumsey, C., & Langrall, C. W. (2016). Promoting mathematical argumentation. *Teaching Children Mathematics*, 22(7), 412-419.
- Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, E. A. (2014). A common measurement system for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation*, 40, 18–35. <https://doi.org/10.1016/j.stueduc.2013.11.005>
- Schmidt, W. H., & Houang, R. T. (2007). Lack of focus in the mathematics curriculum: A symptom or a cause? In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 65–84). Brookings Institution Press.
- Schumacker, R. E. & Lomax, R. G. (2004). *A beginners guide to structural equation modeling*. Lawrence Erlbaum Associates, Inc.
- Shulman, L. (2009). Assessment of teaching or assessment for teaching? In D.H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality*. Sage Publications.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press.
- Trung, T. T., Quỳnh, T. T. X., Uyên, N. P., & Nga, N. T. (2022). Xây dựng và chuẩn hóa công cụ đánh giá năng lực stem của học sinh trung học phổ thông tại thành Phố Hồ Chí Minh (Develop and standardize a stem competency assessment tool for high school students in Ho Chi Minh City). *Ho Chi Minh City University Of Education Journal Of Science*, 19(8), 1255-1270. [https://doi.org/10.54607/hcmue.js.19.8.3408\(2022\)](https://doi.org/10.54607/hcmue.js.19.8.3408(2022))
- Turner, E. E., Roth McDuffie, A., Bennett, A. B., Aguirre, J., Chen, M. K., Foote, M. Q., & Smith, J. E. (2022). Mathematical modeling in the elementary grades: Developing and testing an assessment. *International Journal of Science and Mathematics Education*, 20(7), 1387-1409.

- TÜSİAD. (2014). *STEM Alanında eğitim almış işgücüne yönelik talep ve beklentiler araştırması*. http://tusiad.org/tr/yayinlar/raporlar/item/download/7014_d28ffa2adda423c6d3852cc01c965993 adresinden alındı.
- Ullman, J. B. (2001). Structural equation modeling. In B. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp.653-771). Allyn & Bacon.
- van Borkulo, S., Chytas, C., Drijvers, P., Barendsen, E., & Tolboom, J. (2021, October). Computational thinking in the mathematics classroom: fostering algorithmic thinking and generalization skills using dynamic mathematics software. In *The 16th Workshop in Primary and Secondary Computing Education* (pp. 1-9).
- White, D. W. (2014). What is STEM education and why is it important. *Florida Association of Teacher Educators Journal*, 1(14), 1–9.
- Wong, G. K., Cheung, H. Y., Ching, E. C., & Huen, J. M. (2015, December). School perceptions of coding education in K-12: A large scale quantitative study to inform innovative practices. In *2015 IEEE international conference on teaching, Assessment, and learning for engineering (TALE)* (pp. 5-10). IEEE.
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K.R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide (NCEE 2012-4055)*, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Washington, DC. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/.
- Yıldırım, H., & Gelmez-Burakgazi, S. (2020). Research on STEM education studies in Turkey: A qualitative meta-synthesis study. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 50, 291-314. <https://doi.org/10.9779/pauefd.590319>

Ekler

Sorular

Soru 1. [Algoritmik Düşünme]

İki basamaklı bir sayının 999 ile çarpımından elde edilen sonucu kısa yoldan hesaplamak için alttaki adımlar izlenir.

1.adım: 999 ile çarpılan sayıdan 1 çıkarılır.

2.adım: 999'dan birinci adımda elde edilen sayı çıkarılır.

3.adım: Birinci ve ikinci adımda elde edilen sonuçlar yan yana yazılır.

Örneğin, 15×999 çarpımı

1) $15 - 1 = 14$

2) $999 - 14 = 985$

3) 14985'tir.

28×999 işleminin sonucunu yukarıda verilen kısa yoldaki basamakları

1)
2)

kullanarak bulunuz.

Soru 4. [Kavramlar ve Prensipler]

1'den başlayarak kendisi dışında tam olarak bölündüğü sayıların toplamı yine kendisine eşit olan sayıya "mükemmel sayı" denir.

Örneğin, 6 sayısı kendisi dışında 1, 2 ve 3 sayılarına tam olarak bölündüğü ve $1+2+3=6$ olduğu için 6 sayısı mükemmel sayıdır. (6 sayısı 4 ve 5 sayılarına tam olarak bölünmez).

Ama, 10 sayısı kendisi dışında 1, 2 ve 5 sayılarına tam olarak bölündüğü ama $1+2+5=8$ olduğu için 10 sayısı mükemmel sayı değildir. (10 sayısı 3, 4, 6, 7, 8 ve 9 sayılarına tam olarak bölünmez). Buna göre, altta verilen sayılardan hangisi mükemmel sayıdır?

A) 30 B) 28 C) 15 D) 11

Soru 5. [Örüntüler]

Birinci Sayı	İkinci Sayı	Üçüncü Sayı	Dördüncü Sayı	Beşinci Sayı				Dokuzuncu Sayı
3	2	6	2	2				?

Üstteki sayı dizisini elde etmek için birinci sayı ve ikinci sayı çarpılır. Sonuç bir basamaklı ise o sayı, iki basamaklı ise sonucun birler basamağı dizinin üçüncü sayısı olarak yazılır. Benzer şekilde ikinci ve üçüncü sayı çarpılarak dördüncü sayı bulunur. Bu şekilde dizi devam etmektedir.

Örneğin, ilk olarak birinci sayı olan 3 ve ikinci sayı olan 2 çarpılarak 6 sayısı elde edilir. 6 sayısı bir basamaklı bir sayı olduğu için dizinin üçüncü sayısı 6 olarak yazılır.

Ardından ikinci sayı olan 2 ile üçüncü sayı olan 6 çarpılarak 12 sayısı elde edilir. 12 sayısı iki basamaklı bir sayı olduğu için dizinin dördüncü sayısı 2 olarak yazılır.

Dizi bu şekilde devam ettiğine göre dokuzuncu sayı kaç olur?

- A) 2 B) 3 C) 6 D) 8

Soru 7. [Argümantasyon]

Tablo: Mağazadaki Ürünlerin Fiyat Değişiklikleri

Ürün	Eski Fiyat (TL)	Yeni Fiyat (TL)
Ayakkabı	100	70
Bot	500	400
Ceket	2000	2300
Dizlik	50	40
Elbise	2000	1900

Tabloda listelenen 5 ürünün her birinden birer adet almak için mağazaya gelen Melis fiyatların değiştiğini görüyor. “Bugün şanslı günüm. Alacağım ürünlerden dördü indirimde girmiş. Bir ürünün fiyatı artmış. Daha az ödeme yapacağım” diye düşünüyor. Melis daha az ödeme yapacağı konusunda haklı mıdır?

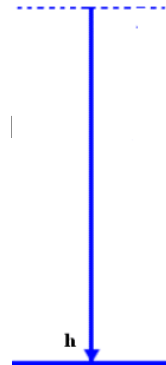
Yanıtınızda “haklı” veya “haklı değil” diye belirterek, nedenini açıklayınız.

Soru 10. [Fen okuryazarlığı]

Bir kuleden bırakılan bir topun ne kadar sürede yere çarptığını bilirse bırakıldığı kulenin yüksekliğini bulabiliriz.

Bu işlem için düşme süresini kendisi ile çarpıp, çıkan sonucu da çarpabiliriz.

Örneğin, 2 saniyede yere düşen bir top $2 \times 2 = 4$, $4 \times 5 = 20$ metre yüksekliğinde bir kuleden bırakılmıştır.



Buna göre, 3 saniyede yere düşen bir top kaç metre yüksekliğinde bir kuleden bırakılmıştır?

- A) 25 B) 27 C) 45 D) 125

Soru 11. [Modelleme]

Farklı ülkelerdeki kaza sayıları ile gelir ve sıcaklık arasındaki ilişki inceleniyor.

Ülkelerin gelirleri arttıkça kaza sayısının azaldığı, sıcaklıkları arttıkça kaza sayısının arttığı görülüyor. Hangi ülkede kaza sayısının daha yüksek olmasını beklersiniz?

	Ülkenin Adı	Gelir (TL)	Sıcaklık (C°)
A)	Astanya	2000	28
B)	Buanda	2000	25
C)	Ceristan	3000	28
D)	Dastarya	3000	25

Soru 13. [Kodlama]

Alttaki satırlarda matematiksel komutlar ve bu komutlara göre yapılması gereken işlemler listelenmiştir. Bir kutuda yer alan sayılardan biri çekiliyor ve o sayıya göre alttaki komutlar izlenerek hangi renk ışık yanacağı belirleniyor.

Satır	Komut
1	Eğer "çekilişte çıkan sayı" > 5
2	Bu sayıyı 2 ile çarp ve 5. satıra git
3	Değilse
4	Bu sayıyı 3 ile çarp ve 5. satıra git
5	Elde ettiğin sayıdan 7 çıkar ve alttaki satırlardan uygun olanına git.
6	Eğer sonuç < 10
7	Ekran kırmızı renk olsun
8	Eğer 9 < sonuç < 15
9	Ekran mavi renk olsun
10	Eğer 14 < sonuç < 20
11	Ekran sarı renk olsun
12	Eğer sonuç > 19

Çekilişte çıkan sayı 4 ise ekran ne renk olur?

- A) Kırmızı
- B) Mavi
- C) Sarı
- D) Yeşil

Soru 16. [Tekno mühendislik]

Bir dondurmacı dondurma fiyatına yapacağı her 1 TL'lik zammın günde 10 tane daha az dondurma satılmasına neden olacağı öngörüyor. Mevcut durumda dondurmayı 4 TL'den satıyor. Günde 100 dondurma satarak 400 TL gelir elde ediyor.

Dondurmacı alttaki tabloyu doldurarak en fazla gelir elde edeceği dondurma fiyatını bulmak istiyor. İlk iki satırı dolduruyor. En fazla gelir elde edeceği dondurma fiyatı kaçtır? Diğer satırları doldurarak bulabilirsiniz.

Dondurma Fiyatı (TL)	Satılan Dondurma Adeti	Gelir (TL)
4 TL	100	$4 \times 100 = 400$ TL
5 TL	90	$5 \times 90 = 450$ TL

- A) 6
- B) 7
- C) 8
- D) 9



Adaptation of STEM Competencies Assessment Framework for 4th-grade Level

Serkan ARIKAN¹, Melek PESEN² & Emine ERKTIN³

Abstract

As observed in the studies encountered in the literature, when students' STEM competencies are evaluated, scores from large-scale mathematics and/or science tests are used instead of scores from interdisciplinary measurement tools specifically developed to measure STEM competencies. In order to develop an assessment tool to determine students' STEM competency scores, firstly, a STEM competencies assessment framework is needed because no assessment framework has been found to assess STEM competencies at the primary school level. For this reason, the current study aimed to adapt the STEM competencies assessment framework to the 4th-grade level, which was developed by Arıkan et al. (2022) and presented validity evidence at the 8th-grade level. In this context, the reliability and validity evidence of the scores obtained from a measurement tool on the assessment framework adapted to the 4th-grade level are presented. The assessment framework presented in the current study was prepared by considering the multidimensional and integrated structure of STEM competencies- science, technology, engineering, and mathematics. The factor structure was evaluated by confirmatory factor analysis using data collected from 4th-grade students. The items in the developed measurement tool were analyzed at the item level using Item Response Theory. The findings show that the adapted STEM assessment framework can be used for 4th graders.

Article Details

Research Article

Received
10/02/2023

Accepted
12/10/2023

Published
15/01/2024

Key words

STEM
assessment
framework,
Non-routine
problems,
Confirmatory
factor analysis,
Item analysis

¹Boğaziçi University, ORCID: 0000-0001-9610-5496, serkan.arikan1@boun.edu.tr

²Boğaziçi University, ORCID: 0000-0003-1167-0707, melek.pesen@boun.edu.tr

³Boğaziçi University, ORCID: 0000-0002-9428-7115, erktin@boun.edu.tr

Suggested Citation:

Arıkan, S., Pesen, M., & Erkin, E., (2024). Adaptation of STEM competencies assessment framework for 4th-grade level. *Pamukkale University Journal of Education [PUJE]*, 60, 201-225. <https://doi.org/10.9779/pauefd.1249861>

Introduction

Importance of STEM Education

The concern about having a place in the global economy has led countries to take action to prioritize STEM (Science, Technology, Engineering, Mathematics) education in their education policies (Kelly & Knowles, 2016). Being successful in STEM fields is linked with being competitive in the labor market and having economic strength and supremacy (Biter et al., 2017; National Research Council, 2005; 2011; National Science and Technology Council, 2011). These efforts have transformed STEM education into a popular field and brought it to the focus of many researchers, politicians, teachers, and administrators (Biter et al., 2017; Brown et al., 2011; Capraro & Corlu, 2013). In Turkey, the need for individuals trained in STEM fields has been emphasized in the literature, various educational reports, and the road maps of the Ministry of National Education (Ministry of National Education [MoNE], 2016; Türk Sanayicileri ve İş İnsanları Derneği [TÜSİAD], 2014; Yıldırım & Gelmez-Burakgazi, 2020).

By its very nature, STEM education reflects an interdisciplinary teaching and learning approach in which science, technology, engineering, and mathematics overlap and integrate (Braggy et al., 2017). Mathematics is crucial for understanding scientific concepts, scientific ways of thinking, and scientific concepts form various contexts for mathematics. Engineering and technology are based on mathematics and its applications. Research shows that the brain performs best when it works in situations where mathematics, science, and related disciplines coexist (Biçer et al., 2017; Freudenrich & Boyd, 2001). STEM education aims to develop students' critical thinking skills so that they can solve problems effectively in all STEM fields (White, 2014).

Assessment of STEM Competencies

When assessment and evaluation are used correctly in teaching, they increase the quality of instruction and contribute to learning (Shulman, 2009). In order to improve the quality of STEM education, STEM competencies should be systematically assessed and monitored (Saxton et al., 2014). Formative and summative assessment tools that can be used to measure STEM competencies are essential for more qualified STEM education. Formative assessment incorporates instant feedback provided at every stage of the learning process. Summative assessment, on the other hand, reveals the skills that students have and provides the opportunity to monitor their development over time. In the literature, there are a limited number of studies on increasing the effectiveness of formative and summative assessment in STEM education (Capraro & Corlu, 2013; Haudek et al., 2011).

The most important problem with assessments of STEM competencies is the need for an interdisciplinary assessment framework. Many studies take students' mathematics or science scores from large-scale exams and use these scores as STEM achievement indicators. However, these assessment tools are not designed to measure students' STEM competencies. The interdisciplinary nature of STEM is missing in these assessment tools. Han et al. (2015) used the Texas Assessment of Knowledge and Skills test to investigate the relationship between STEM education and mathematics achievement. Likewise, Ing (2014) used the mathematics scores of the Longitudinal Study of American Youth Scale to measure STEM achievement. The fact that mathematics and/or science scores were taken from tests that were not particularly developed for STEM competencies shows the lack of a well-defined multidimensional STEM assessment framework. The establishment of a framework for assessing STEM competencies is expected to guide the development and validation of instruments assessing these competencies. There is a need to develop a STEM assessment model that centers on both specific content areas and the interdisciplinary nature of STEM fields (Harwell et al., 2015).

In one of the studies designed to measure STEM competencies, Saxton et al. (2014) worked on the development of a common STEM measurement system. In this study, which emphasized a strong theoretical structure, the importance of the connections between STEM dimensions was underlined. However, the framework presented in that study was not empirically tested, only aiming to conceptualize the assessment system as an initial stage without providing an assessment tool. In another study, Biçer et al. (2017) stated that an integrated STEM assessment model is needed to see students' competencies in the STEM field. They reported the correlation between students' mathematics and science scores on the Texas Assessment of Knowledge and Skills test and stated that this could serve as an adequate STEM assessment model until technology and engineering fields are included in the assessment tools. Harwell et al. (2015) tried to measure mathematics, science and technology skills by using engineering as a tool. Recently, several studies have evaluated high school students on STEM competencies through self-assessment (Nga et al., 2022; Trung et al., 2022). The framework prepared by these studies includes 5 dimensions (information gathering and processing, information management and use, solution implementation, technical security, and community sharing) and 15 behavioral indicators.

In order to assess STEM competencies, Arikan et al. (2022) developed an assessment framework suitable for the 8th-grade level and presented the validity and reliability evidence of a measurement tool prepared based on this framework. In this framework, sub-dimensions related to the fields of science, technology, engineering and mathematics

(physics, chemistry, biology, modeling, technology, and engineering problems, coding, algorithmic thinking, concepts, patterns, and argumentation) were defined and the relationships between them were determined. However, the need for a STEM assessment framework suitable for the 4th-grade level still pertains since it is crucial to measure the STEM competencies of students as they move from primary to secondary school.

STEM Competencies Assessment Framework

Although some disciplines stand out in STEM approaches where interdisciplinary structure is emphasized (English, 2015; Fitzallen, 2015), the emphasis on mathematics in many STEM-related studies is striking. This emphasis is based on the effectiveness of mathematics skills in understanding STEM competencies and predicting STEM-related career choices and the importance of mathematical literacy in solving STEM problems (Enderson & Ritz, 2016). The requirement of a strong mathematical foundation for an integrative STEM approach has been emphasized (National Council of Teachers of Mathematics [NCTM], 2018). Mathematics is central to STEM disciplines as it serves as an important language for other STEM disciplines (Bennett & Ruchti, 2014; English, 2015; Schmidt & Houang, 2007) and as a necessary tool to solve problems encountered in these disciplines (Enderson & Ritz, 2016; English, 2015). Mathematics plays an important role in every STEM education program and "The mathematics that students learn in school includes content and thinking skills that can be used as tools to tackle integrative STEM problems" (NCTM, 2018, p. 3). For all these reasons, the STEM assessment framework also needs to have the foundations to represent STEM problems involving mathematics as a tool. Although some disciplines stand out in STEM approaches where the interdisciplinary structure is apparent (English, 2015; Fitzallen, 2015), the emphasis on mathematics in many STEM-related studies is conspicuous. This emphasis is based on the effectiveness of mathematics skills in understanding STEM competencies and predicting STEM-related career choices and the importance of mathematical literacy in solving STEM problems (Enderson & Ritz, 2016). Mathematics is central to STEM disciplines as it serves as an important language for other STEM disciplines (Bennett & Ruchti, 2014; English, 2015; Schmidt & Houang, 2007) and as a necessary tool to solve problems encountered in these disciplines (Enderson & Ritz, 2016; English, 2015). Mathematics plays an important role in every STEM education programme and the mathematics that students learn in school includes content and thinking skills that can be used as tools to tackle integrative STEM problems (NCTM, 2018). For all these reasons, the STEM assessment framework also needs to have the foundations to represent STEM problems involving mathematics as a tool. The STEM

assessment framework put forward by Arikan et al. (2022) also emphasized an integrative approach but treated mathematics as a tool.

While creating the STEM assessment framework, it is necessary to define the cognitive domain. Reasoning skill, one of the cognitive domains defined in the TIMSS (Trends in International Mathematics and Science Study) 2019 Assessment Framework, is compatible with the nature of STEM. According to TIMSS 2019, reasoning skills are defined as skills that "go beyond solving routine problems to include unfamiliar situations, complex contexts, and multi-step situations" (Mullis & Martin, 2017, p. 22). TIMSS emphasizes unfamiliarity when defining non-routine problems in reasoning skills. In contrast to non-routine problems, problems that students can encounter in any textbook or course and demand them to use a known algorithm are called routine problems. According to these definitions, non-routine problems are problems that require a student's reasoning, are meaningful for students but involve situations that they do not repeatedly encounter in daily life (Milgram, 2007). Another distinctive aspect of non-routine problems is that there is no anticipated and repeatedly taught approach; there may be more than one solution (Elia et al., 2009; Lee et al., 2014; Pantziara et al., 2009; Woodward et al., 2012). The STEM assessment framework by Arikan et al. (2022) also suggested that the problems to be created should be constructed from non-routine problems covering unfamiliar situations.

Sub-dimensions of STEM fields should be defined to measure STEM competencies. The STEM assessment framework developed by Arikan et al. (2022) in accordance with the 8th-grade level states that the mathematics dimension can be measured with algorithmic thinking, concepts, patterns, and argumentation; the science dimension with physics literacy, chemistry literacy, biology literacy; and the technology and engineering dimension with modeling, techno-engineering problems, and coding sub-dimensions.

Taking the need for an assessment framework that can be used to assess STEM competencies at the fourth-grade level into consideration, the aim of the current study was to adapt the STEM competencies assessment framework proposed by Arikan et al. (2022) to the 4th-grade level and to assess the reliability and validity of the scores obtained from a measurement tool developed based on this assessment framework. When adapting the assessment framework, a structure reflecting the multidimensional structure of STEM, to the fourth-grade level, the use of mathematics as a tool, and the focus on non-routine problems were maintained. The research question of this study was "Are the dimensions defined within the scope of the STEM competencies assessment framework adapted to the 4th-grade level and the expected relationships between the dimensions supported by student responses?".

Method

Participants

The participants were 4th-grade students who voluntarily participated in the study. Within the scope of the study, data were collected at two different times: pilot and main study. A total of 95 students from a school in Istanbul participated in the pilot study. 205 students from two different cities and six schools participated in the main study. 43% of the participants in the pilot study were female, 57% of the participants were male. 53% were female and 47% were male in the main study. Ethics committee permissions were obtained for the study.

Data Collection Tool

In this study, the STEM competencies assessment framework proposed by Arıkan et al. (2022) was first adapted to the fourth-grade level, and then a measurement tool was developed based on the sub-dimensions that emerged. The sub-dimensions of the STEM competencies framework, which determines the items created by considering the fourth-grade level, are shown in Table 1. The mathematics dimension consists of algorithmic thinking, concepts, patterns, and argumentation sub-dimensions as suggested by Arıkan et al (2022). At the eighth-grade level, science literacy was divided into three sub-dimensions: physics, chemistry, and biology literacy, on the other hand, at the fourth-grade level, all of them were combined into a single dimension as science literacy. In the technology and engineering dimension, the dimensions of modeling, coding, and techno-engineering problems are based on the proposed structure. As Arıkan et al. (2022) emphasized, in order to measure STEM skills, items were prepared based on non-routine problems that require reasoning skills instead of problems that students are familiar with. These sub-dimensions are explained in detail below. Sample questions representing these dimensions are also given at the end of the article. This study was conducted with the permission of Muğla Sıtkı Koçman University Human Research Ethics Committee (dated 23/03/2018 and numbered 45).

Table 1

Sub-dimensions of the STEM Competences Environment

Mathematics	Science	Technology and Engineering
Algorithmic Thinking	Science Literacy	Modelling
Concepts	Coding	

Patterns

Techno-Engineering
ProblemsArgumentation

Algorithmic Thinking

Algorithmic thinking is defined by Brown (2015) as "the ability to understand, execute, evaluate and create algorithms" (p.1). An algorithm contains step-by-step sequential instructions that are clear, and executable (Brown, 2015). Algorithmic thinking is often associated with computational thinking and thus computer science (van Borkulo et al., 2021).

Concepts

Concepts are generally related to students' knowledge and application of mathematical concepts and principles in the curriculum. In the STEM assessment framework, concepts are related to students' literacy and problem-solving skills within the framework of new definitions that they are not familiar with based on the concepts they ought to know.

Patterns

Pattern recognition involves being able to see the existing order and mathematical structure and being able to derive the necessary inferences and/or rules from this structure (Blanton & Kaput, 2011; Miller, 2019). In pattern recognition questions, students are expected to recognize the number or shape pattern and find the terms that follow the pattern (Arikan et al., 2022).

Argumentation

Argumentation is related to the creation of mathematical statements and arguments in a competent way and the evaluation of the accuracy of the arguments created by others (Rumsey & Langrall, 2016). In the process of mathematical argumentation, the actions of producing new knowledge and persuading others by providing justifications also take place (Rumsey & Langan, 2016). Toulmin (2003) stated that an argument basically consists of a claim, data, and justification. In argumentation problems within the STEM assessment framework, a claim is presented to students, and students are asked to evaluate the claim according to the data given and to justify their decisions mathematically (Arikan et al., 2022).

Science Literacy

Science literacy is important for better analysis and evaluation of scientific texts and situations in modern life and for better reasoning in decision-making processes (Feinstein, 2011). The STEM assessment

framework is based on science literacy skills that can be associated with mathematics in science subjects. In the items prepared, students are expected to analyze the short scientific texts presented to them, to reveal the relationships and to reach conclusions in line with the data given (Arikan et al., 2022).

Modeling

Mathematical models refer to the use of certain operations, relationships, and rules to describe, explain and predict a system (Doerr & English, 2003). Modeling is defined as the translation of scenarios that can be encountered in real life into mathematical expressions (Blum & Ferri, 2009; Turner et al., 2022). Within the framework of STEM assessment, students are expected to determine the best model using the given data for the modeling sub-dimension (Arikan et al., 2022).

Techno-Engineering Problems

Techno-engineering problems are presented to students in the context of optimization problems. In this context, within the framework of STEM assessment, students are expected to consider various possibilities based on the information provided and create the most appropriate response within the constraints presented (Arikan et al., 2022).

Coding

Coding is seen as an important field that requires numerical thinking along with programming skills, prepares students for engineering careers and develops algebraic thinking skills (Miller, 2019; Wong et al., 2015). As a 21st century skill, numerical thinking includes abstraction and the ability to recognize patterns, decompose, and create algorithms (Miller, 2019). Coding in STEM assessment framework dimension was presented to the students with questions requiring the execution of the given code. Students are expected to use the available input, run the code by evaluating the correctness of the expressions and find the desired output in the context of primary school mathematics (Ankara et al., 2022).

Using the STEM assessment framework defined here, applicability was also considered and two items from each dimension were prepared and shown in Table 2.

Table 2

Pilot and Main Study Specification Table

Sub-Dimensions	Items
Algorithmic Thinking	Q1, Q2
Concepts and Principles	Q3, Q4
Patterns	Q5, 56

Argumentation	Q7, Q8
Science Literacy	Q9, Q10
Modeling	Q11, Q12
Coding	Q13, Q15
Techno-Engineering	Q14, Q16
<hr/>	
Total Number of Items	16
<hr/>	

Data Analysis

This section describes the data analysis methods on internal consistency, item analysis, and confirmatory factor analysis. Reliability was reported using Cronbach's alpha value for the internal consistency of the items. If the alpha value is less than 0.70, internal consistency is problematic; 0.70 to 0.80 is acceptable; 0.80 to 0.90 is good; and above 0.90 is excellent (Cicchetti, 1994). SPSS 22.0 was used to calculate the alpha value.

The item analysis of the pilot study was conducted based on the Classical Test Theory (CTT), considering the sample size (Hambleton & Jones, 1993). In order to identify and revise problematic items, if any, item difficulty and item discrimination values were calculated. Item difficulty shows the percentage of correct answers and varies between 0 and 1. A low value indicates that an item is difficult, while a high value indicates that an item is easy for the sample (Crocker & Algina, 2008). Item discrimination shows the degree to which an item distinguishes high and low-achieving students from each other. Item discrimination ranges between -1 and 1 and values higher than 0.30 are desired (Crocker & Algina, 2008). In this study, corrected item-total correlation (R_{it}) was used as an item discrimination index. IRT analyses were performed with ITEMAN 4.4 software.

The item analysis of the main study was performed with Item Response Theory (IRT). While the item parameters of open-ended questions were estimated with the Generalized Partial Credit method, a three-parameter logistic model (b , a , and c values) was used for multiple-choice questions, in which item difficulty, discrimination, and guessing parameter were calculated (Guyer & Thompson, 2014). In IRT, item difficulty, b , is defined as the skill level required to answer the item correctly with a 50% probability when the guessing parameter is zero. Item difficulty generally ranges between -3 and +3, with high values indicating that the question is difficult and low values indicating that the question is easy (Guyer & Thompson, 2014). Discrimination is estimated by calculating the slope of a value at the point corresponding

to the b value in the item characteristic curve. The discrimination value is expected to be high. The guessing parameter, c, reflects the probability of answering the question correctly with a chance factor. It is the probability that students with very low skill levels will answer the question correctly. The expected value for the guessing parameter is the number of options divided by 1. When the guessing parameter is higher than the expected value, it is said that some of the options in the question are easily eliminated and the chance factor comes to the fore in answering the question (Guyer & Thompson, 2014). In open-ended questions scored as incorrect, partially correct, and fully correct (0-1-2), discrimination parameter (a) and two different difficulty parameters (b1 and b2) are estimated. b1 indicates the skill required to move from 0 to 1 point, and b2 indicates the skill required to move from 1 to 2 points. Item Response Theory analyses were conducted with IRTPRO 4.20 software (Cai et al., 2011).

Confirmatory factor analysis was used to determine the compatibility of the model proposed within the STEM assessment framework with the student responses in the prepared measurement tool (Byrne, 1998; Kline, 1998; Schumacker & Lomax, 2004). Confirmatory factor analysis (CFA) is used to provide evidence of construct validity by evaluating the factor structures of cognitive and psychological constructs (Distefano & Hess, 2005). Covariance matrix obtained from the student responses and from the proposed model are expected to be similar (Ullman, 2001). Whether the model and the data are similar or not is decided by the goodness of fit indices. Among these goodness of fit indices, Root Mean Square Error of Approximation (RMSEA) value of 0.08 and below (Byrne, 1998), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values of 0.95 and above (Ullman, 2001) indicate that the model is good. In the current study, CFA was performed using MPLUS 7.4 software.

Findings

Internal Consistency

Cronbach Alpha values for the reliability of the data obtained in the pilot and main study are given in Table 3. Since the reliability values obtained in both studies were around 0.70, the reliability of the data is acceptable.

Table 3

Internal Consistency Values of Test Scores

Study	Number of People	Number of Questions	Cronbach Alpha
Pilot	95	16	.69
Main	205	16	.75

Item Analysis

Pilot Study

Item analysis results for the pilot study are given in Table 4. In the pilot study, 16 questions from 8 sub-dimensions were used. Item difficulty values ranged between 0.26 and 0.88. The average item difficulty was calculated as 0.47. These values show that the difficulty of the test is appropriate for the group. Item discrimination values ranged between 0.07 and 0.51. The average discrimination value was calculated as 0.26. Although these values were positive and there was no negative discrimination value, the discrimination values of six questions were less than 0.20 and the discrimination values of two questions were between 0.20 and 0.30. Due to the purpose of the pilot study, necessary revisions were made by maintaining content validity.

Table 4

Pilot Study Item Analysis Results

Item #	Item Sub-Dimension	Item Format	Item Difficulty (p)	Item Discrimination (Rir)
1	Algorithmic Thinking	OE	.83	.36
2	Algorithmic Thinking	OE	.86	.19*
3	Concepts	MC	.40	.36
4	Concepts	MC	.30	.07*
5	Patterns	MC	.38	.11*
6	Patterns	MC	.88	.08*
7	Argumentation	OE	.44	.32
8	Argumentation	MC	.41	.32
9	Science Literacy	MC	.65	.24*
10	Science Literacy	MC	.76	.51
11	Modeling	MC	.54	.26*
12	Modeling	MC	.37	.10
13	Coding	MC	.26	.30
14	Techno-Engineering	MC	.48	.30
15	Coding	MC	.31	.18*
16	Techno-Engineering	MC	.34	.40

Notes. MC: Multiple Choice; OE: Open-Ended; *Rir < 0.30

Main Study

The item analysis results of the main study are given in Table 5. In the main study, 16 questions from 8 sub-dimensions were used. The item difficulty values (p) of these questions calculated according to the classical test theory vary between 0.27 and 0.80. Item Response Theory

item difficulty values (b) ranged between -2.17 and 2.34. The fact that b values varied in such a wide range can be interpreted that the difficulty distribution of the questions was appropriate for the group. The discrimination values (a) vary between 0.52 and 3.67. The average discrimination value was calculated as 1.56. In general, discrimination values are good. Guessing parameter values (c) vary between 0.14 and 0.24. Nonexistence of value greater than 0.25 indicates that the probability of finding the correct answers by chance was lower than expected. In general, the item parameters were good.

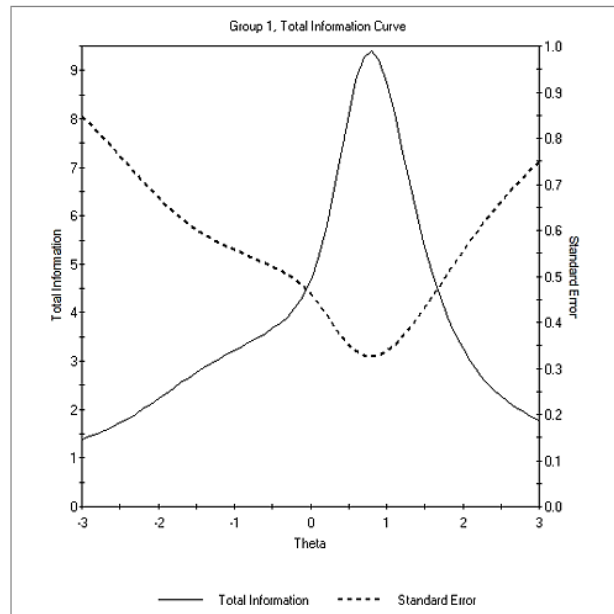
Table 5*Main Study Item Analysis Results*

Item #	Item Sub-Dimension	Item Format	Item Difficulty (P)	Item Difficulty (b)	Item Discrimination (a)	Guessing Parameter (c)
1	Algorithmic Thinking	OE	.69	-.28	1.37	NA
2	Algorithmic Thinking	OE	.76	-.62	1.77	NA
3	Concepts	MC	.33	1.40	1.27	.16
4	Concepts	MC	.27	2.34	.82	.14
5	Patterns	MC	.47	.71	.85	.16
6	Patterns	MC	.80	-2.17	.52	.20
7	Argumentation	OE	.31	.94	1.82	NA
8	Argumentation	MC	.43	.71	1.25	.15
9	Science Literacy	MC	.57	.45	.71	.24
10	Science Literacy	MC	.70	-.56	1.34	.17
11	Modeling	MC	.37	1.81	.65	.16
12	Modeling	MC	.41	1.09	1.13	.19
13	Coding	MC	.43	.86	2.26	.23
14	Techno-Engineering	MC	.40	.80	3.14	.20
15	Coding	MC	.43	.66	3.67	.20
16	Techno-Engineering	MC	.41	.72	2.39	.18

Notes. MC: Multiple Choice; OE: Open-Ended; NA: Not Appropriate

Test Information Function

The test information function for the main study is given in Figure 1. The test information function shows the ability level at which the questions are most discriminating and the student group for which the test works best. The main test had less standard error for the group of students with an ability between -0.50 and 2.00.

Figure 1*Test Information Curve***Confirmatory Factor Analysis**

The findings on how the constructed framework was supported by students' responses were obtained through confirmatory factor analysis. Firstly, the fit of the eight-factor structure of the test with the student responses was analyzed as shown in Figure 2. The values for the eight-factor structure indicate an acceptable level of fit (CFI = 0.924, TLI = 0.905, RMSEA = 0.053). This result shows that the constructs in the developed test were also supported by student responses. The targeted sub-dimensions of the test appear as separate dimensions. When the factor loadings were analyzed, it was seen that the lowest value was 0.306.

As an alternative model, the one-factor model was examined and the fit values obtained showed that a unidimensional structure was not supported (CFI = 0.820, TLI = 0.793, RMSEA = 0.078). Thus, the proposed STEM assessment framework was supported by student responses (see Table 6).

Figure 2
STEM Assessment Framework Model

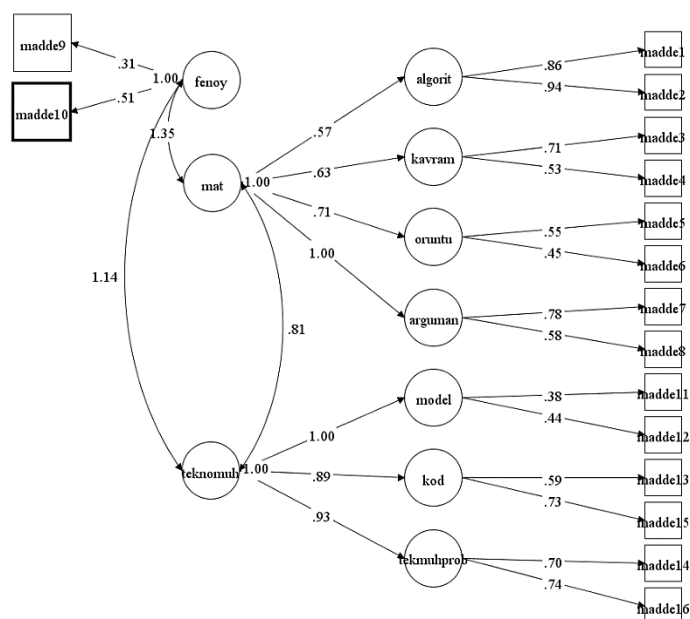


Table 6
Confirmatory Factor Analysis Fit Indices

Study	χ^2	df	χ^2/df	TLI	CFI	RMSEA (90% CI)
Main Study (8 Factors)	150.966 *	96	1.572	.905	.924	.053 (.036; .069)
Alternative model (1 Factor)	234.202 *	104	2.25 2	.793	.820	.078 (.065; .092)

* $p < 0.05$

Discussion

The importance of STEM education and STEM competencies have become increasingly prominent worldwide. The prospective leaders of modern societies are predicted to come from among the students who are academically successful in STEM fields. Being successful in STEM fields helps countries to compete globally in the labor market and to achieve economic strength and dominance (Bicer et al., 2017; National Science and Technology Council, 2011). Many countries aim to increase the number of students who can continue their education in STEM careers (Carter, 2007). This leads to efforts to improve the quality of STEM education in both school curricula and extracurricular studies in order to benefit more from the opportunities of STEM education

(Moreno et al., 2016) and more research on the factors associated with STEM achievement (Biçer et al., 2014). To evaluate the quality of STEM education, examine STEM achievement, determine students' STEM competencies, and examine their development, there is a need for measurement tools that can obtain valid and reliable scores. However, one of the main problems with STEM research is that measurement tools have not been developed based on a well-defined assessment framework to identify and evaluate STEM competencies. Existing research uses students' mathematics or science scores from national or international exams to investigate STEM achievement (Bicer et al., 2014; Bicer et al., 2017; Han et al., 2015; Ing, 2014). Using scores from achievement tests developed for other purposes, such as placement exams, international benchmark exams, or course grades, limits the quality of STEM-related research. There is a STEM competencies assessment framework recently developed by Arıkan et al. (2022) at the eighth-grade level. In order to bring the STEM competencies assessment framework for the primary school level to the literature, the aforementioned assessment framework was adapted to the fourth-grade level in the current study.

In the current study, the STEM assessment framework was empirically tested, and student data supported the proposed structure with 3 main dimensions and 10 sub-dimensions. In adapting this framework, the relationships within and between the dimensions were first theoretically evaluated and then empirically tested. Non-routine problem items were developed to test the structure of the framework. STEM dimensions provided different problem scenarios for students. A multidimensional STEM assessment model focusing on reasoning skills was validated. The model includes algorithmic thinking, concepts, pattern recognition, and argumentation sub-dimensions under the mathematics core dimension. The science core dimension has a science literacy subdimension. The main dimension of technology and engineering consists of modeling, techno-engineering problems, and coding sub-dimensions. Since the obtained data supported the adapted model, this model can be used in other studies to analyze fourth-grade students' STEM competencies. Therefore, the applicability of the framework is not limited to the local context of the framework and the developed test. Since the framework was developed by reviewing the skills identified in international large-scale assessments, it is expected to be widely used.

Conclusion

The importance of the current study lies in determining the relationships between STEM subdimensions, defining each sub-dimension, and creating an assessment framework by sampling items related to them. The developed test can be used by any researcher or

teacher to assess student competencies in the STEM field. Teachers or researchers can develop their own tests using the proposed framework. Since this framework was developed and validated for 4th and 8th-grade subjects based on data from students at these grade levels, the content of the test is generalized to cover these grade-level subjects. It is possible to develop tests based on the same sub-dimensions for student groups at lower or higher levels. The tests that can be developed for high school students using this framework are expected to include items that cover similar skills with respect to the relevant subject dimensions. A similar approach could be used to develop a test for primary school students younger than fourth grade.

It is anticipated that more studies will be conducted on examining the correlations between STEM sub-dimensions using the presented STEM assessment framework. With this framework, longitudinal studies aiming to monitor STEM competencies can be conducted. In addition, new studies focusing on the defined dimensions, measuring them, and revealing the relationships among them and between other factors can contribute to the current literature.

Limitations

The most prominent limitation of this study is the use of two questions for each subdimension. In order not to give 4th grade students a very long test, two questions were preferred for each sub-dimension. In other studies, this structure can be tested with tests containing more items in each sub-dimension. Another limitation is that the samples of the pilot and main study were selected based on convenience sampling. Data were obtained from schools that agreed to participate in the study. Although the data are sufficient for statistical analysis, replicating the findings based on larger random samples from different contexts may provide important evidence for generalizing the findings.

Ethics Committee Approval: *This study was conducted with the permission of Muğla Sıtkı Koçman University Human Research Ethics Committee (dated 23/03/2018 and numbered 45).*

Conflict of Interest: *The authors declare no conflict of interest.*

Author Contribution: *The authors contributed equally to the study.*

Acknowledgments: *This paper is a part of a research project supported by the Scientific and Technological Research Council of Turkey (TUBITAK) with program no. 2218. Any opinions expressed here are those of the authors and do not necessarily reflect the views of the TUBITAK.*

References

- Arikan, S., Erktin, E., & Pesen, M. (2022). Development and validation of a STEM competencies assessment framework. *International Journal of Science and Mathematics Education*, 20, 1-24. <https://doi.org/10.1007/s10763-020-10132-3>
- Bennett, C. A., & Ruchti, W. (2014). Bridging STEM with mathematical practices. *Journal of STEM Teacher Education*, 49(1), 17–28. <https://doi.org/10.30707/JSTE49.1Bennett>.
- Bicer, A., Capraro, R. M., & Capraro, M. M. (2017). Integrated STEM assessment model. *Eurasia Journal of Mathematics Science and Technology Education*, 13(7), 3959–3968. <https://doi.org/10.12973/eurasia.2017.00766a>.
- Bicer, A., Navruz, B., Capraro, R. M., & Capraro, M. M. (2014). STEM schools vs. non-STEM schools: Comparing students mathematics state based test performance. *International Journal of Global Education*, 3(3), 8–18.
- Blanton, M., & Kaput, J. (2011). Functional thinking as a route into algebra in the elementary grades. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 5–23). Berlin: Springer.
- Blum, W., & Ferri, R. B. (2009). Mathematical modelling: Can it be taught and learnt? *Journal of Mathematical Modelling and Application*, 1(1), 45–58.
- Broggy, J., O'Reilly, J., & Erduran, S. (2017). Interdisciplinarity and science education. In B. Akpan & K. Taber (Eds.), *Science Education. New directions in mathematics and science education* (pp. 81–90). Sense Publishers.
- Brown, R., Brown, J., Reardon, K., & Merrill, C. (2011). Understanding STEM: Current perceptions. *Technology and Engineering Teacher*, 70(6), 5–9.
- Brown, W. (2015). *Introduction to algorithmic thinking*. Retrieved from <https://raptor.martincarlisle.com/Introduction%20to%20Algorithmic%20Thinking.doc>.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates, Inc.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Capraro, R. M., & Corlu, M. S. (2013). Changing views on assessment for STEM project-based learning. In R. M. Capraro, M. M. Capraro, & J. Morgan (Eds.), *STEM project-based learning: An integrated Science, Technology, Engineering, and Mathematics (STEM) approach* (2nd ed., pp. 109–118). Rotterdam, The Netherlands: Sense Publishers.
- Carter, L. (2007). Globalization and science education: The implications of science in the new economy. *Journal of Research in Science Teaching*, 45(5), 617–633. <https://doi.org/10.1002/tea.20189>.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating

- normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23, 225-241.
- Doerr, H., & English, L. D. (2003). A modeling perspective on students' mathematical reasoning about data. *Journal for Research in Mathematics Education*, 34(2), 110-136.
- Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education*, 41(5), 605–618. <https://doi.org/10.1007/s11858-009-0184-6>.
- Enderson, M. C., & Ritz, J. (2016). STEM in general education: Does mathematics competence influence course selection. *Journal of Technology Studies*, 42(1), 30–41. <https://doi.org/10.21061/jots.v42i1.a.3>.
- English, L. D. (2015). STEM: Challenges and opportunities for mathematics education. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 4–18). Hobart, Australia: PME.
- Feinstein, N. (2011). Salvaging science literacy. *Science Education*, 95(1), 168-185.
- Fitzallen, N. (2015). STEM education: What does mathematics have to offer? In M. Marshman (Ed.), *Mathematics education in the margins. Proceedings of the 38th annual conference of the Mathematics Education Research Group of Australasia* (pp. 237–244). Sydney, Australia: Mathematics Education Research Group of Australasia (MERGA).
- Freudenrich, C., & Boyd, R. (2001). *How your brain works*. Howstuffworks. <http://www.howstuffworks.com/brain1.htm>
- Guyer, R., & Thompson, N. A. (2014). *User's Manual for Xcalibre item response theory calibration software, version 4.2.2 and later*. Woodbury MN: Assessment Systems Corporation.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>.
- Han, S., Capraro, R., & Capraro, M. M. (2015). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement. *International Journal of Science and Mathematics Education*, 13(5), 1089–1113. <https://doi.org/10.1007/s10763-014-9526-0>.

- Harwell, M., Moreno, M., Phillips, A., Guzey, S. S., Moore, T. J., & Roehrig, G.H. (2015). A study of STEM assessments in engineering, science, and mathematics for elementary and middle school students. *School Science and Mathematics, 115*(2), 66–74. <https://doi.org/10.1111/ssm.12105>.
- Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., & Urban- Lurain, M. (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: Forging a national network. *CBE—Life Sciences Education, 10*(2), 149–155.
- Ing, M. (2014). Can parents influence children's mathematics achievement and persistence in STEM careers? *Journal of Career Development, 41*(2), 87–103. <https://doi.org/10.1177/0894845313481672>.
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(11), 1–11. <https://doi.org/10.1186/s40594-016-0046-z>.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. The Guilford Press.
- Lee, N.H., Yeo, D. J. S., & Hong, S.E. (2014). A metacognitive-based instruction for primary four students to approach non-routine mathematical word problems. *ZDM, 46*(3), 465–480. <https://doi.org/10.1007/s11858-014-0599-6>.
- MEB. (2016). *STEM Eğitimi Raporu*. Ankara: Millî Eğitim Bakanlığı- Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü (YEĞİTEK).
- Milgram, R. J. (2007). What is mathematical proficiency? In AH Schoenfeld (Ed.), *Assessing mathematical proficiency* (pp. 31–58). Cambridge University Press. <https://doi.org/10.1017/CBO9780511755378.007>.
- Miller, J. (2019). STEM education in the primary years to support mathematical thinking: Using coding to identify mathematical structures and patterns. *ZDM, 51*(6), 915–927.
- Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology, 25*(6), 889–897. <https://doi.org/10.1007/s10956-016-9643-3>.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. <http://timssandpirls.bc.edu/timss2019/frameworks/>
- National Council of Teachers of Mathematics. (2018). *Building STEM education on a sound Mathematical Foundation. Position statement*. Retrieved from <https://www.nctm.org/Standards-and-Positions/NCTM-Position Statements/>.
- National Research Council. (2005). *How students learn: Mathematics in the classroom*. The National Academies Press.
- National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. The National Academies Press.

- National Science and Technology Council. (2011). *The federal science, technology engineering, and mathematics (STEM) education portfolio*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/costem_federal_stem_education_portfolio_report.pdf
- Nga, N. T., Quỳnh, T. T. X., Uyên, N. P., & Trung, T.T. (2022). Một số nghiên cứu về năng lực STEM trên thế giới và đề xuất khung năng lực STEM cho học sinh phổ thông tại Việt Nam (An overview study on STEM competencies in the world and propose a STEM competency framework for high school students in Vietnam), *Tạp chí Giáo dục*, 22, 48-53.
- Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics*, 72(1), 39–60. <https://doi.org/10.1007/s10649-009-9181-5>.
- Rumsey, C., & Langrall, C. W. (2016). Promoting mathematical argumentation. *Teaching Children Mathematics*, 22(7), 412-419.
- Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., & Skinner, EA (2014). A common measurement system for K-12 STEM education: Adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking. *Studies in Educational Evaluation*, 40, 18–35. <https://doi.org/10.1016/j.stueduc.2013.11.005>.
- Schmidt, W. H., & Houang, R. T. (2007). Lack of focus in the mathematics curriculum: A symptom or a cause? In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 65–84). Brookings Institution Press.
- Schumacker, R. E. & Lomax, R. G. (2004). *A beginners guide to structural equation modeling*. Lawrence Erlbaum Associates, Inc.
- Shulman, L. (2009). Assessment of teaching or assessment for teaching? In D.H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality*. Sage Publications.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press.
- Trung, T. T., Quỳnh, T. T. X., Uyên, N. P., & Nga, N. T. (2022). Xây dựng và chuẩn hóa công cụ đánh giá năng lực stem của học sinh trung học phổ thông tại thành Phố Hồ Chí Minh (Develop and standardize a stem competency assessment tool for high school students in Ho Chi Minh City). *Ho Chi Minh City University of Education Journal Of Science*, 19(8), 1255-1270. [https://doi.org/10.54607/hcmue.js.19.8.3408\(2022\)](https://doi.org/10.54607/hcmue.js.19.8.3408(2022))
- Turner, E. E., Roth McDuffie, A., Bennett, A. B., Aguirre, J., Chen, M. K., Foote, M. Q., & Smith, J. E. (2022). Mathematical modeling in the elementary grades: Developing and testing an assessment. *International Journal of Science and Mathematics Education*, 20(7), 1387-1409.
- TÜSİAD. (2014). *STEM Alanında eğitim almış işgücüne yönelik talep ve beklentiler araştırması*. http://tusiad.org/tr/yayinlar/raporlar/item/download/7014_d28ffa2adda423c6d3852cc01c965993

- Ullman, J. B. (2001). Structural equation modeling. In B. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp.653-771). Allyn & Bacon.
- van Borkulo, S., Chytas, C., Drijvers, P., Barendsen, E., & Tolboom, J. (2021, October). Computational thinking in the mathematics classroom: fostering algorithmic thinking and generalization skills using dynamic mathematics software. In *The 16th Workshop in Primary and Secondary Computing Education* (pp. 1-9).
- White, D. W. (2014). What is STEM education and why is it important. *Florida Association of Teacher Educators Journal*, 1(14), 1–9.
- Wong, G. K., Cheung, H. Y., Ching, E. C., & Huen, J. M. (2015, December). School perceptions of coding education in K-12: A large scale quantitative study to inform innovative practices. In *2015 IEEE international conference on teaching, Assessment, and learning for engineering (TALE)* (pp. 5-10). IEEE.
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K.R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide (NCEE 2012-4055)*, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Washington, DC. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/.
- Yıldırım, H., & Gelmez-Burakgazi, S. (2020). Research on STEM education studies in Turkey: A qualitative meta-synthesis study. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 50, 291-314. <https://doi.org/10.9779/pauefd.590319>

Appendix

Questions

Question 1. [Algorithmic Thinking]

To calculate the result of multiplying a two-digit number by 999 in a short way, follow the steps below.

Step 1: Subtract 1 from the number multiplied by 999.

Step 2: Subtract the number obtained in the first step from 999.

Step 3: Write the results obtained in steps 1 and 2 side by side.

For instance, for 15×999

1) $15 - 1 = 14$

2) $999 - 14 = 985$

3) 14985'tir.

Find the result of 28×999 using the steps given above.

1)

2)

Question 4. [Concepts and Principles]

Starting from 1, a number is called a "perfect number" if the sum of the numbers it is exactly divided by, other than itself, is equal to itself.

For example, the number 6 is a perfect number because it is exactly divisible by 1, 2 and 3 and $1+2+3=6$ (6 is not exactly divisible by 4 and 5).

However, since the number 10 is exactly divisible by 1, 2 and 5 except itself, but $1+2+5=8$, the number 10 is not a perfect number (the number 10 is not exactly divisible by 3, 4, 6, 7, 8 and 9).

Accordingly, which of the numbers given below is a perfect number?

A) 30 B) 28 C) 15 D) 11

Question 5. [Patterns]

First Number	Second Number	Third Number	Fourth Number	Fifth Number				Ninth Number
3	2	6	2	2				?

To get the number sequence above, multiply the first number by the second number. If the result has one digit, that number is written as the third number of the sequence; if it has two digits, the first digit of the

result is written as the third number of the sequence. Similarly, the second and third numbers are multiplied to get the fourth number. In this way, the sequence continues.

For example, the number 6 is obtained by multiplying the first number 3 and the second number 2. Since 6 is a one-digit number, the third number of the sequence is written as 6.

Then the second number 2 is multiplied by the third number 6 to get the number 12. Since 12 is a two-digit number, the fourth number of the sequence is written as 2.

Since the sequence continues in this way, what is the ninth number?

- A) 2 B) 3 C) 6 D) 8

Question 7. [Argumentation]

Table: Price Changes of Products in the Store

Product	Old Price (TL)	New Price (TL)
Footwear	100	70
Bot	500	400
Jacket	2000	2300
Knee Brace	50	40
Dress	2000	1900

Melis comes to the store to buy one of each of the 5 items listed in the table and sees that the prices have changed. "Today is my lucky day. Four of the products I was going to buy are on sale. The price of one product has increased. I will pay less," she thinks. Is Melis right that she will pay less?

In your answer, indicate "she is right" or "she is not right" and explain why.

Question 10. [Science literacy]

If we know how long it takes a ball dropped from a tower to hit the ground, we can find the height of the tower from which it was dropped.

To do this, we multiply the fall time by itself and multiply



the result by five.

For example, a ball that hits the ground in 2 seconds is dropped from a tower $2 \times 2 = 4$, $4 \times 5 = 20$ meters high.

Accordingly, how many meters high was a ball dropped from a tower that fell in 3 seconds?

- A) 25 B) 27 C) 45 D) 125

Question 11. [Modeling]

The relationship between the number of accidents in different countries and income and temperature is analyzed.

As income increases, the number of accidents decreases, but as temperature increases, the number of accidents increases. In which country would you expect the number of accidents to be higher?

	Country	Income (TL)	Temperature (°C)
A)	Astanya	2000	28
B)	Buanda	2000	25
C)	Ceristan	3000	28
D)	Dastarya	3000	25

Question 13. [Coding]

The following lines list the mathematical commands and the operations to be performed according to these commands. One of the numbers in a box is drawn and the color of the light is determined by following the commands below according to that number.

Line	Command
1	If "number in the draw" > 5
2	Multiply this number by 2 and go to line 5
3	If not
4	Multiply this number by 3 and go to line 5
5	Subtract 7 from the number you get and go to the appropriate row below.
6	If result < 10
7	Let the screen be red

8	If $9 < \text{result} < 15$
9	Let the screen be blue
10	If $14 < \text{result} < 20$
11	Let the screen be yellow
12	If $\text{result} > 19$
13	Let the screen be green.

If the number in the draw is 4, what color will the screen be?

- A) Red
- B) Blue
- C) Yellow
- D) Green

Question 16. [Techno-engineering]

An ice cream shop predicts that every 1 TL increase in the price of ice cream will result in 10 fewer ice creams sold per day. Currently, he sells ice cream at 4 TL. He sells 100 ice creams a day and earns 400 TL.

The ice cream seller wants to find the ice cream price that will generate the most revenue by filling in the table below. He fills in the first two rows. What is the price of ice cream that will generate the most income? You can find it by filling in the other rows.

Ice Cream Price (TL)	Number of Ice Cream Sold	Revenue (TL)
4 TL	100	$4 \times 100 = 400$ TL
5 TL	90	$5 \times 90 = 450$ TL

- A) 6
- B) 7
- C) 8
- D) 9